Protein-protein interactions in the bacterial type VI secretion system

Dr. Asma Ahmad

PhD Thesis

Supervisor: Dr. Mark Thomas

Year of submission 2013

Protein-protein interactions in the bacterial type VI secretion system

Dr. Asma Ashor Alajeli O. Ahmad

PhD Thesis

Medical school

Department of Infection and Immunity

Supervisor: Dr. Mark Thomas

Date of submission 1 October 2013

Abstract

Protein secretion in Gram-negative bacteria plays a vital role in their survival in the environment and their virulence. At least nine secretion systems have been identified in Gram-negative bacteria, including the type VI secretion system (T6SS). The T6SS appears to play an important role in the virulence of many pathogenic bacteria and also in interbacterial competition. The T6SS consists of 13 core protein subunits (TssA-TssM) some of which exhibit structural similarity to components of the bacteriophage T4 contractile tail. The T6SS is predicted to employ an analogous mechanism to inject bacterial effector proteins directly into host eukaryotic cells, competing bacteria or into the extracellular medium.

The interaction between the 13 core components of the T6SS was investigated by using the bacterial adenylate cyclase two-hybrid (BACTH) system. The results of the BACTH analysis of the interactions between these proteins showed that many Tss subunits are self-interacting such as TssH, TssI, TssJ, TssK, TssL and TssM. The results also revealed a number of novel hetero-oligomeric interactions between the different subunits. Furthermore, BACTH interaction results confirmed that TssA self-interacts mainly through its C-terminal domain and also interacts with a number of other T6SS subunit proteins, suggesting that it is central to the assembly of the baseplate complex of the T6SS. The analysis of proteolytic degradation products of TssA support a model in which TssA is organised into two domains linked by an unstructured linker region. Size exclusion chromatography and analytical ultracentrifugation revealed that TssA forms a large oligomer. Negative stain EM showed that TssA forms a ring of 20-25 nm diameter with external projections which resemble the baseplate component of the T4 phage tail complex. Moreover, by constructing a MBP-TssA C-terminal domain fusion the C-terminal domain of TssA was demonstrated to be necessary and sufficient for ring formation. Two-hybrid analysis, SEC and negative stain EM suggest that a related protein, TssA^E, encoded within the *Aeromonas hydrophila* T6SS gene cluster, is functionally orthologous to TssA.

Acknowledgements

First and foremost I would like to express my sincere gratitude to my supervisor Dr. Mark Thomas for all the help, guidance and support that he has given me throughout my PhD. His guidance helped me during the entire time of my research and writing of this thesis.

I would like to thank Mr. Svetomir Tzokov for doing all the EM work, Mrs. Amy Barker for the AUC analysis, Dr Arthur Moir for the N-terminal sequence analysis, Mr. James Ault for the mass spectrometry and Dr. Rosemary Staniforth for the circular dichroism analysis.

I thank Dr. Sravanthi Shastri for her guidance during the initial months of my PhD. I would also like to thank my colleagues in Dr. Thomas' lab, Mr. Daniel Mosby, Miss Helena Spiewak, Mr. Jamie Hall, Mr. Richard Jones, Miss Ruyue Sun, Miss Sarah Batters and Miss Saeedah Aljadani for making the environment in the lab thoroughly enjoyable.

I would like to thank my family, especially my mother and father for always believing in me, for their continuous love and their support in my decisions and without whom I could not have made it here. I dedicate this work to my parents and my sister.

List of abbreviations

AAA$^{(+)}$ - ATPase associated with various cellular activities

AS - ammonium sulfate

Ap^r - ampicillin resistrance

APEC - avian pathogenic *E. coli*

ATP - Adenosine triphosphate

AUC - Analytical Ultracentrifugation

BACTH - Bacterial Adenate Cyclase-Based Two-Hybrid

Bcc - Burkholderia cepaecia complex

BHI - Brain-Heart Infusion broth

Bp - Base pairs

BSA - Bovine Serum Albumin

cAMP - Cyclic adenosine 3', 5'-monophosphate

CF - Cystic fibrosis

Cm - Centimetres

Cm^r - Chloramphenicol resistance

CO - cut off

CTD - C-terminal domain

Da - Daltons

DMSO - Dimethyl sulphoxide

DNA - Deoxyribonucleoside triphosphate

DNAse - Deoxyribonuclease

dNTP - Deoxyribonucleoside triphosphate

Δ - Deletion

ºC - Degrees centigrade

EM - Electron Microscopy

EAEC - enteroaggregative *E. coli*

EDTA - Ethylenediamine tetra-acetic acid

FHA - forkhead associated domain

Gp (ex. Gp19) - Gene Product

GSP - General Secretory Pathway

g - Grams

×g - Gravitational accaleration

PG - Peptidoglycan

His-tag - Histidine tag

H - Hours

Hcp - haemolysin co-regulated protein

IAHP - IcmF-associated homologous proteins

IPTG - Isopropyl β-D-Thiogalactoside

IM - Inner membrane

IMAC - Immobilised metal ion affinity chromatography

IPTG - Isopropyl β-D-thiogalactoside

K - Kilo

Kb - Kilobase pairs

Km - Kanamycin

kDa - Kilodalton

LB - Lysogeny Broth

L - Litres

mol. wt - molecular weight

M - Molar

m - Milli (10^{-3})

ml^{-1} - Per milli litre

MCS - Multiple cloning site(s)

μ - Micro (10^{-6})

min - Minutes

mol - Moles

MOPS - 4-morpholinepropanesulfonic acid

mRNA - Messenger RNA

NTD - N-terminal domain

NBD – nucleotide binding domain

OD - Optical density

Oligo - Oligodeoxyribonucleotide

OM - Outer membrane

ORF - Open reading frame

ori - Origin(s) of DNA replication

% - Percentage

P - Plasmid

p - Promoter

PAGE - Polyacrylamide gel electrophoresis

PCR - Polymerase chain reaction

PAS - protein A Sepharose

PEI - polyethyleneimine

^{r} - Resistant

RNA - Ribonucleic acid

RNAse - Ribonuclease

rpm - Revolutions per minute

Rhs - rearrangement hot spot

sec - Seconds

SDS - Sodium deodecyl sulphate

SDS-PAGE - SDS Polyacrylamide Gel electrophoresis

SEC - Size Exclusion Chromatography

SEC-MALLS - Size Exclusion Chromatography-Multi-Angle Laser Light Scatter

Tris - Tri (hydroxymethyl) methylamine

T1SS - Type 1 Secretion System

T2SS - Type 2 Secretion System

T3SS - Type 3 Secretion System

T4SS - Type 4 Secretion System

T5SS - Type 5 Secretion System

T6SS - Type 6 Secretion System

T7SS - Type 7 Secretion System

T8SS - Type 8 Secretion System

T9SS - Type 9 Secretion System

Tags - T6SS associated genes

Tat - Twin-arginine translocation

TPS - Two-Partner Secretion

T/tss - Type 6 secretion system (gene or protein name)

UV - Ultra violet (light)

Vas - Virulence associated secretion

VgrG - Valine glycine repeat protein

VSVg - G glycoprotein of the vesicular stomatitis virus

V - Volts

VgrG - Valine glycine repeat protein G

v/v - Volume:volume ratio

w/v - Weight:volume ratio

X-gal - 5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Contents

Chapter 1. Introduction

1.1. The *Burkholderia cepacia* complex

The genus *Burkholderia* consists of more than 60 species of non-fermentative, non-spore forming bacilli within the β class of the phylum *Proteobacteriaceae* and therefore they are Gram-negative (Compant *et al*., 2008). The *Burkholderia cepacia* complex (BCC) consists of 17 species with similar phenotypic features but distinct genetic composition (Mahenthiralingam *et al*., 2008, Vanlaere *et al*., 2008). The first member of the BCC was identified in 1950 by W.H. Burkholder, who described it as the causative agent of soft onion root (Burkholder, 1950). It was initially described as a member of the genus *Pseudomonas* due to the close phenotypic resemblance to the pseudomonads, and named *P. cepacia*. Until the early 1970s, different species now known to belong to the BCC were classed under the genus *Pseudomonas*, which belongs to the γ subgroup of proteobacteria. In 1973 considerable differences were demonstrated within the pseudomonad family at the ribosomal RNA homology level (Palleroni *et al*., 1973). Therefore, *P. cepacia* and 6 other *Pseudomonas* species (*Pseudomonas solanacearum*, *Pseudomonas pickettii*, *Pseudomonas gladioli*, *Pseudomonas mallei*, *Pseudomonas pseudomallei*, and *Pseudomonas caryophylli)* were classified within RNA homology group II pseudomonas. Later they were reclassified under a new genus - *Burkholderia* based on their DNA-DNA homology values, cellular lipid and fatty acid composition, 16S rRNA sequences and phenotypic characteristics (Yabuuchi *et al*., 1992). By 1997, it became clear that *B. cepacia* consisted of at least five distinct genomovars (I-V) which are phenotypically similar but genotypically distinct strains (Vandamme *et al*., 1997). Subsequently, these were assigned species designations (*B. cepacia, B. multivorans, B. cenocepacia, B. stabilis* and *B. vietnamiensis*). Since 1997 further studies have resulted in addition of another twelve species to the BCC. Table 1.1 lists the seventeen currently known species that constitute the BCC.

BCC bacteria are widely distributed throughout the environment and commonly found in soil, especially the rhizosphere of plants. Some species have beneficial interactions with plants and provide protection against fungal infections (biopesticidal properties) such as infection of the seeds with *Rhizoctonia solani* and *Pythium* species which cause a condition known as "damping off". Another useful feature of some members of the BCC is the ability to degrade pollutants because of their ability to use a wide variety of carbon sources such as groundwater pollutants. However, they have a dark side as some

Table 1.1 *B. cepacia* complex species and their habitat[a]

Name	Habitat	Reference
B. cepacia	Humans (CF and non-CF), elimination of pollutants from the environment and biological control	(Vandamme *et al*., 1997)
B. multivorans	Humans (CF and non-CF)	(Vandamme *et al*., 1997)
B. cenocepacia	Humans (CF and non-CF), animals, soil, plant, water, and biological control	(Vandamme *et al*., 2003)
B. stabilis	Humans (CF and non-CF), rhizosphere soil, hospital equipment	(Coenye *et al*., 2001)
B. vietnamiensis	Humans (CF and non-CF), elimination of pollutants from the environment and biological control	(Coenye *et al*., 2001)
B. dolosa	Humans (CF)	(Vermis *et al*., 2004)
B. ambifaria	Humans (CF and non-CF) and biological control	(Coenye *et al*., 2001)
B. anthina	Humans (CF and non-CF)	(Vandamme *et al*., 2002)
B. pyrrocinia	Humans (CF and non-CF) and biological control	(Vandamme *et al*., 2002)
B. ubonensis	Hospital-acquired infection	(Vanlaere *et al*., 2008)
B. latens	Humans (CF)	(Vanlaere *et al*., 2008)
B. diffusa	Humans (CF and non-CF), soil, water	(Vanlaere *et al*., 2008)
B. arboris	Humans (CF and non-CF), soil, rhizosphere soil, water	(Vanlaere *et al*., 2008)
B. seminalis	Humans (CF and non-CF), soil, rhizosphere soil, water	(Vanlaere *et al*., 2008)
B. metallica	Humans (CF)	(Vanlaere *et al*., 2008)
B. contaminans	Infections in animals and in humans (CF)	(Vanlaere *et al*., 2009)
B. lata	Forest soil	(Vanlaere *et al*., 2009)

[a]Adapted from (Sousa *et al*., 2010).

species are able to cause opportunistic infections in humans (Parke and Gurian-Sherman, 2001).

BCC are major opportunistic pathogens in immunocompromised patients. The severity of infection with this type of pathogenic bacteria in cystic fibrosis patients can vary considerably between chronic lung infection to ''cepacia syndrome'', which is a combination of necrotizing pneumonia and sepsis leading to a rapid clinical deterioration in lung function and eventually death (Isles *et al*., 1984). Patients with chronic granulomatous disease (CGD) are also at risk of serious BCC infections (Speert, 2001). Cepacia syndrome is the second leading cause of death in CGD patients (Johnston, 2001).

Among BCC species, *B. cenocepacia* is the most common member of the group to cause clinical infection in cystic fibrosis patients, resulting in as many as 70% of cases of BCC infection in cystic fibrosis patients (Figure 1.1). However, recent studies shown that *B. multivorans* has become the most common BCC isolate in the United States and the United Kingdom (Reik *et al*., 2005, Mahenthiralingam *et al*., 2008). *B. multivorans* is the second most frequent BCC pathogen isolated from cystic fibrosis patients (Mahenthiralingam *et al*., 2002, Mahenthiralingam *et al*., 2005). The nucleotide sequences of the genomes of several strains of *B. cenocepacia* have been determined. In one strain, J2315, the genome is 8.06 Mb and consists of three circular chromosomes and one plasmid.

1.2. Virulence determinants of the BCC

A wide variety of virulence determinants may have a role in the ability of some species of BCC to cause disease, such as biofilm formation, iron acquisition, the ability to survive intracellulary and some protein secretion systems. Clinically; BCC species are problematic because they can cause invasive lung disease in immuno-compromised and cystic fibrosis (CF) patients and they are difficult to treat as they exhibit a high level of intrinsic antibiotic resistance. Therefore, research on the virulence determinants is important to identify bacterial processes that might be targeted by a novel vaccine or new antibiotics.
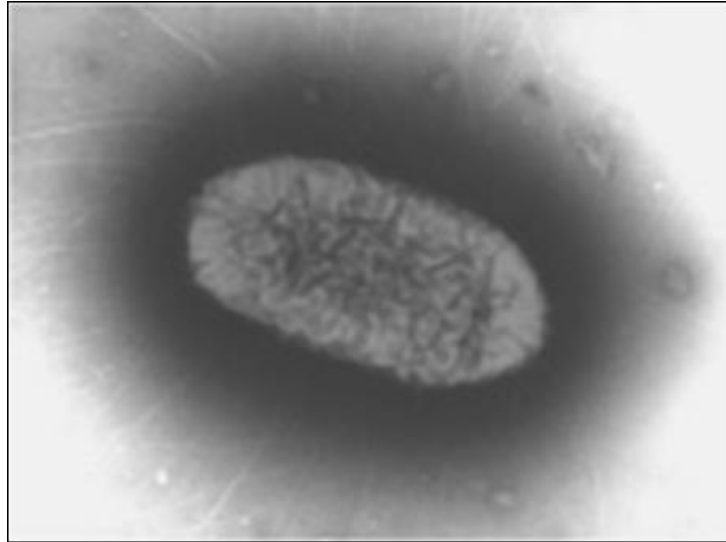
Figure 1.1. Electron micrograph of *B. cenocepacia* strain J2315. Reproduced from http://accessscience.com/content/Burkholderia-cepacia-complex. Image source: Emma Ralph, Eshwar Mahenthiralingam and Philip Sayre, Cardiff School of Biosciences, University of Cardiff, UK.

For instance, biofilm formation in *B. cenocepacia* can protect the bacteria from antibiotics as well as the host immune response (Estrela *et al*., 2009). It is predicted that in the CF lung *B. cenocepacia* live in biofilms, sometimes with *P. aeruginosa* and communicate together by using their quorum-sensing systems which is another virulence determinant of *B. cenocepacia* (Tomlin *et al*., 2001). Another important virulence trait of BCC species is the production of up to four types of siderophores (pyochelin, ornibactin, cepaciachelin, and cepabactin) for iron uptake and chelation (Thomas, 2007). However, most species only produce two of these, ornibactin and pyochelin.

In addition, a number of secretion systems were found to have a role in the virulence of BCC species by secreting effector molecules that disrupt cell function and contribute to causing a disease. In *B. cenocepacia*, the type III secretion system (T3SS) is required for survival in a murine infection model as the bacterial mutant showed attenuated virulence compared to the wild type (Tomich *et al*., 2003). One of the two type IV secretion systems (T4SS) in *B. cenocepacia* contributes to the intracellular survival and replication in macrophages and airway epithelial cells (Sajjan *et al*., 2008). *B. cenocepacia* survive inside the macrophages within a membrane-bound vacuole (BcCV). The type VI secretion system (T6SS) has a very interesting role in the virulence of *B. cenocepacia*. The T6SS was shown to trigger a drastic alteration in the morphology of infected macrophages by inactivation and reduction of the cellular pool of Rac1 and Cdc42, members of Rho family GTPases which are involved in the regulation of intracellular actin dynamics. The T6SS also inhibits the recruitment of the NADPH oxidase complex, including Rac1, to the BcCV membrane (Flannagan *et al*., 2012, Rosales-Reyes *et al*., 2012b). Moreover, cooperation between type II secretion system (T2SS) and T6SS was recently proposed in *B. cenocepacia*. Several proteins are secreted to the extracellular environment by the *B. cenocepacia* T2SS such as the zinc-dependent metalloproteases ZmpA and ZmpB that can degrade several substances including host immunity proteins (Kooi and Sokol, 2009). Although the T2SS cannot inject its effectors directly into the target cell (Section 1.3.2), Rosales-Reyes and colleagues showed that a functional T6SS facilitates the translocation of ZmpA into the cytoplasm of macrophages (Rosales-Reyes *et al*., 2012a).

1.3. Protein secretion systems in Gram-negative bacteria

Protein secretion is defined as the transport of proteins from the interior to the exterior of the cell. It is an important process for bacterial survival as it plays a key role in numerous aspects of the bacterial life cycle such as the mediation of the interaction between bacteria and their eukaryotic hosts, biosynthesis of extracellular organelles such as pili and flagella, and nutrition. The structure of Gram-positive and Gram-negative bacteria differ by the latter having two membranes which make protein secretion a more complicated process. Protein secretion in Gram-negative bacteria involves the transport of the secreted protein across the inner membrane (IM) and outer membrane (OM) as well as passing through the periplasmic space. A further difficulty faced by Gram-negative bacteria is the lack of ATP or energy source in the OM (Thanassi *et al*., 2005). Therefore, either the secretion process across both membranes should occur in a single step or in separate steps in which case the energy required for the transport of the secreted protein from the periplasm across the OM must be acquired at the IM (Thanassi and Hultgren, 2000). In situations where the protein is secreted in two steps, it is translocated across the IM using the Sec or Tat (see below).

Research on protein secretion has identified at least nine specialised protein secretion systems in Gram-negative bacteria. Five of them, i.e. the Type II, -Va, -Vb, -VII, -VIII and -IX secretion systems, use the Sec or Tat pathway for transporting the secreted protein across the IM, whereas the other secretion systems, i.e. the Type I, -III, -IV and Type -VI secretion systems adopt a direct transport mechanism through both membranes without a periplasmic intermediate (Figure 1.2). Despite the involvement of all these secretion systems in the pathogenicity of bacteria, it has been found that the evolution of several of these secretion systems are from bacterial organelles associated with specific cell functions other than virulence. The type I secretion system (T1SS) is believed to be evolved from ATP-binding cassette (ABC) transporters, while the T2SS, T3SS and T4SS are similar to type IV pili, flagella and conjugation machines, respectively (Cascales, 2008).
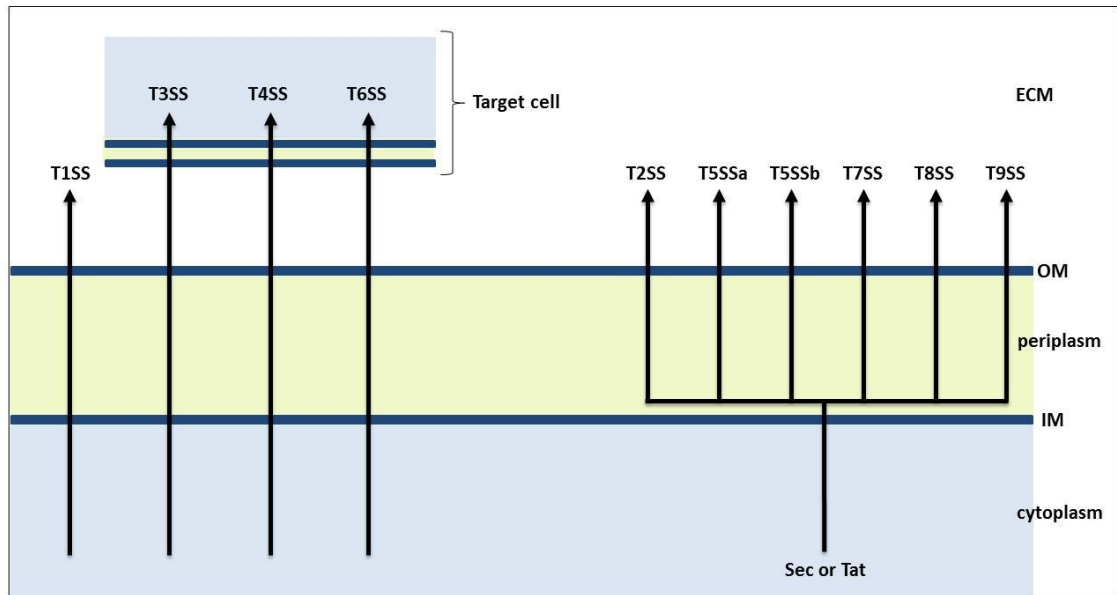
Figure 1.2. Schematic representations of secretion systems of Gram-negative bacteria. The diagram shows that type I, III, IV and VI secretion systems transport their substrate proteins to the extracellular media or to the target cell cytosol in one step, whereas, type II, Va, Vb, VII, VIII and IX are two-step systems where the first step (translocation across the inner membrane) is mediated via Sec or Tat. Abbreviations: IM, bacterial inner membrane; OM, bacterial outer membrane; ECM, extracellular milieu.

1.3.1. The Sec and Tat systems

Translocation of protein across the bacterial inner (or cytoplasmic) membrane to the periplasmic space is controlled by two systems: the Sec (general secretory) and Tat (twin-arginine translocation) pathways. The proteins that are transported by the Sec or Tat pathways have specific N-terminal signal sequences that are used to guide the protein to the Sec or Tat system located in the cytoplasmic membrane. When the protein passes through the inner membrane, the function of N-terminal sequence is completed and therefore it is cleaved off by a signal peptidase (Economou, 1999, Sargent *et al*., 2006).

The Sec system of *E. coli* is constituted of two main parts, a chaperone located in the cytosol named SecB, and a large protein complex in the inner membrane of the bacterial cell referred to as the translocase. The function of SecB chaperone is to hold the proteins in the correct conformation for translocation, and also targeting them to the translocase. The translocase has an ATPase component (SecA) and a membrane protein complex (Driessen *et al*., 1998, Stathopoulos *et al*., 2000). The process of protein translocation begins by selective binding of SecB to the preprotein that needs to be translocated (Driessen *et al*., 1998). It is worth mentioning that SecB chaperone cannot differentiate between secretory and non-secretory proteins, as it does not have specificity for signal peptides (Knoblauch *et al*., 1999, Stathopoulos *et al*., 2000). Subsequently, the SecB-preprotein complex binds to SecA and the preprotein is released from SecB and binds to SecA. SecA undergoes conformational changes after binding of an ATP molecule to the nucleotide binding site. These results in a part of the preprotein and the signal peptide being inserted into the translocation channel and ~ 2.5 kDa of the preprotein is translocated. Hydrolysis of ATP results in reverse of the conformational changes of SecA and release of the bound preprotein. Then the cycle is repeated until all of the protein is translocated (Driessen *et al*., 1998, Stathopoulos *et al*., 2000). The Sec pathway can only transport proteins in their unfolded state as they must fit through the 5-8 Å pore of the Sec system and the SecB chaperone keeps the preprotein in the required conformation (Pugsley, 1993, Van den Berg *et al*., 2004). If a Sec-dependent protein folds prematurely in the cytoplasm, it cannot be translocated via the Sec system. This might result in accumulation of precursors and subsequent cell toxicity. The Sec system is used to transport proteins in Gram-positive bacteria, Gram-negative bacteria, eukaryotes and Archaea (Pohlschroder *et al*., 1997). In Gram-negative

bacteria the secretion of Sec-dependent proteins across the OM must be completed by separate mechanisms which are specialised secretion systems (Driessen *et al*., 1998).

In *E. coli* the Tat system is composed of at least 3 main proteins that form the translocase: TatA, TatB and TatC (Berks, 1996, Settles *et al*., 1997, Berks *et al*., 2000). Proteins exported through the Tat system appear not to require a channel, as proteins are extracted through a layer of TatA by TatC (Fisher *et al*., 2008). The Tat system displays preference for proteins that binds to a cofactor in the cytoplasm such as haem or iron-sulpher clusters, and therefore they are folded before export (Berks *et al*., 2000, Lee *et al*., 2006). However, *in vitro* studies showed that some cofactor-free proteins are also translocated by the Tat system (Berks, 1996, Bruser *et al*., 1998).

The signal peptides of both Sec and Tat systems are short peptide chains of ~ 24 amino acids that are composed of three different regions: the n-region, which is an N-terminal positively charged region; the h-region, which is a hydrophobic helical region, and a c-domain that contains the signal peptidase cleavage site (von Heijne, 1985, Izard and Kendall, 1994, Cristobal *et al*., 1999). Tat system-dependent signal peptides have additional features, the most distinctive feature of which is a conserved S/T-R-R-x-F-L-K motif. The name of the Tat system is related to this conserved motif which contains a dispensable but highly conserved twin-arginine motif (Chaddock *et al*., 1995, Sargent *et al*., 1998).

1.3.2. Type I secretion system

The type I secretion system (T1SS) is Sec-independent and secretes proteins across the IM and OM in one step. The T1SS consists of three components, two components of the IM translocase: an ATP-binding cassette (ABC) transporter and a membrane fusion protein (MFP), in addition to a pore-forming outer membrane-associated protein, the so-called outer membrane (OMP) protein. The secreted substrates have a non-cleavable C-terminal signal sequence (Delepelaire, 2004). The T1SS exports high molecular weight toxins and exoenzymes from the inside to the outside of the bacterial cell in one step. The OMP forms a β-barrel through which proteins are exported. An example of a T1SS is the secretion system for α-haemolysin (HlyA) of pathogenic *E. coli* (Kostakioti *et al*., 2005).

1.3.3. Type II secretion system

The type II secretion system (T2SS), also referred to as the secreton, is a Sec-dependent multi-protein complex composed of 12-15 proteins that span the inner and outer membranes (Voulhoux *et al*., 2001). Apart from one subunit, the outer membrane protein D (i.e. GspD or XcpQ), which is also known as the secretin as it forms a pore in the OM, all the T2SS proteins are associated with the IM and periplasm (Korotkov *et al*., 2011, Van der Meeren *et al*., 2012). GspD is composed of two domains: a highly conserved C-terminal domain and a non-conserved N-terminal domain. The C-terminal domain is integrated into the outer membrane and oligomerises into a ring (Nouwen *et al*., 1999, Thanassi, 2002, Peabody *et al*., 2003). The N-terminal domain is extended into the periplasm, and has many roles, including substrate recognition. In some systems, an additional protein, the lipoprotein GspS, is integrated into the periplasmic side of the OM. This protein is known as a pilotin and it is involved in assembly of the secretin (Hardie *et al*., 1996, Shevchik *et al*., 1997). The current model of the T2SS suggests that it is formed from three sub-assemblies (Douzi *et al*., 2012). A sub-complex associated with the IM called the IM platform (IMP), which composed from GspC, -F, -L and -M proteins; the cytoplasmic traffic ATPase, the second subassembly, is connected to IMP via its interaction with GspL (Bleves *et al*., 1996, Arts *et al*., 2007). In addition, the T2SS contains five proteins GspG, -H, -I, -J and -K which are homologous to the type IV pilin protein PilA and are therefore called pseudopilins (Sauvonnet *et al*., 2000, Korotkov and Hol, 2008). Lastly, the third subassembly is composed of the pore-forming GspD subunit in the OM which forms a channel through which secreted proteins can pass to the extracellular environment (Brok *et al*., 1999, Nouwen *et al*., 1999). It is proposed that GspC recruits secreted proteins which are transferred to the T2SS vestibule. Then the secreted protein is contacted by the tip of a pseudopilus formed from the pseudopilin proteins which grows and pushes it through the GspD pore (Douzi *et al*., 2012). Examples of proteins secreted by the T2SS include hydrolytic enzymes such as proteases, cellulases, pectinases, phospholipases and lipases, as well as some toxins such as cholera toxin (Sandkvist, 2001, Hazes and Frost, 2008).

1.3.4. Type III secretion system

The type III secretion system (T3SS) is Sec-independent and delivers toxin into eukaryotic cells in one step crossing both bacterial membranes and the eukaryotic cell membrane. It consists of at least 25 different proteins, which form two main parts: the basal body, which spans the entire cell envelope, and the needle complex (NC) which protrudes from the outer membrane and is also called the injectisome (Galan and Collmer, 1999, Galan and Wolf-Watz, 2006). The injectisome shares similarity with bacterial flagella and they are believed to share a common ancestor (Hayes *et al.*, 2010). Bacterial-host contact is required for the activation of the T3SS and it is used by many pathogenic bacteria such as *Yersinia* spp., *Salmonella* spp., *Shigella* spp. and enteropathogenic *E. coli*. Also, the T3SS is used by symbiotic bacteria. The T3SS has been shown to deliver toxins, virulence factors and other effector proteins to the extracellular milieu or directly into the host cell (Mota *et al.*, 2005). The T3SS effectors have a specific protein sequence that acts as a signal sequence to direct these effectors to the T3SS. Also, many T3SS effectors require a chaperone, which dissociates from the effector and remains in the bacterial cytoplasm during translocation of the effector into the target cell (Ghosh, 2004). When T3SS effectors are injected into the target cell they become activated by host factors which are not present in the secreting bacterium. Therefore, preventing toxicity to the secreting bacterium (Dean, 2011).

1.3.5. Type IV secretion system

The type IV secretion system (T4SS) is a complex and diverse system which can be divided into two main types:

1.3.5.1. Type IV A secretion system

The type IV A secretion system (T4SSa) is similar to the conjugation system of bacteria. In addition to the secretion of protein effectors, this system can export and import DNA and therefore may facilitate the spread of antibiotic resistance genes (Christie and Cascales, 2005, Hayes *et al.*, 2010). Usually, direct cell-cell contact is required for the activation of T4SS. The T4SSa is largely made up from three subcomplexes, a pilus/adhesin that mediates cell-cell contact, a secretion channel through which protein passes, and finally a receptor for the channel's substrate proteins known as type 4 coupling protein (T4CP). A well-studied example of a T4SSa is the VirB/VirD4 system of the plant pathogen *Agrobacterium tumefaciens* which injects DNA directly into the

host cell. The secretion channel is composed of 11 proteins (VirB1-VirB11) which are known as mating-pair formation (Mpf) proteins. In this system, T4CP is called VirD4 and it delivers the transferred DNA strand to the secretion channel. VirD4 protein may act as a translocase as it has DNA-dependent ATPase activity. VirB4 and VirB11 are also ATPases which may provide the energy required for T4SS assembly and/or secretion (Hayes *et al.*, 2010).

1.3.5.2. Type IV B secretion system

The type IV B secretion system (T4SSb) transports and delivers effector proteins directly into the cytoplasm of target cells (Christie and Vogel, 2000). The T4SSb is present in *Legionella pneumophila* and contains seven proteins that are also present in the T4SSa and two that are also present in the T6SS (DotU (IcmH)=TssL, and IcmF=TssM) (Segal *et al.*, 2005).

1.3.6. Type V secretion system

The type V secretion system (T5SS) is Sec-dependent and can be subdivided into two types based on variations in their mechanism.

1.3.6.1. Type V A secretion system

The type V A secretion system (T5SSa) is also called the autotransporter (AT) system, because the secreted proteins can autonomously cross the outer lipid bilayer. The substrates secreted by the T5SSa are mainly virulence proteins. Autotransporters are composed of 3 main domains: an N-terminal Sec-dependent signal sequence, an internal passenger domain which will be delivered to the exterior of the cell, and a C-terminal β-domain (Kostakioti *et al.*, 2005).

As in other Sec-dependent proteins, the N-terminal signal sequence targets the protein to the Sec system which translocate protein across the inner membrane into the periplasm. Once in the periplasm, the C-terminal region forms a β-barrel in the outer membrane. The β-barrel facilitates the translocation of the passenger domain through outer membrane which can then be cleaved off and released into the extracellular compartment or remain anchored to the cell surface (Henderson *et al.*, 2004, Desvaux *et al.*, 2004). The first described autotransporter was the IgAI protease secreted by *Neisseria gonorrhoeae* (Pohlner *et al.*, 1987).

30

1.3.6.2. Type V B secretion system

The Two-partner secretion (Tps) system, also known as the type V B secretion system (T5SSb), is Sec-dependent, and is composed of two separate polypeptides: TpsA which is the secreted substrate or effector, and TpsB which is the translocator component (Jacob-Dubuisson *et al*., 2001). The Tps system is related to the T5SSa. The difference is that the β-barrel domain of an autotransporter is a separate protein in the Tps system (TpsB). Therefore the Tps is considered as a unique secretion system (Jacob-Dubuisson *et al*., 2004). The TpsA contains an N-proximal domain called the secretion domain with a specific secretion signal. The latter directs the TpsA component towards TpsB which form a channel in the OM for the secretion of TpsA. TpsA physically interacts with TpsB on the periplasmic side of the OM. Examples of proteins secreted by the Tps include an iron acquisition protein, several adhesins and cytolysins such as the high-molecular weight adhesins 1 and 2 of the *Haemophilus influenzae* (Jacob-Dubuisson *et al*., 2001).

1.3.7. Type VII A secretion system

The chaperone-usher pathway secretion system, also known as the type VII secretion system (T7SS), is a Sec-dependent pathway used by bacteria mainly to translocate the proteins which form surface structures such as certain types of adhesive pili. This system is formed from gene cluster contains genes encoding an usher, a chaperone and a pilus subunit (Nuccio and Baumler, 2007). Substrate proteins such as pilin subunits are translocated to the periplasm via the Sec system where they form a complex with the chaperone protein. The function of the chaperone protein is in donor strand complementation that allows the correct folding of the subunits, and simultaneously prevents premature subunit-subunit interactions (Choudhury *et al*., 1999, Sauer *et al*., 1999). Then, the chaperone and usher proteins are involved in the secretion and assembly of pilus subunits into a pilus (Busch and Waksman, 2012).

1.3.8. Type VIII secretion system

The extracellular nucleation-precipitation (ENP) pathway involved in the secretion and assembly of curli, also known as thin aggregative fimbriae, has been proposed to be the type VIII secretion system (T8SS) (Desvaux *et al*., 2009). As a Sec-dependent system, the T8SS is one of the terminal branches of the GSP (Stathopoulos *et al*., 2000). Several pathogenic bacteria display curli on their surfaces that play a role in biofilm formation

and specific interaction with eukaryotic cells. In *E. coli*, curli are made up from two proteins, CsgA and CsgB. CsgG is a lipoprotein located in the OM and required for the secretion of CsgA and CsgB. The secretion mechanism of curli is poorly understood. There are two theories as to how curli components cross the outer membrane using this system, some proposing the CsgG serves as a chaperone stabilizing the two curli components; others suggesting that single CsgG form a channel through which curli components are exported (Loferer *et al*., 1997, Stathopoulos *et al*., 2000).

1.3.9. Type IX secretion system

A secretion system was recently discovered in the phylum *Bacteriodetes*. It was referred to as the PorSS (Porphorymonas secretion system) as it was originally discovered in *Porphorymonas gingivalis*. More recently it has been referred to as the type IX secretion system (T9SS) (McBride and Zhu, 2013, Sato *et al*., 2013). *P. gingivalis* causes periodontal disease and possesses many virulence factors such as lipopolysaccharides, hemagglutinins, fimbriae and proteinases. Gingipain is a secreted proteinase which can break down the periodontal tissue as it has high hydrolytic activities. PorSS is involved in the translocation of the gingipain across the IM and OM into the cell surface or the extracellular environment. A number of proteins make up the PorSS, which include PorK, PorL, PorM, PorN, PorO, PorP, PorQ, PorT, PorU, PorV (PG27, LptO), PorW and Sov. Gingipains are composed of three proteins that possess conserved C-terminal domain (CTD) with common motif, RgpA, RgpB and RgpP. Another 10 potential PorSS secreted effector proteins were identified, all possess the conserved CTD (McBride and Zhu, 2013, Sato *et al*., 2013). Some *P. gingivalis* PorSS proteins share similarity with *Flavobacterium johnsoniae* gliding motility proteins at the amino acid sequence level. Nakane and colleagues demonstrated that the gliding motility of *F. johnsoniae* is mediated through the dynamic movement of the SprB adhesin (Nakane *et al*., 2013). A significant reduction in the motility of *F. johnsoniae* was observed in *sprB* deficient cells suggesting that PorSS may be involved in the gliding motility of this bacterium (Braun *et al*., 2005).

1.4. Type VI secretion system

1.4.1. Discovery of the type VI secretion system

Approximately 10 years elapsed between the first identification of a T6SS component, haemolysin co-regulated protein (Hcp), in 1996 and the identification of the T6SS as protein secretion system in 2006. Williams and colleagues identified Hcp as a novel secreted protein in *Vibrio cholerae.* The absence an N-terminal cleavable signal peptide from this protein indicated a novel mechanism for its secretion (Williams *et al*., 1996).

Further evidence for a previously unidentified secretion system was found in 1997, when Roest and colleagues discovered a chromosomal locus in *Rhizobium leguminosarum* which they termed *imp* (impaired in nitrogen fixation). This locus was required for secretion of a protein and is now known to encode a T6SS. Mutations in this locus resulted in the mutant strain becoming able to nodulate and fix nitrogen on *Pisum sativum* and *Vicia hirsuta* (Roest *et al*., 1997). In the following year, a study by Wang *et al*. reported that the virulence of *P. aeruginosa* in neutropenic mice was induced by a gene locus encoding a serine-threonine protein kinase (STPK). They showed that null mutant strain of this locus exhibit reduced virulence compared to the wild type in neutropenic mice (Wang *et al*., 1998). Another major discovery of T6SS components was made in *V. cholerae* by Das *et al*. in 2003 when they identified a cluster of conserved genes named IcmF-associated homologous proteins (IAHP), because of the similarity between one of the genes in the cluster and the T4SSb *icmF* encoded protein. The only *IAHP* gene which was found to be present in T4SSb at that stage was *icmF*. Later, another gene, *icmH* (*dotU*) also present in T4SSb was identified in IAHP clusters (Das and Chaudhuri, 2003).

In 2002, Pallen and colleagues screened bacterial genomes for the presence of genes encoding forkhead-associated (FHA) domains and surprisingly they found that many species of pathogenic or symbiotic Gram-negative bacteria have homologous gene clusters which contained a gene encoding an FHA domain protein (Pallen *et al*., 2002). It is now known that FHA domain proteins regulate some T6SSs at the post-transcriptional level (Mougous *et al*., 2007, Hsu *et al*., 2009). Also, in 2002 it was convincingly demonstrated that a novel genomic island (*Salmonella* centisome 7 genomic island, *SCI*) in *Salmonella enterica* serovar Typhimurium, that was similar to

icmF-containing gene clusters in other bacteria, was important for the bacteria to invade eukaryotic cells successfully (Folkesson *et al.*, 2002).

Finally, in 2006, the name T6SS was given to the product of IAHP clusters when it was recognised as a novel secretion mechanism by the Mekalanos group, who demonstrated that this system was responsible for cytotoxicity of *V. cholerae* against *Dictyostelium discoideum* and mammalian J774 macrophages (Pukatzki *et al.*, 2006).

1.4.2. Standardized nomenclature for T6SS genes

Many of the T6SS core components were discovered in different Gram-negative bacterial species prior to the identification of the T6SS as a functional secretion system. Therefore organism-specific names of equivalent components exist, making nomenclature confusing. A scheme was proposed by Shalom and colleagues in 2007 who predicted the existence of 13 core components of the T6SS based on the conservation of these genes in T6SS gene clusters. These core subunits were referred to as TssA-TssM (type six secretion) (Table 1.2). In addition, genes that are present in some T6SS clusters, but not others, were referred to as *tag* genes (T6SS-associated-genes). The tag genes encode accessory proteins such as auxiliary components of the secretion system, effectors or regulatory proteins (transcriptional and post-transcriptional) (Shalom *et al.*, 2007, Aschtgen *et al.*, 2010a). This nomenclature will be used in this thesis, and is now widely used in the field alongside some organism-specific designations which still prevail.

Table 1.2 Standardization of nomenclature for type VI secretion systems.

Standardised gene names	S. typhimurium	V. cholerae	APEC	EAEC[a]	E. tarda	R. leguminosarum	P. aeruginosa[b]	B. ceno-cepacia
			Corresponding gene nomenclature in other bacteria					
tssA	*sciA*	*vasJ*	*aec29*	*aaiJ*	*evpK*	*impA*	*hsiA2*	*bscE*
tssB	*sciH*	*vipA*	*aec17*	*aaiA*	*evpA*	*impB*	*hsiB2*	*bscL*
tssC	*sciI*	*vipB*	*aec18*	*aaiB*	*evpB*	*impC/impD*[c]	*hsiC2*	*bscK*
tssD	*sciK/sciM*[c]	*hcp*	*aec16/aec32*[c]	NP	*evpC*	NA	*hcpC*	*bscJ*
tssE	*sciD*	NA	*aec19*	*aaiD*	*evpE*	*impF*	*hsiF2*	*bscI*
tssF	*sciC*	*vasA*	*aec20/aec21*[c]	*aaiE*	*evpF*	*impG*	*hsiG2*	*bscH*
tssG	*sciB*	*vasB*	*aec22*	*aaiF*	*evpG*	*impH*	*hsiH2*	*bscG*
tssH	*sciG*	*vasG*	*aec27*	*aaiP*	*evpH*	NA	*clpV2*	*bdcF*
tssI	*vrgS*	*vgrG*	*aec15*	*aaiG*	*evpI*	NA	*vgrG2*	NP[d]
tssJ	*sciN*	*vasD*	*aec24*	*aaiK*	*evpL*	NP	*lip2*	*bscN*
tssK	*sciO*	*vasE*	*aec25*	*aaiL*	*evpM*	*impJ*	*hsiJ2*	*bscO*
tssL	*sciP*	*vasF*	*aec26*	*aaiN*	*evpN*	*impK*	*dotU2*	*bscP*
tssM	*sciS*	*vasK*	*aec30*	*aaiO*	*evpO*	*impL*	*icmF2*	*bscB*

[a] One of the 3 T6SS present, sciII.
[b] One of the 3 T6SS present, HSI-II.
[c] Two orthologues are present in the same T6SS gene cluster.
[d] Gene is present at another locus outside the T6SS gene cluster.
NA, no alternative name given; NP, no corresponding gene present in the cluster; APEC, avian pathogenic *E. coli*; EAEC, enteroaggregative *E. coli*.

1.4.3. Genomics of the T6SS

T6SS is widely distributed in Gram-negative bacteria in both pathogenic and symbiotic bacteria (Figure 1.3) (Shalom *et al*., 2007, Cascales, 2008). Gene clusters encoding the T6SS have been identified in more than 25% of genome-sequenced bacteria (Bingle *et al*., 2008). It worth nothing that the   type VI secretion system clusters usually exist in pathogenicity islands in various bacterial species such as the EAEC *pheU* island, *S. typhimurium* SCI (*Salmonella* centisome island, now known as *Salmonella* Pathogenicity Island 6 (SPI-6)) and *Pseudomonas aeruginosa* HSI (Hcp-secretion island) (Folkesson *et al*., 2002, Mougous *et al*., 2006, Dudley *et al*., 2006) and in many cases are present on the accessory bacterial chromosome that encodes genes directly or indirectly implicated in virulence or survival inside the host cell (Cascales, 2008).

The number of T6SS-related genes in a T6SS gene cluster ranges from 12 to more than 20 genes, with different composition and organisation between species (Cascales, 2008, Shalom *et al*., 2007). The gene order and orientation of the T6SS genes vary among species. It is worth mentioning that in some cases *tssD* (hcp) and *tssI* (vgrG) are located at other genetic loci. Also Zheng and Leung carried out a comprehensive study to identify T6SS genes in *Edwardsiella tarda*, and they found that 13 out of 16 genes of the *evp* cluster were crucial for the proper export of Hcp-like protein (EvpC), VgrG (EvpI) and EvpP (Zheng and Leung, 2007). These genes are *evpA-evpC*, *evpE-evpI* and *evpK-evpO*, and they correspond to *tssA-tssM* (Table 1.2). Single or multiple copies of T6SS are present in many pathogenic Gram-negative bacteria such as *P. aeruginosa* (3 copies), *Yersinia pestis* (4 copies) and *B. pseudomallei* (6 copies) and many others (Mougous *et al*., 2006, Shalom *et al*., 2007, Suarez *et al*., 2008). Studies on *B. cenocepacia* have shown that it carries one copy of T6SS (Aubert *et al*., 2008, Holden *et al*., 2009).
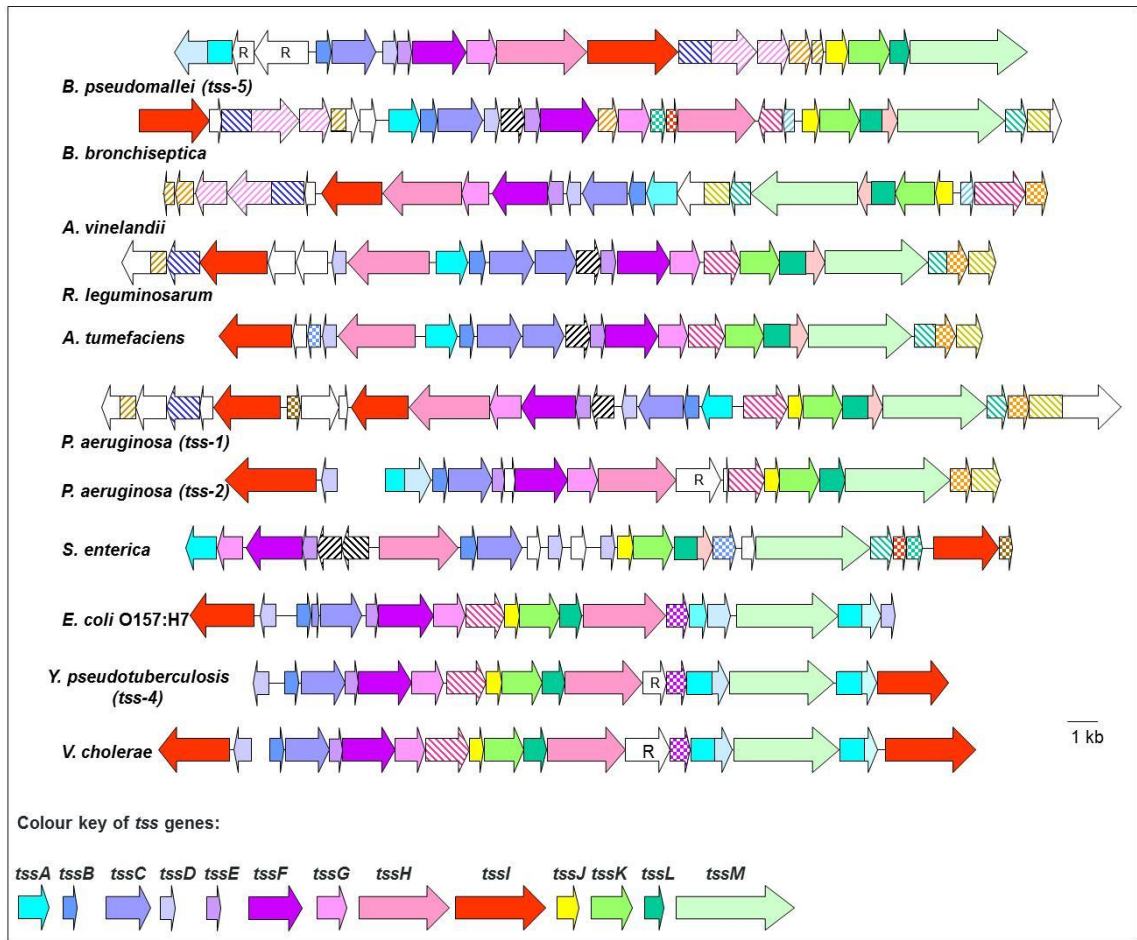
Figure 1.3. Orthologous T6SS gene clusters observed in Gram-negative bacteria.
Genes in each T6SS gene cluster are represented by block arrows. Direction of the arrows
represents gene orientation. Orthologous genes are presented as identical coloured arrows.
Uni-coloured arrows represent 'core' T6SS subunit genes (*tss*), hatched arrows represent
T6SS-associated genes (*tags*). *P.aeruginosa tss1* and *tss2* correspond to *P.aeruginosa*
T6SSs HSI-I and HSI-II (*P.aeruginosa* has 3 T6SS gene clusters). *B. pseudomallei tss5* is
the animal pathogenicity T6SS of *B. pseudomallei* (*B. pseudomallei* has six T6SS gene
clusters), *Y. pseudotuberculosis tss-4* is one of four T6SS gene clusters in this organism.
Other Gram-negative bacteria presented here have a single T6SS gene cluster. Used with
permission (Shalom *et al*., 2007).

1.4.4. The diverse functions of the T6SS

The T6SS has been demonstrated to play versatile functions in Gram-negative bacteria. Some of these functions are described below.

1.4.4.1. Role of T6SS in virulence

The T6SS has been demonstrated to play a role in the virulence of various Gram-negative bacteria. Several studies on Gram-negative pathogens have shown that their virulence is attenuated in mutant strains lacking a functional T6SS. For example, *V. cholerae* is significantly less virulent in some host when the T6SS is absent. Thus, the cytotoxicity of *V. cholerae* cells toward *Dictyostelium* amoebae and mammalian J774 macrophages, which is contact-dependent, was significantly reduced in *vas* gene mutants. Further studies on *V. cholerae* showed that the T6SS was required for actin crosslinking and cytotoxic effects on J774 macrophage cells and *Dictyostelium* (Pukatzki *et al*., 2007). *Aeromonas hydrophila* mutants with a defective T6SS are less cytotoxic toward human cells and less resistant to attack by human macrophages. They are also less virulent in a mouse infection model (Suarez *et al*., 2008). *B. pseudomallei* possess six copies of T6SSs and T6SS-1 was shown to play role in the infection of the macrophages (Shalom *et al*., 2007). T6SS-1 gene cluster homologous was identified in *B. mallei* which is closely related species to *B. pseudomallei* and has been demonstrated to be essential for virulence in a hamster model as ΔT6SS mutants were avirulent (Schell *et al*., 2007). In addition, T6SS-1 in *B. pseudomallei* has been shown to play important role in the virulence, as Δ*hcp* mutants were less pathogenic and have reduced intracellular growth and cytotoxicity in a hamster model of acute melioidosis (Burtnick *et al*., 2011).

Similarly, *S. enterica* serovar Typhi is less cytotoxic but more invasive toward human epithelial cells in Δ*sciG* (*tssH*) mutant compared with the wild type (Wang *et al*., 2011). A study on *S. enterica* serovar Typhimurium by Mulder *et al* showed that disruption of four clusters of non-core genes of T6SS result in impaired intracellular replication and systemic spread in mice. They also found that Δ*sciG* mutant display significant reduction in the ability to replicate in macrophages (Mulder *et al*., 2012). A recent study on the same bacteria aimed to identify the role of T6SS in the infection of the chicken showed that ΔT6SS mutant had impaired ability to colonise the gut and the internal organ of chicken (Pezoa *et al*., 2013).

1.4.4.2. Role of T6SS in antipathogenesis

In addition to the role of the T6SS in pathogenesis, in some cases it plays a role in attenuating the virulence of the bacteria which harbour the T6SS, i.e. antipathogensis. This function enables the host cells and the pathogen to survive together by pushing the relationship away from pathogenesis and towards a commensal interaction, i.e. chronic infection. Several studies were conducted to study the role of T6SS in chronic infections. Mutation in the HSI-1 cluster (one of the T6SS gene clusters in *P. aeruginosa*) made the mutant more virulent in the rat model of chronic pulmonary infection (Potvin *et al*., 2003). Also in 2003, Bladergroen group showed that the mutualistic interaction of *R. leguminosarum* and leguminous plants is highly specific and deletion mutation in *tss* (*imp*) genes enables the bacterium to infect pea plants. The results of the study indicate that the T6SS was involved in defining the host specificity of the *Rhizobium*-plant interaction (Bladergroen *et al*., 2003).

Similarly, the T6SS in *S. enterica* serovar Typhimurium allows it to survive longer in host cells by limiting intracellular bacterial replication and thus attenuates the virulence and favours chronic infection (Parsons and Heffron, 2005). Therefore, the T6SS in *S. enterica* is similar to that in *P. aeruginosa* (HSI-1 cluster) as it can be induced in the chronic phase of infection to help the organism to survive within the host. A recent study on *Helicobacter hepaticus* also suggests that the T6SS may have a role in limiting the intracellular multiplication and virulence (Chow and Mazmanian, 2010).

1.4.4.3. Role of T6SS in interbacterial competition

The T6SS has also been shown to have an important role in competitive interbacterial interactions by contact-dependent injection of T6SS effectors. Mougous and colleagues have identified and characterised three putative T6SS effector proteins Tse1-Tse3 in *P. aeruginosa* (Section 1.6.1). Tse2 protein has lethal effects on eukaryotic cells and bacteria when synthesised intracellularly (Hood *et al*., 2010). Although Tse2 protein was cytotoxic towards eukaryotic and prokaryotic cells when it was ectopically produced inside these cells, further studies demonstrated that eukaryotic cells are not the true target of the Tse2 effector.

A recent study on *Acinetobacter baumannii* demonstrated that the T6SS is involved in inter-bacterial competition in a cell contact-dependent fashion. In this study,

competition assays were carried out to assess the potential of *A. baumannii* strain M2 and derived *tssB⁻* and *tssC⁻* mutants to outcompete *E. coli*. The result showed that there was a significant drop in the number of viable *E. coli* when they were incubated with *A. baumannii*, and this was dependent on the presence of functional TssB and TssC subunits (Carruthers *et al*., 2013). A similar result was obtained in competition assays between *Serratia marcescens* Db10 and *E. coli* MC4100 as there was a significant reduction in the number of viable *E. coli* cells when it was cocultured with *S. marcescens* Db10, whereas no significant reduction in the number of *E. coli* cells was observed when MC4100 was cocultured with a T6SS mutant (Murdoch *et al*., 2011).

Further evidence supporting the role of the T6SS in interbacterial interactions was obtained from studies on *V. cholerae* strain V52. This investigation demonstrated that *V. cholerae* strain V52 containing a deletion in the accessory toxins *hlyA*, *rtxA*, and *hapA* was shown to be highly virulent against many Gram-negative bacteria such as *E. coli*, *S. enterica* and *Citrobacter rodentium*, whereas a mutant strains containing a deletion in *vasK* (*tssM* gene of the T6SS) lost its virulence against these Gram-negative bacteria (MacIntyre *et al*., 2010).

1.5. Structural subunits of the T6SS

Although the existence of the T6SS was first reported only in 2006, a model for the assembly and mechanism of the T6SS is starting to take shape. As discussed earlier, the T6SS contains 13 core components which are predicted to assemble into the T6SS apparatus. Currently, they are classified into three groups. Several studies have revealed the remarkable similarity between some T6SS proteins and bacteriophage contractile tail components, and these proteins are classified as bacteriophage-like components (Figure 1.4). A second group of T6SS proteins are associated with the inner or outer membrane, while the third group include Tss proteins with no clear role. Based on our current knowledge of the T6SS, it has been suggested that T6SS core components are assembled into a macro-complex structure which resembles an inverted bacteriophage tail-like structure embedded in the bacterial cell envelope membrane (Cascales and Cambillau, 2012). Figures 1.4 and 1.5 present the current model of T6SS assembly and mechanism of action.
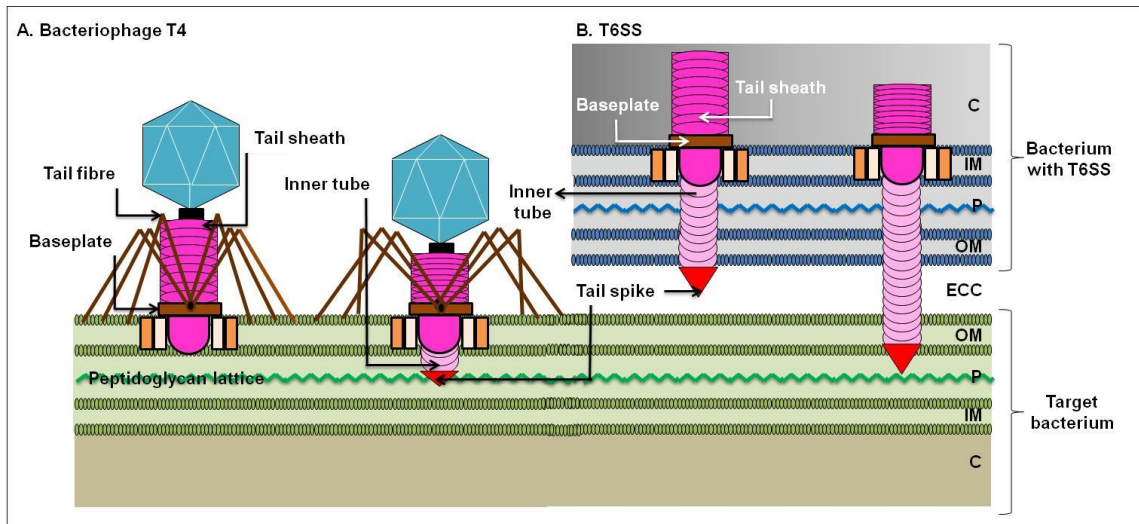
Figure 1.4. Similar injection mechanisms of bacteriophage T4 and the T6SS. Structurally similar proteins in bacteriophage T4 and the T6SS are shown in the same colours. A. The bacteriophage T4 is attached to the surface of the target bacterium. B. A Gram-negative bacterium with T6SS firing at a target bacterium. Abbreviation: C, cytoplasm; IM, inner membrane; P, periplasmic space; OM, outer membrane; ECC, extracellular compartment. Adapted from Cotter (2011).
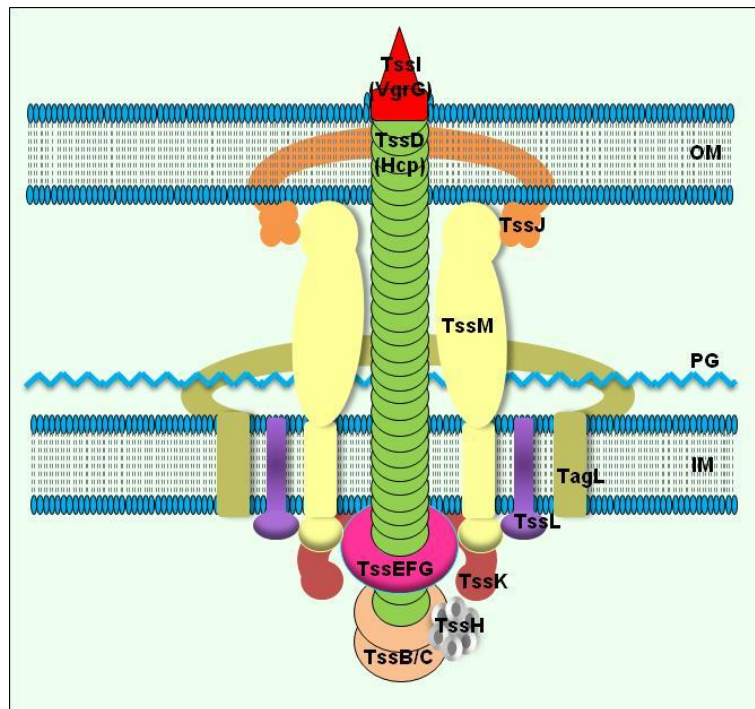
Figure 1.5. Diagrammatic representation of the T6SS assembly. The diagram shows the EAEC Sci-I T6SS which includes TagL with a peptidoglycan-binding domain (PGB) and TssL lacking one (see Section 1.5.3). The location of TssA is unknown (it's predicted to be in the cytoplasm). Adapted from Cascales and Cambillau (2012).

1.5.1. Bacteriophage-like components (the injection apparatus)

1.5.1.1. The T6SS needle tail protein, TssD

TssD (Hcp) is a secreted protein first identified in *V. cholerae* (Williams *et al*., 1996). TssD was detected in the pulmonary secretions of cystic fibrosis patients and TssD-specific antibodies were present in their sera (Mougous *et al*., 2006). In fact, TssD is present in the culture supernatant of all bacteria with a functioning T6SS. Although it is a secreted protein, the absence of TssD results in inactivation of the T6SS (Zheng and Leung, 2007). This led to the conclusion that TssD is not only a secreted protein but it also acts as a T6SS structural component. Another support for this assumption is that secretion of another T6SS component, TssI (VgrG), requires a functional TssD. It appears that TssD and TssI show a mutual dependence for their secretion (Pukatzki *et al*., 2006).

The TssD crystal structure of *P. aeruginosa* TssD1 revealed that TssD monomers assembled into hexameric rings with an outer diameter of 80-90 Å and an inner channel diameter of 35-40 Å that can form nanotubes *in vitro* by stacking on the top of each other (Figure 1.6 and 1.7). The TssD nanotube is large enough to allow a small folded protein or an unfolded or partially folded protein to pass through it (Mougous *et al*., 2006, Ballister *et al*., 2008). As TssD has been found in the culture supernatant and periplasm of bacterial cells containing an active T6SS, this suggests that the TssD tube assembles at the inner membrane and then extends through the outer membrane. However, assembly of the TssD tube has not yet been directly proven *in vivo*.

The TssD monomer shares a common evolutionary origin with the *E. coli* bacteriophage T4 tail tube protein gp19 and has structural similarities to λ phage protein gpV, which constitutes the bacteriophage tail tube (Leiman *et al*., 2009, Pell *et al*., 2009). Therefore, TssD is predicted to form a phage-like tail tube component of the T6SS. The crystal structure of the TssD monomer from *P. aeruginosa* (HSI-1 cluster) consists of an identical fold of two antiparallel β-sheets with a β-hairpin extension (Figure 1.7) (Mougous *et al*., 2006).
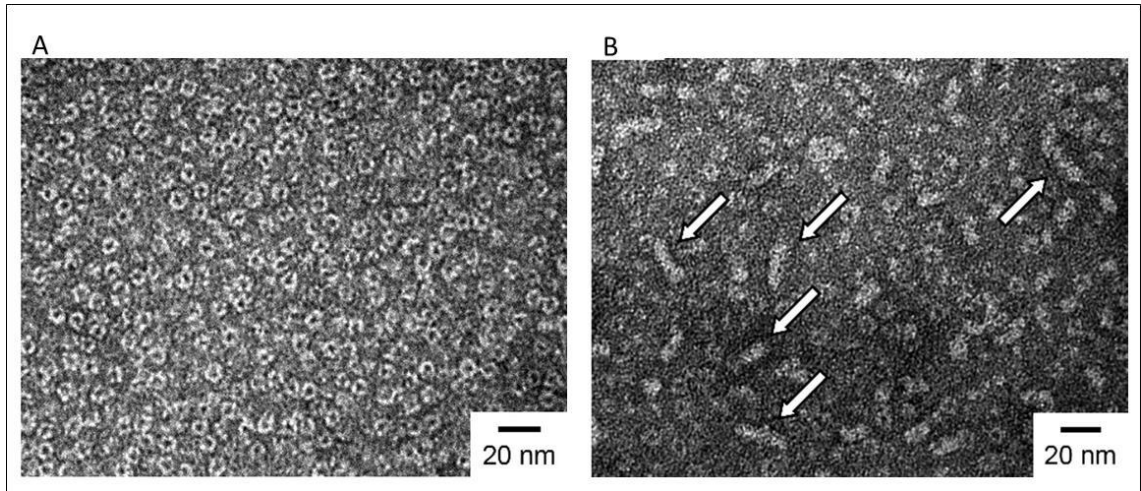
Figure 1.6. Negative stained EM images of TssD proteins from *P. aeruginosa* PAO1. A. EM micrograph of TssD showing TssD rings. B. Refolded purified TssD protein after denaturation in 8 M urea. Arrows indicate polymeric structures of TssD. Adapted from Leiman *et al* (2009).
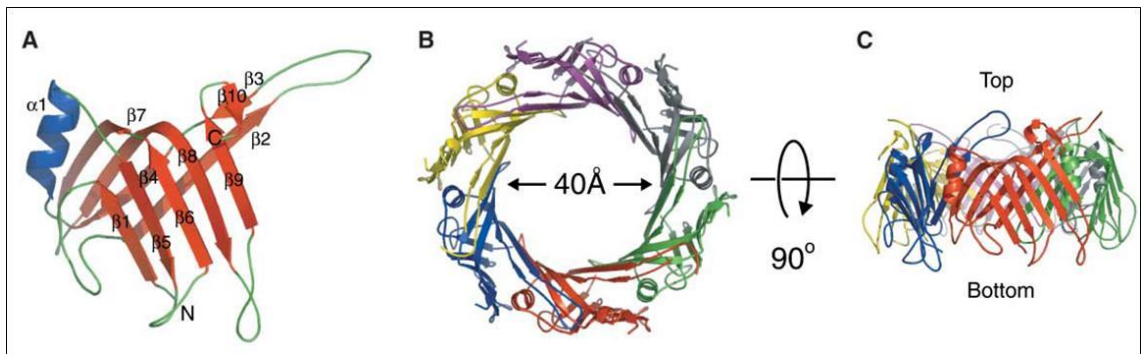
Figure 1.7. The crystal structure of TssD protein. A. Ribbon illustration of the secondary structure of a TssD monomer. B. Top view of the TssD hexameric structure showing the lumen with an internal diameter of 40 Å. Each of the six subunits is coloured differently. C. Side view of the TssD hexameric structure (chain B in the Protein Data Bank, code 1Y12). Adapted from Mougous *et al* (2006).

There is some evidence that TssD may also function as an effector protein. It was shown that ectopic expression of *A. hydrophila* TssD within eukaryotic cells leads to induction of apoptosis (Satchell, 2009). Evidence that TssD can function as a chaperone and effector recognition was revealed in very recent study in *P. aeruginosa*. In this study, TssD was shown to specifically bind to T6SS-effector molecules (the three Tse proteins, but mainly Tse2). The interaction of TssD with Tse2 was demonstrated to be required for the stabilisation and export of the Tse2 (Silverman *et al.*, 2013).

1.5.1.2. The T6SS tail spike protein, TssI

TssI, also known as VgrG (valine-glycine repeats), constitutes a family of proteins containing conserved N-terminal and central regions but showing high sequence divergence at their C-termini. It has been found that the two conserved domains are similar to the bacteriophage T4 gp27 protein and the C-terminal domain of bacteriophage T4 gp5, respectively, which form the bacteriophage T4 tail spike (Leiman *et al.*, 2009). The crystal structure of the conserved N-terminal and central regions of a TssI protein from uropathogenic *E. coli* strain CFT073 that includes the regions corresponding to gp27 and gp5 revealed that TssI consists of four domains, two β-sandwiches, an αβ domain and a domain corresponding to the oligosaccharide/oligonucleotide-binding (OB)-fold domain of gp5 (Figure 1.8). In bacteriophage T4, gp5 is composed of three domains connected together by long linkers: the oligosaccharide/oligonucleotide-binding (OB)-fold domain, the middle domain termed the lysozyme domain, and the C-terminal domain, whereas gp27 consists of two similar domains named tube domains (Leiman *et al.*, 2009). However, the lysozyme domain is not present in the gp5-like region of TssI.

The T4 bacteriophage tail spike consists of trimers of gp27 and gp5, i.e. $(gp27.gp5)_3$, which functions as a puncturing tool towards the target cell that facilitates insertion of the bacteriophage tail tube and subsequent injection of the viral DNA into the host cell (Kanamaru *et al.*, 2002). As the structures of the $(gp27.gp5)_3$ complex and the TssI trimer are analogous, the current model of the T6SS suggests that the TssI trimer is located on the top of the TssD tail tube and functions as a puncturing device, piercing the membrane of the target cells (Figure 1.4). The presence of two β-sandwich domains per TssI monomer gives a six fold symmetry to the TssI trimer which presumably facilitates interaction with the hexameric TssD ring. The cylinder which is formed by

the gp27 trimer complex has an internal diameter of 30 Å, whereas the gp5 trimer complex forms a prism of β-helices with an inner diameter 33 Å at the base and 25 Å at the tip resulting in a tapering structure like a needle (Pukatzki *et al*., 2007). However, the internal diameter of the trimeric TssI channel is too small to allow proteins through it (Hachani *et al*., 2011). Based on this observation, one can speculate that the TssI must be detached from the TssD tubular structure to allow proteins to be secreted through the tail tube (Silverman *et al*., 2012). The location of the TssI trimer suggests that outgrowth of the TssD tail tube pushes out the TssI trimer tip (Cascales and Cambillau, 2012). Similar to TssD, TssI has been found in the culture supernatant of bacteria with a functioning T6SS. Therefore, the presence of TssD and TssI in culture supernatant is the hallmark of a functioning T6SS. As discussed above, TssD and TssI show mutual dependence, since TssD is not present in the culture supernatant of *tssI* cells and *vice versa*. This confirms the role of TssI as an essential component of the T6SS.

Some TssI (VgrG) proteins carry C-terminal extensions and these are referred to as evolved VgrGs (Figure 1.9). In several cases that have been studied the C-terminal extension of the evolved VgrG has been shown to function as an effector domain that is injected into the target cell cytosol (Section 1.6.1) (Pukatzki *et al*., 2007). Moreover, many T6SS-containing bacteria contain multiple *tssI* genes, even where only one T6SS gene cluster is present. Several of these may encode evolved VgrGs containing different effector moieties.

The absence from TssI of the lysozyme domain which is present in phage T4 gp5 protein and is involved in the destruction and lysis of the bacterial cell wall is puzzling. However, it was observed that the gp5 protein of vibriophage KVP40, which is another type of T4 phage, also lacks the lysozyme domain, and it is predicted to encode this lysozyme domain somewhere else in its genome (Rossmann *et al*., 2004). It is possible that the T6SS utilises a similar mechanism although the nature of such a lysozyme is unknown. However, it was recently shown that *P. aeruginosa* non-evolved VgrG1 can deliver peptidoglycan hydrolase effectors (Tse1 and Tse3) into the periplasm of target Gram-negative bacterial cells resulting in destruction and lysis of the bacterial cell wall (Section 1.6.1) (Hood *et al*., 2010, Russell *et al*., 2011). Therefore, in such situation, the requirement for a lysozyme component in the T6SS tail-spike no longer exists. It is also
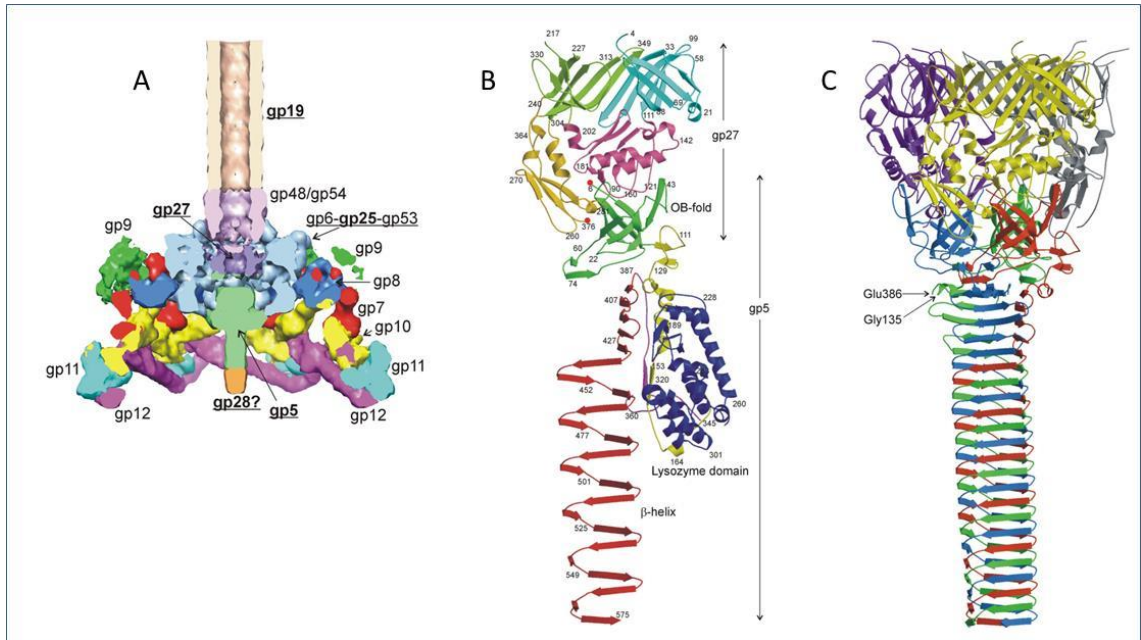
Figure 1.8. The crystal structure of TssI encoded by the *E. coli* CFT073 c3393, and homology to the bacteriophage T4 gp27 and gp5. A. the bacteriophage T4 baseplate, showing the location of the tail spike proteins gp27 and gp5. The gene products homologous to T6SS proteins are underlined. B. Crystal structure of bacteriophage gp27-gp5 monomer. Domains are represented in different colours. C. Structure of the TssI trimer, containing the gp27-like domain and the needle like membrane puncturing domain which is homologous to the gp5 domain from bacteriophage T4. The β-helical region is a prediction. PDB ID 2P52. Adapted from Leiman *et al* (2009).
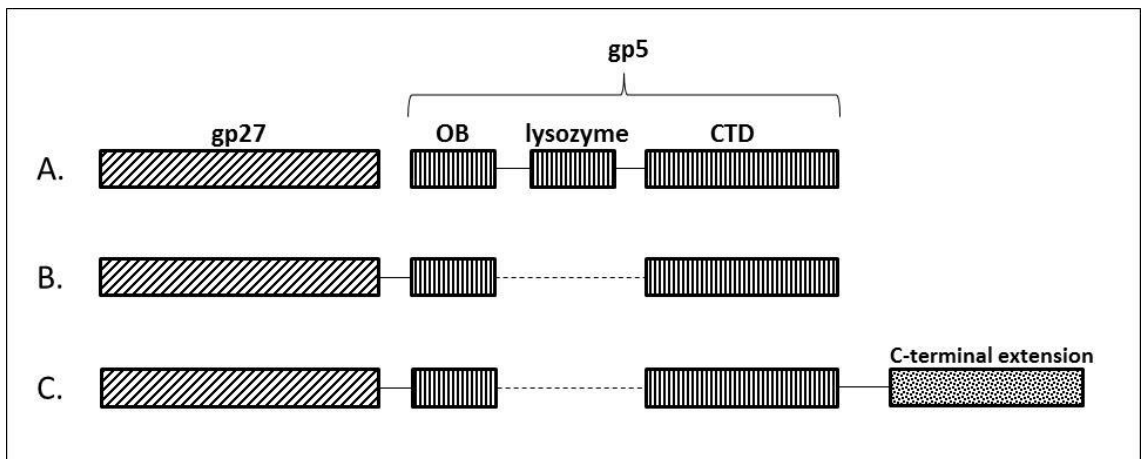
Figure 1.9. Schematic representation of the gp27 and gp5 subunits of the bacteriophage T4 tail spike compared to core and evolved VgrG subunits of the T6SS. The homologous domains are shaded in the same pattern. A. gp27 and gp5 subunits of the bacteriophage T4 showing the individual domains of gp5. B. Core VgrG of T6SS. C. Evolved VgrG with additional domain at C-terminus an "extended extension" (effector domain). OB, oligonucleotide/oligosaccharide-binding fold; CTD, C-terminal domain.

noteworthy that some T6SS gene clusters, such as the one in *B. cenocepacia*, encode a predicted endolysin, TagV (M. Thomas personal communication).

Although the current model of T6SS suggests the formation of a needle apparatus via assembly of the TssD tube with a tip composed of a TssI trimer on top, and more recently it was proposed that a small PAAR (proline-alanine-alanine-arginine) repeat protein sits on the top of the tip of the TssI trimer to "sharpen" it (Section 1.5.5), such a hypothesis has not been experimentally tested (Shneider *et al*., 2013, Filloux, 2013).

1.5.1.3. The T6SS needle sheath proteins, TssB and TssC

The *tssB* and *tssC* genes are always found adjacent to each other and in the same orientation in T6SS gene clusters indicating that they might be functionally related. Accordingly, it has been demonstrated that TssB and TssC interact and form microtubules which were shown to disassembled by TssH into smaller sub-complexes (Bonemann *et al*., 2009, Hachani *et al*., 2011, Lossi *et al*., 2013). A two-hybrid study on the type VI secretion-like system of *Francisella tularensis* demonstrated that IglA (TssB) and IglB (TssC) interact and form a complex. Strikingly, they also found, that the same interaction is conserved between IglA and IglB proteins and corresponding T6SS proteins from other pathogenic bacteria (Broms *et al*., 2009). The interaction of TssB and TssC was subsequently demonstrated in several organisms such as *V. cholerae*, *B. cenocepacia* and *P. aeruginosa* (Bonemann *et al*., 2009, Aubert *et al*., 2010, Lossi *et al*., 2013).

Based on the structure of contractile phage tails, the current model for assembly of the T6SS suggests that TssB and TssC form a contractile sheath which surrounds the TssD tail tube (Figure 1.4 and 1.5). The cylindrical complexes formed by these proteins were estimated to have an inner diameter of 100 Å and an external diameter of 300 Å by electron microscopy (Figure 1.10), which is large enough to accommodate the TssD conduits (Mougous *et al*., 2006, Ballister *et al*., 2008, Bonemann *et al*., 2009). The TssB/TssC complex was shown to exist in two different structural conformations which are predicted to be the extended and contracted forms of the sheath (Basler *et al*., 2012).

In bacteriophage T4, the tail sheath protein gp18 also forms a cylindrical structure with an inner diameter of 110 Å and an external diameter of 250 Å which can accommodate
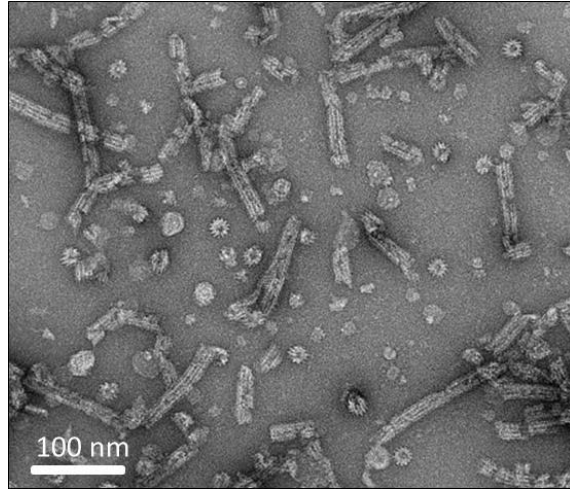
Figure 1.10. Negative stained EM images of TssB/TssC tubular complexes. The micrograph shows the tubular and ring form of TssB/TssC complex. Adapated from Lossi *et al*. (2013).

the gp19 tail tube. The contraction of the T4 bacteriophage tail sheath, which surrounds the viral tail tube, provides the energy required for injection of the tail needle into the bacterial cell envelope (Aksyuk *et al*., 2009). Despite the fact that there no amino acid sequence similarities between the TssB/TssC complex and gp18, they appear to have a similar structural organisation. Therefore, the TssB/TssC tubular complex of the T6SS is predicted to function in a similar way to the T4 bacteriophage tail sheath. In support of this hypothesis, a recent fractionation study on *F. novicida* showed that TssD, TssB and TssC can be found in multiple fractions which suggest that these proteins are part of a large complex (de Bruin *et al*., 2011). Also, it was demonstrated that TssB and TssC proteins are localised in the cytoplasm and the outer membrane (Bonemann *et al*., 2009, Aubert *et al*., 2010). Based on these observations, it is predicted that the shrinkage of TssB/TssC tubules might provide the energy for the needle tail and spike apparatus (i.e. TssD and TssI) to be injected into target cells (Leiman *et al*., 2009, Bonemann *et al*., 2010). A recent study on *V. cholerae* demonstrate that TssB/TssC tubules are formed *in vivo* and they undergo continues cycle of elongation, contraction and disassembly. These tubules are positioned near-perpendicular to the IM. Interestingly, TssH had been shown to recycle the TssB/TssC tubules after effective contraction and also to prevent the spontaneous assembly of non-productive contracted tubules (Kapitein *et al*., 2013).

1.5.1.4. The needle hub protein, TssE

TssE is the only one of the 13 core components of the T6SS which was shown to have a significant (~ 40%) amino acid sequence similarity to a phage protein, the T4 bacteriophage baseplate protein gp25 (Shalom *et al*., 2007, Leiman *et al*., 2009, Lossi *et al*., 2011). In the bacteriophage T4, gp25 is a constituent of the baseplate (Kostyuchenko *et al*., 2003, Yap *et al*., 2010). The interactions of gp25 with other components of the bacteriophage baseplate depend on the state of the injection apparatus. In the resting state where the baseplate is comprised of six wedge-shaped protein complexes that form a hexagon, gp25 interacts with the (gp27-gp5)$_3$ complex and in the active state of the injection apparatus where the wedges reorientate to form a six-pointed star-like structure, it interacts with the tail tube (Kostyuchenko *et al*., 2003, Kostyuchenko *et al*., 2005, Leiman *et al*., 2010). Taking gp25 protein of bacteriophage T4 as a plausible model for TssE, TssE is predicted to contribute to the assembly of the T6SS baseplate by interacting with TssI in the resting state and TssD in the active state (Leiman *et al*., 2009, Lossi *et al*., 2011). Figure 1.11 presents a model of the TssE

structure based on similarity with the gp25-like phage protein GSU0986 of *Geobacter sulfurreducens*.

1.5.2. The sheath recycling protein, TssH

TssH (ClpV) is a member of a group of proteins termed AAA$^+$, which all have an AAA domain responsible for ATP hydrolysis and binding. AAA$^+$ proteins (ATPase associated with various cellular activities) play different roles in various cellular processes such as protein refolding and degradation, and DNA replication and repair (Ogura and Wilkinson, 2001, Sauer *et al*., 2004). TssH is a ClpB homologue, which is conserved in all T6SSs and plays an essential role in the secretion of TssD and TssI (Bonemann *et al*., 2009). ClpB forms hexameric rings and binds ATP via its conserved AAA$^+$ domain and uses the energy of ATP hydrolysis to drive unfolding or disassembly of proteins (Lee *et al*., 2003). The domain structure of ClpV is highly similar to that of ClpB protein (Figure 1.12), despite the fact that they only share 30-35% amino acid sequence homology. ClpV proteins have four distinct domains, an N-terminal domain, two AAA$^+$ domains, each containing Walker A and Walker B boxes involved in nucleotide binding and hydrolysis, and a middle domain inserted into the first AAA$^+$ domain. The two AAA$^+$ domains are highly conserved among ClpV and ClpB proteins, whereas the N-terminal and the middle domains are variable in length and sequence (Schlieker *et al*., 2005).

TssH was shown to interact with TssB/TssC tubules via its N-terminal domain which binds specifically to an N-terminal alpha helix of TssC. The C-terminal AAA$^+$ domain of TssH provides the required energy for the dissociation of the contracted TssB/TssC tubules. Furthermore, the N-terminal domain is important for TssB/TssC binding, and the TssC will only bind to the TssH N-terminal domain when in complex form with TssB (Bonemann *et al*., 2009). In the TssH of *V. cholerae*, a cleft in the N-terminal domain is formed by an α-helix, and is where TssC binds (Pietrosiuk *et al*., 2011). TssH prevents the accumulation of TssB/TssC tubule in the cytoplasm which might interfere with the translocation of TssB or TssC monomers into the periplasm to form the predicted tail sheath (Bonemann *et al*., 2009, Pietrosiuk *et al*., 2011). As disussed before, TssH had been shown to have a dual role as it recycle the TssB/TssC tubules and prevent the spontaneous assembly of non-productive contracted tubules (Kapitein *et al*., 2013).
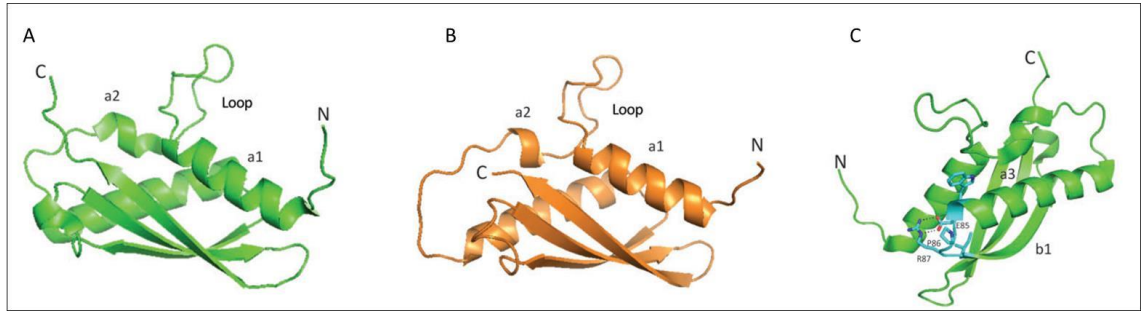
Figure 1.11. Predicted model of TssE of *P. aeruginosa*. Bioinformatics tools were employed to predict the structure of TssE (gp25-like protein) based on the crystal structure of the gp25-like phage protein encoded by *GSU0986* of *G. sulfurreducens*. A. Ribbon representation of GSU0986, demonstrates the loop conformation between the α1 and α2 helices. B. Ribbon model representation of TssE based on the homology with GSU0986. C. Ribbon representation of the GSU0986 illustrating some identical residues between it and TssE. Adapted from Lossi *et al* (2011).
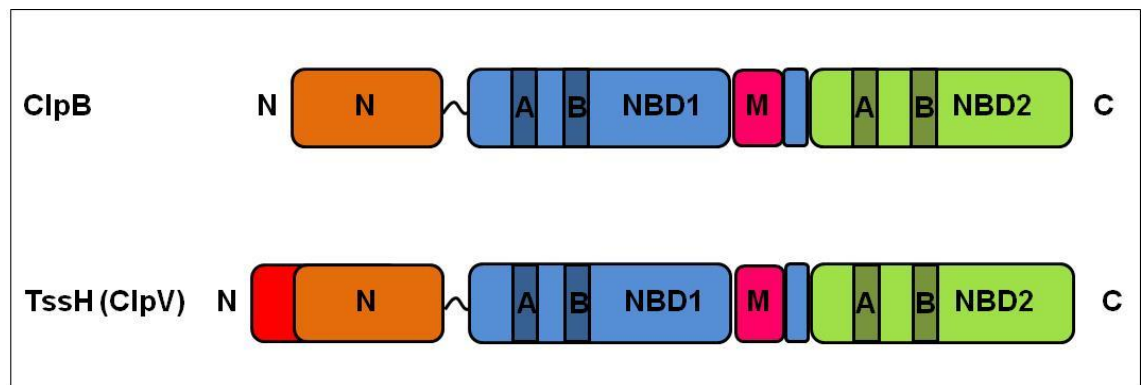
Figure 1.12. Structure similarities between ClpB and TssH (ClpV). The N-domain is shown in brown. In TssH the N-domain has an N-terminal extension shown in red. The two highly similar AAA$^+$ domains are called NBD1 and NBD2, and are shown in blue and green, respectively. A and B denote Walker A and Walker B boxes. The middle domain (M) is shown in deep pink.

1.5.3. The transmembrane-spanning complex (TssJ, TssL and TssM)

A recent study on one of the three T6SSs in EAEC, Sci-1, identified a complex consisting of four membrane-associated proteins: TssL, TssM, TssJ and TagL (Aschtgen *et al*., 2010a). Localisation studies have shown that TssM and TssL are inserted into the IM (Ma *et al*., 2009b), whereas TssJ is an OM lipoprotein which is anchored in the OM (Aschtgen *et al*., 2008). TagL is also embedded in the IM (Aschtgen *et al*., 2010a). TssM was shown to interact with TssL and TssJ. Therefore, TssM links the IM and the OM (Zheng and Leung, 2007, Ma *et al*., 2009b, Felisberto-Rodrigues *et al*., 2011). Sequence analysis of T4SSb and T6SS gene clusters reveal that two of the membrane components are shared between these two systems: TssL (DotU/IcmH) and TssM (IcmF) (Das and Chaudhuri, 2003, Nagai and Kubori, 2011). In T4SSb these two genes encode putative inner membrane proteins which interact with each other and form a complex that stabilise the other core components of the apparatus (Sexton *et al*., 2004, VanRheenen *et al*., 2004).

TssM is embedded in the IM by three TMDs. The N-terminus of TssM is located in the cytoplasm and contains a few residues, followed by two TMDs connected by a short periplasmic loop. A small cytoplasmic domain ($\sim$30 kDa) is located between the second and third TMDs. This cytoplasmic domain carries conserved Walker A and B motifs, suggesting its function in ATP-binding and hydrolysis (Ma *et al*., 2009b). Mutagenesis studies have revealed conflicting data about the importance of the ATPase function of TssM in different organisms. The inactivation of the Walker A motif in *E. tarda* had no effect on the secretion of TssD and TssI proteins (Zheng and Leung, 2007), whereas in *A. tumefaciens* the inactivation of Walker A motif completely abolishes the secretion of Hcp (Ma *et al*., 2009b). However, some TssM orthologues lack the walker box motifs. In addition, many TssM are predicted to have only one TMD (Cascales and Cambillau, 2012). A large C-terminal periplasmic domain ($\sim$80 kDa) follows the last TMD, and can be divided into two subdomains: an N-terminal α-helical domain and a C-terminal β-strand domain (Felisberto-Rodrigues *et al*., 2011). TssM interacts with TssJ through its periplasmic domain (Zheng and Leung, 2007, Felisberto-Rodrigues *et al*., 2011).

TssL is anchored in the IM by a single C-terminal TMD. The main part of TssL is located in the cytoplasm and adopts a particular shape similar to a hook (Aschtgen *et al*., 2012, Durand *et al*., 2012). This has been suggested to recruit the T6SS secreted

proteins to the secretion apparatus (Cascales and Cambillau, 2012). The crystal structure of the TssL cytoplasmic domain of EAEC demonstrated that this region is constituted of two bundles of three helices which are connected by two short helices. Also, this domain was shown to form a dimer (Durand *et al.*, 2012). In many T6SS clusters, the encoded TssL subunit contains an additional domain at the C-terminus which is located in the periplasmic space. This C-terminal periplasmic domain contains a periplasmic OmpA-like domain which carries a PGB motif. When the peptidoglycan binding (PGB) motif is absent from TssL, it is usually carried by a T6SS-associated protein (Tag), i.e. TagL in the case of the EAEC Sci-1 T6SS, which interacts directly with TssL (Section 1.5.5) (Aschtgen *et al.*, 2010a, Aschtgen *et al.*, 2010b).

TssJ is inserted into the OM through the N-terminal acylated cysteine (Aschtgen *et al.*, 2008). The crystal structure of TssJ showed that it contains a β-sandwich fold with four-stranded β-sheets (Felisberto-Rodrigues *et al.*, 2011). A recent study on *Serratia marcescens* showed that TssJ lipoprotein oligomerises (Rao *et al.*, 2011). Therefore, it is possible that the TssL-TssM-TssJ sub-complex assembles as a ring-like structure to form a channel spanning the IM and the OM of the bacterial cell envelope, similar to T3SS and T4SS membrane-spanning complexes (Fronzes *et al.*, 2009, Worrall *et al.*, 2011).

1.5.4. T6SS cytoplasmic and predicted cytoplasmic components (TssA, TssF, TssG and TssK)

1.5.4.1. TssK

TssK is a trimeric cytoplasmic protein that has been shown to play an important role in the assembly of the T6SS (Figure 1.13). The two subassembly complexes (phage tail-like and trans-envelope) of the T6SS seem to be connected by TssK as it interacts with a number of T6SS components from the two groups. TssK in EAEC was shown to interact with TssA, TssC, TssD and TssL. Also, TssK was found to play a role in the sheath assembly (Zoued *et al.*, 2013). The *tssK* gene was found to be organised in tandem with *tssL* gene in more than 70% of T6SS gene clusters which suggests an interaction between the two proteins or a functional relationship (Boyer *et al.*, 2009). TssK protein has also been found to be present as a component of the inner membrane proteome of *P. aeruginosa* which provides further evidence for the association of TssK with the membrane complex (Casabona *et al.*, 2013). Also, TssK in *V. cholerae* was

shown to be associated with the T6SS sheath as it co-purified with the isolated sheath adding further evidence for the prediction of TssK function as a connector between the membrane and phage-like subassemblies (Basler *et al*., 2012).

1.5.4.2. TssA, TssF and TssG

There is little published literature on TssA, TssF and TssG. Primary sequence analysis or secondary structure predictions of TssA, TssF and TssG showed no obvious similarities with bacteriophage T4 subunits. Nevertheless, Cascales and Cambillau predicted that these subunits are soluble cytoplasmic proteins which might have structural similarities with the bacteriophage T4 proteins (Cascales and Cambillau, 2012).

TssA was discovered in *S. enterica* serovar Typhimurium in 2002 and referred to as 'SciA' (Folkesson *et al*., 2002). Alignment of *B. cenocepacia* TssA with TssA orthologues present in other Gram-negative bacteria showed that the N-terminal region of TssA proteins is conserved whereas the C-terminal region varies between different organisms. TssA proteins can be divided into two classes, based on divergence within their C-terminal regions (M.S. Thomas, personal communication). The first class is known as the SciA class (TssA^{S}), which includes TssA proteins with a C-terminal region similar to that of SciA such as TssA of *B. cenocepacia*. The second class, termed the enteric class (TssA^{E}), is encoded by T6SS gene clusters of a number of pathogenic bacteria that can cause enteric infections such as *V. cholerae* and *A. hydrophila*. In addition, TssA^{E} is also found in T6SS clusters of some non-enteric pathogens such as T6SS-5 of *B. pseudomallei* and T6SS-2 (HSI-II) of *P. aeruginosa*.

In many T6SS gene clusters encoding TssA^{E} orthologues a second *tssA*-like gene is also present. The latter gene encodes a protein with a similar N-terminal region to TssA^{S} and TssA^{E} but the C-terminal region is distinct from those of TssA^{S} and TssA^{E}, and it contains a region encoding a predicted TMD located midway along the length of the polypeptide chain (Figure 1.14). In *V. cholerae*, it has been shown that whereas the TssA^{E} orthologue, VCA0119 (VasJ), is essential for T6SS function, the other *tssA*-like gene, VCA0121 (VasL), is not (Zheng *et al*., 2011). Therefore, it can be defined as a Tag (T6SS-associated gene). It will be referred to here as TalT (<u>T</u>ssA-<u>l</u>ike <u>T</u>ag). Where *tssA^{E}* and TalT occur together in a T6SS gene cluster, *tssA^{E}* is usually located
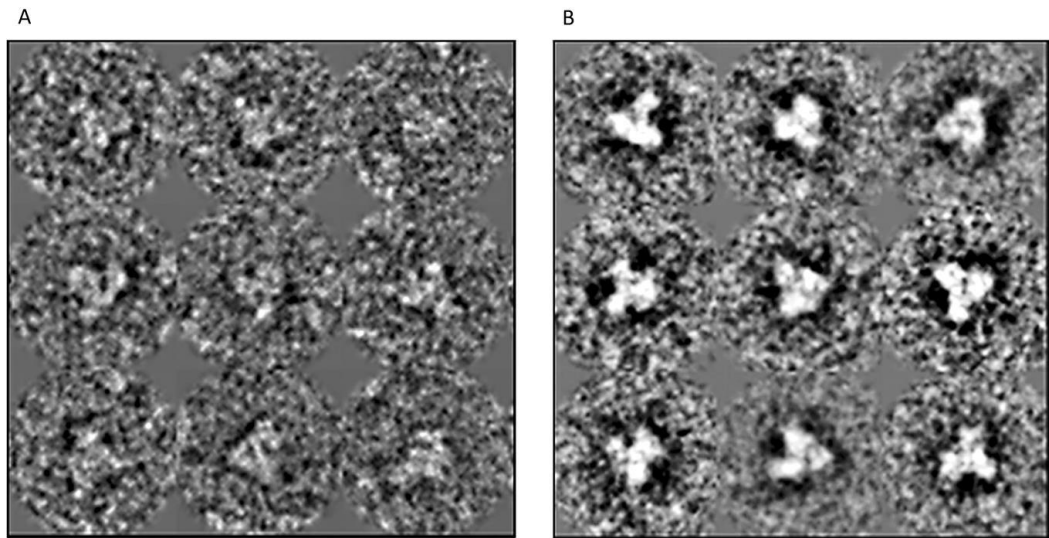
Figure 1.13. Negative stained EM images of TssK of EAEC. The micrograph shows the tri-armed structure of TssK. A. Negative stain EM of TssK. B. EM averaging of raw particles of TssK. Adapted from Zoued *et al* .(2013).
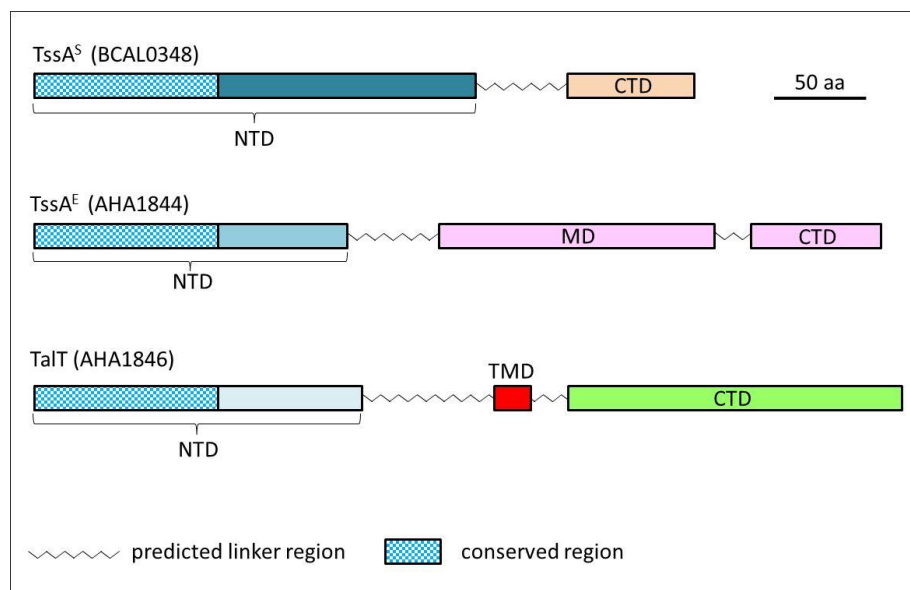
Figure 1.14. Predicted domain organisations of the two TssA classes and TalT. The conserved N-terminal region present in both classes of TssA and TalT is represented by the chequered fill colour. Regions shown in uniform colour are conserved within a TssA class or TalT. Each type of TssA and TalT contains a predicted N-terminal domain (NTD) and a C-terminal domain (CTD) separated by a linker region (zigzag motif). TssA^E contains an additional domain (middle domain, MD) separated from the CTD by a short linker region. TalT orthologues are predicted to contain a trans-membrane domain (TMD) between the NTD and CTD (M. Thomas, personal communication).

55

immediately upstream of *tssM* and *talT* is located immediately downstream of *tssM* (Figure 1.15).

The secondary structure and domain organisation of TssA^S, TssA^E and TalT was predicted by amino acid sequence alignments and the PSIPRED program (Figures 1.16, 1.17, 1.18, 1.19 and 1.20). In all three groups of proteins the N-terminal 100-130 amino acids are conserved (Figure 1.16). This is followed by a non-conserved helical region connected to the group-specific C-terminal region by a non-conserved random coil linker region (which includes a TMD in the case of TalT). In the work described here, for simplicity, the entire region located N-terminal from the predicted linker sequence will be referred to as NTD (N-terminal domain). The TssA^S region located C-terminal to the linker consists of 60-75 amino acids and is referred to here as TssA$^S_{CTD}$ (Figure 1.20A). The longer TssA^E C-terminal region can be subdivided into at least two domains based on the presence of non-conserved regions that are predicted to form random coils. Thus, TssA^E appears to have a middle domain of ~ 150 amino acids and a C-terminal domain of 80-120 amino acids which is conserved over a 70 amino acid stretch (Figure 1.20B). Therefore, TssA^S and TssA^E have predicted CTDs of similar size but distinct sequence. The C-terminal region of TalT orthologues is approximately 120 amino acids in length and is conserved within this group of proteins (Figure 1.20C). It bears no obvious similarity to the C-terminal domain of TssA^S or the middle or C-terminal domains of TssA^E.

As most of the work in this study was carried out on the TssA subunit of *B. cenocepacia*, a TssA^S member, for simplicity it will be referred to as TssA throughout this thesis.
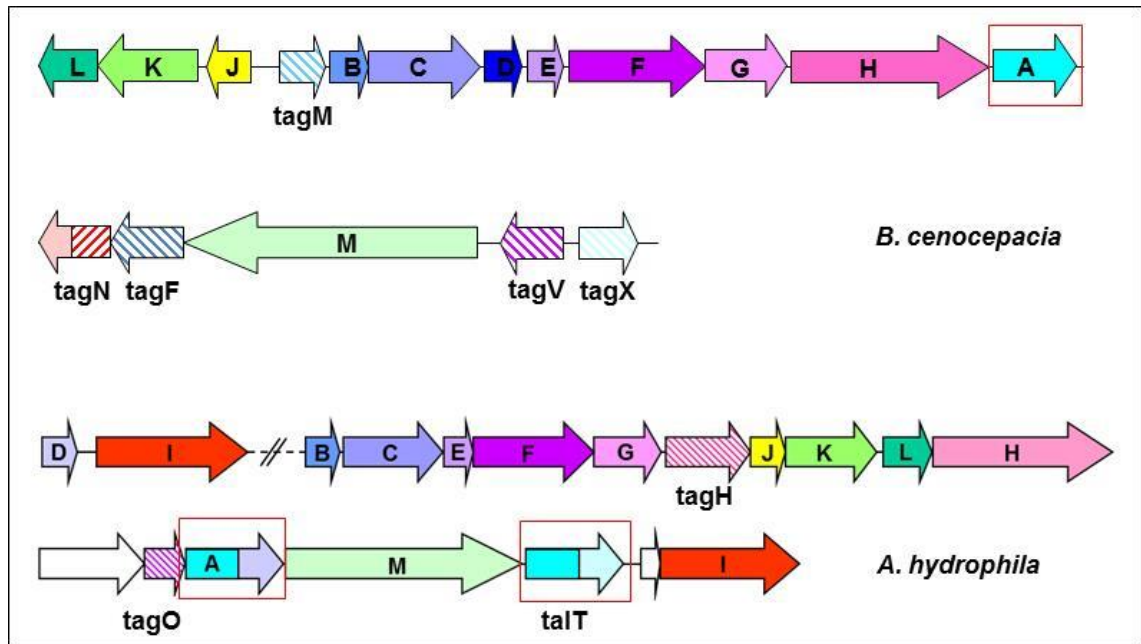
Figure 1.15. Organisation of T6SS gene clusters of *B. cenocepacia* and *A. hydrophila*. *tss* genes encoding the 13 T6SS core components are represented as uniformly coloured block arrows whereas the *tag* genes (*tss* associated genes) are hatched. The orientation of the block arrows indicates the direction of transcription of the genes within the cluster. *B. cenocepacia* *tssI* orthologues are located elsewhere on the *B. cenocepacia* genome. The *B. cenocepacia* T6SS gene cluster (BCAL0337-BCAL0352) contains a single *tssA* gene, encoding a SciA class TssA (BCAL0348). The *A. hydrophila* T6SS gene cluster (AHA1826-AHA1848) contains a *tssA* gene located upstream of *tssM* that encodes an Enteric class TssA (AHA1844). In addition, a *tssA*-like *tag* (*talT*, AHA1846) is located downstream of *tssM*.

```
BCAL0348     1 --------------------MPIN-------LPEILTPISEASP-SGDDILFSNEEAIQDARRYDDPTL
PA0082       1 -------------------MID-------VPVILAAVSPDSP-CGDDLEYDAALELERIA-QGQPER
BB0799       1 --------------MPRMTTNVE-------FADILKTLTPSIP-CGEDLEYDADLQLQQAA-VGRSEQ
SciA         1 -------------------VD--------MSITKPVSAEQP-CGPDPEYDPEYLLLFSRA-APQAEA
ImpA         1 -------------------MN-------VREIIDPLEQNHP-CGENVRSNTAFREIYYRIKDARNAA
YPO0499      1 -------------MI------------LKRVDNNMSDIFPQALFGVEYDPAIGEIESISQLDESA
AHA1844      1 ----------------------MSYQHPWCARILTSLPDEQI-RCAVIADEPRWDIYVETEIV------
Aec29        1 --------------------MAVDLRNPDVWLAHILENLPEDKL-SAALDDGNADWEFVDSEIV------
ECA3433      1 -------------------MHAHPWCKRIQTSLPDERL-RAAVIADDPLWBKVETEIV------
YPO3602      1 ----------------MRHVDITSQHPWHVLIQAPLSPEQT-AQALADDDPQWEYIDGQMV------
BPSS1493     1 -----MSERRPPGGAAARARMPID---VEALAVLGRTDIDSAYP-ACADVRADARFDALHAEIA------
PA1656       1 -------------------MTYSSKLSSHYLELAEQPISKDVF-ACEDVRFSSEYBAIENEIG------
AHA1846      1 -----------------------------MSQNQQGQQAIK-VGRDPRMLPEYEALRAEIN------
Z0249        1 MCIPTRKITRSAVGCSASLVCQTH---CTEEAGMNSNVLTQTIV-TGSDPRGLPLFSAIREEIN------
Aec31        1 --MWIPKITRSPAVCSACSVYRR---CIKGIVMSNYALTQTIV-TGSDPRALPEBSAIREEIN------
ECA3431      1 -------------------------MQDAQQAIK-VGRDPRMLPEFIALRAEIN------
YPO1489      1 -------------------------VAGNTERLIK-TGGDPRALALFTAIRGEIN------
EC042_4550   1 -------------------------MTSEVKIK-TGGDPRSLPDYAALRDEIS------

BCAL0348    43 DQG-----EWVTEIKEADWGFVVDHAGELLRTRTKDLRIIAVILTEALALEDGITGLTEGYALIEGICREF
PA0082      41 QMG-----DAVLPAEPPEWPRVRALASEL-FGRSKDLRVANLLQSNVALDGIDGIADGILLVRELIGQY
BB0799      47 QFG-----ATIIPAQAPDWRAVERLAIGL-LERTRDLRIIAYILRAWTEMRGIPGYAEGVTIAGTEQY
SciA        39 QYG-----DFVSTPEAVNWPEIERDARRL-LTRSKDIRVLILILRSRIQQACAQGIAEMITQIAEISVIW
ImpA        41 RTAERSIIPGETITIAPAWHDVSNLCIQILSSKSKDIEVLAIAEAQLRLRCFSGLHIVYVATVSLIDKH
YPO0499     42 DPL-----SRPHEPPQINWHHISEQANKL-LEQCFDLRVMLWFIRANIHKGISIYGFMRINAQTQDT
AHA1844     40 -KL------GSLAHSQVDLNAVAEACIGLLESRTKDMRVLAQLLRCLQPAKATPLGAAISLIEAWIQAY
Aec29       44 -KL------GSLAHSQIDIPELQRRGIMILASEIKDFRLLAHLLRTLQHAGD---ILLASRVIAQYTEHY
ECA3433     39 -KL------GSLAHNQVDLNVVAGYCITLLESKTKDMRVLVQLLRCLQHPAKATPFSTAIMLIDSWIESY
YPO3602     45 -KL------GSLAHATINIDDIQQQAMALLSQKSKDFRIVVHILRTLQHGGQPDEIMLAMSLITEYIQLF
BPSS1493    56 -KL------ASPGASGQIDWRAATHLAAELIRELGKDILVGCYLAGALLQTGGAAGIRCGLEIVGDLIVERH
PA1656      44 -KA------LSLHENGQTDWLKILENSEALLRAQSKDLRIVAAILTWALYQRESFPGILAGIGLIERICSRH
AHA1846     32 -KI------SHASRPEIDWQRIHQLASLITFEKIGVDIQTAIIYFLLARSRLQGISGFTECCEFIANLIVTQ
Z0249       61 -KA------SHPSQPEINWKLVESLAIAITFKANGVDLHTATIYYLLARTRTQGIAGFCEGAELIAAMVSHD
Aec31       59 -KA------NHPSQPEINWKLVESLAISIFKANGVDLHTATIYYLLARTRNQGIAGFCEGAELIAAMIIHE
ECA3431     29 -KI------SHASRPDIDWMLVHDMATTIFEKQGVDIQTAILYFLLARSRLAGITGFTESCEFIANLIVTQ
YPO1489     30 -KI------AHPARPDIDWVRVEQLCIALFRQNGVBIQTAVDFILARTHTVGIAGLCEGIELIAGLYCHQ
EC042_4550  28 -KI------THPARPDIDWRYVETLCIRLYEHNGVBIQTASIYYIIARMISIGINEGIALIVALTRHH

BCAL0348   108 WDTFHPLPEDDDIEH-------RLGNVAWLSGITAE-----LIRAVPLT-DGASNAFSTLDWEVAQHVAQ
PA0082     105 WDGVVPLLDADDD-------NDPTFRINAITGLVAEPLLQLVWAIPLVR-SR---AFGPVNIRAALNAA-
BB0799     111 WDAVHPMLASGGE-------EDPVPRVNALASLGDPQGCVRGVRSACLL-DD---VHGRLSIRDAEAL--
SciA       103 SDAIHPQLLTGEGASAESIEDASIARSNAIAALLDHEGVMADIRGITLS-NS---AALRLQVRDVEFAL-
ImpA       111 EDAIHSIGDGDVEE--------RFAPFAGINGVGGEGTLIQAIRLTSLIPGGKFAQFSLWDFQISQRPN-
YPO0499    106 DVVIYPQSEEPPL------NSGHAAALGWISTAQ----CIAELKTARLTEEH---PYIILQDLITTEALPN
AHA1844    103 WLLAWPGNASQKQ---------RIMV--------------------------------QIVKRFEG
Aec29      104 WTCAAPQNMAHKK---------RFAA--------------------------------QVIKRFES
ECA3433    102 WVTAWPASPVQKQ---------KIMI--------------------------------QIVRRFEG
YPO3602    108 WTTAWPQNPLHKR---------RFAQ--------------------------------QIIKRFIS
BPSS1493   120 WDAMSPPVSRMRA---------RRCALQWLVDRVDAMHD----AGAAAC-GGACSAELVAQIRAAARR--
PA1656     108 WVEIHPLKAR--T---------RAAAIAWLVPRLEQALN----ENVPIK-EQ---LPLFRRIAEHLEK--
AHA1846     95 WESFWPPVHQERA---------RIEMIDWFIARISD----VIRQY-QISH--------EDKRLIYRCER
Z0249      124 WDKFWPQGG--PA---------RTEMIDWFNSITGN-----IIRQQISFAE--------SDIIPLIYRTER
Aec31      122 WDKFWPQSG--PA---------RTEMIDWFNTRTGN-----IIRQQVSFSE--------NDISLIYRTER
ECA3431     92 WDNFWPPVHQERA---------RIEMIDWFIARISE----VIRQY-AISH--------EHKRLVYRCER
YPO1489     93 WRGIWPPQT--HA---------RVALIAWLSDILQQ-----VWRTMTLCY--------GDIAQVYRAPR
EC042_4550  91 WSVMWPLNT--HA---------RLEIITGLFNILQK-----TIRAMPPDDR-------DNLPLIYQTET
```

Figure 1.16. Alignment of the N-terminal regions of TssA^S, TssA^E, TalT orthologues. Alignment was carried out between TssA^S orthologues (red font) present in *B. cenocepacia* strain J2315 (BCAL0348), *P. aeruginosa* strain PAO1 (PA0082), *Bordetella bronchiseptica* strain RB50 (BB0799), *S. enterica* serovar Typhimurium str. LT2 (SciA), *R. leguminosarum* bv. viciae 3841 (ImpA) and *Y. pestis* strain CO92 (YPO0499); TssA^E orthologues (green font) present in *A. hydrophila* subsp. hydrophila ATCC 7966 (AHA1844), *E. coli* (Aec29), *Pectobacterium atrosepticum* SCRI1043 (ECA3433), *Yersinia pestis* CO92 (YPO3602), *Burkholderia pseudomallei* K96243 (BPSS1493) and *P. aeruginosa* PA01 (PA1656) and TalT orthologues present in *A. hydrophila* subsp. hydrophila ATCC 7966 (AHA1846), *E. coli* O157:H7 (Z0249), avian pathogenic *E. coli* BEN2908 (Aec31), *Erwinia carotovora* subsp. atroseptica (ECA3431), *Yersinia pestis* CO92 (YPO1489) and *E. coli* (Ec042_4550); and TalT orthologues (blue font) present in *A. hydrophila* subsp. hydrophila ATCC 7966 (AHA1846), *E. coli* O157:H7 (Z0249), avian pathogenic *E. coli* BEN2908 (Aec31), *Erwinia carotovora* subsp. atroseptica (ECA3431), *Yersinia pestis* CO92 (YPO1489) and *E. coli* (Ec042_4550). Alignment was carried out using Clustal Omega and shading was conducted using BoxShade program. Amino acids identical or similar in 50% of the sequences are shown in white font and shaded by black or grey, respectively. The red arrow indicates the C-terminal end of the N-terminal homologous regions.

```
BCAL0348  302  GIQNRAQAVDQTRAVARYFRQTEPHSPVAYLADKAAEWADMPLHKWLESVVKDDGSLSHI
PA0082    278  EIANREDVLRQLDRLLEYYVRHEPSSPVPVLLKRAKTLVTADFAEIVRNLIPDGISQFE-
BB0799    280  RIATREDAMAMLAKVSAYFETHEPSHPAPYLIRRVQQLIPLDFHDIIRNLAPQGIAQFE-
SciA      290  PIRDRNDALERLRCIRRWFESSEPSSPTIPLLRQAERLVGKRFSEVINEIPVELLEKWD-
ImpA      285  MIRSREEAFELLIAVARYFRRTEPHSPISMSIETLVRRGRMDFFELLAELLPEQQLRNA-
YPO0499   268  SIRSRQEIIMVLDRLLEYFQHYEPSHPAPIFIRRTKEMIGMDFYSIVEEILPEAVITLK-

BCAL0348  362  RELLGVRPDEQS----
PA0082    337  -TLRGPESE-------
BB0799    339  -AWTAREAASQS----
SciA      349  ----ALE---------
ImpA      344  -VLTAAGIQPVADKGG
YPO0499   327  -QFTCKTNLPFR----
```

Figure 1.17. Alignment of the C-terminal region of TssA[S] orthologues. Alignment was carried out between TssA[S] orthologues present in *B. cenocepacia* strain J2315 (BCAL0348), *P. aeruginosa* strain PAO1 (PA0082), *Bordetella bronchiseptica* strain RB50 (BB0799), *S. enterica* serovar Typhimurium str. LT2 (SciA), *R. leguminosarum* bv. viciae 3841 (ImpA) and *Y. pestis* strain CO92 (YPO0499). Alignment was carried out using Clustal Omega and shading was conducted using BoxShade program. Amino acids identical or similar in 50% of the sequences are shown in white font and shaded by black or grey, respectively.

```
AHA1844    228  SSNDRAWRQTQL---KVAELLIERQPEVAVGYRLRRHAVWAGITAVPMSGAGN-KTPLAP
Aec29      230  NHDDKAWRDTLL---KVAAILCERQPESPQGYRLRRHALWQTITSTPQAESDG-RTPLAA
ECA3433    220  SSDERSWRQTQL---NVAALLVERHPDSPMGYRLRRHAIWSGIATPPMSSKGH-KTQLAP
YPO3602    217  NTSDRAWRQTLLI---KMADLLSEQQPDAAIGFRLRRHAVWGALTAPPMAQSDG-RTPLAA
BPSS1493   328  ---ERALADALAQLHCVATAFAQADWADARGFRLRRVACWSSVCALPEIDAENGRTRIAA
PA1656     224  HKAMRAQQEAAR---PICAWWIKQKATDLRALRLNRTMIWLPIESMPERNAEQ-VTALRG

AHA1844    284  MSADMVDEYRAAVNAPDQ-GLWQRIEQSLTLAPYWFEGHRLSAEVAEKLGTGA--VAQAI
Aec29      286  FPVDMMDDYLARINADY-ALWQQVEKSLLLAPYWLDGHYLSAQTAIRLGVKQ--VADAI
ECA3433    276  VSSDRVDEYQSALAQADI-ALWERIEQSLVLAPYWFDGHMLSASVASRLGHSS--VATAM
YPO3602    273  VSADRTADYLARLANADL-PLWHQVEQSLTLAPYWLDGHVLSAQIALQLGYDA--VAQAI
BPSS1493   385  PSASIVGAAKSIDCDGEPVAAVRFAEAHAQAFPLWLDLQRIAARALARAGGDCADARREV
PA1656     280  VPADKIKSYQERFAQGLYADILVEIEASLARAPFWFDGQRLVWECLQGINAIQ--AMREV

AHA1844    341  AEELGTFIQRLPALRELAFSDGSPFLSPECSRWLQPAKG--G---S--------------
Aec29      343  RDEVTDFIARLPALINILFNDRTPFVSEQTKQWLASSGSVNQ---TI--------------
ECA3433    333  AEELSAFIQRLPELRELAFSDGAPFLIGKCSQWLQSSQPVRG---G--------------
YPO3602    330  RDELSVFIARIPALKTLFFTDMTPFLSSESAAWLQQDANHQG---R--------------
BPSS1493   445  ETAVRALLARLPGLDALTEADGTPEADDATRAWLGELGAPVVAADAVSPSSLPLSPRPSP
PA1656     338  EMHFALLIQRLPGLVELRFHDGSAFADAATRGWISAHVMPHLQNDSA-------------

AHA1844    382  -------AG----------I-GEAGLAEEVAQRHGEQGIAAAIAILDERIAQLKEPRDRF
Aec29      386  -------VP----------VVQTDEEIQAAKACFDENGLEAALRYLE-NIPE-GDPRHQF
ECA3433    376  -------GG----------A-RQDDLATEAASCRDEKGIGAAMLILDERVRFLKEPRDRF
YPO3602    373  -------SR----------TIEQ----DEIWQCYQQQGLEAALQMIN-RQPQQSEPRDRF
BPSS1493   505  PERSSPMAGEPARAPGDACGASADDAVDRACAFAASGQLDLALHAIQHAIDRATSAEQRL
PA1656     385  PRKVETVAL----------QAEWDVALDEVQPVLRKDGLKAAVQVIKQGVKRAIGGRARF

AHA1844    424  HALLVQAELIAQEGMEALARQHYQHLWQEASRLGLSHWEPGLVNRLESLAAPLSK-----
Aec29      427  HRQFFGAQLIEEAGVQLAQQQYRMLFRTGLHMMLSEWEPSLLKALEQKLTAEQ------
ECA3433    418  YAELVILADLLABEGMKSLAAQHYQHMWQESQOLGIMQWEPGMVSRVERLAASRKK-----
YPO3602    411  YHQLLSAQLFEKAGITALAQQHYHSILLVGQQLQLSEWEPALIALITEKQRQLKP-----
BPSS1493   565  RARVRLCELARDHWPHEVPEAFARGVIEPIRRHDILAWNPELALDGLSAAYAL--LIRRD
PA1656     435  FWQLSLARLCFLAKKYELAKTQLESLDHQLHESGIHAWEPDLALDVLHILHSCCELLPQN

AHA1844    478  --------------------------
Aec29      480  --------------------------
ECA3433    472  --------------------------
YPO3602    465  --------------------------
BPSS1493   623  RE-SAHARTVLDEIASVDAARAMRLST
PA1656     495  HAVRERKEDIYRRLCHLDLEVVLE---
```

Figure 1.18. Alignment of the C-terminal region of TssA[E] orthologues. Alignment was carried out between TssA[E] orthologues present in *A. hydrophila* subsp. hydrophila ATCC 7966 (AHA1844), avian pathogenic *E. coli* BEN2908 (Aec29), *Pectobacterium atrosepticum* SCRI1043 (ECA3433), *Yersinia pestis* CO92 (YPO3602), *Burkholderia pseudomallei* K96243 (BPSS1493) and *P. aeruginosa* PA01 (PA1656). Alignment was carried out using Clustal Omega and shading was conducted using BoxShade program. Amino acids identical or similar in 50% of the sequences are shown in white font and shaded by black or grey, respectively. The predicted middle domain is underlined by the red line and the predicted C-terminal domain is underlined by the green line.

```
AHA1846      353 LYPDSIAVKAIDKQWQAQLGGLQGDVISVPTYLDARAGVDALLDQLIELERQRK-TVTIS
Z0249        351 RWPESIQQQQASTQWNEALKTRAQSSPQLRGWLQTRQDLHAFADLVQR--EKE-GLTLS
Aec31        345 LWPESIQQQEASRMWSNILRNRAQASPQMKGWQQARQNLRDFADLMMKKETEKQ-GFTLS
EcA3431      355 LYPDSIAVKEIERQWQIILERQQGDEPKTLGYEQARARVNDTLQQLIELERQRR-TVTIS
YPO1489      331 LWPTNQEVKRINALWQQQRAAGAAPLVELKHYALAQDRLRQLAERINSLDEKKRGYVTVS
Ec042_4550   334 LFPDNPRVQEIVDNWQKSVRSRALPEEAMAGWNEGMTRLQQLAERINRLDEQRGKYVTVS
```

```
AHA1846      412 YLKSQLYEVQRNLMQNIPFSLRLGELEARKASQEPITAAELKSIENDIKALNIRLYQLQQ
Z0249        408 YIKNVIWQAERGLGQETPVESLITQYQDARAQKQN-TDALEKQINRPLEGVLSRWLLLKN
Aec31        404 YIKTVTWQAERLLNQETPLEYLLTQYQETRTQKQD-TQALEKEINRLDGLLSRWLLLKN
EcA3431      414 YLKSKLYDMQKDIISDVPFGIRLRELEARKIKSQSLTPAELRGVEDIRAFNIRLYRLQQ
YPO1489      391 ELKSAVEAIQQPIAQAVPLEELLRQYQEQLESGQTPSSAIRQQIDSRFSQLINRYAILVV
Ec042_4550   394 ELKTEVEGIMQAFNRHIPAEEQLRRYDEVRNQNSS--EQQQKKVENGLVELVSRYWVLTQ
```

```
AHA1846      472 GASGS--------------------------
Z0249        467 NTIPTIKKALNFNNIHEYKGVLNGEFNLFNTKW
Aec31        463 TGQDMATDNR-TEPVHPTH--------------
EcA3431      474 SNSAS--------------------------
YPO1489      451 PPGVQ--------------------------
Ec042_4550   452 GDMK--------------------------
```

Figure 1.19. Alignment of the C-terminal region of TalT orthologues. Alignment was carried out between TaiT orthologues present in *A. hydrophila* subsp. hydrophila ATCC 7966 (AHA1846), *E. coli* O157:H7 (Z0249), *E. coli* (Aec31), *Erwinia carotovora* subsp. atroseptica (ECA3431), *Yersinia pestis* CO92 (YPO1489) and *E. coli* (EC042_4550). Alignment was carried out using Clustal Omega and shading was conducted using BoxShade program. Amino acids identical or similar in 50% of the sequences are shown in white font and shaded by black or grey, respectively.

A. TssA^S

A. TssA^{S}

Conf: ▊▊▊▊▊▊▊▊▊▊▊▊▊▊▊▊▊▊▊▊▊▊▊▊▊▊▊▊▊▊▊▊▊▊▊▊▊▊

Pred: —◯ H1 ⬭——————————◯ H2 ⬭——

Pred: CCCCHHHHHCCCCCCCCCCCCCCCCCCHHHHHHHHHHHCCCCC

AA: MPINLPELLTPISEASPSGDDLLFSNEFDAIQDARRYDDP

          10   20   30   40

Conf: ▊▊

Pred: ——————————◯ H3 ⬭—◯ H4 ⬭

Pred: CCCCCCCCCCCCCCCHHHHHHHHHHHHHCCCHHHHHHHHHH

AA: TLDQGEWVTEIKEADWGFVVDHAGELLRTRTKDLRLAVWL

          50   60   70   80

Conf: ▊▊

Pred: ◯ ⬭———◯ H5 ⬭———————◯

Pred: HHHHHHHCCCCHHHHHHHHHHHHHHHHHCCCCCCCCCCCCH

AA: TEALALEDGITGLTEGYALLEGLCREFWDTFHPLPEDDDI

          90   100   110   120

Conf: ▊▊

Pred: ◯ H6/7 ⬭————◯ H8 ⬭

Pred: HHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCHHHHHHHHH

AA: EHRLGNVAWLSGRTAELLRAVPLTDGASNAFSTLDWEVAQ

          130   140   150   160

Conf: ▊▊

Pred: ◯ ⬭—◯ H8b ⬭—◯ H9 ⬭—◯

Pred: HHHHHHCCCCCHHHHHCCCCCHHHHHHHHHHCCCHHHHHHH

AA: HVAQSIKRDPEHADDIARGKPSIEQIDASRRVTSIAFYTA

          170   180   190   200

Conf: ▊▊

Pred: ◯ H10 ⬭————◯ H11 ⬭

Pred: HHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCHHHHHHHHHHH

AA: LLANLKAFEFALDAFEERLVERAGDSAPSFRQARDAFETV

          210   220   230   240

Conf: ▊▊

Pred: ⬭———◯ H11a ⬭————

Pred: HHHHHHHHHHHCCCCCCCCCCHHHHCCCCCCCCCCCCCCCCC

AA: YRLAERFAREQGYTGSAPHTQAVQQAQPERIEPVFGQPIQ

          250   260   270   280

63

```
Conf: [■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■]
Pred: ─────────────────────────────────────○▭▭▭▭▭▭ H12 ▭▭▭▭▭▭
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHH
  AA: TEETHVQQQTASRPPVTQTIAGIQNRAQAVDQLRAVARYF
              290       300       310       320

Conf: [■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■]
Pred: ─○─────○▭▭ H13 ▭▭○─────○▭▭ H14 ▭▭○──○▭ H15 ▭
Pred: HHCCCCCCHHHHHHHHHHHHCCCHHHHHHHHHCCCHHHHHH
  AA: RQTEPHSPVAYLADKAAEWADMPLHKWLESVVKDDGSLSH
              330       340       350       360

Conf: [■■■■■■■■■■■■■■■]
Pred: ▭▭▭▭▭▭○────────
Pred: HHHHHCCCCCCCC
  AA: IRELLGVRPDEQS
              370
```

B. TssA^E


```
Conf: [bar chart]
Pred:           ⊂===H1===⊃                    ⊂======H2======⊃
Pred: CCCCCCCHHHHCCCCCCCCCCCCCCCCCCHHHHHHHHHHHH
  AA: MSYQHPWCARLLTSLPDEQIRGAVLADEPRWDYVETELVK
                 .         .         .         .
                10        20        30        40

Conf: [bar chart]
Pred:              ⊂=====H3=====⊃     ⊂=====H4=====⊃
Pred: CCCCCCCCCCCHHHHHHHHHHHHHCCCCHHHHHHHHHHHHHC
  AA: LGSLAHSQVDLNAVAEACLGLLESRTKDMRVLAQLLRCLQ
                 .         .         .         .
                50        60        70        80

Conf: [bar chart]
Pred:        ⊂==========H5==========⊃      ⊂====H6====⊃
Pred: CCCCHHHHHHHHHHHHHHHHHHHHHCCCCCCCHHHHHHHHH
  AA: HPAKATPLGAAISLLEAWIQAYWLLAWPGNASQKQRLMVQ
                 .         .         ↑.         .
                90       100       110       120

Conf: [bar chart]
Pred: ⊂==================H6==================⊃
Pred: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCC
  AA: IVKRFEGALPRICESASAAELAQLLAQAEQLERVWLAQCP
                 .         .         .         .
               130       140       150       160

Conf: [bar chart]
Pred:        ⊂==========H7==========⊃
Pred: CCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCC
  AA: DKGELLDPLVMGLKRAQRQQLAQAEANAAGQPQSSGAAAA
                 .         .         .         .
               170       180       190       200

Conf: [bar chart]
Pred:                                     ⊂===H8===⊃
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHH
  AA: GSPASVASTASCAGAMVLSGSAGVDVDSSNDRAWRQTQLK
                 .         .         .         .
               210       220       230       240

Conf: [bar chart]
Pred: ⊂===H9===⊃    ⊂==========H9==========⊃
Pred: HHHHHHHHCCCCHHHHHHHHHHHHHHHHHCCCCCCCCCCCCCC
  AA: VAELLIERQPEVAVGYRLRRHAVWAGITAVPMSGAGNKTP
                 .         .         .         .
               250       260       270       280
```

C. TalT


```
Conf: [████████████████████████████████████████]
Pred:                    ○━━━━━━━━━━━━━○
                                H1
Pred: CCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHCCCCCCCC
  AA: MSQNQQGQQALKVGRDPRMLPEYEALRAEINKLSHASRPE
                  10        20        30        40

Conf: [████████████████████████████████████████]
Pred:       ○━━━━━━━━━━━○  ○━━━━━━━━━━━━○ ○━━━━━
                 H2              H3
Pred: CCHHHHHHHHHHHHHCCCHHHHHHHHHHHHHHHHCCHHHH
  AA: VDWQRIHQLASLIFEKHGVDLQTAIYFTLARSRLQGLSGF
                  50        60        70        80

Conf: [████████████████████████████████████████]
Pred: ━━━━━━━━━━━━━○      ○━━━━━━━━━━━━━━━━
            H4                    H5
Pred: HHHHHHHHHHHHHCCCCCCCCCCCHHHHHHHHHHHHHHHHH
  AA: TEGCEFLANLIVTQWESFWPPVHQERARIEMLDWFIARIS
                  90       100       110       120

Conf: [████████████████████████████████████████]
Pred: ━━━━━━━━━━  ○━━━━━━━━━━━━━━━━━━━━━○     ○┐
                          H6
Pred: HHHHHHHCCCCHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCH
  AA: DVIRQYQISHEDKRLIYRCERALQLISEKLHNADLSRIPR
                 130       140       150       160

Conf: [████████████████████████████████████████]
Pred: ━━━━━━━━━━
              H7
Pred: HHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCC
  AA: VENLVHFIEGYTHLFDETEIVIVSDEPGLKEDLQIPPMVF
                 170       180       190       200

Conf: [████████████████████████████████████████]
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
  AA: FKSDMESDHGSTVSASTPHLPQGSILVGCEKGQVKPTVLK
                 210       220       230       240

Conf: [████████████████████████████████████████]
Pred:            ○━━━━━━━○   ○━━━━━━━━○  ○┐
                    H8              H9
Pred: CCCCCCCCCHHHHHHHHHHHCCCCCCHHHHHHHHHHCCCCHH
  AA: IEQHRRQRPAWFWFGCGLLSCALPVMGWLGWQQHQQEKTV
                 250       260       270       280
```

```
Conf: [||||||||||||||||||||||||||||||||||||||||||]

Pred:      H10              H11    H12

Pred: HHHHHCCCCCCCCCCCCCCHHHHHHHHHHHHHCHHHHCCCCHHH
  AA: AAQRLVQPAAELPRALSYDDIRQARIVLGEQTLQNMESDL
              |         |         |         |
             290       300       310       320


Conf: [||||||||||||||||||||||||||||||||||||||||||]

Pred:        H13              H14

Pred: HHHHHHHHHHHHCCHHHHHHHHHHHHHHHHHHHHHCCCCHHH
  AA: VARYQNQLTRLEQTSPLYWYRYGEGLRNSLQMLYPDSLAV
              |         |         |         |
             330       340       350       360


Conf: [||||||||||||||||||||||||||||||||||||||||||]

Pred:        H15              H16

Pred: HHHHHHHHHHHHCCCCCCCCCCHHHHHHHHHHHHHHHHHHH
  AA: KALDKQWQAQLGGLQGDVISVPTYLDARAGVDALLDQLLE
              |         |         |         |
             370       380       390       400


Conf: [||||||||||||||||||||||||||||||||||||||||||]

Pred:              H17           H18

Pred: HHHHCCCCCHHHHHHHHHHHHHHHHHHCCCCHHHHHHHHHHH
  AA: LERQRKTVTISYLKSQLYEVQKNLMQNIPFSLRLGELEAR
              |         |         |         |
             410       420       430       440


Conf: [||||||||||||||||||||||||||||||||||||||||]

Pred:                  H19

Pred: HHCCCCCCHHHHHHHHHHHHHHHHHHHHHHCCCCCC
  AA: KASQEPITAAELKSLENDLKALNIRLYQLQQGASGS
              |         |         |
             450       460       470
```

Legend:

- helix

- strand

- coil

Conf: - confidence prediction

Pred: predicted secondary structure

AA: target sequence
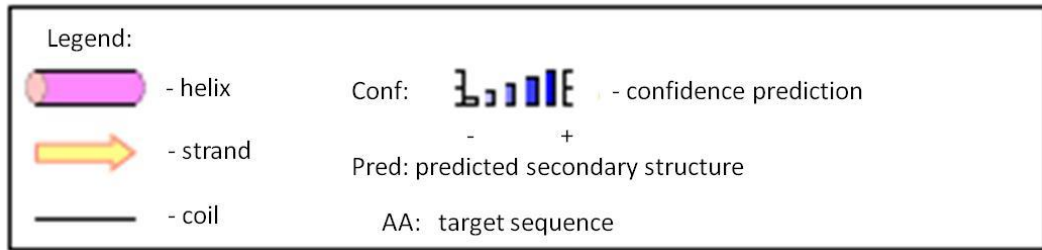
Figure 1.20. PSIPRED secondary structure predictions of *B. cenocepacia* TssA^S (BCAL0348), *A. hydrophila* TssA^E (AHA1844) and *A. hydrophila* TalT (AHA1846). The amino acid sequence of a representative of each of the two classes of TssA and TalT was submitted to the PSIPRED program (v3.3) to predict the secondary structure. Regions highlighted with a red box are the predicted linker regions. The red arrow indicates the position of the C-terminal end of conserved part of the NTD. A. The secondary structure prediction of *B. cenocepacia* TssA^S (BCAL0348) shows an unstructured peptide chain (linker region) of ~ 55 amino acids separating a large N-terminal region of 11 α-helices from a C-terminal region of 4 α-helices. α-helices are numbered according to their position relative to the N-terminus. The region corresponding to H6/H7 is folded into two helices (H6 and H7) in other members of the TssA^S class and the region folded into the H8b helix is not predicted to form a helix in other members of the TssA^S class. The H11a helix is a region with low confidence of prediction and is not detected when an older version of the PSIPRED program was used (results not shown). B. Secondary structure prediction of *A. hydrophila* TssA^E (AHA1844) shows a predicted random coil (linker region) of ~ 42 amino acids separating an N-terminal region of 7 α-helices from a large C-terminal region of 12 α-helices. The region folded as H14 is probably part of the middle domain-CTD linker because it is not strongly conserved in other TssA^E members. The H15 region is also not strongly conserved (Figure 1.18). Therefore a second linker region extending from position 351 to 388 divides the C-terminal region into two domains, the middle domain (MD) containing 7 α-helices and the CTD containing 4 or 5 α-helices. α-helices are numbered according to their position relative to the N-terminus. C. Secondary structure prediction of *A. hydrophila* TalT (AHA1846) shows an long unstructured peptide chain (linker region) ~74 amino acids separating an N-terminal region of 7 α-helices from a predicted TMD (underlined by the green bracket) and a large C-terminal region of 10 α-helices. The region extending C-terminal from the TMD to postion 352 is not well conserved among TalT orthologues and has been omitted from the TalT CTD alignment (Figure 1.19) α-helices are numbered according to their position relative to the N-terminus.

1.5.5. T6SS accessory proteins

A number of proteins have been identified that serve as auxiliary subunits of the T6SS. Many of these are referred to as Tags (Section 1.4.2) (Shalom *et al*., 2007). Tags are not present in all organisms that encode a T6SS, and in those that do contain *tag* genes, their number and nature varies. The most studied Tag is TagL as it was demonstrated to carry a specific domain, the peptidoglycan-binding motif of the OmpA/MotB/Pal family of proteins which anchors such proteins to the bacterial cell wall (Aschtgen *et al*., 2010b). EAEC TagL was shown to interact with the PG *in vivo* and *in vitro* (Aschtgen *et al*., 2010a). In many bacteria possessing a T6SS, a peptidoglycan-binding motif is carried by the C-terminal periplasmic domain of TssL ("evolved TssL"), whereas in a few species such as EAEC and uropathogenic *E. coli* (UPEC) a PG-binding motif is carried by TagL (Aschtgen *et al*., 2010a, Aschtgen *et al*., 2010b). The PG-binding domain is also located at the C-terminus of other Tag proteins such as TagN and TagP and in the central region of TagW. TagP is similar to TssM, as they share a conserved N-terminal domain, but the C-terminal periplasmic domain of TssM is replaced by the PG-binding domain in TagP. Only a few bacteria contain TagN, such as *B. cenocepacia*. TagW is a large polypeptide which was found in some species of *Vibrio* such as *V. parahaemolyticus* (Aschtgen *et al*., 2010b). The proposed function of the PG-binding motif is to stabilise the T6SS in the bacterial cell wall. It should be noted that some bacteria with a T6SS lack an evolved TssL and a Tag containing a PG-binding domain (Aschtgen *et al*., 2010b). It is not clear how the T6SS in these organisms is anchored to the PG.

TagJ has also been observed in *P. aeruginosa* HSI-I, *S. typhi* (where it is also known as SciE), *B. pertussis*, *V. parahaemolyticus, Serratia marcescens* and *Rhizobium* (ImpE) (Shalom *et al* 2007; Mark Thomas personal communication). This protein is present in 30% of bacteria that contain a T6SS (Lossi *et al*., 2012). TagJ in *P. aeruginosa* is specific to H1-T6SS and it was shown to interact with an N-terminal α-helix in TssB encoded by H1-T6SS forming a subcomplex which might mediate the integration of TssB into the T6SS (Lossi *et al*., 2012). Also, tagJ of *S. marcescens* was shown to interact with TssB from *S. marcescens*. Interestingly, the H1-T6SS encoded TssB has an N-terminal helix that the other two TssB subunits encoded by H2- and H3-T6SS do not have. Also the TssB subunit in bacteria with T6SS gene clusters deficient in the TagJ-like subunit often lack the N-terminal α-helix region. Moreover, the interaction of TagJ

with TssB1 was shown to be specific as TagJ does not interact with TssB2 (Lossi *et al.*, 2012).

A very recent study to identify additional components associated with the T6SS tail spike that correspond to orthologues in the bacteriophage T4 tail spike revealed an accessory subunit. The study identified a PAAR protein which forms a sharp extension on the tip of the TssI (VgrG) spike. Furthermore, the study showed that the PAAR proteins are essential for the function of the T6SS in *Acinetobacter baylyi*, which encodes 3 PAAR proteins, as a triple deletion mutant that removes all three PAAR genes result in a significant reduction in TssD (Hcp) secretion and ~ a 10,000 fold decrease in T6SS-dependent killing of *E. coli*. A similar result was obtained in *V. cholerae* which has two PAAR proteins. The authors suggest that PAAR proteins may contribute to the T6SS assembly or they may play a role in the translocation of the VgrG into the host cells (Shneider *et al.*, 2013). PAAR proteins also play a role in Rhs effector secretion (Section 1.6.6).

1.6. Effector proteins of the T6SS

Despite intensive research on the T6SS, only a few secreted effector proteins have been identified so far. The first identified effector proteins of the T6SS were VgrG derivatives which showed a dual role as structural subunits and as secreted effector proteins of the T6SS (Section 1.5.1.2). Some evidence was also provided that the TssD subunit also has a secondary role as an effector (Section 1.5.1.1). Some small and large effector proteins were subsequently identified which are not subunits of the T6SS (see below). In several cases, immunity proteins have been identified to specific VgrG and non-VgrG effectors encoded adjacent to the effector gene in the producing bacterium.

1.6.1. Evolved TssIs (VgrGs)

The T6SS in *V. cholerae* has been shown to secrete three VgrG (TssI) proteins: VgrG1, VgrG2 and VgrG3, of which VgrG1 and VgrG3 are evolved derivatives (Pukatzki *et al.*, 2007). A well-studied evolved VgrG is VgrG1. The C-terminal domain of VgrG1, which leads to actin cross-linking, is homologous to the actin crosslinking domain (ACD) of RtxA toxin of *V. cholerae* (Sheahan *et al.*, 2004, Ma *et al.*, 2009a). The C-terminal region of VgrG3 (VCA0123) in *V. cholerae* has a peptidoglycan-binding domain and exhibits peptidoglycan hydrolase activity. Interestingly, *V. cholerae*

protects itself from the toxic effect of VgrG3 by production of type six secretion antitoxin B (TsaB) a product of the downstream gene *tsaB*. Therefore, VgrG3 and TsaB represent a toxin-antitoxin pair (Brooks *et al*., 2013). VgrG2 has no C-terminal extension and therefore constitutes a core VgrG subunit (Pukatzki *et al*., 2007). Interestingly, it was demonstrated that inactivation of VgrG2 prevents the secretion of TssD, VgrG1 and VgrG3 and reduces the virulence of *V. cholerae* towards *Dictyostelium* amoebae and J774 macrophages (Pukatzki *et al*., 2006). The C-terminal vegetative insecticidal protein-2 (VIP-2) domain, which known is for its actin ADP-ribosylating activity, of VgrG1 in *A. hydrophila* was also shown to induce host cell cytotoxicity by modifying the cytoskeleton of eukaryotic cells (Suarez *et al*., 2010). These observations have led to the assumption that the non-evolved VgrG can target bacterial cells whereas evolved VgrG can target eukaryotic cells.

1.6.2. Non-VgrG-derived peptidoglycan hydrolase effectors

Hood and co-workers identified three non-VgrG effector proteins Tse1-Tse3 (type VI secretion exported 1-3) secreted by the haemolysin co-regulated protein secretion island (HSI-I)-encoded T6SS (HI-T6SS) of *P. aeruginosa* and demonstrated that Tse2 is a cytoplasmic toxin component of a toxin-immunity system. Tse2 can arrest the growth of many prokaryotic and eukaryotic organisms, but it was shown to be specifically targeted against the bacterial cells (Hood *et al*., 2010). Subsequent studies in the same lab showed that Tse1 and Tse3 are lytic enzymes that break down the peptidoglycan layer of the targeted (prey) bacterium by amidase and muramidase activity, respectively (Russell *et al*., 2011). In addition to Tse1, Tse2 and Tse3, *P. aeruginosa* synthesizes three immunity proteins, T6SS immunity 1, 2 and 3 (Tsi1, Tsi2 and Tsi3), which antagonise the effect of Tse1, Tse2 and Tse3, respectively. Tsi1 and Tsi3 are located in the periplasm, while Tsi2 is located in the cytoplasm. Therefore, *P. aeruginosa* protects itself from the lytic effect of Tse1 and Tse3 and the toxic effect of Tse2, injected by the sister cells. These immunity proteins seem to be limited to *P. aeruginosa* and hence the species can use H1-T6SS to attack closely related bacterial species which are in competition with *P. aeruginosa* for niches (Russell *et al*., 2011, Li *et al*., 2012).

In 2012 Russell *et al*. developed a heuristic approach for the screening of T6SS substrates. Using this approach, they undertook a study on *B. thailandensis* and identified BTH_I0068 and BTH_I0069 as an amidase-immunity pair analogous to the *P.*

aeruginosa Tse1-Tsi1 pair. They suggested the name Tae2 (type VI amidase effector) and Tai2 (type VI amidase immunity) for BTH_I0068 and BTH_I0069, respectively, and for other members of this family of proteins (Russell *et al*., 2012). The heuristic approach was used to identify T6SS amidase EI pairs in 193 phylogenetically distinct Gram-negative bacteria possessing a T6SS. 51 effectors from the 419 predicted EI pairs identified in the large scale screen were divided into four families (Tse1, Tae2, Tae3 and Tae4) based on their primary sequence homology. The Tae3 and Tae4 families include effectors present in *S. typhi* and *S. typhimurium* respectively which are able to degrade peptidoglycan and have corresponding immunity proteins (Tai3 and Tai4). The study by Russell *et al* also identified Tae3-Tai3 EI pairs in *P. fluorescens* (Russell *et al*., 2012). Tae4-Tai4 EI pairs were also identified in *Enterobacter cloacae* (Zhang *et al*., 2013). The four amidase EI pair families have distinct cleavage specificities against the peptidoglycan of Gram-negative bacteria which provides a useful measure against the structural variability of peptidoglycan (Russell *et al*., 2012).

Two T6SS-dependent effector proteins (Ssp1 and Ssp2) in *S. marcescens* were identified and confirmed as antibacterial toxins, along with their cognate immunity proteins (Rap1a and Rap2a). The Ssp1 and Ssp2 proteins contain a domain of unknown function (DUF4285) which is present in both eukaryotic and prokaryotic proteins. Ssp1 and Ssp2 were shown to be periplasmic-acting and were therefore presumed to target the bacterial cell wall (English *et al*., 2012). In fact, these proteins were recognised as being members of the Tae4 family of amidase effectors while the Rap proteins were identified as member of the Tai4 family of immunity proteins (English *et al*., 2012, Russell *et al*., 2012) .

A more recent study identified a new group of peptidoglycan hydrolases that cleave the carbohydrate backbone of PG (i.e. peptidoglycan glycosidases or muramidases) rather than the peptide component targeted by the amidases. They identified three families of Tge proteins, Tge1-Tge3, (type VI secretion glycoside hydrolase effectors) along with corresponding immunity proteins (Tgi1-Tgi3) (Whitney *et al*., 2013).

1.6.3. T6SS lipase effectors

A recent study identified a divergent superfamily of bacterial phospholipases which includes enzymes with phospholipase A1 and A2 activity. These enzymes are T6SS effectors that mediate antagonistic bacterial interactions (Russell *et al*., 2013). Bioinformatics analysis identified 377 putative lipases which can be divided into five different families called type VI lipase effectors 1-5 (Tle1-5) (Russell *et al*., 2013). The first four families (Tle1-4) share the GXSXG motif which is widespread in esterases and lipases, while the Tle5 family have the dual HXKXXXXD motifs present in phospholipase D enzymes (Aloulou *et al*., 2012). A member of the GXSXG family was identified in *B. thailandensis* (BTH_I2698, Tle1[BT]) (Russell *et al*., 2012). A subsequent study showed that the BTH-I2699 protein of *B. thailandensis* is an immunity protein that protects the host against Tle1[BT] and was named as type VI secretion lipase immunity 1 (Tli1[BT]) (Russell *et al*., 2013). A Tle2 lipase enzyme was identified in *V. cholerae* (VC1418, Tle2[VC]) (Russell *et al*., 2013). Analysis of the biochemical activity of Tle1[BT] and Tle2[VC] demonstrated that they act specifically as a phospholipase A2 (PLA2) and a phospholipase A1 (PLA1), respectively (Russell *et al*., 2013). The Tle2[VC] has also been investigated by Dong group, who reffered to it as TseL (Dong *et al*., 2013). TseL was shown to act as an effector in prokaryotic and eukaryotic cells. The killing effect *V. cholerae* towards *Dictyostelium discoideum* amoebae was demonstrated to require the action of both TseL and VasX (Section 1.6.4) as they function synergistically (Dong *et al*., 2013). A Tle5 family member was identified in *P. aeruginosa* as PldA (PA3487) and named as Tle5[PA] (Russell *et al*., 2013). Tle5[PA] was shown to specify a eukaryotic-like phospholipase D activity (Wilderman *et al*., 2001). *P. aeruginosa* expresses PA3488 protein in the periplasm which acts as a cognate immunity protein for Tle5[PA]. Therefore, it was assigned the name Tli5[PA] (Russell *et al*., 2013).

1.6.4. VasX

VasX (VCA0020) was identified in *V. cholerae* as a virulence factor regulated by the transcriptional regulatory protein VasH and contributes toward the killing of *D. discoideum* (Miyata *et al*., 2011). A recent study on T6SS effector and immunity proteins in *V. cholerae* identified the function of VasX as a T6SS-dependent effector protein that targets the bacterial cell wall. Moreover, the study identified the corresponding immunity proteins, TsiV2, for VasX required for the protection against

the attack of sister cells. A *vasX* and *vgrG3* double mutant as well as a *vasX vgrG3 tseL* triple mutant were unable to secrete TssD which suggest that they play a role in the assembly of a functional T6SS (Dong *et al*., 2013).

1.6.5. EvpP

Zheng and Leung identified a small protein that is secreted by the T6SS. In their comprehensive study on the T6SS components in *E. tarda,* they found that EvpP secretion is dependent on 13 Evp proteins, including Hcp (TssD) and VgrG (TssI), whereas an *evpP* mutant is able to secrete TssD and TssI proteins but not EvpP. Importantly, the *evpP* mutant was shown to be non-virulent (Zheng and Leung, 2007). These authors also showed that EvpP interacts with TssD inside and outside the bacterial cell which suggests that this effector passes through the TssD tube.

1.6.6. Rhs

The *rhs* genes are a family of horizontally acquired genes which was first identified in *E. coli* and shown to be widely distributed among Gram-negative bacteria. In *E. coli* it was observed that the recombination of distinct *rhs* genes leads to chromosomal rearrangement, and they were therefore called <u>r</u>earrangement <u>h</u>ot <u>s</u>pot (rhs) (Lin *et al*., 1984). The size of these genes ranges from 2-12 kilobase pairs (kb) and they consist of long core sequences which are GC-rich and highly conserved and short tip sequences which are low in GC content and poorly conserved. Similarly, Rhs proteins are predicted to include two domains: an N-terminal large core domain which contains a various number of tyrosine-aspartate (YD) repeats and a C-terminal small tip domain (Jackson *et al*., 2009). The *rhsT* gene of *P. aeruginosa*, which encodes a toxic protein, is induced by contact with phagocytic cells. Translocation of RhsT into phagocytic cells induces inflammasome-dependent cell death. As RhsT protein lacks the N-terminal signal sequence, is predicted to be secreted by a Sec-independent secretion system (Kung *et al*., 2012). A recent study in *Dickeya dadantii* 3937 identified two Rhs proteins (RhsA and RhsB) that carry nuclease domains involved in the degradation of target cell DNA and therefore inhibit the growth of neighbouring bacteria. The study also identified the role of the T6SS in secretion of Rhs proteins by construction of a mutant lacking *tssI* genes from $\Delta rhsA/rhsB^+$ inhibitor cells. The mutant lost the ability to outcompete other bacteria. A similar result was also observed with a mutant lacking the

tssD gene. These results indicate that Rhs proteins are effector proteins secreted via the T6SS (Koskiniemi *et al*., 2013).

The newly identified T6SS PAAR proteins which bind to the tip of VgrG and serve as the membrane piercing tool were suggested to play a role in the secretion of RhsA and RhsB. Shneider *et al* identified PAAR repeat regions in RhsA and RhsB and predicted that these two T6SS nuclease effectors bind to PAAR proteins and therefore can be secreted into target cells (Shneider *et al*., 2013).

1.7. Mechanism of effector protein secretion

Although information is being rapidly generated concerning effectors of the T6SS, it is still unclear by what mechanism non-VgrG effectors are recognised and secreted by this system. Based on the T6SS subunit structure information, the current model of the T6SS as an inverted bacteriophage tail on the surface of the Gram-negative bacteria translocates the VgrG-based effectors located at the tip of this structure into target cells (Ma *et al*., 2009a, Basler *et al*., 2012). For non-VgrG effectors, other mechanisms must be used in order to translocate these effectors into target cells. For instance, the Tse proteins of *P. aeruginosa* (Hood *et al*., 2010, Russell *et al*., 2011) are not VgrG orthologues. A different secretion mechanism is proposed for these small effectors based on the fact that the inner channel of the TssD tube ($\sim$ 4 nm) can allow translocation of folded proteins with a molecular weight less than 50 kDa (Mougous *et al*., 2006, Silverman *et al*., 2012). Therefore, the known T6SS amidase effectors of *P. aeruginosa* and *S. marcescens* can be transported by this mechanism as their sizes would allow them to fit into the Hcp channel (Dong *et al*., 2013). A recent study on *P. aeruginosa* demonstrates a direct and specific interaction of Tse2 and TssD1 of the H1-T6SS which helps in the stability and export of the effector protein. Also, Tse1 and Tse3 were shown to require a direct interaction with TssD1 for their secretion. Therefore, TssD was proposed to function as a chaperone and receptor for T6SS effectors (Silverman *et al*., 2013).

However, the molecular weights of some known *V. cholerae* T6SS effectors such as VasX or TseL are much larger. Therefore, these effectors cannot be transported through the lumen of the TssD tube in their folded state (Dong *et al*., 2013). Dong *et al*. demonstrated a direct interaction between the TseL and VgrG3 effectors of *V. cholerae*

and also showed that secretion of TseL depended on the presence of VgrG3. Therefore, they suggested another secretion mechanism for secretion of T6SS effectors that does not require transit through the lumen of the TssD tube. They proposed that other T6SS effectors might bind to a cognate VgrG orthologue or certain heterotrimers of VgrG and are secreted in the form of an effector-VgrG complex (Dong *et al*., 2013). A recent study by Shneider and colleagues proposed two new mechanisms for the secretion of effectors by the T6SS that depend on PAAR protein. In some cases they suggest that the effector may attach to this protein by non-covalent binding, whereas in other cases the effector may be fused to the PAAR protein as an N- or C-terminal domain extension (Shneider *et al*., 2013).

1.8. Aims and Objectives

1. To identify protein-protein interactions between the *tss* gene products of T6SS from *B. cenocepacia* and *A. hydrophila* by the bacterial adenylate cyclase two-hybrid system and perform confirmatory biochemical experiments.

2. To characterise the T6SS subunit protein TssA.

Chapter 2. Materials and Methods

2.1. Bacterial strains and plasmids

The bacterial strains and plasmids used in this study are shown in Tables 2.1 and 2.2

Table 2.1. Bacterial strains used in this study

Strain	Genotype	Reference
E. coli		
JM83	F⁻ *ara* Δ (*lac-proA*B) *rpsL* (Sm^r) Φ80 d*lacZ*ΔM15 Hsd$^+$	**(Yanisch-Perron *et al*., 1985)**
C41(λDE3)	Spontaneous mutants of BL21(λDE3)	(Miroux *et al*., 1996)
C43(λDE3)	Spontaneous mutants of BL21(λDE3)	(Miroux *et al*., 1996)
MC1061	hsdR *araD*139 Δ(*ara-leu*)$_{7697}$ Δ*lac*X74 *galU galK rpsL*(Sm^r)	(Casadaban and Cohen, 1980)
XL1 blue	*recA*1 *endA*1 *gyrA96 thi*-1 *hsdR17* (r$_k$⁻ m$_k$⁺) *sup*E44 *relA1 lac* [F' *proAB lac*I^{q}ZΔM15::Tn10(Tc^r)]	Stratagene
NEB Express	*fhuA2 [lon] ompT gal sulA11 (mcr-73::miniTn10 (Tc^S)2 [dcm] (zgb-210::Tn10 (Tc^S) endA1 Δ(mcrC-mrr)114::IS10*	Enhanced BL21 derivative created by New England Biolabs.
BTH101	F⁻ *cya-99 araD139 galE15 galK16 hsdR2 mcrA1 mcrB1 rpsL1* (Sm^r)	karimova *et al.,* 1998
B. cenocepacia		
715j	CF isolate, prototroph	(McKevitt *et al*., 1989; Darling *et al*., 1998)
H111	CF isolate, prototroph	(Romling *et al*., 1994; Gotschlich *et al*., 2001)
J2315	CF isolate from patient in Edinburgh Hospital, ET12 lineage	**(Isles *et al*., 1984, Holden *et al*., 2009)**

Abbreviations used within this table: Sm^r, streptomycin resistant; Tc^r, tetracycline resistant; Tc^s, tetracycline sensitive and CF, Cystic Fibrosis.

Table 2.2. Plasmids used in this study

Plasmid	Description	Source
pET14b	*E. coli*-specific expression vector for producing N-terminal His-tagged protein	Novagen
pET14b-His$_6$.TssA	pET14b containing *tssA* cloned between the *Nde*I and *Bam*HI sites (Ap^r)	(S.Shastri, 2011)
pET14b-TssA	pET14b containing *tssA* cloned between the *Xba*I and *Bam*HI sites (Ap^r)	This study
pET14b-His$_{10}$.TssA	pET14b containing *tssA* and 10 histidine tag cloned between the *Nco*I and *Nde*I site (Ap^r)	This study
pET14b-TssA.His$_{10}$	pET14b containing *tssA* cloned between the *Xba*I and *Bam*HI sites (Ap^r)	This study
pET14b-His$_6$.TssA.NTD	pET14b containing *tssA.NTD* cloned between the *Nde*I and sites *Bam*HI (Ap^r)	This study
pET14b-His$_6$.TssA.CTD	pET14b containing ~250 bp the C-terminal end of TssA cloned between the *Nde*I and *Bam*HI sites (Ap^r)	(S.Shastri, 2011)
pET14bΔT7	*E. coli*-specific expression vector with deleted T7 promoter	This study
pET14b-TssD	pET14b containing *tssD* cloned between the *Nde*I and *Bam*HI sites (Ap^r)	(S.Shastri, 2011)
pET14b-His$_6$.TssD	pET14b containing *tssD* cloned between the *Nco*I (*Psi*I) and *Bam*HI sites (Ap^r)	(S.Shastri, 2011)
pET14b-TssE	pET14b containing *tssE* cloned between the *Nco*I (*Bsp*HI) and *Xho*I sites (Ap^r)	(S.Shastri, 2011)
pACYCDuet-1	Expression vector containing two MCS each proceeded by a T7 promoter/*lac* operator and RBS. p15A origin of replication.	Novagen
pACYCDuet-TssA	pACYCDuet-1 containing *tssA* cloned between the *Bam*HI and *Nco*I sites (Cm^r)	This study
pACYCDuet-linkerHis1.TssA	pACYCDuet-1 containing *tssA* cloned between the *Bam*HI and *Nco*I sites (Cm^r)	This study
pACYCDuet-linkerHis2.TssA	pACYCDuet-1 containing *tssA.CTD* cloned between the *Bam*HI and *Kpn*I sites (Cm^r)	This study
pACYCDuet-His$_6$.TssA.CTD-S.tag	pACYCDuet-1 containing *tssA* cloned between the *Nco*I and *Bam*HI sites (Cm^r)	This study
pGS301	pTrc99A containing *B. pseudomallei fur* gene (Ap^R)	(Lowe *et al.*, 2001)

pACYCDuet-His$_{10}$.TssA	pACYCDuet-1 containing *tssA* cloned between the *Nco*I and *Bam*HI sites (Cm^r)	This study
pACYCDuet-His$_6$.TssE	pACYCDuet-1 containing *tssE* cloned between the *Bam*HI and *Hin*dIII sites (Cm^r)	This study
pACYCDuet-His$_6$.TssE-ΔT7.2	pACYCDuet-His$_6$.TssE containing deletion of the second T7lac promoter	This study
pACYCDuet-His$_6$.TssE-TssA	pACYCDuet-His$_6$.TssE containing *tssA* cloned in second MCS between the *Nde*I and *Bgl*II sites (Cm^r)	This study
pACYCDuet-TssE.His$_6$-TssA	pACYCDuet-TssA containing DNA fragment encoding TssE with C-terminal His-tag cloned between *Bsp*HI/*Nco*I and *Bam*HI sites (Cm^r)	This study
pACYCDuet-VSVg.TssD	pACYCDuet-1 containing DNA fragment encoding TssD with an N-terminal VSV-G tag cloned between the *Nco*I and *Bgl*II sites (Cm^r)	This study
pACYCDuet-VSVg.TssD-linkerHis1.TssA	pACYCDuet-1 containing DNA encoding TssD with N-terminal VSV-G tag cloned between the *Nco*I and *Bam*HI sites and TssA with a hexahistidine tag in the linker region, cloned between the *Nde*I and *Bgl*II sites, and (Cm^r)	This study
pACYCD-VSVg.TssD$_m$	pACYCDuet-1 containing DNA encoding TssD$_m$ with an N-terminal VSV-G tag cloned between the *Nco*I and *Bgl*II sites (Cm^r)	This study
pACYCD-VSVg.TssD$_m$-linkerHis1.TssA	pACYCDuet-1 containing DNA encoding TssD$_m$ with N-terminal VSV-G tag cloned between the *Nco*I and *Bam*HI sites and TssA with a hexahistidine tag in the linker region, cloned between the *Nde*I and *Bgl*II sites, and (Cm^r)	This study
pUC19-gp27gp5	pUC19 containing DNA fragment encoding gp27gp5-like region of TssI BCAM0148 cloned between the *sma*I (Ap^r)	This study
pACYCDuet-gp27gp5	pACYCDuet-1 containing DNA fragment encoding gp27gp5-like region of TssI BCAM0148 cloned between the *Nco*I and *Bam*HI sites (Cm^r)	This study
pACYCDuet-VSVg.TssH	pACYCDuet-1 containing DNA fragment encoding TssH with an N-terminal VSV-G tag cloned between the *Nco*I and *Bgl*II sites (Cm^r)	This study
pACYCDuet-VSVg.TssH-linkerHis1.TssA	pACYCDuet-1 containing DNA encoding TssH with N-terminal VSV-G tag cloned between the *Nco*I and *Bam*HI sites and TssA with a hexahistidine tag in the linker region, cloned between the *Nde*I and *Bgl*II sites, and (Cm^r)	This study This study

pACYCDuet-VSVg.TssI-linkerHis1.TssA	pACYCDuet-1 containing DNA encoding TssI BCAM0148 with N-terminal VSV-G tag cloned between the *Nco*I and *Bam*HI sites and TssA with a hexahistidine tag in the linker region, cloned between the *Nde*I and *Bgl*II sites (Cm^r)	This study
pMAL-c5X	*E. coli* vector containing ampicillin resistance genes (*bla*) and *malE* gene (encoding MBP) under control of the *tac* promoter (Amp^r)	New England Biolabs
pMAL-c5X-TssA.CTD	pMALc5X containing a DNA fragment encoding TssA.CTD cloned between the *Nde*I and *Bam*HI sites (Amp^r)	This study
pUT18	Cloning and expression vector; encodes the T18 fragment (amino acids 225–399 of CyaA) with a MCS at the N-terminal end of T18 (Ap^r)	Karimova *et al*., 1998)
pUT18C	Cloning and expression vector; encodes the T18 fragment (amino acids 225–399 of CyaA) with a MCS at the C-terminal end of T18 (Ap^r)	Karimova *et al*., 1998)
pKT25	Cloning and expression vector, encodes the T25 fragment (amino acids 1–224 of CyaA) with a MCS at the C-terminal end of T25 (Km^r)	Karimova *et al*., 1998)
pKNT25	Cloning and expression vector; encodes the T25 fragment (amino acids 1–224 of CyaA) with a MCS at the N-terminal end of T25 (Km^r)	Karimova *et al*., 1998)
pKT25-Zip	A derivative of pKT25 in which the leucine zipper of GCN4 (1) is genetically fused in frame to the T25 fragment (Km^r)	Karimova *et al*., 1998)
pUT18C-Zip	A derivative of pUT18C in which the leucine zipper of GCN4 (1) is genetically fused in frame to the T18 fragment (Ap^r)	Karimova *et al*., 1998)

Abbreviations used within this table: Tp^r, trimethoprim resistant; Km^r, Kanamycin resistant; Ap^r, Ampicillin resistant; Cm^r, Chloremphenicol resistant.

2.2. Bacteriological techniques

2.2.1. Bacteriological media

The following bacterial culture media were used routinely throughout this project

Luria-Bertani (LB) broth

Tryptone (Difco)	10 g
Yeast extract (Difco)	5 g
NaCl	10 g
H_2O	made up to 1 litre and autoclaved

Luria-Bertani (LB) Agar

LB agar plates were made by adding 15 g/litre of agar to LB medium before autoclaving. When the agar has cooled to 50-60°C appropriate antibiotics and supplements were added and poured into 90 mm Petri-dishes. The agar plates were left to set and either used immediately after drying or stored at room 4°C future use.

MacConkey-maltose agar

40 g of MacConkey agar base (Difco) was dissolved in 1 litre of distilled water and autoclaved. A stock solution of glucose-free maltose (20% in water) was sterilized by filtration. For MacConkey-maltose agar, maltose (1% final concentration) as well as antibiotics (ampicillin at 100μg/ml and kanamycin at 50μg/ml) was added to the autoclaved MacConkey medium just before pouring the plates. IPTG (isopropyl-β-D-thiogalactopyranoside, 0.5 mM) can be included in the medium to induce full expression of the hybrid proteins.

Brain-heart infusion (BHI) broth

43 g of BHI broth powder was added to 1 litre of distilled water. The solution was mixed well to dissolve the powder and was then distributed into aliquots and sterilized by autoclaving at 15 lb sq.in^{-1} for 20 minutes.

Auto-induction (AI) medium

This medium was used for inducing protein overproduction from vectors that employ theT7 promoter system without the need for IPTG. anhydrous D-glucose, lactose (Fisher Scientific UK) and glycerol (Fisher Scientific UK) were added to a final

concentration of 0.05%, 0.2% and 1%, respectively to 150 ml of LB broth in a glass bottle and topped up to 200 mls with LB broth.

2.2.2. Media Supplements
IPTG

Isopropyl β-D-thiogalactoside was dissolved to 1 M in distilled H_2O, filter sterilized and stored at –20°C.

Carbon sources

Maltose was prepared as a 20% (w/v) solution in deionized water and filter sterilized.

Antibiotics

To ensure that only strains containing the correct plasmid were present on plates or in the medium, antibiotics were added whenever appropriate.

Antibiotic stock solutions

Ampicillin (Ap) – 100 mg ml^{-1} in distilled water, filter sterilized and stored at -20°C.
Kanamycin (Km) – 50 mg ml^{-1} in distilled water, filter sterilized and stored at -20°C.
Chloramphenicol (Cm) – 25 mg ml^{-1} in 100% ethanol, and stored at -20°C. Antibiotic were used at 1000-fold dilution in media.

2.2.3 Maintenance of bacterial strains (glycerol stocks)

0.7 ml of an overnight culture grown in LB was mixed with 0.3 ml sterile 50% glycerol in a cryogenic vial, to give a final concentration of ~15 % glycerol. The vial containing the culture was vortexed briefly and stored frozen at –80 °C.

2.3. DNA preparation and purification techniques

2.3.1. Genomic DNA preparation

Genomic DNA from *B. cenocepacia* strains 715j, H111 and J2315 and *A. hydrophila* strain ATCC 7966 were prepared using the Nexttec DNA isolation system according to the manufacturer's instructions. Purified genomic DNA was dissolved in 100 µl of EB and stored at -20°C.

2.3.2. Crude boiled lysate preparation of chromosomal DNA

A crude genomic DNA preparation was prepared by suspending a loop of bacteria in 200 µl of TE buffer contained in tube and the suspension was heated at 100°C for 5 minutes in boiling water bath. The suspension was then centrifuged for 1 minute at 13,200 rpm and the supernatant was transferred into a new microcentrifuge tube.
TE buffer was 10 mM Tris-HCl (pH 8.0); 1 mM EDTA.

2.3.3. Plasmid DNA preparation

Small scale plasmid DNA preparations (plasmid minipreps) were generally performed using the alkaline lysis procedure incorporating phenol chloroform step. However, when DNA was required for nucleotide sequence determination, the minicolumns procedures was used.

2.3.3.1. Alkaline lysis followed by phenol-chloroform method
Stock solutions for miniprep procedure
Solution 1

- 50 mM glucose
- 25 mM Tris-HCl (pH 8.0)
- 10 mM EDTA (pH 8.0)

This was autoclaved and stored at 4 °C. This solution was prepared using sterile (autoclaved) stock solutions of 1 M Tris-HCl (pH 8.0) and 0.5 M EDTA (pH 8.0). 1 M Tris-HCl (pH 8.0) was made by adding 12.1 g Tris base in 70 ml water and adjusted to (pH 8.0) using HCl on magnetic stirrer. This was then made up to 100 ml by adding water, then autoclaved. EDTA (pH 8.0) was prepared by adding 3 g of EDTA in 70 ml water, adjusting to (pH 8.0) using 10 M NaOH and then making the volume up to 100 ml by adding water following which the solution was autoclaved.

Solution II

- 0.2 M NaOH
- 1% (w/v) sodium dodecyl sulphate (SDS)

This solution was stored at room temperature without autoclaving.

Solution III

- 60 ml 5M potassium acetate
- 11.5 ml acetic acid, glacial
- 28.5 ml H_2O (>18 MΩ)

The solution was autoclaved and stored at 4 °C.

Miniprep procedure

Bacterial cultures were inoculated into 2 ml LB, and incubated overnight in the shaker at 37^oC. 1.5 ml of each culture was poured into a microcentrifuge tube, and the cells were collected by centrifugation at 13,000 rpm for 5 minutes in a bench top microcentrifuge. The supernatant was discarded and the cell pellet thoroughly resuspended in 100 μl solution I and kept on ice for 5-10 min. Then, 200 μl solution II was added and the microcentrifuge tubes were inverted 3-4 times gently and stored on ice for 5 min. 150 μl solution III was added and the tubes were shaken following which they were kept on ice for a further 5 min. After this time the tubes were centrifuged at 13,000 rpm for 5 minutes in a bench top microcentrifuge.

Following centrifugation, the supernatant was transferred to a clean microcentrifuge tube. 400 μl of phenol-chloroform was added and the sample vortexed briefly and centrifuged at 13,000 rpm for 3 minutes in a bench top microcentrifuge. The upper, aqueous, phase was then transferred to microcentrifuge tubes, and the nucleic acids were precipitated using 2 volumes (~ 800 μl) absolute ethanol.

The sample was vortexed briefly and left on the bench to precipitate the nucleic acids for at least 5 min. Following this, the sample was centrifuged at 13,000 rpm for 5 minutes in a bench top microcentrifuge. After that the ethanol was removed leaving the

nucleic acid to which pellet and 1ml of 70% ethanol was added. After brief vortexing, the samples were re-centrifuged for 5 min, the 70% ethanol was completely removed from the microcentrifuge tube, and it was then left open on the bench to air dry for about half hour. The dried nucleic acid pellet consisting mainly of plasmid DNA and stable RNA was resuspended in 50 μl water and analysed on an agarose gel.

In order to remove the contaminating RNA from the sample, 1 μl of 1 mg ml^{-1} RNAse A was added to the 50 μl plasmid solution and this was incubated at 37 °C for 1 hour.

2.3.3.2. Alkaline lysis followed by minicolumn method

Plasmid extractions from cells (mini preps) were carried out using the Qiagen mini prep kit per the manufacturer's instruction (only for DNA sequencing). DNA purification after digestions, or post gel extraction was also carried out using the Qiagen PCR purification kit.

2.3.4. DNA amplification by the polymerase chain reaction (PCR) for cloning

PCRs contained the following components (stock concentrations shown in parenthesis):

- 3 μl genomic DNA/boiled lysate
- 5 μl KOD 10x buffer
- 4 μl MgSO$_4$ (25 mM)
- 5 μl dNTPs (2 mM each dNTPs)
- 3 μl 10 μM forward primer
- 3 μl 10 μM reverse primer
- 2.5 μl DMSO
- 0.5 μl KOD DNA polymerase (1u/μl)
- 24 μl distilled water

50 μl PCR sample made up in 0.5 ml PCR tube.

2.3.4.1. PCR regime

- 95°C Temperature step for 5 minutes. After 3 minutes, cycler was paused and KOD DNA polymerase was added. This "hot start" is to increase the specificity of the primers for their specific target sequences on the genomic DNA.

After the initial temperature step, the following cycle was repeated 30 times.

- Denaturation at 95°C for 30 seconds
- Annealing for 30 seconds. Temperature varies depending on the length and G+C content of the primer. The annealing temperature was calculated using the formula ((A+T) x 2) + ((G+C) x 4)-5°C.
- Elongation (polymerisation) at 72°C. The duration of step depended on the size of DNA fragment being amplified. In order to polymerise 1 kilobase (kb) of DNA, 1 minute at 72°C was allowed.

2.3.4.2. DNA amplification by PCR for recombinant plasmid screening ("colony PCR")

Screening of potential clones to determine if they contained the correct DNA insert was conveniently performed by PCR screening. The PCR was carried out as described in section 2.3.4, but the purified genomic DNA or boiled cell lysate was replaced by a small quantity of cells from transformant colonies. Individual colonies were picked using a sterile toothpick which was used to inoculate a PCR tube followed by patching onto an agar plate containing the appropriate antibiotic. Taq polymerase master mix containing primers specific for the gene inserted, or regions flanking the cloning site in the plasmid, was then added to each PCR tube containing colonies to be screened along with positive and negative controls.

2.4. DNA manipulation techniques

2.4.1. DNA sequencing

DNA polymerase can introduce mutation in the amplified DNA sequences, all plasmid constructs containing PCR-amplified inserts were sequenced with appropriate primers to check the sequence of the entire insert. Sequence changes were confirmed as strain to strain sequence variations or the result of mutations by sequencing the PCR product of a separate amplification. Plasmid DNA for sequencing was prepared using the QIAprep Spin Miniprep Kit (Qiagen). DNA sequencing was carried out by Lark technologies at the University of Sheffield Medical School Core Sequencing Facility.

2.4.2. Restriction endonuclease digestion of DNA

Restriction digestions were performed using enzymes and buffers purchased from Promega corporation. Reactions consisted of the following components in a final volume of 50 μl.

DNA   10 μl for PCR product or 15 μl plasmid miniprep

Restriction enzyme buffer (10x)   5 μl

Bovine serum albumin (BSA)   0.5 μl (recommended)

Enzyme   1 μl

H$_2$O up to   50 μl

For diagnostic restriction digestion the final volume of the reaction was 30 μl. Volumes of the buffer and water were adjusted accordingly. Digestions were left for 2 hours at 37°C after which the digested DNA was purified using the Qiagen PCR purification protocol, if it was to be used for subsequent enzymatic manipulation such as DNA ligation.

2.4.3. Annealing complementary oligonucleotides

For the purpose of cloning very small DNA fragments complementary single-stranded oligonucleotides were annealed that generated double-stranded DNA fragment with restriction site-compatible sticky ends. To do this 45 μl each of 100 pmol ml^{-1} single stranded complementary oligonucleotides in milliQ water and 10 μl of ×10 annealing buffer (10 mM MgCl$_2$, 200 mM Tris-Cl (pH 8.0) in milliQ water) were combined into 0.5 ml microcentrifuge tube and incubated at 90 ºC for 10 minutes and then left at room temperature to cool for 30 minutes. 1 μl of this mixture of annealed oligonucleotides was used in the ligation reaction.

2.4.4. Ligations

Ligations were performed using Promega T4 DNA ligase and ligation buffer. The following were added to a microcentrifuge tube:

- 10 μl digested plasmid DNA
- 10 μl digested PCR product
- μl ligase buffer (10x)
- 1 μl DNA ligase (3u/μl)

Made up to 30 µl with distilled water. Ligations were incubated at room temperature overnight.

A vector control (no PCR product or ligase) and ligation control (no PCR product) were included to judge the effectiveness of the ligations and the digestions.

2.4.5. Transformation of *E. coli* with plasmid DNA
2.4.5.1. Preparation of competent cells

To prepare competent cells, the method of Hanahan was used (1985) 0.5 ml of fresh overnight culture of *E. coli* was inoculated into 50 ml LB broth and grown at $37^{o}C$ with shaking to exponential phase (OD_{600}=0.3-0.5). The culture was chilled on ice for 10-15 min, and bacteria were harvested by centrifugation at 4,000 rpm for 10 min at $4^{o}C$. The supernatant was discarded and the cell pellets were resuspended by moderate vortexing in 16 ml ice cold RF1 solution. The cells were incubated on ice for 30 minutes prior to centrifugation for 10 minutes at 4,000 rpm, at 4 °C. The cell pellets were resuspended in 4 ml of solution RF2 and kept on ice for 15 min. The resuspended competent cells were aliquoted into microcentrifuge tubes and frozen immediately at $-80^{o}C$. As prepared, competent cells were stored for a year without significant loss of transformation frequency.

Solution RF1

KCl	3.73g
$MnCl_2.4H_2O$	4.95g
Potassium acetate	1.47g
$CaCl_2.2H_2O$	0.75g
Glycerol	76 ml

The above were dissolved in ~300 ml H_2O (>18 MΩ), and the resultant solution was adjusted to pH 5.8 using 0.2 M acetic acid. The final volume was then made up to 500 ml. The solution was sterilised by filtration through a nitrocellulose 0.45 µm membrane, and stored at 4 ˚C.

Solution RF2

- Solution A: 0.5 M MOPS, (pH 6.8)
- Solution B: 10 mM KCl

                            75 mM $CaCl_2.2H_2O$

                            15% (w/v) glycerol

Both solutions A and B were made up in >18 MΩ H_2O, and stored at 4 ˚C. 0.2 ml of solution A and 9.8 ml of solution B were mixed together on the day of use.

2.4.5.2. Transformation procedure

1 μl of plasmid DNA or 15 μl of ligation reaction to be used for transformation was chilled on ice for ~10 minutes prior to addition of 100 μl competent cells to each sample. A cell control without addition of any DNA was also prepared. The samples were then incubated on ice for 30 minutes whereupon the tubes were heat-shocked at 42°C for 2.5 minutes and returned to ice for 5 minutes. 1 ml LB broth was then added to each mixture and the tubes were incubated at 37°C for 1 hour to allow expression of antibiotic resistance genes carried by the plasmid.

100 μl of each transformed culture was spread on selective agar medium and the plates were incubated at 30 or 37°C as appropriate and for the recommended duration. The remainder of the transformation mixture was kept in the fridge at 4°C for a few days in case more transformants were required. In such cases the remainder of the culture was harvested by centrifugation at 13,000 rpm for 30 seconds in a microcentrifuge. Most of the supernatant was drained from each tube, leaving ~ 100 μl broth into which the cell pellets were resuspended and then spread onto a separate agar plate.

2.4.6. Agarose gel electrophoresis

Gel electrophoresis was used qualitatively to check DNA preparations, and quantitatively for determination of relative insert and vector concentrations in cloning reactions. It was also used to separate and purify specific DNA fragments for cloning. All agarose gels were prepared in 1xTAE buffer and heated to boiling for 1-2 min in a microwave oven (normally 0.8% (w/v) agarose was used. However, a higher percentage (1.0-1.5%) was used to resolve bands of <500 bp). The molten agarose was cooled and poured into the casting tray of a mini-gel tank (Bio-Rad minigel

system), a comb appropriate for the number and volume of samples was inserted to form the DNA loading wells and the agarose left to set. After removing of the gel comb and sealing tape, the gel was placed in the gel tank and submerged in 1xTAE buffer and the sample; containing 2 µl of 2X DNA loading solution, 8 µl water and 5 µl DNA, was loaded into each well. A marker lane was always included which contained either a supercoiled DNA ladder (Promega) or Q-Step 4 Quantitative DNA ladder marker (Yorkshire Bio). Electrophoresis was carried out at 80 V for at least 60 min. Following electrophoresis the gel was transferred to a tray, covered with ethidium bromide (0.1µg/µl) and shaken for 15-20 minutes to allow intercalation of the ethidium bromide into the DNA within the gel. After rinsing the gel with tap water, the gel was visualized on a UV trans-illuminator and photographed.

50 X TAE buffer

- 242 g   Tris-base
- 57.1 ml   glacial acetic acid
- 37.2 g   di-sodium EDTA dihydrate

- H_2O (>18 MΩ) to 1 L

The final pH of 50 xTAE is 8.

10x DNA Loading Dye:

50 % (v/v) glycerol in milliQ H_2O, 0.1 % (w/v) bromophenol blue. Diluted to 2x using further milliQ water when required.

2.5. Protein purification and analysis techniques

2.5.1. Protein analysis techniques

2.5.1.1. Growth of bacterial cultures and induction of protein overproduction

E. coli BL21(λDE3) cells harbouring plasmids expressing cloned target genes were grown at 37 ˚C in LB broth to an OD_{600} of 0.5, in the presence of the appropriate antibiotic. 1 ml of the culture was taken immediately before addition of IPTG and left to grow for a further 3 hours. IPTG was added to the rest of the culture to a final concentration of 1 mM to induce expression of the cloned gene. Cells were incubated for a further 3 hours. Samples were taken from the pre- and post-induction culture for

analysis by SDS-PAGE. 100 µl of each sample was centrifuged at 13,200 rpm for 3 min, the supernatant was discarded and the cell pellet was resuspended in 50 µl of lysis buffer and mixed with 50 µl of 2X Laemmli sample buffer and boiled for 10 minutes to denature and solubilise the proteins. Laemmli sample buffer consisted of:

Laemmli sample buffer 2X:
- 250 mM Tris-HCl (pH 6.8)
- 10 % SDS
- 20 % glycerol
- % Bromophenol blue

The above components were made upto 99 ml with H_2O. For each 9.9 ml, 0.1 ml β-mercaptoethanol was added.

2.5.1.2. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)
Gel preparation

SDS polyacrylamide gels were prepared according to the procedure described by Laemmli (1970). The resolving gel was poured between two glass plates separated by 0.75 mm spacer upto ~ 2.0 cm from the top to allow for the stacking gel. The 10 ml resolving gel can be used to pour 2 gels. A few drops of butanol were added to the top of the poured resolving gel to maintain a level surface before polymerisation. The gel was then left to polymerise for 20-30 minutes and the butanol was then washed off the surface using water.

12 % resolving gel (10 ml)
- 4.3 ml H_2O (>18 MΩ)
- 3 ml   40 % acrylamide:bis-acrylamide (Fisher scientific)
- 2.5 ml   1.5 M Tris-HCl (pH 8.8)
- 100 µl   10 % (w/v) SDS
- 100 µl   10 % ammonium persulphate
- 5 µl   TEMED (Sigma)

15 % resolving gel (10 ml)

- 3.5 ml H_2O (>18 MΩ)
- 3.75 ml   40 % acrylamide:bis-acrylamide (Fisher scientific)
- 2.5 ml   1.5 M Tris-HCl (pH 8.8)
- 100 μl   10 % (w/v) SDS
- 100 μl   10 % ammonium persulphate
- 5 μl   TEMED (Sigma)

The stacking gel was applied over the resolving gel and the gel comb was immediately inserted in the gel. After polymerisation, the comb was removed and the gel was placed into the gel tank filled with 1X SDS running (reservoir) buffer.

5 % stacking gel (4.5 ml)

- 3.2 ml H_2O (>18 MΩ)
- 625 μl   40 % acrylamide:bis-acrylamide (Fisher scientific)
- 625 μl   1 M Tris-HCl (pH 6.8)
- 50 μl   10 % (w/v) SDS
- 50 μl   10 % ammonium persulphate
- 5 μl   TEMED (Sigma)

10 X SDS running buffer

- 30.3 g   Tris base
- 144 g   glycine
- 10 g   SDS
- Up to 1 L H_2O (>18 MΩ)

2.5.1.3. Electrophoresis of the gel

The protein samples for SDS analysis were prepared as described above. Samples were 10-15 μl loaded in each well and a protein molecular weight marker was loaded in one well. Then the SDS-PA gel was electrophoresed at 120 V for approximately 2 hrs (until the bromophenol blue dye had migrated to the bottom of the gel). When the gel was to be used for western blotting, a prestained molecular weight standard was used.

2.5.1.4. Staining of polyacrylamide gels

Gels were routinely stained with Coomassie brilliant blue R-250 on a gently shaking platform for ~2 hrs or overnight at room temperature.

Staining solution

- 40%   Methanol
- 10% (v/v)   Glacial acetic acid
- 50% dH_2O
- 0.2%   Coomassie brilliant blue R-250

2.5.1.5. Destaining of polyacrylamide gels

Gels were destained by 3-4 washes (~100 ml) on a gently shaking platform for 1-2 hrs at room temperature until the gel background became clear.

Destaining solution

- 10%   Acetic acid
- 40%   Methanol
- 50% dH_2O

2.5.2. Solubility test of the overproduced protein

To test the solubility of the overproduced proteins, 250 ml of BL21(λDE3) liquid culture containing an expression vector such as pET14b were grown and induced as described in section 2.5.1.1. The cell culture was then centrifuged at 12,500 xg for 20 minutes at 4˚C in a 6 x 250 rotor and the cell pellet was resuspended in 25 ml of buffer containing:

Wash buffer

- 25 mM Tris
- 130 mM NaCl
- 2 mM EDTA

Resuspended cells were re-centrifuged at 12,500 xg for 20 minutes at 4˚C and 1 g of cell pellet was resuspended in 5ml of lysis buffer:

Lysis buffer

- 50 mM Tris-HCl (pH 8.0)
- 2 mM EDTA (not used if protein was subsequently being purified by IMAC)
- 200 mM NaCl
- 10 % glycerol
- 10 mM imidazole (only used if protein was subsequently being purified by IMAC).

To the cell suspension 0.1 ml of freshly made 10 mg ml^{-1} lysozyme was added with stirring for 20-30 minutes at 4˚C. While stirring, PMSF was added to a final concentration of 25 µg ml^{-1} from a freshly made 2.5 mg ml^{-1} stock solution in ethanol or acetone and sodium deoxycholate was added to a final concentration of 500 µg ml^{-1} from a 25 mg ml^{-1} solution, and the suspension was left stirring for a further 20-30 minutes.

Following this the BL21(λDE3) cells were subjected to sonication for 4-5 times for 30 seconds with 2 minute intervals on ice in between each sonication. 50 µl of this crude suspension was removed and boiled with 50 µl of 2 x Laemmli buffer for 10 minutes. The remaining culture was centrifuged for 30 minutes at 20,000 rpm and 50 µl of supernatant was removed and boiled with 50 µl of 2 x Laemmli buffer for 10 minutes. The four samples (uninduced cells, induced cells, crude lysate, and cleared lysate) were then analysed by SDS PAGE using a Bio Rad mini protean II system at 120 volts. If the desired protein was found to be soluble, then the growth and induction procedure was scaled up to 1 litre. The washed cell pellet was stored frozen at -20 ˚C and a small amount was processed to obtain the soluble fraction on the day of protein purification. Only 5 ml of cleared cell lysate was used each time.

2.5.3. Immobilized metal affinity chromatography (IMAC)

The soluble cell lysate containing the protein of interest was prepared as described in Section 2.5.2. A 1 ml Hi-Trap nickel column from GE Healthcare was attached to an AKTA protein purification system and equilibrated with 10 ml of lysis buffer. The lysed cell supernatant containing the target protein was sterilised by filtration through a 0.22 µm filter using 10 ml syringe and then applied to the nickel column at a flow rate of 1

ml/min. The flow-through from the column was collected and in some situations the flow-through was recycled for further 2 times. After the entire sample had been loaded, the column was washed with at least 20 ml of lysis buffer. The target protein was eluted by a 10-500 mM imidazole gradient applied over the column in a 50 ml volume elution buffer which contained:

- 50 mM Tris-HCl (pH 8.0)
- 200 mM NaCl
- 10 % glycerol

The elution of the target protein was monitored at 280nm, due to absorbance by tryptophan and tyrosine side chains. The eluted protein was collected in 1 ml fractions and 50 μl of each fraction was taken and boiled for 10 minutes with 50 μl of 2 x Laemmli buffer. 15 μl of each sample was then analysed by SDS PAGE. The protein concentrations of the samples corresponding to the UV peak were also determined using the Bradford protein assay.

Regenerating the nickel resin column

The 1 ml nickel affinity column was washed with 10 ml 20% ethanol after finishing the protein purification and stored at 4ºC for future use. In order to completely remove residual protein material from the column, it was washed with 5 ml 50 mM EDTA (pH 8.0) followed by 5 ml 500 mM guanidine hydrochloride that completely strip the column by removing all the nickel along with any residual proteins. The column was then washed with water several times and then re-charged with 5 ml 100 mM NiCl.

2.5.4. Anion exchange chromatography

Following IMAC, the protein solution was loaded into a dialysis bag with appropriate molecular wt. cut off and the bag was sealed on both sides. The protein solution was dialysed against 100 volumes of 20 mM Tris (pH 8.0) for 2 h then the dialysis buffer was changed and dialysis continued with stirring at 4ºC overnight. The dialysed partially purified protein was applied onto an anion exchange chromatography column for further protein purification. The chromatography matrix of the Q column employed has positively charged moieties covalently linked to it which will interact with negatively charged proteins. A mono Q column from GE-Healthcare was used for anion exchange chromatography. The column was equilibrated with low salt buffer before

applying the dialysed partially purified protein to the column at a flow rate of 1 ml min^{-1}. The flow-through from the column was collected and then the column was washed with 20 ml of low salt buffer:

Low salt buffer for anion exchange chromatography
- 50 mM Tris-HCl (pH 8.0)
- 20 mM NaCl
- 10 % glycerol

The target protein was eluted with a 20-1000 mM NaCl gradient applied to the column in a 30 ml volume of buffer containing 50 mM Tris-HCl (pH 8.0) and 10% glycerol. Elution was followed at 280nm. Samples from each 1 ml fraction was then analysed by SDS PAGE to check the purity of the eluted protein.

2.5.5. Amylose column chromatography of MBP fusion proteins

To purify maltose-binding fusion protein (MBP), the plasmid encoding the gene of interest (pMAL c5x) was transformed into the NEB Express cells (New England Biolabs) and the culture was processed as described in Section 2.5.2. The only difference was the use of column buffer instead of lysis buffer.

Column buffer:
- 20 mM Tris-HCl (pH 7.4)
- 200 mM NaCl
- 1.0 mM EDTA
- 1.0 mM sodium azide
- 11 mM β-mercaptoethanol

The above components were dissolved in 800 ml milliQ water, topped up to 1L with water and autoclaved for 20 minutes.

5 ml of the soluble fraction of the cell lysate containing the MBP fusion protein was applied onto a 2 ml amylose column (New England Bio Labs) pre-equilibrated with 10 ml column buffer. The flow-through was collected and the column washed with 20 ml column buffer. Then, the MBP fusion protein was eluted with the column buffer

containing 10 mM maltose. Ten 3 ml fractions were collected and analysed by SDS-PAGE.

Regenerating the amylose resin

After each use the amylose column was regenerated with the following sequence of washes:

Water: 3x column volumes

0.1% SDS: 3x column volumes

Water: 1x column volume

Column buffer: 3x column volumes

Following this, the amylose column was stored at 4ºC for future use.

2.5.6. Size exclusion chromatography (gel filtration)

Size exclusion chromatography (SEC) was carried out using a Superose 6 or 12 column (GE healthcare), according to the desired separation capacity of the column, by attaching it to an AKTA protein purification system. The column was equilibrated with the following buffer:

- 50 mM Tris-HCl (pH 8.0)
- 200 mM NaCl
- 10 % glycerol

This was followed by loading the column with 100 μl of concentrated protein sample via a 100 μl superloop and the flow rate was adjusted to 1 ml min^{-1} across the column. The flow-through from the column was monitored at 280nm. Samples collected from the size exclusion column were analysed by SDS PAGE.

The SEC column was also calibrated by known molecular weight protein standards when the aim of the SEC was to estimate the size of the target protein. 2 mg of each protein standard (ribonuclease (13.7 kDa), carbonic anhydrase (29 kDa), ovalbumin (43 kDa), conalbumin (75 kDa), alcohol Dehydrogenase (149.5 kDa), thyroglobulin (669 kDa), apoferritin (443 kDa) and the polysaccharide blue dextran (2000 kDa) (GE Healthcare)) were mixed and dissolved in equilibration buffer and concentrated to 100

μl before applying to the SEC column via a 100 μl superloop. The elution time of the proteins was used to create a calibration curve using graphical software. From this curve it was possible to estimate the approximate mass of target proteins eluted from the SEC column by measuring the time at which they elute.

2.5.7. Bradford assay

The protein concentration in a sample was determined by the Bradford assay (Bradford, 1976). A stock solution of 10 mg ml^{-1} bovine serum albumin (BSA) was prepared by dissolving 50 mg of BSA in 5 ml of MilliQ H$_2$O. A sample of BSA at 1 mg ml^{-1} was prepared by dilution of the stock solution. The concentration of 1 mg ml^{-1} BSA sample was confirmed by measuring the sample at OD$_{280}$ in a quartz cuvette. At OD$_{280}$ the absorbance of a sample that is 1 mg/ml should be 0.667. This 1 mg ml^{-1} sample was used to prepare series of protein concentration of standards at 0.2, 0.4, 0.6, 0.8 and 1.0 mg ml^{-1} in water. A serial dilution of the unknown protein sample was also prepared. 20 μl of each protein concentration standard and unknown protein sample was added to 980 μl of Bradford reagent, mixed thoroughly in a microcentrfuge tube and left for 5 minutes before being transferred to a cuvette to measure the OD$_{595}$. A sample without any added protein was included as a blank. The measurements of the standards were used to create a calibration curve from which the protein concentration in the test sample was calculated. Dilutions of the protein sample were used if the absorbance of the original protein sample fell outside of the ranges used in the calibration curve.

2.5.8. Proteolytic cleavage of MBP from MBP fusion protein

 20 μl fusion protein at 1 mg ml^{-1} was mixed with 1 μl factor Xa diluted to 200 μg ml^{-1}. Another 5 μl of the fusion protein was placed in separate micro-centrifuge tube without factor Xa (negative control). As a positive control, 10 μl of MBP5-paramyosin-ΔSal (NEB) was mixed with 0.5 μl of diluted factor Xa. The three tubes were incubated at room temperature and 5 μl samples were taken at 2, 4, 8 and 24 hours from the reaction tube. The four test samples, the negative control and the positive control (both controls at 24 h) were mixed with an equal volume of 2x Laemmli buffer and boiled for 10 minutes. All samples were analysed by SDS-PAGE.

After determining the optimal time for the cleavage, the experiment was scaled up. The cleavage was carried out using factor Xa reaction buffer (recommended by NEB) which

contained final concentrations of 20 mM Tris-HCl (pH 8.0), 100 mM NaCl and 2 mM CaCl$_2$ in 500 ml milliQ water. Filter sterilised using a cellulose nitrate membrane filter (0.2μm, Whatman). A small amount of the undigested fusion was saved as reference. Samples should be checked for complete cleavage by SDS-PAGE.

2.5.9. Thrombin cleavage of His-tagged protein
Small scale optimisation

The thrombin cleavage kit was purchased from Novagen. The thrombin enzyme was first diluted in thrombin dilution/storage buffer. The Novagen protocol was used to estimate the appropriate range of enzyme: target protein. Serial dilutions of thrombin in thrombin dilution buffer were prepared as 1:25, 1:50, 1:100 and 1:200 containing ~ 0.04, 0.02, 0.01, and 0.005 U enzyme per μl, respectively. The following components were added to five microcentrifuge tubes:

- 5 μl 10X thrombin cleavage/capture buffer
- 10 μg target protein
- 1 μl diluted thrombin (each tube received 1 μl of a different enzyme dilution. The fifth tube received 1 μl dilution buffer only as a negative control)
- deionized water to make up the total volume to 50 μl

The five micro-centrifuge tubes were incubated at room temperature and 10 μl samples were taken from each tube after 2, 4, 8 and 16 h incubation. The samples were mixed with an equal volume of 2x Laemmli buffer and boiled for 10 minutes. The extent of cleavage was determined by analysing the samples by SDS-PAGE.

The reaction was scaled up proportionately using the appropriate enzyme:target protein ratio determined from the pilot experiment.

Biotinylated thrombin capture

Streptavidin agarose was used to quantitatively remove the biotinylated thrombin following the proteolytic cleavage of the target protein. For each unit of enzyme, 16 μl settled resin (32 μl of the 50% slurry) was used. The Streptavidin agarose beads slurry was resuspended gently by inverting the bottle several times. The desired volume of

agarose beads was added to the reaction mixture and incubated at room temperature for 30 minutes with gentle shaking. Then the mixture was transferred to the sample cup of a Spin Filter (supplied with the kit) and centrifuged at 500 rpm for 5 minutes. The filtrate in the collection tube contained the proteolytically released target protein, free of biotinylated thrombin.

2.5.10. Nickel affinity pulldown

At least three micro-centrifuge tubes were loaded with ~ 200 μl of nickel resin slurry and centrifuged for one minute at 500 rpm before the supernatant were discarded. 1 ml of low salt lysis buffer (50 mM Tris-HCl (pH 8.0), 50 mM NaCl and 10% glycerol) was added to each tube and mixed gently by inverting the tubes several times. The tubes were centrifuged again and the supernatant was replaced with 1 ml of fresh low salt lysis buffer followed by re-centrifugation and removal of the supernatant. This wash step was repeated two more times. To the resulting pellet of resin, 200 μl of protein sample in low salt lysis buffer (50 mM Tris-HCl (pH8.0), 50 mM NaCl, 10% glycerol and 10 mM imidazole) was added and incubated at room temperature with gentle mixing on a rotating wheel for 2 hours at room temperature. The resin was collected by centrifugation at 500 rpm for one minute and the supernatant was removed. Beads were washed by addition of 500 μl of fresh lysis buffer followed by centrifugation at 500 rpm for one minute and removal of the resulting supernatant. The wash step was then repeated twice. Bound proteins were eluted by the addition of 100 μl lysis buffer containing 500 mM imidazole and gentle mixing on a rotating wheel at room temperature for 30 minutes. The samples were centrifuged at 500 rpm and the supernatant was collected. Samples of the eluted proteins were boiled with an equal volume of 2x Laemmli buffer and analysed by SDS-PAGE and western blotting.

2.5.11. Co-immunoprecipitation procedure

The EZview Red Protein A Affinity Gel beads (Sigma-Aldrich) was used to carry out co-immunoprecipitation (Co-IP) experiments. The soluble fraction of cell lysate containing the target proteins was prepared as described in Section 2.5.2. 1 ml of the soluble fraction was pipetted into a 1.5 ml micro-centrifuge tube and mixed with 1-10 μl of an appropriate dilution of antibody. The mixture was then vortexed briefly and incubated at 2-8 °C with thorough, gentle mixing for 1 hour.

The bottle of EZview Red Protein A Affinity Gel beads was carefully mixed and aliquots of 30 µl of the 50% slurry were pipetted into a 1.5 ml micro-centrifuge tube held on ice. The beads were washed and calibrated in low salt lysis buffer (50 mM Tris-HCl (pH8.0), 50 mM NaCl, 10% glycerol and 10 mM imidazole) by adding 0.75 ml of the buffer to each tube followed by vortexing and centrifugation for 30 seconds at 8,200 rpm. The supernatants were removed and the tubes containing the pellet of beads set on ice. The wash step was repeated two more times and then the washed bead pellets were returned to ice. The cell lysate plus antibody mixture was transferred into the tube of washed protein A beads, mixed briefly and left to stir for 1 hour at 2-8 °C to allow antibody-antigen complexes to bind to the protein A on beads. The mixture was then centrifuged for 30 seconds at 8,200 rpm, the supernatant was removed with a pipette and the tube with the bead pellet was placed on ice. The bead pellets were washed with 0.75 ml of low salt lysis buffer by vortexing and incubated with thorough, gentle mixing at 2-8 °C for 5 minutes. After brief centrifugation, the supernatants were removed and the tubes containing the bead pellets set on ice. The wash step was repeated two more times. A 25 µl of lysis buffer was added to each bead pellet and vortexed briefly. An equal volume of 2x Laemmli buffer was then added to the samples which were boiled for 10 minutes to elute antibody-antigen complexes. For analysis of immunoprecipitated proteins by SDS-PAGE, the boiled samples were vortexed, and centrifuged for 30 seconds, whereupon the supernatants were transferred into clean micro-centrifuge tubes for subsequent analysis.

2.5.12. Western blot analysis

Protein samples were electrophoresed in SDS-PA gel along with EZ-Run Prestained Rec protein ladder as molecular weight markers (section 2.5.1.2). A piece of polyvinylidene fluoride (PVDF) membrane (GE Healthcare: Amersham Hybond™-P) of the same size of the gel was soaked in 100% methanol for 15 seconds and then washed in water for 2 minutes. The PVDF membrane, the polyacrylamide gel, two pieces of filter paper (Whatman 3M) the same size as the gel and 2 sponge pads were saturated in 1x transfer buffer (for the recipe see below). To assemble the Western blot stack, the gel was laid on filter paper set on a sponge pad and the PVDF membrane was placed on the top of the gel, then the second filter paper was placed on the membrane followed by the second sponge pad. Before adding the second filter paper on top, it was ensured that no air bubbles were present between the layers. The stack was placed

inside a transfer tank (BioRad) containing 1x western transfer buffer with 10% methanol and transfer was carried out at 100 V for 60 minutes. An ice cassette was placed inside the transfer tank for the duration of the transfer to keep it cool. The membrane was then blocked in Tris-buffered saline containing 0.05% (v/v) Tween® 20 (TBS-T) with 5% semi-skimmed milk powder at room temperature for 60 minutes. The membrane was then washed with TBS-T for 10 minutes on a shaker. This was repeated a further 2 times. The membrane was then probed with the appropriate primary antibody at the required dilution in TBS-T containing 5% (w/v) semi-skimmed milk powder. For better results the solution was poured into a 50 ml centrifuge tube, the membrane was inserted (protein side facing towards the inside of the tube) and the membrane incubated with the primary antibody overnight at 4°C with gentle agitation. To remove the unbound antibody, three successive room temperature washes (10 min each) were performed with TBS-T. The membrane was then probed with the appropriate secondary antibody at the required dilution for 60 minutes at room temperature in TBS-T containing 5% (w/v) semi-skimmed milk powder with gentle agitation. The membrane was then washed with TBS-T for 10 minutes by gentle agitation. The washing step was repeated a further 2 times. 1 ml of detection reagent (EZ-ECL) was made by combining 0.5 ml EZ-ECL solution A and 0.5 ml EZ-ECL solution B. This was applied to the PVDF membrane and incubated for 5 minutes at room temperature. The excess detection reagent was removed and the Western blot was imaged using the Bio-Rad molecular imager ChemiDoc™ XRS+.

10X TBS (pH 7.4)

- 15 g   Tris-base
- 40 g   NaCl
- 1 g   KCl
- H$_2$O (>18 MΩ) to 1 L

The final pH of 10 xTBS was adjusted to 7.4 with HCl

10X western transfer buffer (pH 8.3)

- 30.3 g Tris base
- 144 g glycine
- H$_2$O (>18 MΩ) to 1 L

The final pH of 10 x transfer buffer was adjusted to 8.3.

2.5.13. Ammonium sulfate precipitation

When high concentration of ammonium sulfate (AS) is present, proteins tend to aggregate and precipitate. In order to determine the suitable AS concentration to precipitate the target protein, different concentrations of AS were tested (5%, 10%, 15%, 20% and 30%). Before precipitation of the target protein with AS, nucleic acids was precipitated first by polyethyleneimine (PEI) (each target protein was precipitated by addition of 40 µl of PEI per ml of sample) and the protein resuspended in an equal volume of lysis buffer containing 1 M NaCl. Then the treated cell lysate containing the target protein was divided equally into five tubes and each tube received a certain concentration of AS and rotated gently for 10-15 minutes. The solution was centrifuged at 10,000 xg for 10 minutes, the supernatant was removed and the precipitate resuspended in 2 pellet volumes of buffer. Samples were analysed by SDS-PAGE.

2.5.14. Preparation of inclusion bodies

One litre of IPTG induced bacterial culture prepared as described in Section 2.5.1.1 was centrifuged at 9,000 rpm for 15 minutes, and the pellet resuspended in 20 ml 2% sodium deoxycholate solution. The suspension was then sonicated until the viscosity had decreased. Then the sonicated suspension was centrifuged at 30,000 x g for 30 minutes at 4 ˚C in a 6 x 250 rotor, to collect the inclusion bodies. The pellet, containing the inclusion bodies, was washed three times with wash buffer. After each wash the inclusion bodies were collected by centrifugation at 30,000 x g for 30 minutes at 4˚C. Then pellet was weighed and each 1 g of cell pellet resuspended in a 5 ml buffer containing 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, 10 mM imidazole, 10% glycerol and 8 M urea, and incubated for 1 hour at 4˚C with gentle shaking. A final clarifying centrifugation was carried out at 30,000 x g, for 30 minutes at 4 ˚C, and the supernatant, containing the solubilised inclusion body protein, was retained for analysis by SDS-PAGE.

2.6. Bacterial two-hybrid assay

The bacterial adenylate cyclase two-hybrid (BACTH) system was employed to study protein-protein interactions in the T6SS. The principle of the BACTH system relies on the interaction-mediated reconstitution of adenylate cyclase activity in *E. coli*. The BACTH assay utilises the fact that the adenylate cyclase domain from *Bordetella pertussis* Cya A protein consists of two complementary fragments, T25 and T18. When these fragments are physically separated, they are inactive. However, when they are fused to interacting polypeptides, the interaction of these hybrid proteins results in functional reconstitution between T18 and T25 which leads to cAMP synthesis. Cyclic AMP produced from the interacting fragments binds to the cAMP receptor protein (CRP) (also known as the catabolite activator protein (CAP)). The cAMP/CRP complex activates the promoters of reporter genes, including genes of the *lac* and *mal* operon required for lactose and maltose catabolism. The bacteria become able to utilise lactose or maltose as the unique carbon source, therefore they are easily distinguished on indicator media.

In order to generate fusion proteins, *tss* genes were inserted into the BACTH vectors pKT25 and pKNT25, which are kanamycin resistant and encode the T25 fragment with the multiple cloning site located at the C-terminal and the N-terminal coding regions of the T25 fragment, respectively, and pUT18C and pUT18, which are ampicillin resistant and encode the T18 fragment with the multiple cloning site located at the C-terminal and the N-terminal coding regions of the T18 fragment, respectively. The different antibiotic resistance-genes allow selection for two compatible BACTH plasmids within one individual *E. coli* cell.

The *E. coli cya⁻* strain BTH101 was co-transformed with the recombinant fusion plasmids and was plated on MacConkey-maltose agar (Section 2.2.1) containing both antibiotics. Plates were incubated at 30°C and transformant colonies observed after 72 and 120 hours incubation. The plate should contain reasonable number of individual colonies (less than 500 colonies). A Cya$^+$ phenotype yields purple *E. coli* colonies, and a Cya$^-$ phenotype gave rise to white colonies. A Cya$^+$ phenotype was scored according to the colour of the colonies compared to the positive control. All BACTH combinations were repeated three times to confirm the results.

2.7. Electron microscopy (EM) grid preparation and staining

The negative staining procedure for the protein samples was carried out using a uranyl formate stain. 0.75 % (w/v) uranyl formate stain was prepared by adding 37.5 mg uranyl formate powder to 5 ml boiling water, which was left to stir for 5 minutes. Then ~ 8 μl of 5 M NaOH was added and the mixture stirred for another 5 minutes. The uranyl formate solution was filter sterilised using a 0.22 μm syringe filter unit, and stored in a foil wrapped tube at 4°C.

Carbon-coated grids were glow discharged for 20-30 seconds inside the coating machine (Cressington 208carbon), whereupon 3-5 μl of protein sample (the concentration of protein samples which submitted to EM was about 0.5 mg/ml. Three dilution 1:10, 1:50 and 1:100 were prepared and examined for each sample) was applied to the discharged carbon grid for 1 minute. Excess protein sample was removed by blotting of the grid on clean filter paper. The grid was then washed twice using distilled water and once using uranyl formate stain, with the grid being blotted on filter paper between each wash. Then the grid was submerged in uranyl formate stain for 20 seconds and blotted to remove excess stain before drying it using a small vacuum pump. The prepared carbon grid was examined by the electron microscopy and stored in a grid box at room temperature for future use.

Chapter 3. Investigation of protein-protein interactions that occur within the *Burkholderia cenocepacia* type VI secretion system using the bacterial adenylate cyclase two-hybrid system (BACTH)

3.1. Introduction

The core components of the T6SS were identified by a bioinformatics analysis by Shalom and colleagues who implicated 13 proteins that were named TssA-TssM (Shalom *et al*., 2007). In support of this study, Zheng and Leung carried out a comprehensive mutagenesis study to identify T6SS subunit genes in *Edwardsiella tarda* and they found that 13 out of 16 genes in the *evp* T6SS cluster are crucial for the proper export of Hcp (EvpC), VgrG (EvpI) and a putative effector, EvpP (Zheng and Leung, 2007). These genes are *evpA-evpC*, *evpE-evpI* and *evpK-evpO*, and they correspond to *tssA-tssM*. These thirteen components of T6SS are proposed to interact in some way to assemble the T6SS complex. The bacterial adenylate cyclase two-hybrid (BACTH) system was previously used to identify and characterise protein-protein interactions between the TssA-TssG subunits of the T6SS of *B. cenocepacia* (S.Shastri, 2011). The aim of the work described in this chapter was to identify protein-protein interactions occurring between the TssH-TssM subunits of the T6SS, and between these subunits and TssA-TssG.

The principle of the BACTH system relies on the interaction-mediated reconstitution of bacterial adenylate cyclase activity in *E. coli*. It provides all the advantages of working with *E. coli*. The BACTH assay utilises the fact that the adenylate cyclase domain of the *Bordetella pertussis* cya A protein can be expressed as two complementary fragments, T25 and T18 (Karimova *et al*., 1998). The interaction between these two fragments leads to the synthesis of cAMP (Ladant, 1988). When these fragments are physically separated, they are inactive (Ladant and Ullmann, 1999). Based on these facts, the BACTH system was designed in an endogenous adenylate cyclase-deficient *E. coli* strain (Δ*cya*). If T25 and T18 fragments are fused to interacting polypeptides, X and Y, the interaction of these hybrid proteins will result in functional reconstitution of adenylate cyclase activity which leads to cAMP synthesis (Karimova *et al*., 2000). Cyclic AMP produced from the interacting fragments binds to the *E. coli* cAMP receptor protein (CRP), also called catabolite activator protein (CAP). The cAMP/CAP complex activates promoters of various genes involved in the metabolism of alternative carbon sources, some of which like the *lac* and *mal* operons required for lactose and maltose catabolism, act as reporter genes for the assay. The ability of the bacteria to utilise lactose or maltose as the unique carbon source can be easily distinguished on indicator or selective media (Figure 3.1).

The BACTH system consists of two pairs of compatible vectors. The advantage of the different plasmids is to allow construction of fusion proteins at the N- and C-termini of the T25 and T18 CyaA fragments. pKNT25 and pKT25 are plasmids that specify kanamycin resistance and encode the T25 fragment. The multiple cloning site (MCS) is engineered at the N-terminal of coding region of T25 in pKNT25 and C-terminal coding region of T25 in pKT25 to facilitate the insertion of the gene of interest in frame with the T25 coding sequence either at the N- or C-terminus, respectively. pUT18 and pUT18C express ampicillin resistance and encode the T18 fragment. The MCS is located upstream of the T18 coding sequence in pUT18 and downstream of this sequence in pUT18C. Therefore, the fusion protein can be created at the N- or C-terminus of the T18 fragment, respectively (Figure 3.2). The pUT18 and pUT18C plasmids are derived from ColE1 and are therefore compatible with the p15A derivatives pKT25 and pKNT25.

In addition to these four plasmids, the BACTH system provides two positive control plasmids pKT25-zip and pUT18C-zip which encode the leucine zipper region of the yeast transcription regulatory protein GCN4 inserted into the *Kpn*I site within the MCS of pKT25, and between the *Kpn*I and *Eco*RI sites within the MCS of pUT18C, resulting in production of T25 and T18 fused to a long (35 aa) leucine zipper (Karimova *et al*., 1998). These fusion proteins can interact due to the dimerization of the fused leucine zipper motifs resulting in strong Cya$^+$ phenotype when they are used to co-transform a Δ*cya E. coli* strain. The combinations of plasmids, both for controls and screening for interactions, used in this study are shown in Table 3.1.

The BACTH system employs two alternative *E. coli* reporter strains, which are the non-reverting adenylate cyclase deficient (*cya*) mutants BTH101 and DHM. The BTH101 strain was used in this study.
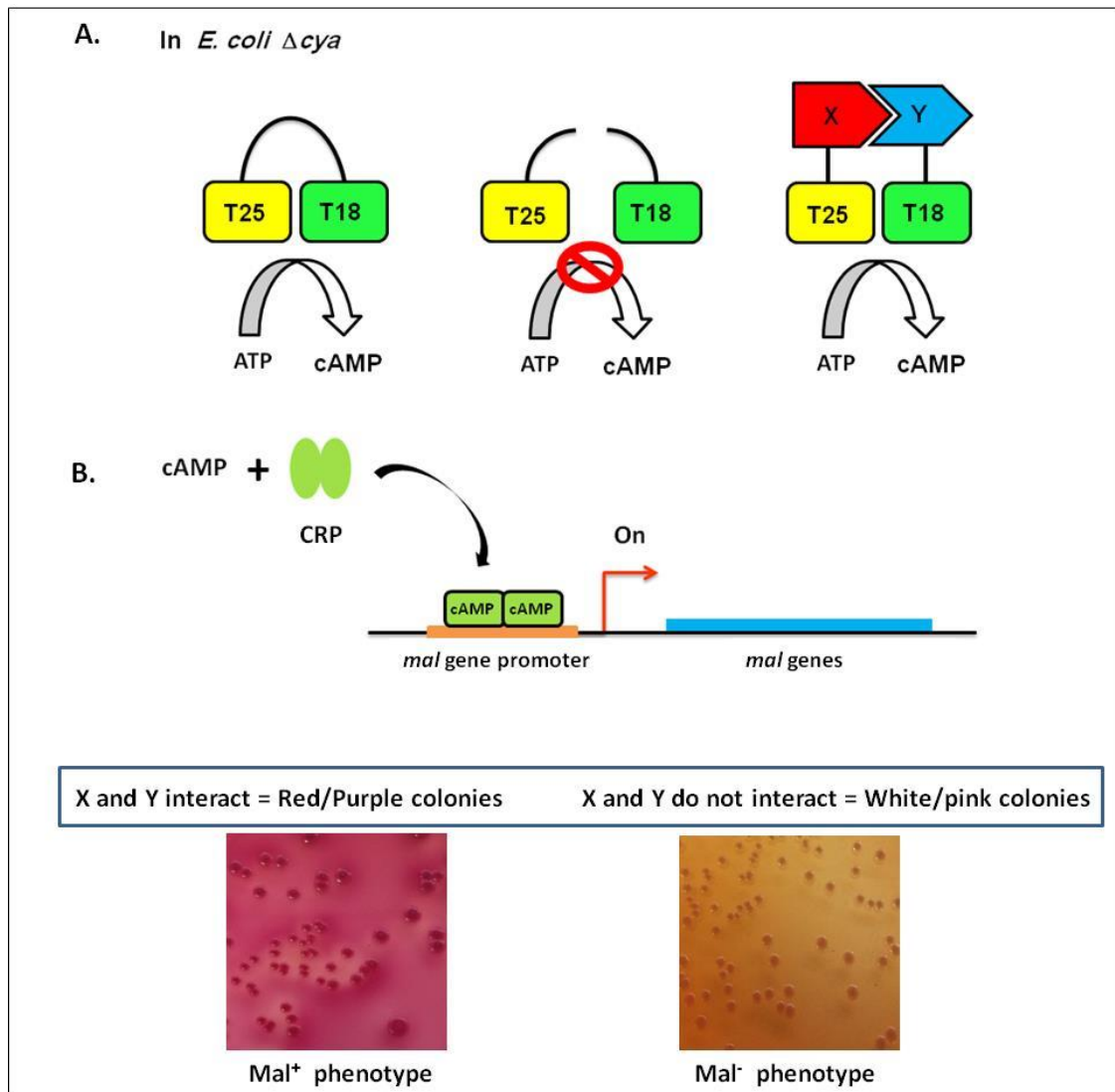
Figure 3.1 Principle of the BACTH system. A. Activation of the BACTH system. Interaction of T25 and T18 leads to production of cAMP. If the T25 and T18 components are physically separated, this will lead to the cessation of cAMP production. When proteins X and Y fused to the T25 and T18 components interact, the T25 and T18 domains are brought together which resumes cAMP production. B. cAMP produced by the reconstituted enzyme binds to catabolite activator protein (CRP) and turns on the expression of the *mal* operon involved in maltose catabolism. This facilitates the utilisation of maltose by the *E.coli* Δ*cya* host strain, which results in a Mal+ phenotype which is distinguishable when the bacteria are grown on indicator media such as MacConkey-maltose agar shown here.
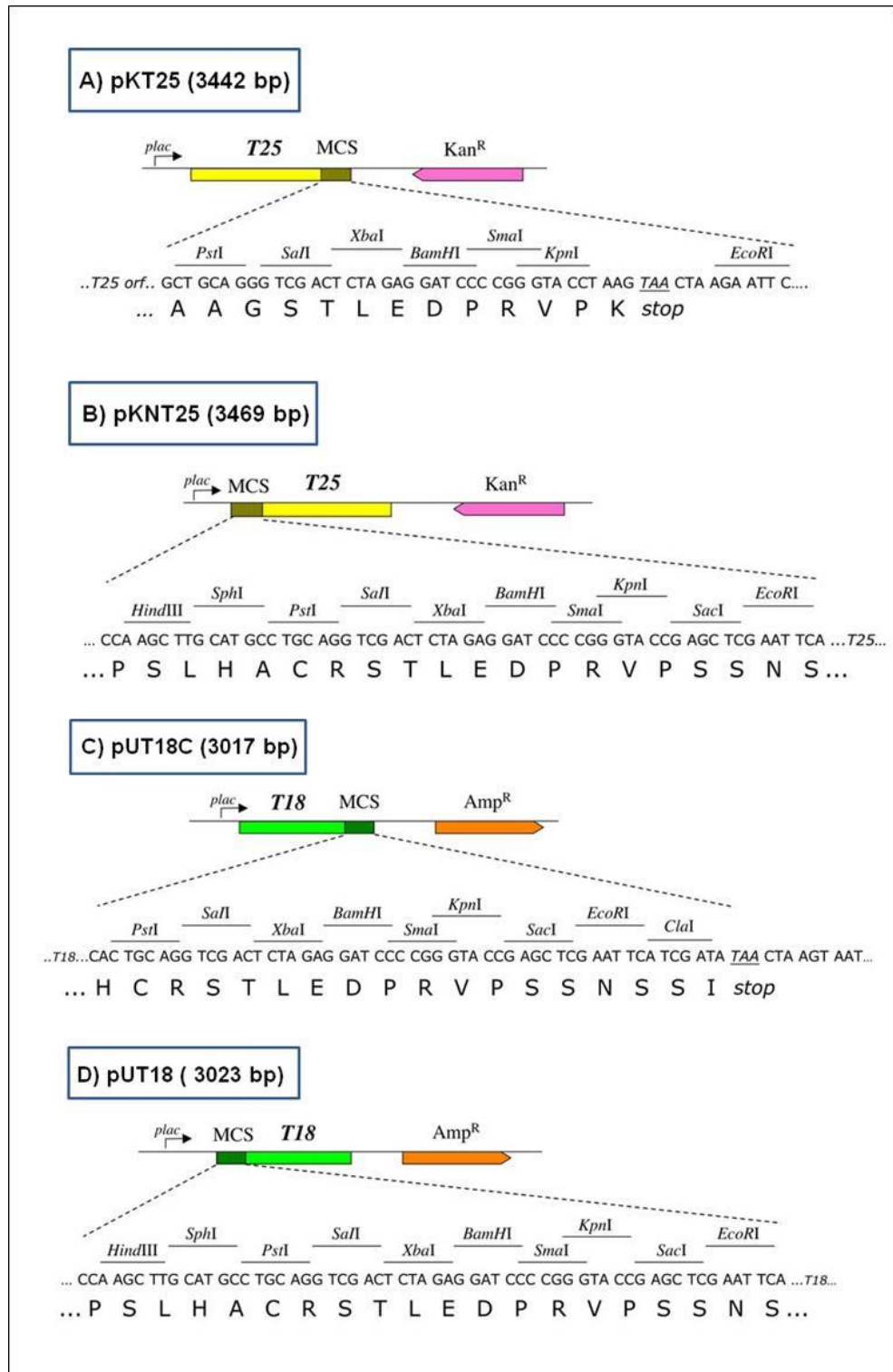
Figure 3.2 Diagram of the BACTH plasmids showing the MCS sequences. Plasmids are represented in linear form below which is presented the nucleotide sequences of the MCS of each vector showing the T18 or T25 reading frame. A) pKT25 = T25 coding sequence located upstream from the MCS; B) pKNT25 = T25 coding sequence located downstream from the MCS; C) pUT18C = T18 coding sequence located upstream from the MCS; D) pUT18 = T18 coding sequence located downstream of the MCS. Adapted from BACTH system manual (Euromedex, 2000) designed by Karimova *et al*. (1998).

Table 3.1 BACTH system plasmid combinations used to analyse protein-protein interactions

Plasmid encoding first *tss* gene	Plasmid encoding second *tss* gene	Expected Mal phenotype
A		
pKT25-gene X	pUT18-gene Y	Unknown
pKT25-gene X	pUT18C-gene Y	Unknown
pKNT25-gene X	pUT18-gene Y	Unknown
pKNT25-gene X	pUT18C-gene Y	Unknown
pUT18-gene X	pKT25-gene Y	Unknown
pUT18C-gene X	pKT25-gene Y	Unknown
pUT18-gene X	pKNT25-gene Y	Unknown
pUT18C-gene X	pKNT25-gene Y	Unknown
B		
pKT25-gene X	pUT18-gene X	Unknown
pKT25-gene X	pUT18C-gene X	Unknown
pKNT25-gene X	pUT18-gene X	Unknown
pKNT25-gene X	pUT18C-gene X	Unknown
C		
pKT25-gene X or Y	pUT18	Negative
pKT25-gene X or Y	pUT18C	Negative
pKNT25-gene X or Y	pUT18	Negative
pKNT25-gene X or Y	pUT18C	Negative
pUT18-gene X or Y	pKT25	Negative
pUT18C-gene X or Y	pKT25	Negative
pUT18-gene X or Y	pKNT25	Negative
pUT18C-gene X or Y	pKNT25	Negative
pKT25-zip	pUT18-zip	Strong positive
D		
pKT25	pUT18	Negative
pKT25	pUT18C	Negative
pKNT25	pUT18	Negative
pKNT25	pUT18C	Negative

A Plasmid combinations utilised to test hetero-oligomeric protein-protein interactions.
B Plasmid combinations to test for homo-oligomeric protein interactions.
C Control plasmid combinations.
D Empty plasmid combinations.

3.2. Cloning of the *Burkholderia cenocepacia tssH-tssM* genes into BACTH plasmids

PCR amplification of the *tssH* to *tssM* genes was carried out from chromosomal DNA of *B. cenocepacia* H111 using pairs of primers designed specifically to amplify each gene. In each primer, a cleavage site for a restriction enzyme was included to allow cloning of the amplicon into the MCS of each of the BACTH plasmids (see appendix 1 for the primer sequences).

In pKT25 and pUT18C, as mentioned above, the MCS is located at the 3' end of the T25 and T18 coding sequences, respectively, to allow construction of in-frame fusions at the C-terminal end of the T25 and T18 polypeptides. Therefore the primers were designed to ensure that the *tss* gene was in frame with the T18 and T25 coding sequence, and the stop codon of the *tss* gene was retained. On the other hand, in pKNT25 and pUT18, the T25 and T18 coding sequences are located downstream from the MCS. Therefore, each *tss* gene was cloned into these vectors so that it was in-frame with the translation initiation codon located upstream of the MCS and also with the T18 and T25 coding sequence downstream of the MCS, i.e. the *tss* gene stop codon was excluded. For this reason, every *tss* gene was amplified with two different pairs of primers with the same forward primer being used but the reverse primer that was used to produce C-terminal fusions contained a stop codon at the end of the *tss* gene whereas the reverse primer used to construct the N-terminal fusion was without a stop codon.

As shown in Figure 1.15 there is no copy of the *tssI* gene in the *B. cenocepacia* T6SS gene cluster, whereas there are 10 copies distributed around the genome. However, in several other members of the BCC, one or more copies of *tssI* are present in the corresponding T6SS gene cluster. BCAM0148 was chosen as the TssI orthologue for this analysis based on its similarity to the two *tssI* genes located in the corresponding T6SS gene cluster of *B. lata* strain 383 (a member of the BCC). BLASTP searches with each of them as the query, revealed BCAM0148 as the most homologous one in *B. cenocepacia* J2315.

Some Tss proteins contain N-terminal signal sequences or transmembrane domains (TMD) which may result in some protein domains being targeted to the periplasm. To ensure that the proteins remained in the cytoplasm, which is essential for the BACTH

assay, these regions were excluded from the amplified *tss* gene by making the forward primer complementary to sequences located downstream from the signal peptide and N-terminal TMD coding sequences and the reverse primer complementary to sequences located upstream of C-terminal TMD coding regions. This strategy was followed in the case of *tssJ*, which encodes an N-terminal signal peptide, and *tssL*, which encodes a C-terminal TMD (Durand *et al.*, 2012). In the case of the *tssM* gene, 3 TMD regions can be identified in the encoded protein, two near the N-terminus of the gene (which was avoided by placing the forward primer downstream of the sequence coding for the second TMD) and one in the middle of the protein. Therefore, TssM was analysed as two separate fragments, the cytoplemic NTD and the periplemic CTD, which were fused separately to T18 and T25.

Using chromosomal DNA of *B. cenocepacia* H111 strain the proofreading enzyme, KOD Hot Start DNA polymerase: was used to amplify the TssH, TssJ, TssK, TssL, TssM$_{NTD}$, and TssM$_{CTD}$ coding sequences. For amplification of *tssI*, no product was obtained from the standard PCR. Therefore, optimised PCR using a range of temperatures and different concentrations of magnesium sulphate was tried. From optimised PCR, again no product was obtained for *tssI*. Due to potential sequence differences between the *tssI* primers that were based on the J2315 genome sequence, and H111 *tssI* sequence, amplification of *tssI* gene was repeated using chromosomal DNA from *B. cenocepacia* strains 715j and J2315. The correct product of *tssI* gene was obtained from both strains. Therefore, it was decided to use the PCR product of *tssI* derived from 715j in this study.

DNA from each amplified *tss* gene was digested with two restriction enzymes corresponding to the restriction sites included in the forward and reverse primers. Simultaneously, DNA from each BACTH plasmid (obtained by plasmid mini preparation from the host *E. coli* strain MC1061 or JM83) was digested with the same pairs of restriction enzymes used to cut the *tss* genes. The ligation of *tss* genes with the BACTH plasmids were carried out as detailed in Section 2.4.4. *E. coli* strain MC1061 or JM83 was transformed with the ligation of the BACTH plasmids with *tss* genes. Transformants harbouring pKT25-tss and pKNT25-tss clones were selected on LB plates supplemented with kanamycin, while those containing pUT18C-tss and pUT18-tss clones were selected on LB plates supplemented with ampicillin. The transformants

were screened by colony PCR and/or manual plasmid mini-preparations to identify plasmids with required DNA inserts.

To confirm that the *tss* genes were cloned successfully into the BACTH plasmids, the correct sized clones for each gene-plasmid combination were digested with the appropriate restriction endonucleases and examined on an agarose gel. If the released fragment was the expected size, then the column plasmid mini-preparation procedure was performed to prepare higher quality DNA for nucleotide sequence determination. The nucleotide sequence of the insert was determined with the appropriate vector primers to confirm that no mutation in the gene of interest was introduced during PCR, and also to ensure that the gene was in frame with the T25 and T18 coding sequences. For the large inserts (*tssI, tssH,* and *tssM*$_{CTD}$) sequencing was completed using primer that annealed to sequences within the insert: tssINFOR, tssINREV, tssHINFOR, tssHINREV, tssM2INFOR, and tssM2INREV, respectively. 27 out of 28 possible clones were successfully obtained. pUT18-TssH was not constructed despite several attempts.

Appendix 2 shows a comparison of the nucleotide sequences of the *tss* genes in J2315 with the corresponding gene derived from H111 or 715j. The amino acid sequences of J2315 and H111 or 715j Tss proteins with the amino acid changes highlighted in different colours are shown in Appendix 3. Differences in nucleotide sequences between *tss* genes derived from J2315 and either H111 or 715j were deemed to be due to strain variation if they were observed in independent amplifications of the *tss* gene. If the mutation was present in one clone but not another, it was concluded to be due to a PCR-induced error and the clone was discarded.

3.3. Investigation into the homo-oligomerisation of TssH, TssI, TssJ, TssK, TssL, TssM$_{NTD}$ and TssM$_{CTD}$
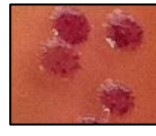
To assess the possibility of TssH, TssI, TssJ, TssK, TssL, TssM$_{NTD}$ and TssM$_{CTD}$ self-interaction all possible self-combinations of T25-Tss or Tss-T25 and T18-Tss or Tss-T18 fusion proteins were tested using the BACTH system (26 combinations). The analysis of TssH-TssH interactions revealed that both tested combinations of TssH fusion proteins gave rise to bright red/purple colonies on MacConkey-maltose plates after 72 hours incubation (figure 3.3A). The other two possible combinations were not tested because pUT18-TssH not constructed. The strong maltose-positive phenotype generated by the two fusion protein combinations suggests that TssH self-interacts to form an oligomer. Table 3.2 describes the system used to score the maltose phenotypes presented in the figures.

In the case of TssI, all four fusion protein combinations gave rise to a Mal$^+$ phenotype after 72 hours incubation (Figure 3.3B). The T25-TssI with T18-TssI combination showed a strong maltose-positive phenotype after 72 hours incubation. On the other hand, T25-TssI in combination with TssI-T18 required 120 hours to give rise to a strong maltose-positive phenotype. The other two combinations of TssI fusion proteins gave a weaker Mal$^+$ phenotype with patchy red colonies on MacConkey-maltose plates (colonies showed red (Mal$^+$) and white (Mal$^-$) areas which are termed "patchy"). These results are consistent with previous observation that TssI is organised as oligomer (Leiman *et al*., 2009, Hachani *et al*., 2011).

Regarding the analysis of TssJ for self-interaction, it was found that all fusion protein combinations yielded a maltose-negative phenotype, suggesting that TssJ does not form oligomers.

Three combinations of TssK fusion proteins gave rise to a Mal$^+$ phenotype (Figure 3.4A), whereas the combination of TssK-T25 with TssK-T18 gave rise to a Mal$^-$ phenotype. These results suggest that TssK also self-interacts.

Table 3.2 Description of the phenotypes observed in the BACTH assay and the classifications attributed to them.

Description of colony phenotype	Attributed rating *
Very strong Mal$^+$ phenotype giving rise to dark red/purple coloured colonies with purple halo around the colonies.	+++++
Strong Mal$^+$ phenotype giving rise to red/purple coloured colonies.	++++
Mal$^+$ phenotype giving rise to moderate red/purple coloured colonies.	+++
Mal$^+$ phenotype giving rise to weak red/purple coloured colonies	++
Mal$^+$ phenotype giving rise to very weak red/purple coloured colonies	+
Colonies showed red (Mal$^+$) and white (Mal$^-$) areas ("patches").	P

*Colony phenotype observed when *E. coli* strain BTH101 was transformed with compatible combinations of the BACTH plasmids and allowed to grow at 30°C for up to 120 h on MacConkey agar containing 1% maltose and the appropriate antibiotics.

Figure 3.3 Analysis of TssH and TssI self-interaction using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of TssH or TssI to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combination of TssH fusion proteins that gave rise to a maltose-positive phenotype. B. Combination of TssI fusion proteins that yielded a maltose-positive phenotype.

123

It was found that one combination of TssL fusion proteins, T25-TssL with T18-TssL, gave rise to red/purple colonies on MacConkey-maltose plates (Figure 3.4B). Also, one combination of TssM$_{CTD}$ fusion proteins (T25-TssM$_{CTD}$ with T18-TssM$_{CTD}$) gave rise to red/purple colonies on MacConkey-maltose plates (Figure 3.4C), whereas all the combinations of TssM$_{NTD}$ fusion protein gave rise to a maltose-negative phenotype. These results suggest that TssH, -I, -K, -L and TssM$_{CTD}$ self-interact to form higher order complexes.

Figure 3.4 Analysis of TssK, TssL and TssM$_{CTD}$ self-interaction using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of TssK, TssL or TssM$_{CTD}$ to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combination of TssK fusion proteins that gave rise to a maltose-positive phenotype. B. Combination of TssL fusion proteins that yielded a maltose-positive phenotype. C. Combination of TssM$_{CTD}$ fusion proteins that yielded a maltose-positive phenotype.

3.4. Investigation into interactions between different T6SS subunits

The BACTH system was used to screen for protein-protein interactions that occur between different T6SS subunits. As an analysis of the interactions between TssA to TssG has already been carried out, this analysis screened for interactions between TssH to TssM on the one hand, and TssA to TssM on the other. To do this, a pairwise interaction analysis was performed between Tss-T25 or T25-Tss and Tss-T18 or T18-Tss fusion proteins in all possible combinations.

3.4.1. Investigation of the interaction between TssH and TssA-TssG and TssI-TssM

3.4.1.1. Investigation of the interaction between TssH and TssA

For each Tss protein combination, eight combinations of fusion proteins were screened except for combinations involving TssH, where it was only possible to screen six fusion protein combinations due to the unavailability of the pUT18-tssH plasmid. Interestingly, two pairwise combinations of TssH and TssA fusion proteins gave rise to a strong Mal$^+$ phenotype indicating that TssH and TssA interact with each other. The Mal$^+$ phenotype was observed only when the C-terminal ends of both TssH and TssA in the fusion protein were free (Figure 3.5A). All other combinations of TssH and TssA fusion proteins gave rise to a Mal$^-$ phenotype.

3.4.1.2. Investigation of the interaction between TssH and TssB

Three out of six combinations of TssH and TssB fusion proteins gave rise to Mal$^+$ phenotype. In one of these combinations, T25-TssH and T18-TssB, a weak Mal$^+$ phenotype was observed. The appearance of the colonies on MacConkey-maltose plates was pink/red with a patchy appearance. The result of the reverse combination (T25-TssB and T18-TssH) was similar. A stronger Mal$^+$ phenotype was generated by the combination of TssB-T25 and T18-TssH, although again with patchy pink/red colonies (Figure 3.5B). A maltose-negative phenotype resulted from the remaining combinations of TssH and TssB fusion proteins.

3.4.1.3. Investigation of interaction between TssH and TssC

All combinations of TssH and TssC fusion proteins in the BACTH system gave rise to a Mal$^-$ phenotype.

3.4.1.4. Investigation of the interaction between TssH and TssD

Of the six possible TssH and TssD protein fusion combinations that could be screened, two yielded patchy pink/red coloured colonies (Figure 3.6A). It was observed that in order for these two proteins to interact, the C-terminal end of both proteins should be free (i.e. unfused).

3.4.1.5. Investigation of interaction between TssH and TssE

A very weak Mal^+ phenotype was observed when the plasmid encoding the TssE-T25 fusion was combined with the plasmid expressing the T18-TssH fusion in the same cell. Colonies exhibited a Mal^+ sectors within the colonies after 120 hours incubation. The rest of the fusion protein combinations gave rise to a Mal^- phenotype (Figure 3.6B).
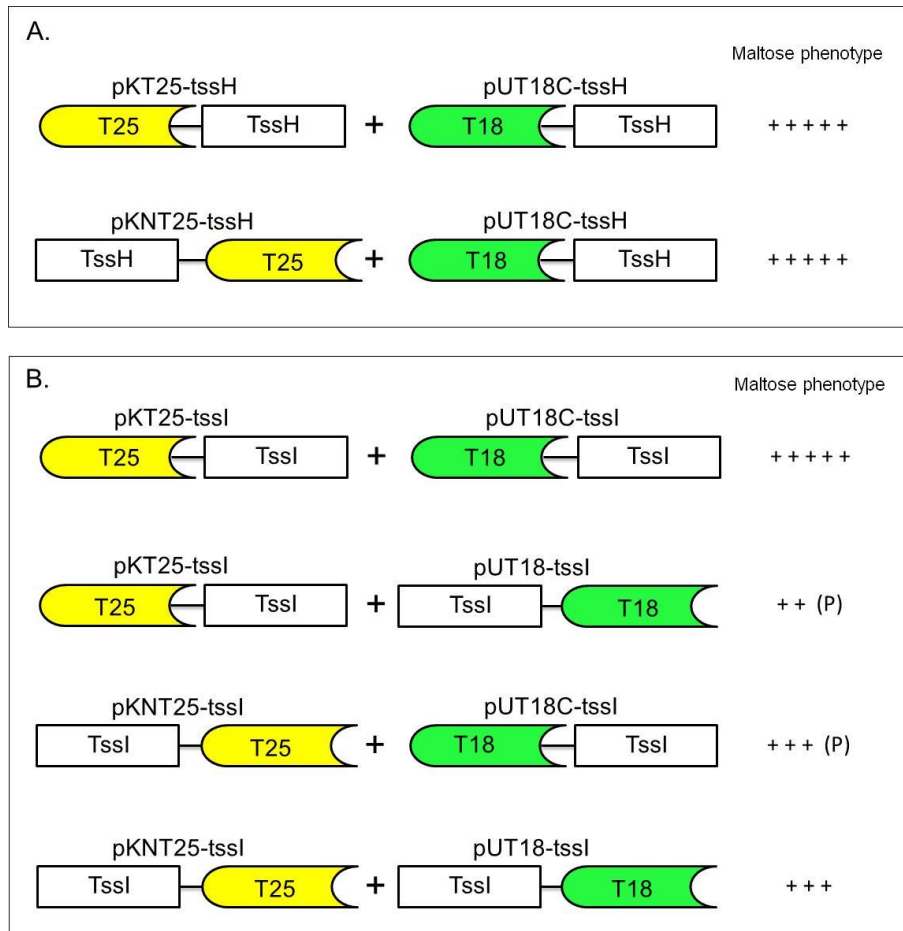
Figure 3.5 Analysis of TssH and TssA or TssB interactions using the BACTH assay.
Pairwise combinations of compatible BACTH plasmids encoding fusions of TssH and TssA or TssB to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combination of TssH and TssA fusion proteins that gave rise to a maltose-positive phenotype. B. Combination of TssH and TssB fusion proteins that yielded a maltose-positive phenotype.

3.4.1.6. Investigation of the interaction between TssH and TssF

Figure 3.6C shows that Mal$^+$ phenotypes were observed in three pairwise combinations of plasmids encoding TssH and TssF fusion proteins. A Mal$^+$ phenotype was found when either T25-TssH or TssH-T25 was combined with T18-TssF. Also, TssF-T25 in combination with T18-TssH gave rise to a Mal$^+$ phenotype. In all three cases, the interactions gave rise to mal$^+$ colonies with, a "patchy" appearance. All other combinations gave rise to a Mal$^-$ phenotype.

3.4.1.7. Investigation of the interaction between TssH and TssG

Mal$^+$ phenotypes were not observed when plasmids encoding TssH fusion proteins were combined with those encoding TssG fusion proteins for any of the plasmid combinations tested.

3.4.1.8. Investigation of the interaction between TssH and TssI

A moderate Mal$^+$ phenotype was obtained for two combinations of fusion proteins when the C-terminal ends of both TssH and TssI were free. These results suggest that the C-terminal end of both TssH and TssI is important for the putative interaction between the two fusion proteins. In addition, a weak Mal$^+$ phenotype was observed after 120 hours incubation when the TssI-T25 and T18-TssH fusion proteins were co-expressed. In all three combinations that gave rise to a Mal$^+$ phenotype, the colonies exhibited the patchy morphology (Figure 3.7). All remaining combinations gave rise to Mal$^-$ phenotype.
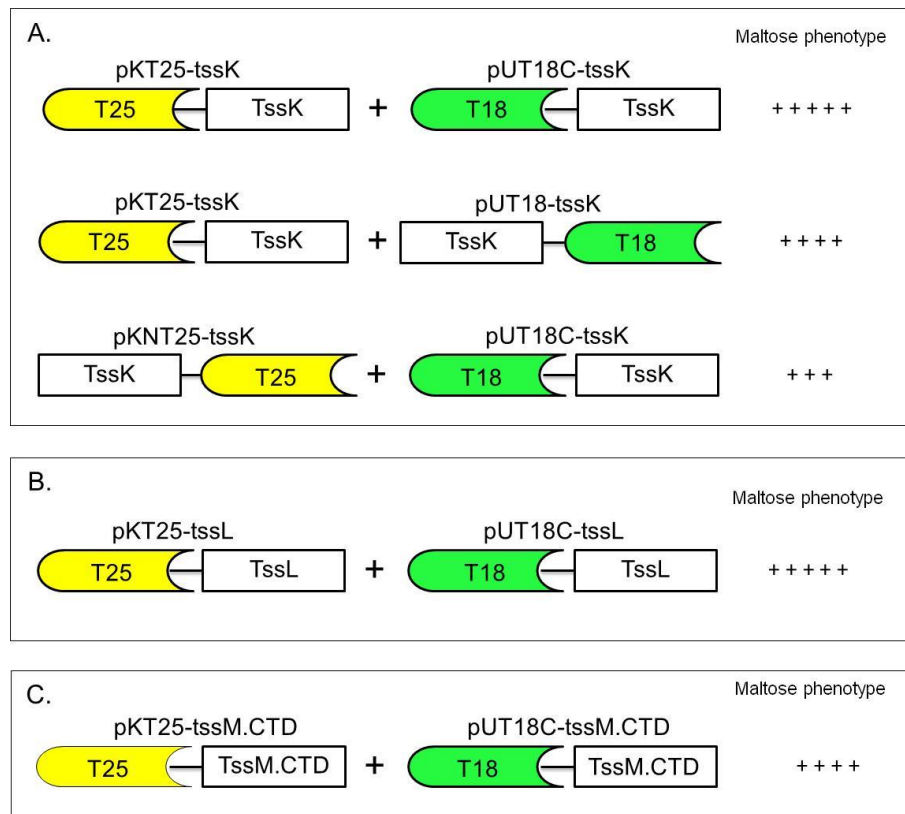
Figure 3.6. Analysis of TssH and TssD or TssE or TssF interactions using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of TssH and TssA or TssB to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combination of TssH and TssD fusion proteins that gave rise to a maltose-positive phenotype. B. Combination of TssH and TssE fusion proteins that yielded a maltose-positive phenotype. C. Combination of TssH and TssF fusion proteins that yielded a maltose-positive phenotype.

3.4.1.9. Investigation of the interactions between TssH and TssJ

Mal⁻ phenotypes was observed when plasmids encoding TssH were combined with those encoding TssJ in all plasmid combinations used. Therefore, no evidence of interaction between these two proteins was obtained using the BACTH system.

3.4.1.10. Investigation of the interactions between TssH and TssK

Mal⁺ phenotypes were observed when plasmids encoding TssH were combined with those encoding TssK for three of the six combinations tested (Figure 3.8A). T25-TssH in combination with T18-TssK, and T25-TssK combined with T18-TssH both produced a strong Mal⁺ phenotype with patchy red colonies. Colonies co-expressing the TssK-T25 and T18-TssH fusion proteins exhibited red patches after 120 hours of incubation.

3.4.1.11. Investigation of the interactions between TssH and TssL

Probing for interactions between TssH and TssL revealed a strong Mal⁺ phenotype between T25-TssH and T18-TssL with pink/red patchy colonies on MacConkey-maltose plates. Interestingly, the other possible combination where the C-terminus of each of the Tss proteins is free (T25-TssL and T18-TssH) gave rise to a weak Mal⁺ phenotype with a patchy appearance after 120 hours incubation (Figure 3.8B).

3.4.1.12. Investigation of interactions between TssH and TssM

The BACTH assay of TssH-T18 and -T25 fusion proteins in combination with $TssM_{CTD}$ and $TssM_{NTD}$ fusion proteins in all twelve possible combinations gave rise to Mal⁻ phenotype for all pairwise combinations. These results indicate that either TssH does not interact with either of the two TssM domains or such interactions were not detected using BACTH assay.
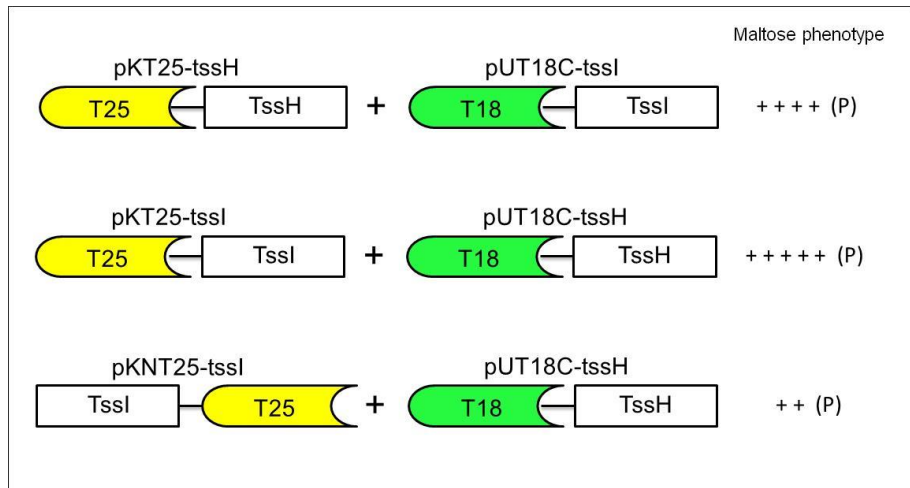
Figure 3.7. Analysis of TssH and TssI interactions using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of TssH and TssI to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation.
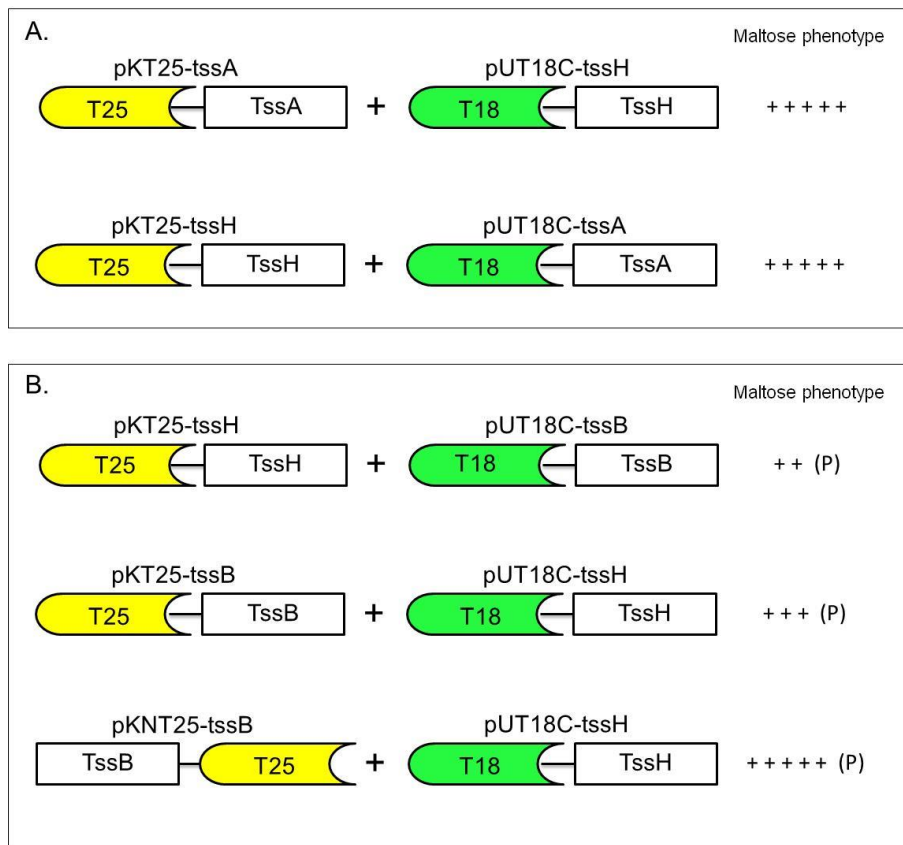
Figure 3.8. Analysis of TssH and TssK or TssL interactions using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of TssH and TssK or TssL to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combination of TssH and TssK fusion proteins that gave rise to a maltose-positive phenotype. B. Combination of TssH and TssL fusion proteins that yielded a maltose-positive phenotype.

3.4.2. Investigation of interactions between TssI and TssA-TssG and TssJ-TssM

3.4.2.1. Investigation of interactions between TssI and TssA

Strong Mal$^+$ phenotypes were observed when plasmids encoding TssI fusions were combined with those encoding TssA fusions for three out of the eight combinations tested (Figure 3.9A). T25-TssI in combination with T18-TssA, and T25-TssA combined with T18-TssI both gave rise to deep purple *E. coli* colonies after 72 hours incubation. These results indicate that both proteins require a free C-terminus to interact or it could simply be that T18 and T25 are too far apart in some of the other combinations. Also, transformant colonies harbouring the TssI-T25 and T18-TssA fusion protein combination turned red and exhibited the patchy phenotype after 120 hours of incubation, but the reverse combination was negative.

3.4.2.2. Investigation of interactions between TssI and TssB

Screening of TssI and TssB protein interactions using the BACTH system showed that two out of eight pairwise combinations yielded a maltose-positive phenotype (Figure 3.9B). When TssI was fused to the C-terminus of T18 and combined with TssB fused to the N- or C-terminus of T25 this gave rise to pink/red colonies with a patchy appearance after 120 hours incubation.

3.4.2.3. Investigation of interactions between TssI and TssC

A very weak Mal$^+$ phenotype was observed after 120 hours incubation of colonies containing the T25-TssI and T18-TssC fusion protein combination (Figure 3.9C). All other compatible combinations gave rise to a Mal$^-$ phenotype.
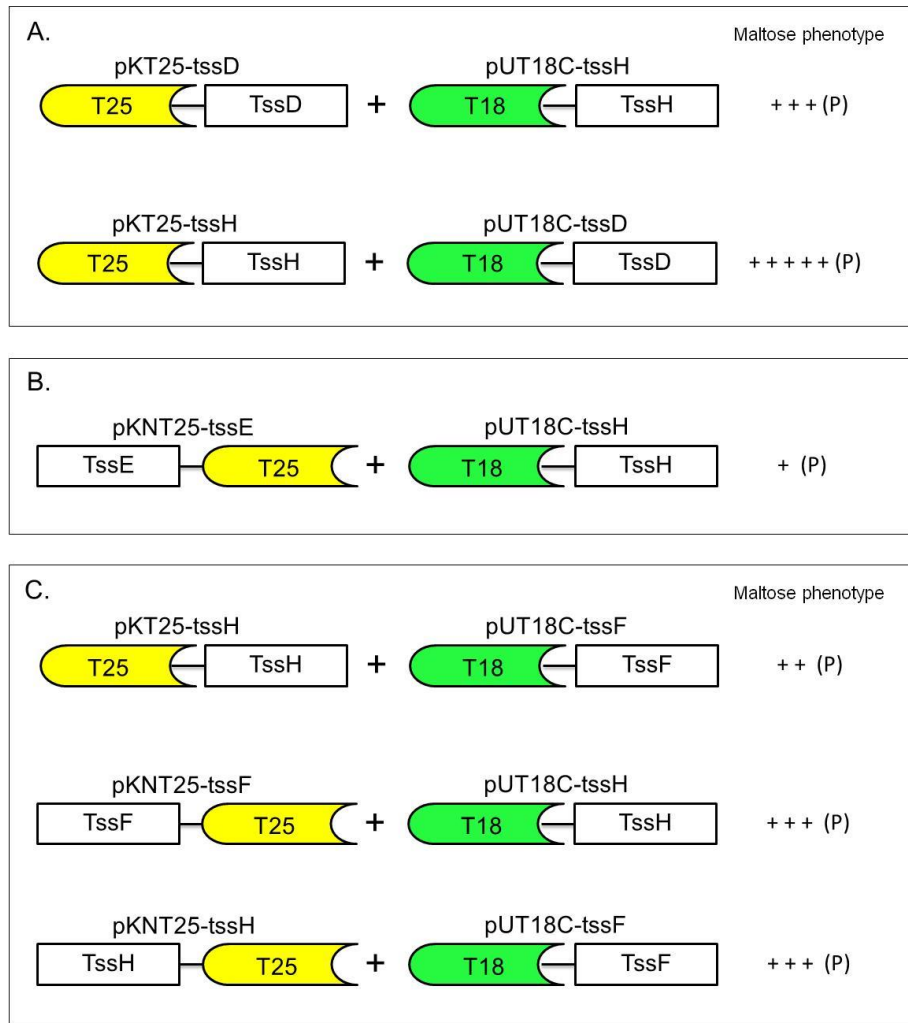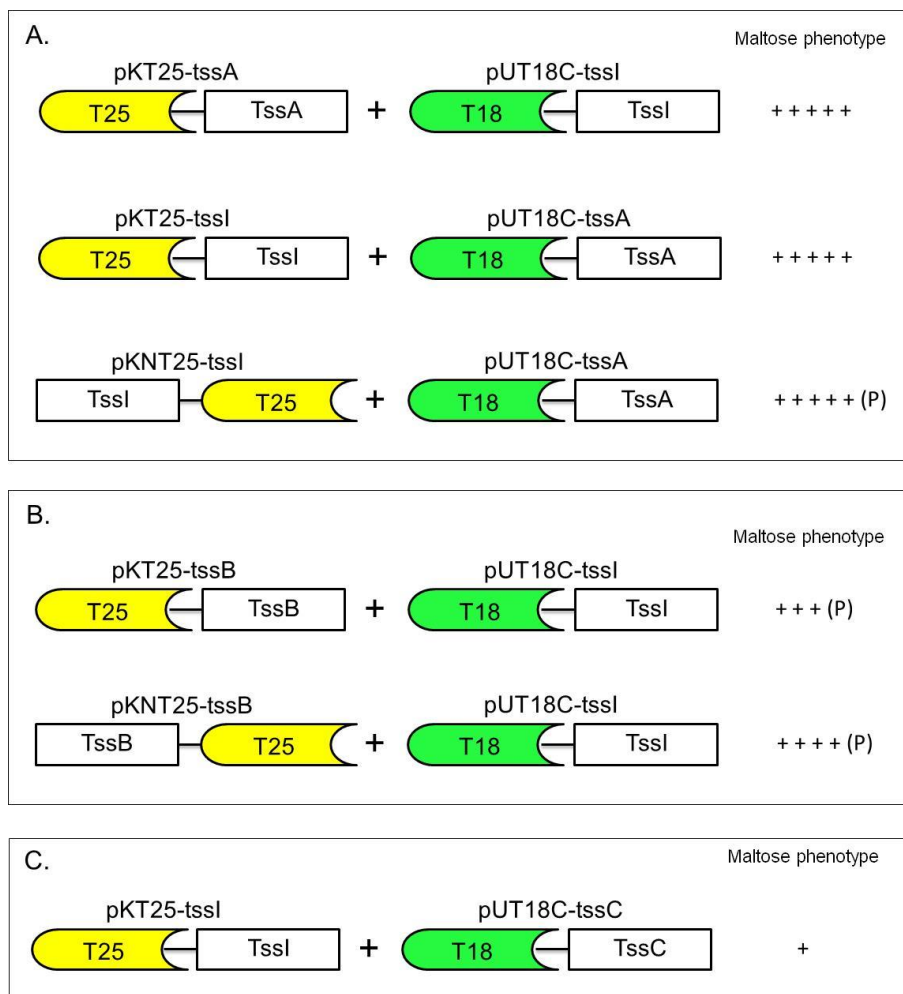
Figure 3.9. Analysis of TssI and TssA or TssB or TssC interactions using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of TssI and TssA or TssB or TssC to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combination of TssI and TssA fusion proteins that gave rise to a maltose-positive phenotype. B. Combination of TssI and TssB fusion proteins that yielded a maltose-positive phenotype. C. Combination of TssI and TssC fusion proteins that yielded a maltose-positive phenotype.

3.4.2.4. Investigation of interactions between TssI and TssD

A strong Mal$^+$ phenotype was generated when T25-TssD and T18-TssI were both present in the same cell, giving rise to pink/red patchy colonies on MacConkey-maltose plates. The other possible combination where TssI and TssD are located at the C-terminus of both fusion proteins (i.e. T25-TssI and T18-TssD) also yielded a Mal$^+$ phenotype with patchy appearance after 120 hours incubation, although the phenotype was not as strong as the T25-TssD and T18-TssI combination (Figure 3.10A).

3.4.2.5. Investigation of interactions between TssI and TssE

BACTH encoding TssI fusion proteins were examined in pairwise combinations with TssE fusion proteins for possible TssE-TssI protein-protein interactions. Out of eight possible combinations, only one combination, TssE-T25 with T18-TssI, gave rise to a Mal$^+$ phenotype, with colonies displaying a pink/red patchy appearance (Figure 3.10B).

3.4.2.6. Investigation of interactions between TssI and TssF

Three TssI and TssF fusion protein combinations gave rise to Mal$^+$ phenotype (Figure 3.10C). A patchy Mal$^+$ phenotype was observed for colonies containing the T25-TssF and T18-TssI fusion proteins on MacConkey-maltose plates. In contrast, the reverse combination (T25-TssI and T18-TssF) gave rise to a Mal$^-$ phenotype. TssI-T25 and T18-TssF, and the analogous combination (TssF-T25 and T18-TssI) both yielded a Mal$^+$ phenotype with uneven pink/red colour of the colonies (patchy phenotype). A maltose-negative phenotype was observed for the remaining five combinations between TssI and TssF fusion proteins.

3.4.2.7. Investigation of interactions between TssI and TssG

Mal$^-$ phenotypes were observed in all 8 pairwise combinations between plasmids encoding TssI and TssG fusion proteins.

3.4.2.8. Investigation of interactions between TssI and TssJ

Of the eight possible combinations of TssI and TssJ fusion proteins, only one combination (TssJ-T25 and T18-TssI) yielded a moderate Mal$^+$ phenotype with pink/red patchy colonies (Figure 3.11A).

3.4.2.9. Investigation of interactions between TssI and TssK

Two combinations of TssI and TssK fusion proteins gave rise to a weak maltose-positive phenotype with pink/red patchy colonies (T25-TssK with T18-TssI, and T25-TssI with T18-TssK). In both fusion protein combinations the C-terminus of TssK and TssI is free (Figure 3.11B). No other fusion protein combinations yielded a Mal$^+$ phenotype.

3.4.2.10. Investigation of interactions between TssI and TssL

Three combinations of TssI and TssL fusion proteins yielded a maltose-positive phenotype with patch pink/red colonies. T25-TssI with T18-TssL and the analogous combination (i.e. T25-TssL with T18-TssI) gave rise to a moderate Mal$^+$ phenotype after 72 hours incubation. A weak maltose-positive phenotype was observed after 120 hours incubation when BTH101 was co-transformed with plasmids expressing the TssI-T25 and T18-TssL fusion proteins (Figure 3.11C). Other combinations led to a Mal$^-$ phenotype.
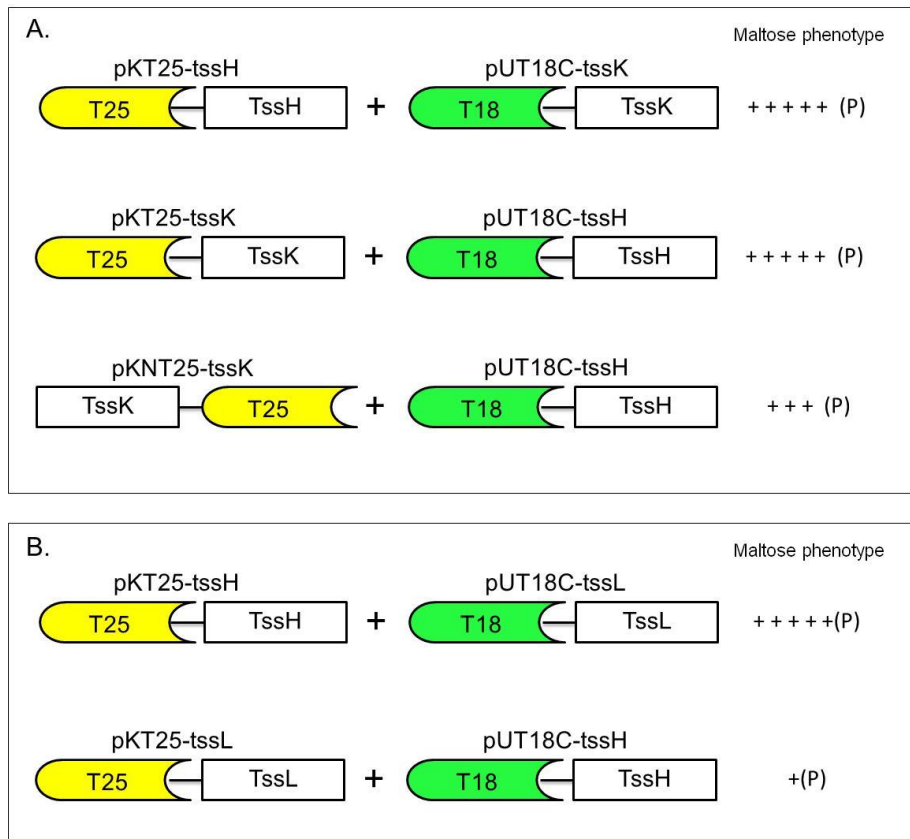
Figure 3.10. Analysis of TssI and TssD or TssE or TssF interactions using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of TssI and TssA or TssB to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combination of TssI and TssD fusion proteins that gave rise to a maltose-positive phenotype. B. Combination of TssI and TssE fusion proteins that yielded a maltose-positive phenotype. C. Combination of TssI and TssF fusion proteins that yielded a maltose-positive phenotype.
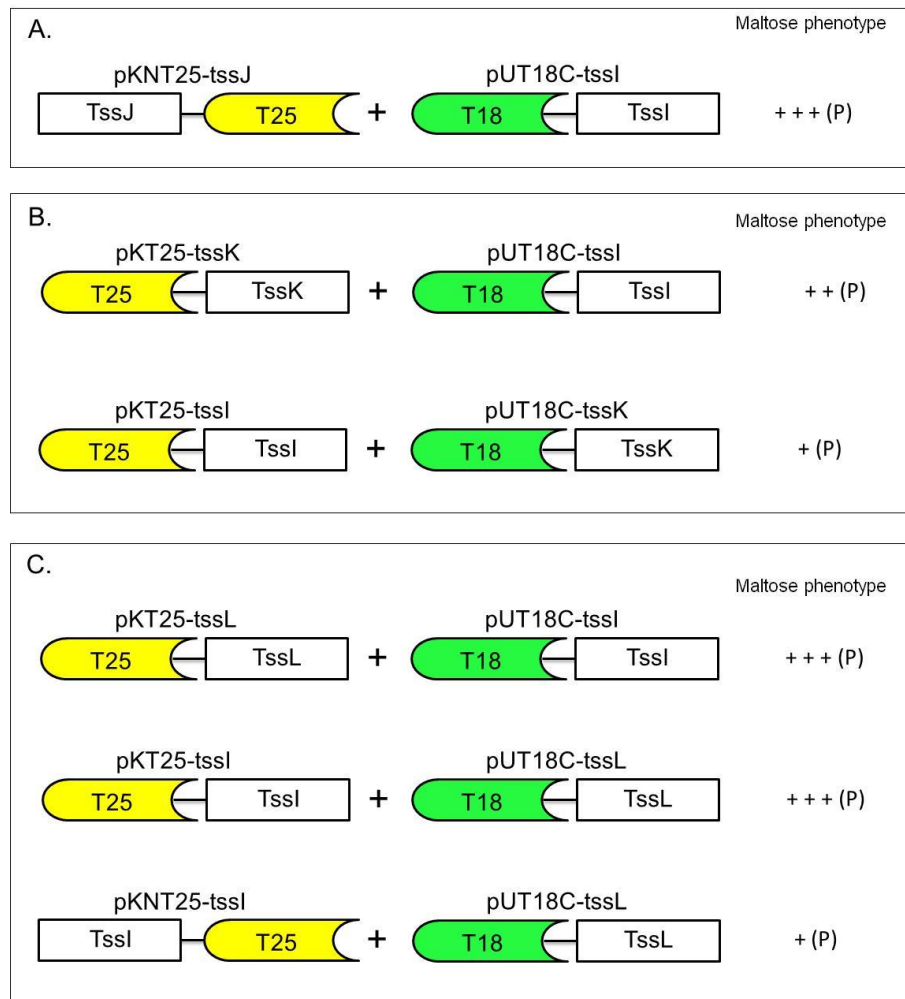
Figure 3.11. Analysis of interactions of TssI with TssJ, TssK and TssL using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of TssI and TssJ or TssK or TssL to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combination of TssI and TssJ fusion proteins that gave rise to a maltose-positive phenotype. B. Combination of TssI and TssK fusion proteins that yielded a maltose-positive phenotype. C. Combination of TssI and TssL fusion proteins that yielded a maltose-positive phenotype.

3.4.2.11. Investigation of interactions between TssI and TssM

None of the 16 possible combinations of plasmids encoding TssI fusion proteins with those encoding $TssM_{NTD}$ or $TssM_{CTD}$ fusion proteins gave use to a Mal^+ phenotype.

3.4.3. Investigation of interactions between TssJ and TssA-TssG and TssK-TssM

All of the 80 possible pairwise BACH fusion protein combinations involving TssJ and TssA-TssG or TssJ and TssK-TssM gave use to a maltose-negative phenotype.

3.4.4. Investigation of interactions between TssK and TssA-TssG and TssL-TssM

The investigation of interactions between TssK fusion proteins and fusion proteins involving TssA-TssG and TssL-TssM using the BACTH system suggested that TssK interacted with four T6SS proteins in addition to TssH (section 3.4.1.10) and TssI (section 3.4.2.9). The details of these positive combinations are given in the section below. The BACTH assay yielded a Mal^- phenotype between TssK, on the one hand, and TssB, TssC, TssE, TssG, TssJ, $TssM_{NTD}$ and $TssM_{CTD}$ on the other.

3.4.4.1. Investigation of the interaction between TssK and TssA

Pairwise interaction analysis of the eight combinations of TssK and TssA fusion proteins showed that one combination, T25-TssK in combination with T18-TssA, yielded colonies with red/purple patches on MacConkey- maltose plates. The analogous combination with the C-termini of TssA and TssK both free (i.e. T25-TssA combined with T18-TssK) also gave rise to a similar result (Figure 3.12A). The other BACTH fusion protein combinations gave rise to a maltose-negative phenotype.

3.4.4.2. Investigation of the interaction between TssK and TssD

All compatible TssK and TssD plasmid combinations were examined by the BACTH assay. A Mal^+ phenotype was detected when T25-TssK was combined withT18-TssD in which the C-terminus of both encoded Tss proteins was free (Figure 3.12B). This combination revealed red patchy colonies on MacConkey-maltose agar. The other combination where both Tss proteins were located at the C-terminus of the fusion proteins and the combination of TssK-T2S with T18-TssD yielded weak maltose-positive phenotypes after 120 hours incubation. The remaining combinations resulted in a maltose-negative phenotype.

3.4.4.3. Investigation of the interaction between TssK and TssF

Figure 3.13A illustrates the combinations of TssK and TssF fusion proteins that were observed to give rise to a maltose-positive phenotype. When TssK, fused to either the N- or C- terminal end of T25, was combined with TssF fused at C-terminal end of T18, it resulted in pink/red patchy colonies on MacConkey-maltose agar. Similar results were obtained when TssF-T25 or T25-TssF were combined with T18-TssK. All other combinations gave rise to Mal$^-$ phenotype.

3.4.4.4. Investigation of the interaction between TssK and TssL

BTH101 competent cells were transformed with the all eight pairwise combinations of TssK and TssL fusion proteins. When TssK, fused at N- or C-terminus of T25 (i.e. pKNT25-TssK or pKT25-TssK), was combined with TssL fused at the C-terminus of T18 (i.e. pUT18C-TssL), it resulted in a maltose-positive phenotype with patchy pink/red coloured colonies on MacConkey-maltose agar (Figure 3.13B). A maltose-negative phenotype was observed with the remaining combinations.
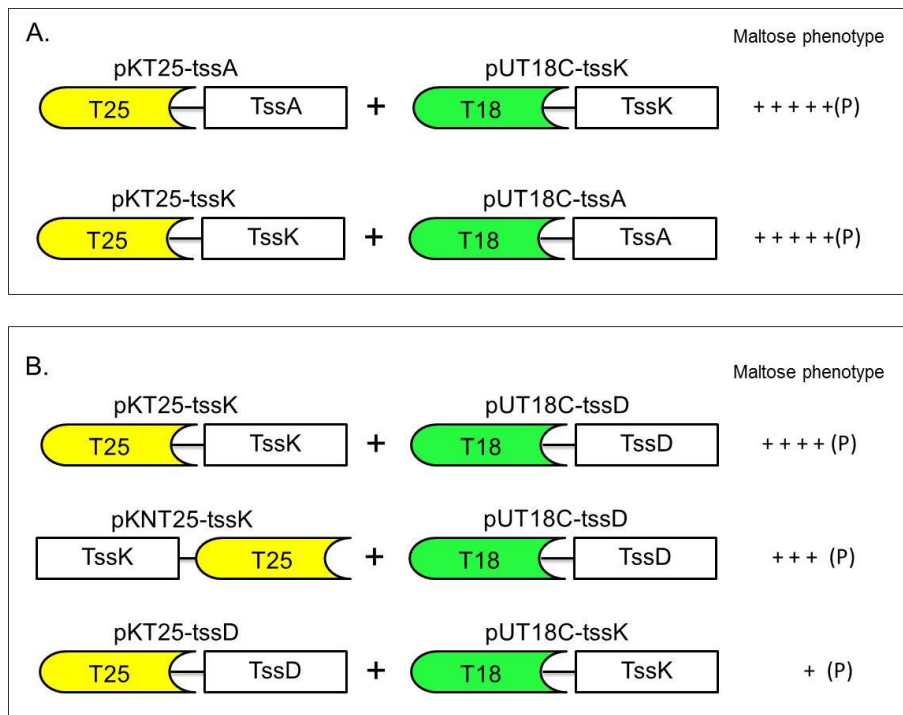
Figure 3.12. Analysis of TssK and TssA or TssD interactions using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of TssK and TssA or TssD to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combination of TssK and TssA fusion proteins that gave rise to a maltose-positive phenotype. B. Combination of TssK and TssD fusion proteins that yielded a maltose-positive phenotype.
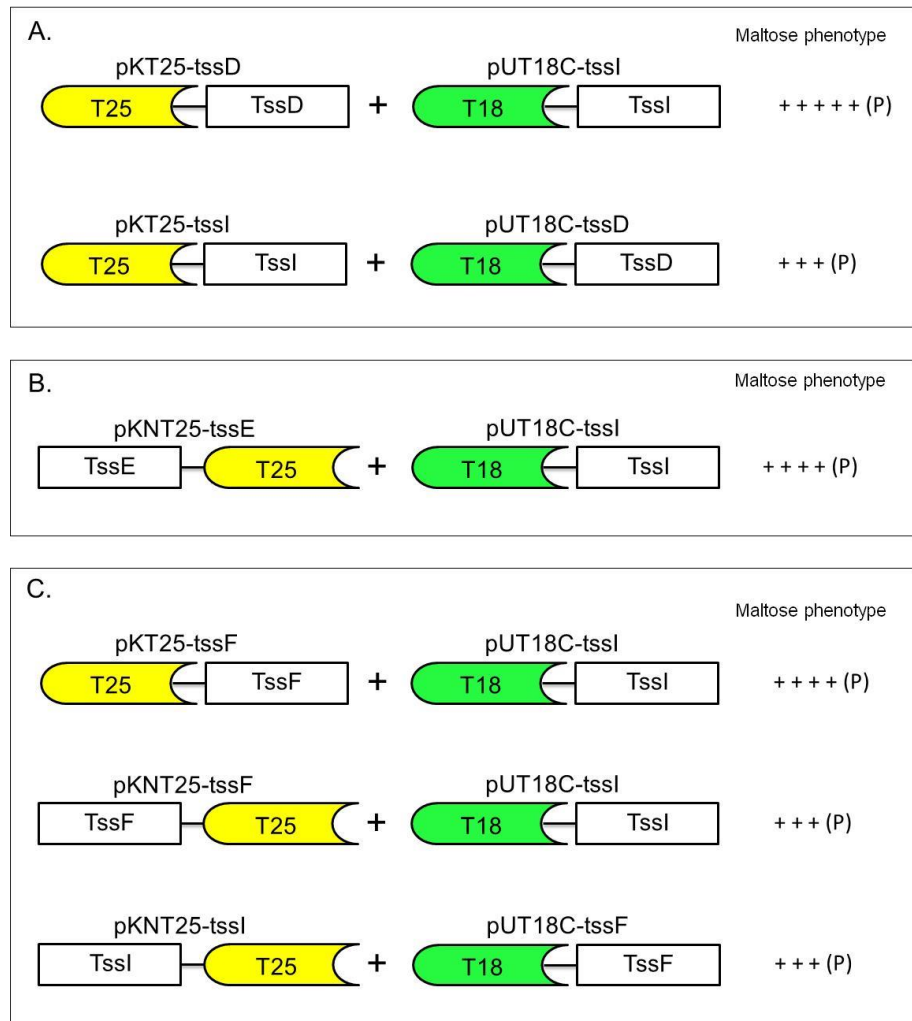
Figure 3.13. Analysis of TssK and TssF or TssL interactions using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of TssK and TssF or TssL to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combination of TssK and TssF fusion proteins that gave rise to a maltose-positive phenotype. B. Combination of TssK and TssL fusion proteins that yielded a maltose-positive phenotype.

3.4.5. Investigation of interactions between TssL and TssA-TssG and TssM

3.4.5.1. Investigation of interactions between TssL and TssA

All four TssL fusion proteins were tested for interactions with all four TssA fusion proteins. It was observed that in the two fusion protein combinations where both proteins are located at the C-terminus of T18 and T25, a strong Mal$^+$ phenotype with pink/red patchy patterned colonies was obtained, i.e. T25-TssL with T18-TssA and T25-TssA with T18-TssL (Figure 3.14A).

3.4.5.2. Investigation of interactions between TssL and TssB

BTH101 was co-transformed with eight pairwise combinations of TssL and TssB fusion proteins. Combinations of TssB fused to either the C- or N-terminus of T25 with TssL fused at C-terminal end of T18 yielded a weak maltose-positive phenotype with patchy colonies (Figure 3.14B). All other combinations gave rise to a maltose-negative phenotype.

3.4.5.3. Investigation of interactions of TssL with TssC, TssE and TssG

BACTH assays of TssL fusion proteins with TssC, TssE and TssG fusion derivatives gave rise to a maltose-negative phenotype in all possible combinations.

3.4.5.4. Investigation of interactions of TssL with TssD

TssL and TssD fusion proteins were examined for potential interactions using BACTH assay. Of the eight pairwise combinations screened, colonies with a weak maltose-positive patchy phenotype were observed when T25-TssD was combined with T18-TssL (Figure 3.14C). All other combination gave rise to a maltose-negative phenotype.

3.4.5.5. Investigation of interactions between TssL and TssF

Screening for potential interactions between TssL and TssF using the BACTH system showed that three fusion protein combinations gave rise to colonies with patchy Mal$^+$ phenotypes (Figure 3.15A). Thus, when the T25-TssL fusion protein was combined with T18-TssF, it gave rise to a weak maltose-positive phenotype. Also, the alternative plasmid combination where TssF and TssL are encoded at the C-termini of the fusion proteins (i.e. pKT25-tssF combined with pUT18C-tssL) gave rise to a maltose-positive phenotype, but it was obviously stronger. The combination that gave rise to a maltose-positive phenotype was TssF-T25 with T18-TssL which resulted in colonies with a

weak maltose- positive phenotype on MacConkey agar. All other combinations of TssF and TssL fusion proteins were negative.

3.4.5.6. Investigation of interactions between TssL and TssM

A strong Mal$^+$ phenotype was observed after 72 hours incubation when TssL was fused to the C-terminal end of T25 and combined with the cytoplasmic domain of TssM (TssM$_{NTD}$) fused to the C-terminal end of T18. The analogous combination with both Tss domains located at the C-termini of the fusion proteins (i.e. T25-TssM$_{NTD}$ with T18-TssL) yielded a similar result (Figure 3.15B). In both cases, the colonies had a patchy appearance. The remaining fourteen combinations gave rise to Mal$^-$ phenotype.

3.4.6. Investigation of interactions of TssM$_{NTD}$ and TssM$_{CTD}$ with TssA-TssG

Each of the four TssM$_{NTD}$ and four TssM$_{CTD}$ fusion proteins were tested for possible interactions with each quartet of TssA, TssB, TssC, TssD, TssE, TssF and TssG fusion protein derivatives. All 112 pairwise combinations gave rise to a maltose-negative phenotype.

3.4.6.1. Investigation of interactions of TssM$_{NTD}$ with TssM$_{CTD}$

A maltose-positive phenotype was not observed for any of the fusion protein combinations of TssM$_{NTD}$ with TssM$_{CTD}$.
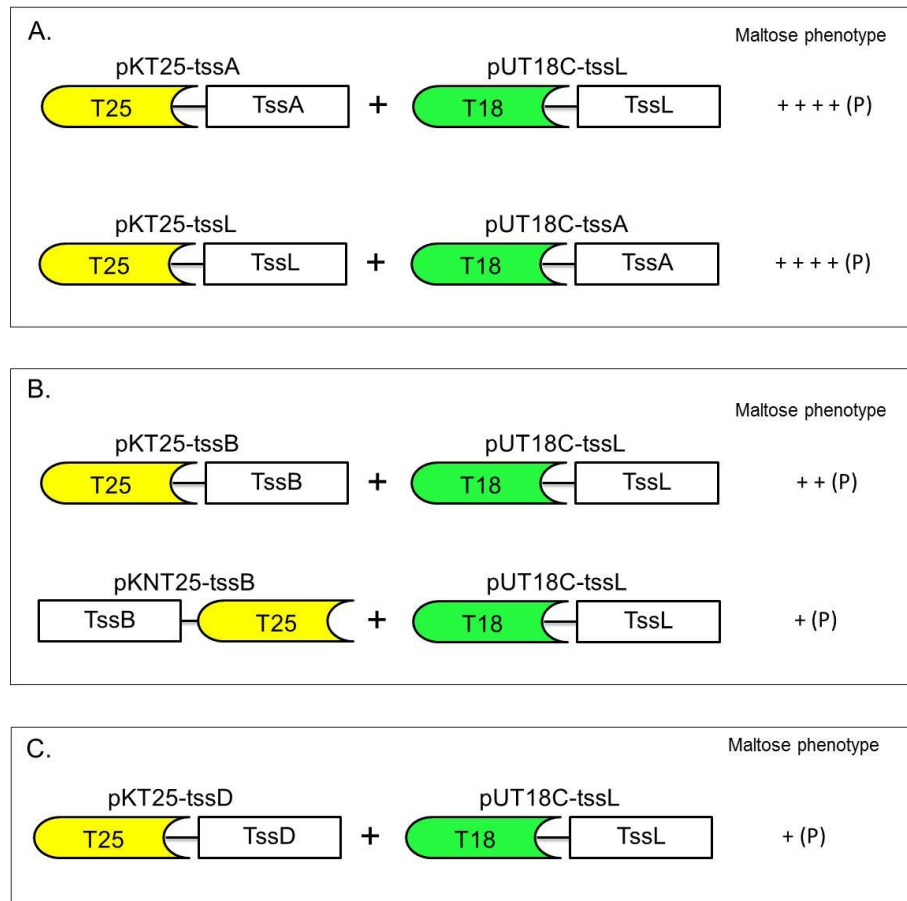
Figure 3.14. Analysis of interaction of TssL with TssA, TssB and TssD using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of TssL and TssA or TssB or TssD to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combination of TssL and TssA fusion proteins that gave rise to a maltose-positive phenotype. B. Combination of TssL and TssB fusion proteins that yielded a maltose-positive phenotype. C. Combination of TssL and TssD fusion proteins that yielded a maltose-positive phenotype.
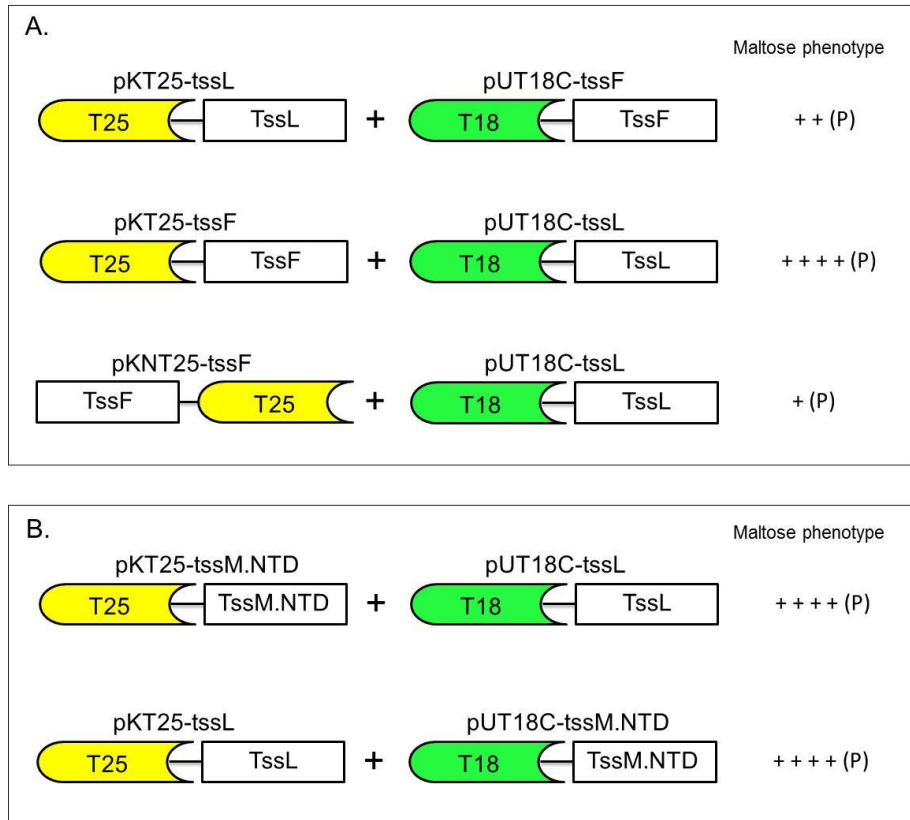
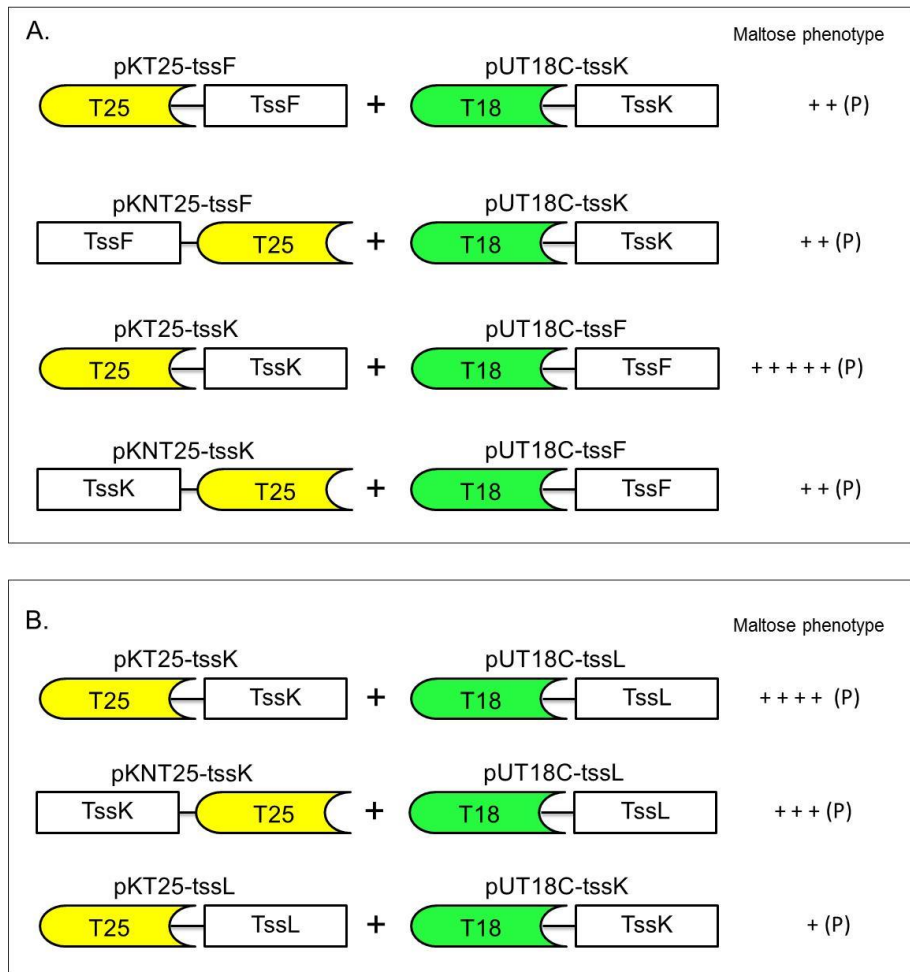**Figure 3.15. Analysis of interaction of TssL with TssF and TssM$_{NTD}$ using the BACTH assay.** Pairwise combinations of compatible BACTH plasmids encoding fusions of TssL and TssF or TssM$_{NTD}$ to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combination of TssL and TssF fusion proteins that gave rise to a maltose-positive phenotype. B. Combination of TssL and TssM$_{NTD}$ fusion proteins that yielded a maltose-positive phenotype.

3.5. Investigation of interactions between TssA domains and other T6SS subunits

Two-hybrid assays have previously indicated that TssA interacts with itself, TssB, TssD, TssE and TssF (S.Shastri, 2011). Here, we have also provided evidence that TssA interacts with TssH, TssI, TssK and TssL. Therefore, TssA potentially interacts with 8 other T6SS subunits and so is likely to be central to the assembly of the T6SS. For this reason, we focussed our subsequent effort on this subunit.

Amino acid sequence alignments and the PSIPRED secondary structure prediction program have suggested that the N-terminal 250 amino acids of TssA may fold into a domain composed of 11-12 α-helices (H1-H11) which is separated by an unstructured ~ 55 amino acid peptide chain from a shorter C-terminal domain of 70 amino acids which is folded into 4 α-helices (H12-H15) (Section 1.5.4.2). DNA encoding the predicted TssA N- and C- terminal domains (i.e. $TssA_{NTD}$ which include 1-256 amino acids and $TssA_{CTD}$ which include 293-373 amino acids) has been previously cloned into all four BACTH vectors and pairwise combinations of these plasmids with compatible plasmids expressing TssA, TssB, TssD, TssE, and TssF fusion proteins were used to screen for the role of each TssA domain in interactions with these subunits (S.Shastri, 2011). To complement these experiments, BACTH assays were performed on the two TssA domains in combination with full-length TssH, TssI, TssK and TssL fusion proteins, as full-length TssA fusion derivatives yielded a Mal^+ phenotype in combination with these subunits. Furthermore, as it had become clear during the course of this work that the previously constructed pKT25-$TssA_{CTD}$ plasmid was defective, this plasmid was reconstructed and the two-hybrid assays between both TssA domains and TssA, TssD, TssE and TssF (and their domains where appropriate) were repeated. As a comparison, BACTH assays were also repeated between full-length TssA and its domains with TssA, TssA domains, TssD, TssE and TssF.

3.5.1. Analysis of TssA domain interactions with TssA and its domains

To confirm previous observations regarding the oligomerisation of TssA and the role of the CTD in the process (S.Shastri, 2011), all four plasmids encoding full-length TssA fusion proteins and all four plasmids encoding $TssA_{CTD}$ fusion proteins were tested in the BACTH assay in all possible pairwise combinations. All four possible pairwise combinations of full-length TssA fusion proteins, all eight possible pairwise combinations of full-length TssA fusion proteins with $TssA_{CTD}$ fusion proteins, and all

four possible pairwise combinations of TssA$_{CTD}$ fusion proteins gave rise to colonies exhibiting a strong Mal$^+$ phenotype with a uniform deep red/purple colour (Figure 3.16A, B and C). All these results are in agreement with those observed previously (S.Shastri, 2011). We also confirmed that when TssA was fused either at the C- or N-terminus of T25 and combined with T18-TssA$_{NTD}$, it gave rise to weak Mal$^+$ phenotype (Figure 3.16D). Other combinations of TssA$_{NTD}$ fusion proteins with full-length TssA revealed a maltose-negative phenotype. From these results we can conclude that TssA self-interacts to form a homo-oligomer of unknown stoichiometry and that the CTD plays an important role in this process.
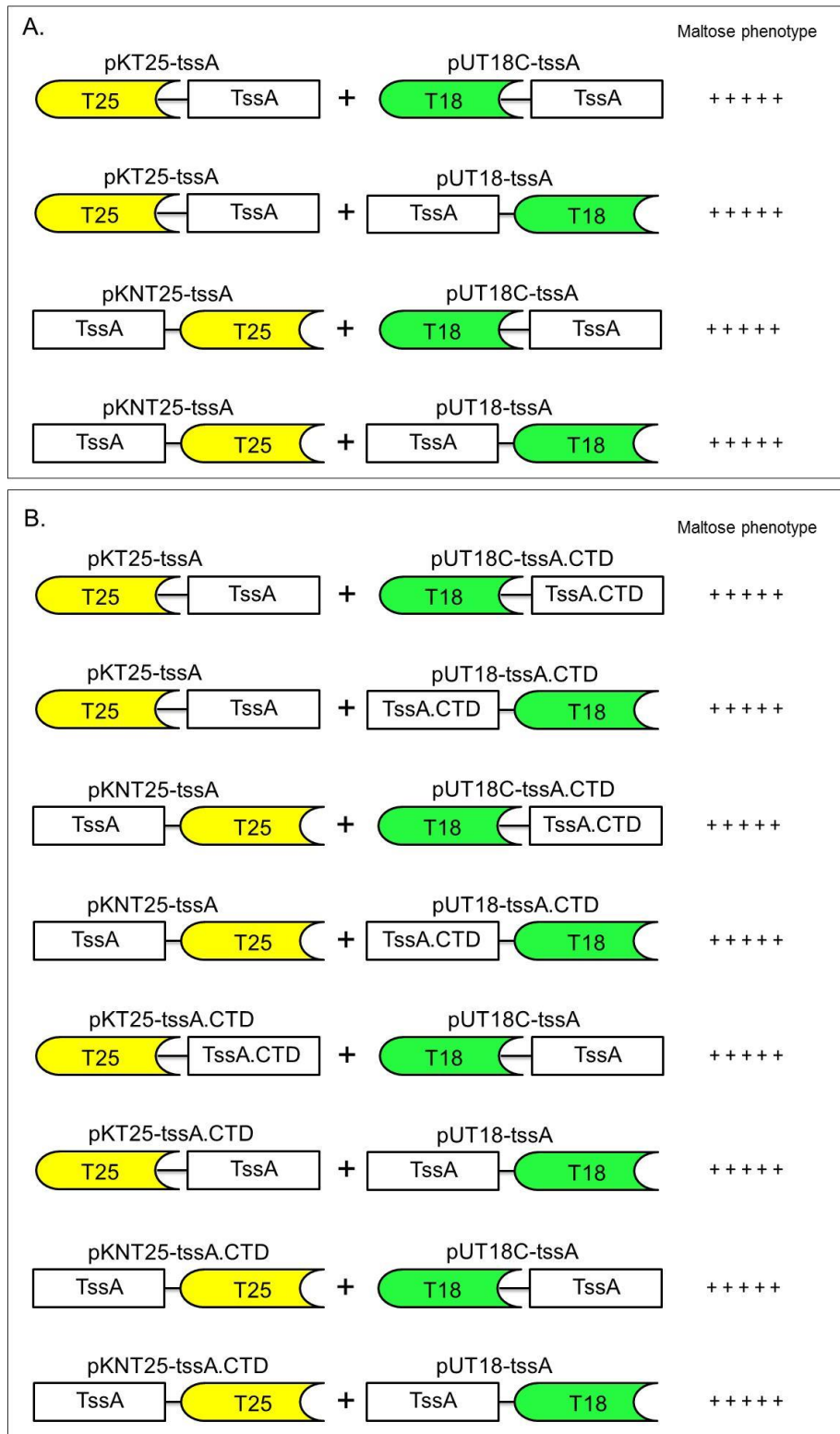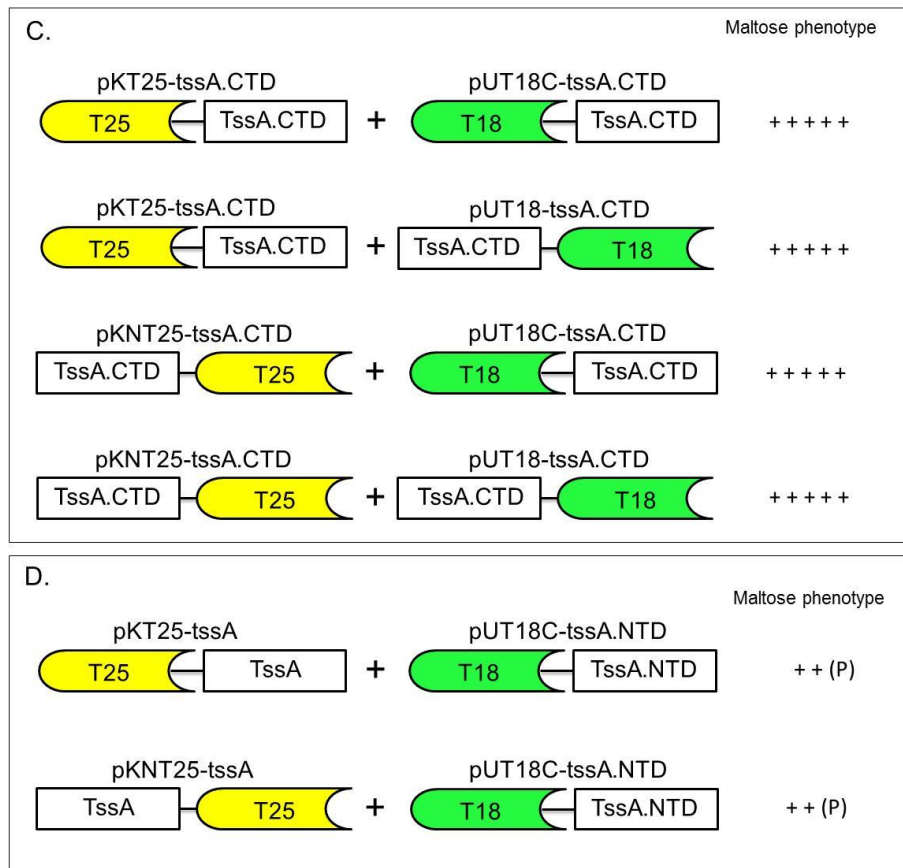
Figure 3.16. Analysis of interaction between TssA and TssA, TssA$_{CTD}$, and TssA$_{NTD}$, and between TssA$_{CTD}$ and itself using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of TssA with TssA, TssA$_{CTD}$ or TssA$_{NTD}$, and TssA$_{CTD}$ with TssA$_{CTD}$ to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combination of TssA and TssA fusion proteins that gave rise to a maltose-positive phenotype. B. Combination of TssA and TssA$_{CTD}$ fusion proteins that yielded a maltose-positive phenotype. C. Combination of TssA$_{CTD}$ and TssA$_{CTD}$ fusion proteins that yielded a maltose-positive phenotype. D. Combination of TssA and TssA$_{NTD}$ fusion proteins that yielded a maltose-positive phenotype.

3.5.2. Interaction of TssA and its domains with TssD

Two out of eight combinations of TssA and TssD fusion proteins yielded a maltose-positive phenotype. In both cases, the C-terminal ends of TssA and TssD were free (Figure 3.17A). These results are in accordance with previously reported observations (S.Shastri, 2011). Testing combinations of TssD with TssA domains revealed a maltose-positive phenotype for one combination, $TssA_{CTD}$-T25 with T18-TssD (Figure 3.17B).

3.5.3. Interaction of TssA and its domains with TssE

Co-transformation of *E. coli* BTH101 with all possible pairwise plasmid combinations encoding full-length TssA and TssE fusion proteins was carried out. Only one combination of fusion proteins, TssE-T25 with T18-TssA, gave rise to colonies with a patchy maltose-positive phenotype (Figure 3.17C) in agreement with (S.Shastri, 2011). However, in contrast to previous observations, no combinations of $TssA_{CTD}$ or $TssA_{NTD}$ with TssE gave rise to a Mal$^+$ phenotype. Previously, it was observed that both pairwise combinations of $TssA_{CTD}$ and TssE, where both proteins were located at the N-terminus of their respective fusion proteins, were positive in the two-hybrid assay (S.Shastri, 2011).

3.5.4. Interaction of TssA and its domains with TssF

Co-transformation of *E. coli* BTH101 with all possible pairwise plasmid combinations encoding TssA or its domains and TssF fusion proteins was carried out. Four combinations of full-length fusion proteins gave rise to a maltose-positive phenotype with pink/red patched colonies (Figure 3.17D) in agreement with (S.Shastri, 2011). Also, two combinations of $TssA_{CTD}$ with TssF gave rise to a maltose-positive phenotype (Figure 3.17E).
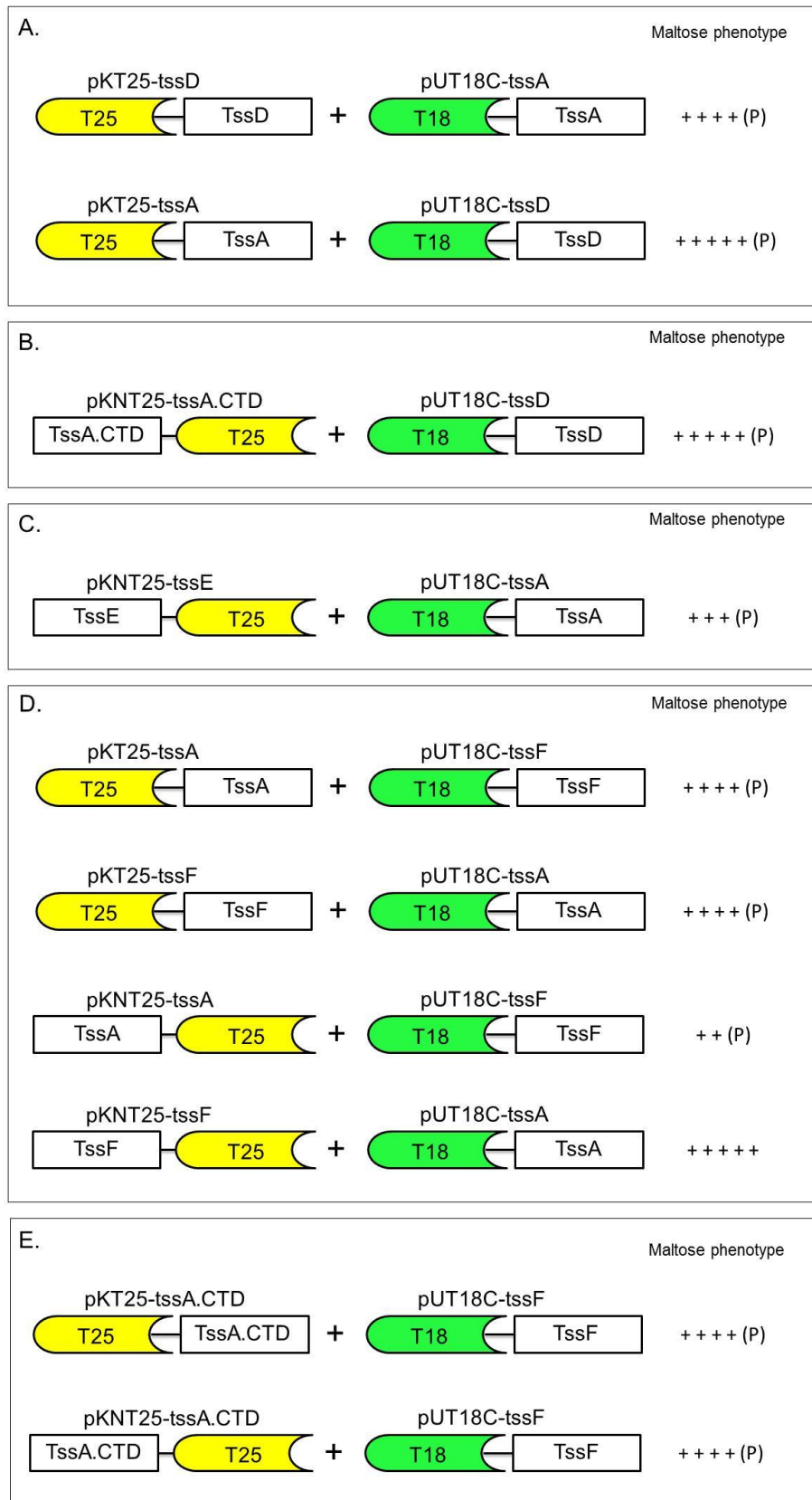
Figure 3.17. Combinations of BACTH plasmids containing TssA or its domains with TssD, TssE and TssF which yielded a maltose-positive phenotype in the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of TssA or TssA$_{CTD}$ with TssD, TssE or TssF to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combinations of TssA with TssD which revealed maltose-positive phenotypes. B. Combination of TssA$_{CTD}$ and TssD fusion proteins which yielded maltose-positive phenotype. C. Combination of TssA and TssE fusion proteins which gave rise to maltose-positive phenotype. D. Combination of TssA and TssF fusion proteins which gave rise to maltose-positive phenotype. E. Combination of TssA$_{CTD}$ and TssF fusion proteins which gave rise to maltose-positive phenotype.

3.5.5. Interaction of TssA and its domains with TssH and its domains

As illustrated in Figure 3.5A, two combinations of TssA and TssH fusion proteins yielded a maltose-positive phenotype. The pKT25 and pUT18C plasmid combinations were employed to determine which domain of TssA is involved in the interaction with TssH. The results show that both combinations of $TssA_{CTD}$ and TssH fusion proteins tested gave rise to a Mal$^+$ phenotype (Figure 3.18A), whereas both combinations of $TssA_{NTD}$ and TssH fusion proteins led to Mal$^-$ phenotype (result not shown). Therefore, $TssA_{CTD}$ is responsible for the observed interaction between TssA and TssH.

Three different regions of *tssH* were previously amplified and cloned into the BACTH vectors pKT25 and pUT18C (Jones.R.A, 2013). These regions encoded (i) the N-domain, (ii) NBD1 together with the M domain, and (iii) NBD2. To determine which domain of TssH interacts with TssA, the BACTH assay was then carried out on combinations of TssA and its domains with individual TssH domains where the Tss moiety is located at the C-terminus of both paired fusion proteins. When the T25-$TssH_{N\text{-}domain}$ fusion was combined with T18-TssA, it resulted in a strong maltose-positive phenotype after 120 hours incubation, whereas the combination of T25-TssA with T18-$TssH_{N\text{-}domain}$ yielded a weak maltose-positive phenotype after the same period of time (Figure 3.18B). Both combinations of $TssH_{NBD1/M}$ with TssA gave rise to a maltose-negative phenotype (results not shown). Interestingly, both combinations of $TssH_{NBD2}$ with TssA led to a strong Mal$^+$ phenotype after 72 hours incubation (Figure 3.18B). This suggests that the N-domain and NBD2 are involved in interactions with TssA.

The BACTH analysis of $TssH_{NBD2}$ with $TssA_{CTD}$ resulted in a weak maltose-positive phenotype with patched pink/red colonies in both possible fusion protein combinations, and with $TssA_{NTD}$ a weak maltose phenotype was also obtained with one fusion protein combination (Figure 3.19A). Combinations of the $TssH_{N\text{-}domain}$ with TssA domains gave rise to a weak to moderate Mal$^+$ phenotype in three of the four possible combinations with the combination of T25-$TssH_{N\text{-}domain}$ and T18-$TssA_{CTD}$ giving the strongest maltose-positive phenotype (Figure 3.19B). These results suggest that both domains of TssA may be involved in interactions with the NBD2 and N-domain of TssH.

3.5.6. Screening for interactions between TssA domains and TssI

TssA was shown to interact with TssI in three pairwise combinations of two-hybrid plasmids (Section 3.4.2.1). To determine which domains of TssA are involved in these interactions, $TssA_{NTD}$ and $TssA_{CTD}$ fusion proteins were expressed in combination with TssI fusion proteins in the BACTH system. A strong Mal^+ phenotype resulted when $T25\text{-}TssA_{CTD}$ was combined with T18-TssI, and very weak maltose-positive phenotype occurred when $T25\text{-}TssA_{NTD}$ was combined with T18-TssI (Figure 3.20A). A maltose-negative phenotype was observed when $T18\text{-}TssA_{NTD}$ or $T18\text{-}TssA_{CTD}$ fusion protein was combined with T25-TssI or TssI-T25.

3.5.7. Screening for interactions between TssA domains and TssK

Figure 3.12A shows the combinations of full-length TssA and TssK hybrid proteins which yielded a maltose-positive phenotype in the BACTH assay. Further BACTH assays were carried out with plasmids encoding TssA domains to determine which domains are involved in interactions with TssK. For these experiments only pKT25 and pUT18C derivatives were employed. The results show that when $T25\text{-}TssA_{CTD}$ was combined with 18-TssK, colonies with a patchy Mal^+ phenotype were observed which required 120 hours incubation to become apparent (Figure 3.20B). Other combinations of plasmids expressing TssA domains and TssK hybrid proteins gave rise to Mal^- phenotype.

3.5.8. Screening for interactions between TssA domains and TssL

Two out of eight plasmid combinations expressing TssA and TssL hybrid proteins gave rise to a maltose-positive phenotype (Section 3.4.5.1). However, the analogous plasmid combinations expressing $TssA_{NTD}$ or $TssA_{CTD}$ with TssK (i.e. only the pKT25 and pUT18C derivatives were tested) yielded a maltose-negative phenotype in all four combinations (results not shown).
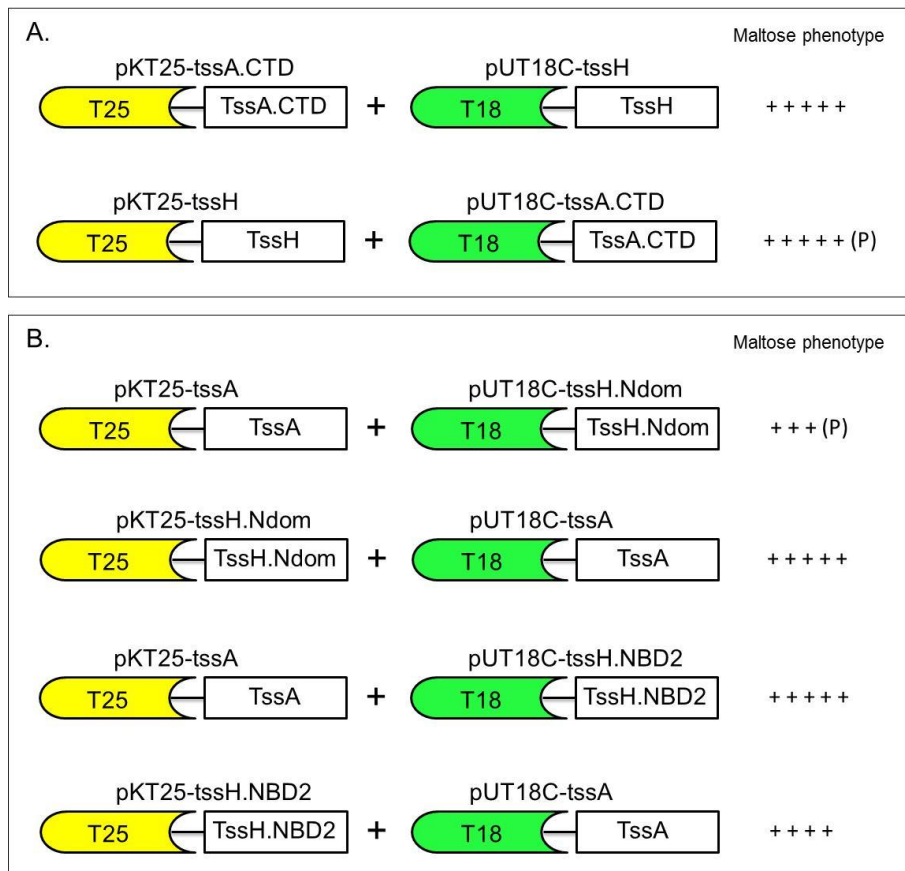
Figure 3.18. Analysis of interactions between TssH and TssA$_{CTD}$, and between TssA and TssH$_{NBD2}$ or TssH$_{Ndom}$ using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of TssH with TssA$_{CTD}$ and TssA with TssH$_{NBD2}$ or TssH$_{Ndom}$ to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combination of TssH and TssA$_{CTD}$ fusion proteins that gave rise to a maltose-positive phenotype. B. Combination of TssA and TssH$_{NBD2}$ or TssH$_{Ndom}$ fusion proteins that yielded a maltose-positive phenotype.
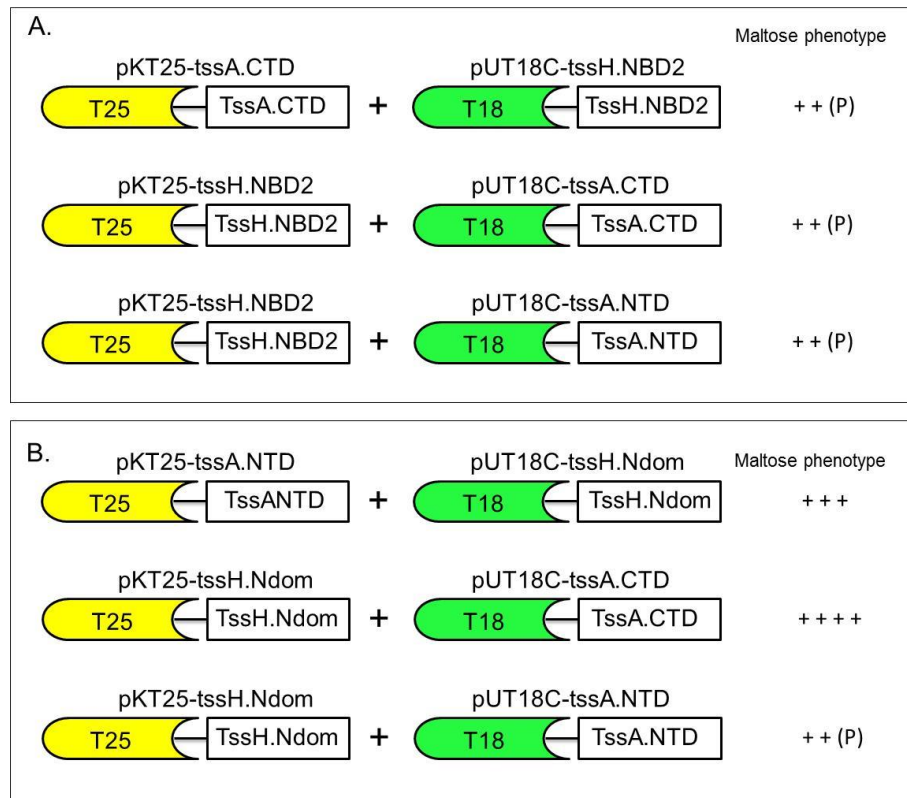
Figure 3.19. Analysis of TssA domains and TssH domains interactions using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of $TssA_{NTD}$ or $TssA_{CTD}$ and $TssH_{NBD2}$ or $TssH_{Ndom}$ to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combination of TssA domain and $TssH_{NBD2}$ fusion proteins that gave rise to a maltose-positive phenotype. B. Combination of TssA domains and $TssH_{Ndom}$ fusion proteins that yielded a maltose-positive phenotype.

3.6. Investigation of interactions between individual TssF domains and other Tss proteins

It was previously shown by BACTH analysis that TssF self-interacts (S.Shastri, 2011). The same study also showed that TssF interacts with TssA, TssB, TssC, TssD and TssE. As TssF is a large protein (amino acids), it was decided to identify the region(s) within TssF which are responsible for the observed interactions of TssF by constructing BACTH plasmids in which the N-terminal 372 codons and the C-terminal 246 codons of TssF were fused to the T18 and T25 coding sequences (S.Shastri, 2011). These were referred to as $TssF_{NTD}$ and $TssF_{CTD}$ accordingly, although these regions do not necessarily correspond to single domains. As TssF was also found to interact with TssI, TssK and TssL in this work (Sections 3.4.2.6, 3.4.4.3 and 3.4.5.1, respectively), $TssF_{NTD}$ and $TssF_{CTD}$ were tested with fusion protein combinations of TssI, TssK and TssL which showed a maltose-positive phenotype in combination with full-length TssF.

It was observed that when $TssF_{NTD}$ is fused in frame at the C- or N-terminal end of T25 and combined with T18-TssI, it gave rise to pink/red colonies with a patchy morphology (Figure 3.21). On the other hand, both $TssF_{CTD}$ combinations with TssI showed a maltose-negative phenotype. These results indicate that $TssF_{NTD}$ is involved in the interaction with TssI. All BACTH combinations of $TssF_{NTD}$ and $TssF_{CTD}$ with TssK and TssL fusion proteins gave rise to a Mal⁻ phenotype (result not shown).
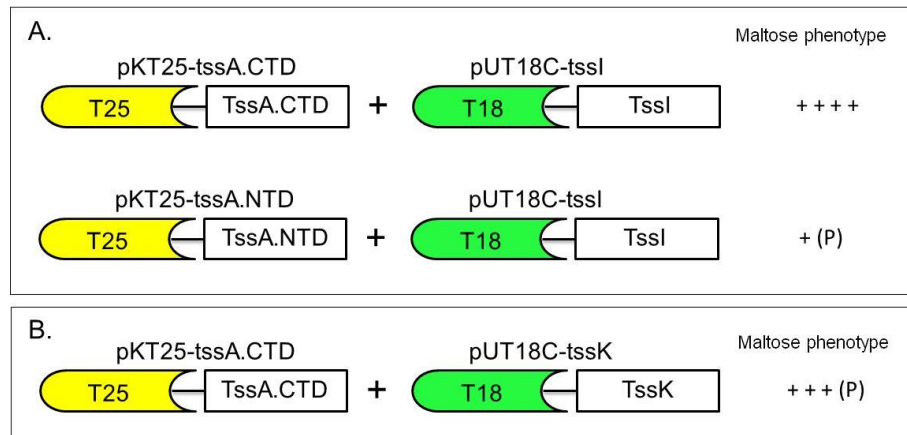
Figure 3.20. Analysis of interaction between TssA domains and TssI or TssK using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of TssA$_{NTD}$ or TssA$_{CTD}$ and TssI or TssK to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combination of TssA domain and TssI fusion proteins that gave rise to a maltose-positive phenotype. B. Combination of TssA$_{CTD}$ and TssK fusion proteins that yielded a maltose-positive phenotype.
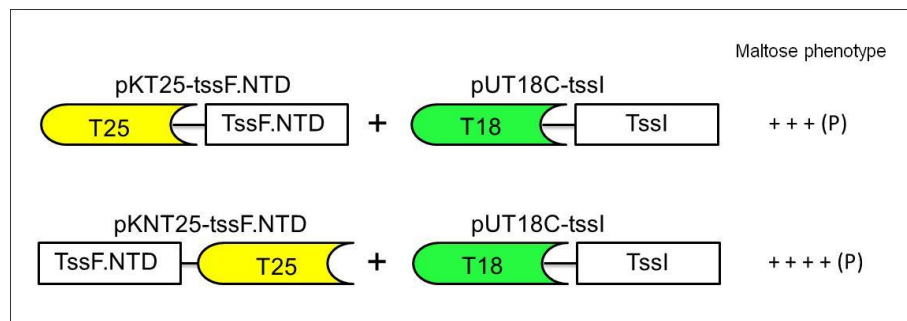
Figure 3.21. Analysis of TssF$_{NTD}$ and TssI interactions using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of TssF$_{NTD}$ and TssI to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation.

3.7. Investigation of interactions between individual TssI domains and other Tss proteins

Recent studies showed that TssI shares structural homology with the gp5 and gp27 proteins that make up the bacteriophage T4 tail spike responsible for puncturing the target cell membrane (Pukatzki *et al.*, 2007, Leiman *et al.*, 2009). The needle-like structure formed by TssI is proposed to be located at the tip of a tube formed from stacked hexameric rings of TssD (Mougous *et al.*, 2006). Multi-sequence alignments of amino acid sequences of a variety of TssI from different Gram-negative bacteria, including BCAM0148 of *B. cenocepacia*, as shown in Figure 3.25, along with a study conducted by Leiman *et al.* (2009) clarify the core regions of TssI corresponding to the bacteriophage T4 gp5 and gp27 proteins. To identify the region(s) of TssI involved in interactions with TssA, TssD and TssH, DNA encoding the gp27- and gp5-like domains of TssI and the fused gp27.gp5 region were cloned into pKT25 and pUT18C for the use in the BACTH system.

3.7.1. Analysis of interactions between the gp27.gp5 region of TssI with itself and with full-length TssI

The BACTH assay results of full-length TssI (Section 3.3) were consistent with the known ability of the homologous bacteriophage T4 proteins gp27 and gp5 to form a trimeric $(gp27.gp5)_3$ complex. To test whether the gp27.gp5-like moiety interact with full-length TssI, all four possible combinations of pKT25-gp27.gp5 or pUT18C-gp27.gp5 with compatible plasmids expressing TssI fusion proteins were tested by BACTH assays and the results showed that they all gave rise to deep purple colonies on MacConkey-maltose plates after 72 hours incubation (Figure 3.23A). Furthermore, the gp27.gp5-like moiety of TssI combinations with itself (i.e. pKT25-gp27.gp5 with pUT18C-gp27.gp5) gave rise to a strong Mal^+ phenotype after 72 hours incubation on MacConkey-maltose agar (Figure 3.23B).

3.7.2. Analysis of the interaction between gp27.gp5 and full-length TssA

Results from BACTH experiments showed that when T25-gp27.gp5 was co-expressed with T18-TssA, deep purple colonies formed where both proteins were located at the C-terminal end of the two-hybrid fusions, indicating gp27.gp5 is able to interact with TssA (Figure 3.24A). However, when TssA was located at the N-terminus of the fusions no

interaction was detectable. These results suggest that the C-terminal region of TssA should be free in order for the two fusion proteins to interact.

3.7.3. Analysis of the interaction the between the gp27.gp5- like region of TssI and TssD

TssD co-expression with the gp27.gp5 like region TssI resulted in a Mal$^+$ phenotype in three of the four possible fusion protein combinations tested (Figure 3.24B). A weak maltose-positive phenotype was obtained after 72 hours incubation when T25-TssD was combined with T18-gp27.gp5. The Mal$^+$ phenotype of this combination become stronger after 120 hours incubation. It was also observed that when TssD was fused at the C- or N-terminal end of T25 and co-expressed with the gp27.gp5 region of TssI fused at the C-terminal end of T18, it resulted in patchy pink/red colonies after 120 hours incubation.

3.7.4. Analysis of the interaction between the gp27.gp5-like region of TssI and TssH

The BACTH assay of TssH with the gp27.gp5-like region of TssI resulted in a Mal$^+$ phenotype, but only in one of the four possible plasmid combinations, i.e. T25.TssH combined with T18-gp27.gp5 (Figure 3.24C). This combination resulted in a Mal$^+$ phenotype with pink/red colonies after 72 hours incubation, suggesting a potential *in vivo* physical interaction between the gp27.gp5-like region of TssI and TssH.

3.7.5. Analysis of the interaction between the separate gp27- and gp5-like regions of TssI with full-length TssI and the gp27- and gp5-like regions

The ability of the separate gp27- and gp5-like regions of TssI to interact with full-length TssI and its constituent regions was probed using the BACTH assay. The gp27-like region gave rise to a maltose-positive phenotype in combination with full-length TssI in three of four pairwise combinations tested (Figure 3.25A). In each case 120 h incubation was required to score colonies as maltose-positive and the colonies exhibited the patchy phenotype. When fusion proteins containing the gp27-like region were combined with those containing the entire gp27.gp5 homologous region, a maltose-positive phenotype was observed in both combinations tested (Figure 3.25B). Again, this was not observable until 120 h incubation and the colonies exhibited the patchy phenotype. On the other hand, the BACTH assays involving combinations of the gp5-

like region with TssI or the gp27.gp5-like region showed a weak Mal$^+$ phenotype in only one pairwise combination with gp27.gp5-like region after 120 hours incubation (Figure 3.25C) and a Mal$^-$ phenotype in all pairwise combinations with TssI. In addition, testing all pairwise combinations of the gp27 and gp5 fusion proteins with each other gave rise to a maltose-negative phenotype (result not shown).

```
Aec15       1  -----------------MSLKGLRFTLEVDGQEPDTFAVVNFRLIQNQSYPEVVSVDVAS
c1888       1  -----------------MSLKGLRFTLEVDGQEPDTFAVVNFRLIQNQSYPEVVSVDVAS
Z0267       1  -----------------MS-TGLRFTLEVDGLPPDAFAVVSFHLNQSLSSLFSLAVSIVS
AHA1119     1  ----------------MADSTGLQFTKVGALPESTEVVAEFALDEGLNRPFNLRTELAS
PA1511      1  ----------------MRQRDLKFTFVVGE-GKLAEDVVEFELEEALCEPERLNLKLAS
VP1394      1  -----------------MVNDVEFKFEVPG-CGHEFRVESFQVNEELSKPFHLSLSLLS
PA0095      1  --------------MALAQQTRLVRVDSPLG-AEVLQLQRMEGREELGRPEAYELELIS
BB0793      1  -------------METILAMAERIVRALTPLP-PQALQFRSMHGHEGLSALYEFEVDLLA
BCAL1294    1  -------MQMSEIASFFQLQSNRLFTIETPLKGRSDLVLVDFHCTEGLSQNFEMHVRLAS
YPO0970     1  ----------MNTTLPTLRFDSHHKLVVRDSTAAVDVLGFEGHESLSQPFCYDIQFTS
c3393       1  ----------MEALMNVQFFDLAHHKLKIRGLQSPVDVLTFEEGREQLSTPFRYDIQFTS
BCAM0148    1  ----MDAHSMIGALTGGTIQQERLLKLDTPLGSNVLIPHRVFG-QSRIGRDYLFMVDCVS
BPSS1503    1  MPSSPRHDAPASRANAAPSANARRFTFASDAYDPATFDVVDINGRDAISQPYRFEITLVS

Aec15      44  DSFM-QTAEMLLEKNATLTIWQGVI---------PLRYVTGVVAGFGMQENNGWQMR---
c1888      44  DSFM-QTAEMLLEKNATLTIWQGVI---------PQRYVTGVVAGFGMQENNGWQMR---
Z0267      43  QQFLSLEFQQILDKMAYLTIWQGDD---------VQRRVKGVVTWFELGENDKNQKL---
AHA1119    45  AQPD-LDFGAVLDQPCELLVWYNGE---------LQRRVCGVVSDFAQGDSGFRRTR---
PA1511     43  DKNA-IDFRQVLDQPGTFTIWQDGR---------PARYVHGIVSHFTQGSSGFRRTR---
VP1394     42  LDPD-ISFDSLIRKAGTLTIYGQGLS--------AARIFNGVVNEVRYLGTGRRFSR---
PA0095     45  ENPD-LPLDGLIGKPASLAIELHD---------GSRRHFHGIVAACSQGSGNGQFAS---
BB0793     47  ATHT-LELKSLIGKPVSLEIETG---------GAPRYLSGQATRCALVGREGDSARQYV
BCAL1294   54  QDKN-IELKKLIGQPVTITIQLTDALA-----SSEERYFHGYVAAFSHLDTGGFAM---
YPO0970    50  ADKA-IDPATMLMHDASLTIAAPMAEAFGVTVQQTQRVIQGVVTGFKRLSASKEECR---
c3393      50  SDKA-IAPESVLMQDGAFSITAPPVQ--GMPVQTALRTLHGVITGFKHLSSSQDEAR---
BCAM0148   56  TSND-VQLKALISQPITTWIQQTDK---------SYLPHHGYIHTARKLGVDGGLAC---
BPSS1503   61  RQLR-IDFAKMLSRGATLAILPPFG-------EAGTTRYAGVIAEFEQKKRFRDFTV---

Aec15      91  YHLRTEPPLWRCGLRRNFRIFQ-QDIRTISATILNENGVTE---WTPLFYEDHPAREFC
c1888      91  YHLRTEPPLWRCGLRRNFRIFQ-QDIRTISAMILNENGVTE---WTPLFYEDHPAREFC
Z0267      91  YSMKVCPPLWRTGLRQNFRIFQ-NEDLASIIGTILQENGVTE---WSPLFSEPHPSREFC
AHA1119    92  YQILVQPALWRLSLRQNSRIFQ-AQKPDEIISILLQEHGITD---YAEALKNEHAKREYC
PA1511     90  YELLLEPQIARLELCCNWRIFQ-EKSVPEIIQALLKEHRVLD---YEQRIYHEHLPREYC
VP1394     90  YQLILVPQAWFLSQRQDCRIFQ-OKSAKDIITEVLDDGSVTD---YRELSGIYPKEYA
PA0095     92  YQVTLRPWLWLLTRTSDCRIFQ-NQKYPDIIKQVFRDLGFSD---FEDALSRSYREWEYC
BB0793     96  YRVTMRPWLWYLTQTSDKIIFQ-OMSVVDVIRQVLADYPFP----VEYRLAGSYRRWEYC
BCAL1294  105  YSATILPWLWMLSRRQDIRIFQ-EENTEAILARVFREYGKLAS--EERLSKGTQNRSYC
YPO0970   106  YELSLQPRIALLSRSHONGIIQ-DMSVPQIVEKLTRERHDMRGQDFVETLAREYPRREQV
c3393     104  YEVRLEPRMALLTRSRONAIYQ-NQTVPQIVEKITRERHQMRGQDVENIKSEYPAREQV
BCAM0148  103  YQLSFSSWIHFLKFRRDQRHWQ-DKYVDAIIADVFDAHPQARG-MYREFSQPLPSRSYC
BPSS1503  110  YRATLVPRLWRLSLYKASDVYLNEQTIPDIVKRVTRAASFGSR-DFRMRHGGGYRKRSEV

Aec15     147  VQYGESDLAFIARLWAEEGIFFFERFAADSPEQKITLCDDVAGLSQA--------GETPF
c1888     147  VQYGESDLAFIARLWAEEGIFFFERFAADSPEQKITLCDDVAGLSQA--------GETPF
Z0267     147  VQYGETDYDFICRMAAEEGIFFYEEHAQKSTDQSLVLCDTVLYLPES--------FEIPW
AHA1119   148  VQYRETDLDFVNRLAAEEGMFYEHEFBAG--KHRIVFADDAAALTAG--------PETFE
PA1511    146  VQAGDSDHYLHDRLAFEEGLVYYFRFDEHRFTLVCSDRLYVQERIAG--------GPVLE
VP1394    146  LQYRESDLHFVQRMMAEHGMWYYEDHTDSNFTMEIVDSNDAIAPLVSSPLNASYIGPIVY
PA0095    148  VQYRETSFDFVSRLMEQEGIYYWFRHEQSRHILVISDAYGAHQSPPNYA-----SVPYY
BB0793    151  VQYQETDFAFVSRLMEHEGIYYWFRHESGRHTLVITDDITQHDECPGAA------QLPYF
BCAL1294  162  TQYRETDIAFVERLMREDGLFYYFEHAKDGFKLIIADNSVTAKPIDGRS------PSTQY
YPO0970   165  MQYGEDDLTFIRRLIAEVGIWFRFTALPKLNIDVVEFYDDQRFYQQG-------LTLQAV
c3393     163  MQYGEDDLTFVSRLISEVGIWFRFATFARLKIEVYFYDDQSGYERF-------LTLPLR
BCAM0148  161  RQD-ETDWNFVHRLFESEGLYGVFRQAQDGKSHTLVMTDRLQTLEPLAP-------ATVE
BPSS1503  169  CQYDESHLDFVSRWFMEKEGLYYYFEEHFGRHETLVVDDRRHQPGPADDLA-----LRYLP

Aec15     199  NPDTSAGAETECVSMFRYEAHVRPSSVQSQDYTFKVPDWPGMYEQQGESLNGFLEQY---
c1888     199  NPDTSAGAETECVSMFRYEAHVRPSSVQSQDYTFKVPDWPGMYEQQGESLNGFLEQY---
Z0267     199  NPNTRTEVSTLCISQFRYSAQLRPSSVVTKDYTFKRPGWAGRFDQEGQHQDYFRTQY---
AHA1119   198  NLGNRSLEQGPYVRQFHYREAVRPSDVELKDYSFKTPAYGLSRKKVGVELTHQRDTY---
PA1511    198  SAQPEGDNPQPVLHSERYSENVRTARQTQRDYSFKRFTYDQEHHLAGEALEHFGSSY---
VP1394    206  HADSGGVADREHISDLELVNRVRTGQVTYTDYNFEQPKIPQEMTHAG----DLDQDL---
PA0095    202  PPTLEQRERDHFY-DTHMAREVQSGSLSLNDYDFQRPGAR-LEVRSNIARSHAAADY---
BB0793    205  GEDRATVPQEQYVSQTWQVAEETPDGFATVDYDFEKKPAAS-LDAQSSNPGAYEPGGL---
BCAL1294  216  NQG-EAMDNLAVTITSFQAQRQLAPDTVGLRTEDYKIPHARRFVSGGTEVNQGDVPSY---
YPO0970   218  PPSGMHDSGMESVWDLSSAHQVVSTGDYNYRTATADMTAG-ADITRGDTTTYG---
c3393     216  HPSGLFDGETEAVWGLNTAYSVVEKSVSTRDYNYREATAEMTTGQHDATGGDNTTYG---
BCAM0148  213  FSRAGVHGEVDALTQWAGTRTLQSVLLSTRTFDLKNFSTPANPKATSVPTMANQGDLPSQ
BPSS1503  224  ATSLDAGIEADRVQAFTCRATPLPREVVLFDFNHRKAELSLEVRERVARDGVGERVS---
```

164

```
Aec15      256 -EIFDYPGRYK---------DEQHCKDETLYRMESLRSDAEKATGQSNSPKLWPGTWETL
c1888      256 -EIFDYPGRYK---------DEQHCKDETLYRMESLRSDAEKATGQSNSPKLWPGTRETL
Z0267      256 -EVYDYPGRFK--------G-AHGQNEARWQMDGWRNNAEVARGTSRSPEIWPGRRIVL
AHA1119    255 -QHFDFPGRYK---------EDPTGKAPAQHRLDAIRNDAVAGSGKSNSAALLPGQTESL
PA1511     255 -ERYDYPGRYK---------RSGAGRPESESRIRCHRRDARVASVSGDDPRLIPGHAEAL
VP1394     259 -KQFDYPGRYV---------DPVMGQVRTTEWVFEHIVDNQQVEASSDVMRLASGYSENI
PA0095     257 -PLYDYPGEY---------VQSQDEQYARTRIEALQARYERVRLRGRARGLGSGHLEKL
BB0793     261 -QVYEWLGGY---------TEPDQGERYSRIRLEALQAHGESVTGACNVRAFAPGYLETL
BCAL1294   272 -EIYDYLGEHG-------FADSDRGEELARFRTEALAAHSKTFVGVITGRRVMPCRYEEL
YPO0970    274 -EAYHYADNYLTAGSEGREPSESGAFYARLRHERYLNNQARFAGVANAAALSPGQELKV
c3393      273 -EAYHYADNFLQQGDK---EAAESGAFYAIIRHERYLNEQAILKGQSTSSLIMPGLEIKV
BCAM0148   273 AEVYEYTGGYTYP-------EQERGDHLGKIRMEEWESRAKRYDGVGGVRRIDAGRRETL
BPSS1503   281 -SDEHFH------------TKDEGQRYAKLRAEALVCEGRRFAGESTAAGLRAGRFEAL

Aec15      306 TGHPQKM----LNREWQVVQSILSGDQEQALHGSQGRG-------------------------
c1888      306 TGHPQKM----LNREWQVVQSILSGDQEQALHGSQGRG-------------------------
Z0267      305 TGHPQAN----LNREWQVVASELHGEQEQAVPGRRGSG-------------------------
AHA1119    305 TEHPNGS----LNTDWQVVHIRHTGLQEQALEEEGGSGP------------------------
PA1511     305 EGHPRAD----FNAWVRPVRVVHRGTQYAGQEEESADAPL-----------------------
VP1394     309 SDHPRSE----INRDYIMLSVMHTGDEQVHEDEASGMP-------------------------
PA0095     307 SGYPRED----QNREYLVVGAEYRVVQEL---YETGGGGV-----------------------
BB0793     311 RNHPRPA-----ENRQYLIAQAHYRIQEGG---YASG--AE----------------------
BCAL1294   324 DDHYDHASTKQEDRQFLLITVSHTGTNN----YEPG--EG----------------------
YPO0970    333 TGNDVPA---QFGKGVIVTRITSHARRDR-------------------------------
c3393      329 QGDDAPA----VFRKGVLITGVTTSSARDR-----------------------------
BCAM0148   326 TGHPEHDQDAADHREFAVIEVEWIIVNNVPLSGHEANYPHSLYSAVKQADADDPDKLFTVSHDDG
BPSS1503   327 SGHYRQD----FDGRYLVTAITHRGSQAHLLFPDLDAPFGATP-------------------

Aec15      340 --TILGNQLEVIPADRT--WRPRLQS-----KPKVDGPQSAIVTGPAGEEIF--CDEHGRVR
c1888      340 --TILGNQLEVIPADRT--WRPRQQS-----KPKVDGPQSAIVTGPAGEEIF--CDEHGRVR
Z0267      339 --TILDNHEAVIPADRT--WRPQPLL-----KPLVDGPQSAVVTGPAGEEIF--CDEHGRVR
AHA1119    340 --TVYHNEESVKASTT--WRARIGSPEAPHKPMVDGPQIVTVVGPDGEEIY--CDEHGRVK
PA1511     341 -GVSYDLRAELVPEDVE--WREAPLP-----RPRIDGPQIATVVGPAGEEIH--CDEWGRVK
VP1394     344 --TIYYNQETCIPRDVV--FKAPKLA-----APVVDGPQTAVVVGPAGEEIY--TDKLGRIK
PA0095     340 -GAQFESELDCIDAGQA--VRELPTTP----LPIVRGPQTAVVVGPKGEEIW--TDQYGRVK
BB0793     342 -DAVEDIDGRVIPATVP--FRVARATP----VPRTHGPQTATVVGQAGEEIY--TDEYGRVK
BCAL1294   358 -TSAYSCSETCIRKKIP--VREAFEGE----RPTIVGPQTAIVVGPKGEEIY--TDSLGRVK
YPO0970    359 ---SYEVHEEAIPYSEEYCTRPILIN-----KPKVAGTLPARVTSTTVNDTYGHIDKDGRYR
c3393      355 ---SYELTETAIPYSERYGVRPALIP-----RPVVAGTLPARVTSTVKNDIYAHIDKDGRYR
BCAM0148   386 STGFYRVAIEAQRTSVP--YRSPLEH-----RKPEAKLESAIVAGPKGQQAY--TDSLNRIK
BPSS1503   366 GEPIYRAEFEAIAADLQ--YRPPRTTP----KPRAACVVSAIVDGEGGGKLA-ELDEYGQYK

Aec15      391 VKFHWDRYNPATEASSCWVRVSQAWAGPGEGNIAIPRVGQEVIVDFLNGDPDQPIIMGRT
c1888      391 VKFHWDRYHGMTEESSCWVRVSQAWAGPGEGNIAIPRVGQEVIVDFLNGDPDQPLVMGRT
Z0267      390 VKFNWDRYNPSNQDSSCWIRVAQAWAGTGEGNIAIPRVGQEVIVDFLNGDPDQPIIMGRT
AHA1119    396 IQFPWDRYGSSNDQSSCWVRVSQGWAGGQYGMIAIPRIGHEVIVSFLEGDPDQPIITGRT
PA1511     393 VQFPWDREGRHDEFSTCWIRVAQNWAGADFGHIAIPRIGQEVIVDYLDGDCDQPIVTGRT
VP1394     395 IQFHWDRYGNNDEHASCWIRVSQSMAAPTWGAVYLPRIGHEVVVTFLEGDPDRPLVTGAV
PA0095     393 VHFHWDRYGKKNENSSCWIRVSVRVSSPWAGGEGGIQLPRVGDEVIVDFIGGYPDRPIVIGRV
BB0793     395 VHFHWDRYGKKNENSSCWIRVSVRVSSPWAGGEGGIQLPRVGDEVIVDFIGGYPDRPIVIGRV
BCAL1294   411 VQFHWDRLGKRDQGSSCWVRVSQPWAGSGEGMIQIPRIGDEVVVSFLDGNPDRPLVTSRV
YPO0970    413 VNLMEDRDSWEQGYESIWVRQARPVAGDSYGLHLPLLAGTEVATAFEDGNPDRPYIASVL
c3393      409 VNLDFDRDTWKPGYESLWVRQSRPVAGDTYGLHLPLLAGTEVSIAFEEGNPDRPYIAGVK
BCAM0148   444 VLFIWDRHNEGDERASCWVRVAQSDTGDGIGGVHMPRAGEELLIGHIGNDIDRPIALHRV
BPSS1503   421 VRFPEAHTAHPANKASARIRVATPVAGDDRCMHLPLLKRTEVKIAFDGGDPDRPVIVGAV

Aec15      451 YHEDNRSPGSLPGTKTQMTIRSKTYKGSGFN---ELRFEDATGGEQVYIHAQKNMDTEVL-------
c1888      451 YHEDNRSPGDLPGTKTQMTIRSKTYKGSGFN---ELRFEDATDKEQVYIHAQKNMDTEVL-------
Z0267      450 YHQENRTPGSLPGTKTQMTIRSKTYKGSGFN---ELKFEDATGKEQVYIHAQKNMNTEVL-------
AHA1119    456 FHATNRPPYELPANKTRTVLRTETHQGEFN---ELRFEDQAGKEEIYIHGQKDLNVLE-------
PA1511     453 YRATNRPPYSLPDHKILSTIKSKEYKGSRAN---ELRIDDTTAQISAALMSDHGASALHL-------
VP1394     455 YNGLHFPPYSLPENKTRTTFRTQIHKGTGYN---ELEDEANQEEVVIHAQKDMSTKVL-------
PA0095     453 YNAEQTVPYELPANATQSGMKSRSSKGGTPANFNEIRMEDKKGAEQLFIHAEKNQDIEVE-------
BB0793     455 YNASNMPPWDLPGNATQSGFLSRS-KNGDRGTANALMFEDSAGAERIWLHAEERDMDCEVE-------
BCAL1294   471 YNSQNMPPWALPANATQSGILTRSTKTGNVNTANAIRFEDKKGEEEVWLHAEKDQRIEVE-------
YPO0970    473 HDSAHG------DHVTISNYKRNVLRTPSNN---KLRLEDERGKEHIKLSTEYGGKSQFN-------
c3393      469 HDSAHT------DHVTIQNYKRNVLRTPANN---KIRLDDERGKEHIKVSTEYGGKSQFN-------
BCAM0148   504 YNGAAKPRWH--SNALLSGYRSQEFSGSGYN---QMVMDDSTGQNRVHLYSSSTNAALHL-------
BPSS1503   481 -------PNSSHRSVVTRRNPAEHRILTEHN---QIYMKDGSGA-ATWLHAPNNHIGIGAVGPGDGL
```

165

```
Aec15     508 ------NNRTTDVKADHTETIGNDQKITVGLGQTVNVGSKKEG--GHDQKVTVAN-----
c1888     508 ------NDRTTDVKHDHTETIGNDQKITVKGQTVQVGTRKEG--GHDQSITVAN-----
Z0267     507 ------NNRTTDVINNHAEKIGNNQAITVNNQILNIGVNQIQTVGVNQVETVGS-----
AHA1119   513 ------NDAAWHIKHDEHRDIDNERVTRIIKANDHLTVEGEKRDQIKADYSLTVDASLHQK
PA1511    510 ------GYLTHPRPEGGKP---RGEGFELRTDEHGAVRAAKGLLISTEQLRAGAGHLDR
VP1394    512 ------NNRYRDIGQDEFLKVARHQTNEVIGDHKETIDGHKTTQVNSTFIETVEQ-----
PA0095    513 ------NDEHWVGHDRTKTIDHDETVHVIHDRTETVDNNETITIGVDRIEKVGNNEKIS
BB0793    514 ------ANESHTVDGNRTTLIGGNDTITVGTRTTTIDGLDTETFNAGAIRTVTGEVSET
BCAL1294  531 ------HDESHWVGNDRTKNIDHDETVHVGHDRTETVDNNETIHVGDRIETVGNN----
YPO0970   524 ------LGHLVD-----NEKQPRGEGFELRTDSFGAIRAEKGIFISADGQAKAQGQVLEM
c3393     520 ------LGHLVD-----AGKEQRGEGFELRTDLWGAVRAKKGIFISADAQDKAQGQVLEM
BCAM0148  559 GYLIDQNDNARGGYLGGFDISSEAYGALRAGRGLFVTTHPVASQPLDASLATRQLNGSA
BPSS1503  541 ALLTSGNKFDFSIGNAYSFSGGLKCSISLGGNTDVYVGVRNSLDVSANFLTTIQGNLRWM

Aec15     555 ---DQHITIKNDRHKVVNNQTSKVTG---------------------TDTEEVVK-KQ
c1888     555 ---DRRITVRNDQTLKVTNDRTVSVSHDDGLYVR--------------NDRRVTVKGKQ
Z0267     556 ---NQIIKVGSNQVEKVGIIRALTVGVAYQTTVG-------------GIMNTSVALLQ
AHA1119   567 LGQSLLMEAGSEVHHKAGMKIVMEAGAELTLKVGGSFVKLDASGVTLSGGSIKMNSGGSP
PA1511    561 GVVVQVIEAALKLAREIGDYAGEHQGVGHDAAPQQT---------LQEAVRDLGHGAND
VP1394    561 ---DVTVTYNANETQYVKNNSDLEIGDNRTTKIG--------------KNDDLDVGENS
PA0095    567 IGANRTEDVGSNETISIGVDRTEKVGSNEKISIG--------------ANRTEDVGNDE
BB0793    568 TTGNETRTFNGDVTETINGVETRTVNGDWKETIT--------------GEMTETRTGDE
BCAL1294  581 --------------------------------------------------------E
YPO0970   573 QPAISLIKSAQEQMEAISADAQTATASPANLQYQ-----------ISLLQQNLTELKQA
c3393     569 QPALARISAALVTMESIAASAQQAQILAADVSRQ-----------QTLLQQKIEQLQQE
BCAM0148  619 NILDGLSQASETSKAESLKDGHDTFKAFADATQHSESGATGGGGVTAGGGTGDASAFKQP
BPSS1503  597 LPGSRSFEINDSASTLIQTLHKQSATGAIRLSAGQDASALLQKQLDKLKGTVRKFMIVSG

Aec15     589 SIKIGDNYELK------------------VEHG-TNIISGDSIELICGQ-GESGTCSIKL
c1888     597 EHKTTGNHVSL------------------VEGKHSLVVKGDLARKVSGALGIKVDGEIVL
Z0267     598 SSQVGLHKSLM------------------VGMGYSVNVGNNVTFSVGKTMKENTGQTAVY
AHA1119   627 GSGSGGWGGKAP------------------IQPGNVEVPTPPPATLPAPAIHKSYESMAPL
PA1511    611 ESGKSNGGKPA------------------IALSGPAGIAAATPASLTLAAGEHVDSVARQ
VP1394    603 NITVGASKSSD------------------IGADDNQTVGGNLTVSVKCNTSYKADGATQI
PA0095    612 TISIGANRSES------------------VGNNETISIGADRSEVGANETIDIGGNQST
BB0793    613 TRTVTGAVTET------------------ITGDVTQTIGAVTQTQTGAHDITITGDQTS
BCAL1294  582 TLTVGGNRNET------------------VEGMENLLIALTSTEVGLAKALTIGGAYTV
YPO0970   621 VLLLSAPKGIA------------------VSSGEHLQISASENLIATACKNADVSVGKNF
c3393     617 VILGSAPKGVA------------------VVSGEDMQISASDNLTLTACKQLDIGVQKDF
BCAM0148  679 VILMAAPAGIG------------------ISTQKSTHVAADQHVNIVSCQSTHVATGKSL
BPSS1503  657 LANACSAAAAAGLIKGGGKLADLPWAGFGISAAQFAGATGVSTALMATSRTLLANVAKLQ

Aec15     629 EK------------------------------------TGKIIIR-GTEFLFEAT-GPVDI
c1888     639 ES------------------------------------SSRISLKVGGSFVIHS-GGVDI
Z0267     640 SA------------------------------------GEHILELCCGKARLVLTKDGSIFL
AHA1119   669 AKPCPAAPPPVPPPPAGGPPAPPPAPPQLAGGPMMPPPPPVAATGCGKKSAKKWATVEI
PA1511    653 NQQ-------------------------------VTAGQKIVVINAGSDIGIFAQGGEIRQ
VP1394    645 IS--------------------------------------GDKIVLKTGGSSLVMNSDGSIKL
PA0095    654 SIG-------------------------------------KNESRSVGQCRDTSVGKDDSLD
BB0793    655 SIT-------------------------------------GAISHTVTGAYTQTVTDPVTVN
BCAL1294  624 TVA-------------------------------------GAINTAACLASAEEVGLSKTTM
YPO0970   663 FIGVG-----------------------------------NTLSVFVRKLGIKIIANQCRITV
c3393     659 TLAAG-----------------------------------KQLSLYSR-AGAKIFSSHNDIDI
BCAM0148  721 VASIT-----------------------------------EKLSIFVQNACMKIFAAKGKIEV
BPSS1503  717 EALPLVADLSLGKQGIALAAKNLTHATRMSLTVDGVSWSTHAKGPGAIGAAMSVGKGRWG

Aec15     652 KCKDIHLNG--------------------------------------------------
c1888     663 VGPKINLNSGGSPGT-----------------------PVPTLQPTVLKTPGGEKSGDGSD
Z0267     665 NCTHIHLEGESDVNG---------------------DAPVIN----------------
AHA1119   729 SKADEALRYYSAQGYTLLNNYLRDRPYKQREAIDTLLSRSYLNDEPTSASEFDQAMKAYV
PA1511    682 ITHQGPMLLQAQKND-----------------IRLEAEQSVEVSASQQHVLVTAKEHI
VP1394    670 SGSSITIEG-------------------------------------------------
PA0095    679 VGKSFTLNAG------------------------------------------------
BB0793    680 ANTSMTVVTPSWTVS-------------------------------------------
BCAL1294  649 VGKTYTVTAG------------------------------------------------
YPO0970   691 QAQNDLMELLARKAIT------------------------------------------
c3393     686 QAQGGNITTWSTQDTY------------------------------------------
BCAM0148  749 QTHADNVEVTAQKSLLLAS--------------------------VTEKVQASAQQEI
BPSS1503  777 VEAAKHAHVHASDTLLFAVPADPTTQFDLKDLIGLRRDLDECMKDIADLEADISENEVLS
```

```
Aec15         -------------------------------------------------------------
c1888     701 SGEENEDPGGSGLAGSGGGDRGDDEDEPEKYTLQFHFTDDDGIPYS--------------
Z0267     686 --------------WNCGATQPVPDAPVPKDLPPGMPDMRQF-----------------
AHA1119   789 ADVEAGLAKLPASPDLDFVYRGLALDKPELAALKDQFTGVGNIIVEPGFMSTSPDKAWVN
PA1511    723 TLMCGGAYLTLKGGNIELGMPGNFVVKAAKHSHVGPAHASTSFNAWDSTPFDDRYVLRDE
VP1394    679 -------------------SDKVVVKGCNVAIN-------------------------
PA0095    689 ---------------------DSITLVTCAAS---IRMKKDGSIVISGKNITIDGSGAIN
BB0793    695 ----SASQQAFWTANTLRGTPARLTVVGAAADFWGVRQQVYGGINSQWSTVKIDLAAFKN
BCAL1294  659 --------------------DRIERTCKAS---IVLESSGHITISGTSIDILG-----
YPO0970   707 ----------------ITSTEDEIKITAKKKITLNAGGSYITLDENRIESGTAGEYLTK
c3393     702 ----------------ISSGK-KLVVTAQDELTLICGGGYIKIKSGNVEIGGPGKLLVK
BCAM0148  781 LLTSGGAYIRLKGGDIEIHAPGKIDIKGQHAFSGPARMDVTHPAFKDLPTRRLMLNTMA
BPSS1503  837 TDQNTFGVSALIPTPPSPAGALAAVAIKVKQAKLVELKAKQKLVALKVDNLQQKFAKHVQ

Aec15         -------------------------------------------------------------
c1888     747 ----EIRYIA-----FFEDGAQIRGETDKDGYTEVFSRTNDANVEIKLLTNDYYIFEVNC
Z0267         -------------------------------------------------------------
AHA1119   849 DTLLRIRLPAGHGGRLLCEAAHFKGEAEMLFPTQTRLRVDRVVSSMSGDFDSLLNTIPTS
PA1511    783 ATLEPLPNIAVEVIRGDCGVVKLMTDSQGRLPKQQHLAMDPVQIRILGKGSHNSDTESNT
VP1394        -------------------------------------------------------------
PA0095    725 -------VKADKNVVVKCRKILQN----------------------------------
BB0793    751 GSNGFESGQAGAQIKAICAQIKTGGAAVISRVINLFT---------------------
BCAL1294  690 ---------SDAVKVDCKTVDLN----------------------------------
YPO0970   750 --------AGYYGRQEKANKPEDFPS----VAPETTEPTSHFTFS--------------
c3393     744 NAGIKKAGSGSMQGVMKSFEPESFDENFIISNPASGKPIANQKYEIHMPDGSVIAGITDA
BCAM0148  841 SPSATRVVPVGMPYKLYADGALVKQGVFDKTGQLPIDHQVTTQKYTLEMANGVKHEIPVP
BPSS1503  897 HLSAVRMSASDAQLGFKCNRLVATADGVTLAHAQGKAKLDVREAKIGVTAGKSSVELDEG

Aec15         --------------------------------------------------
c1888     798 NEHQ----------------------------------------------
Z0267         --------------------------------------------------
AHA1119   909 DNASDNRSRIKRLIEVSVL-------------------------------
PA1511    843 --------------------------------------------------
VP1394        --------------------------------------------------
PA0095        --------------------------------------------------
BB0793        --------------------------------------------------
BCAL1294      --------------------------------------------------
YPO0970       --------------------------------------------------
c3393     804 MGYSSLVASKASDDLKIILKK-----------------------------
BCAM0148  901 GEYRDAENGKLANQGFQYHEGAPDDGAAPDRATHRQIYNELLSPSSEA---
BPSS1503  957 KIAAGCGSASLKLGSDGAIDVSATDVKLNGTNVKLNGSASLKFDGQLIQLG
```

Figure 3.22. Amino acid sequence alignment of TssI subunits from various Gram-negative bacteria. Amino acid sequence alignment of TssI from various bacteria including BCAM0148. gp27 and gp5 homology regions used for BACTH experiments are enclosed with blue and red boxes, respectively. Residues in white font with black shading indicate identical amino acids at the corresponding position in ≥ 50% of sequences shown. Residues in white font with grey shading indicate similar amino acids at the corresponding position.

Figure 3.23. Analysis of interaction of the gp27.gp5-like region of TssI with itself and with full-length TssI using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of the gp27.gp5-like region of TssI and itself or full-length TssI to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combinations of the gp27.gp5-like region and TssI fusion proteins that gave rise to a maltose-positive phenotype. B. Combination of gp27.gp5 fusion proteins that gave rise to a maltose-positive phenotype.
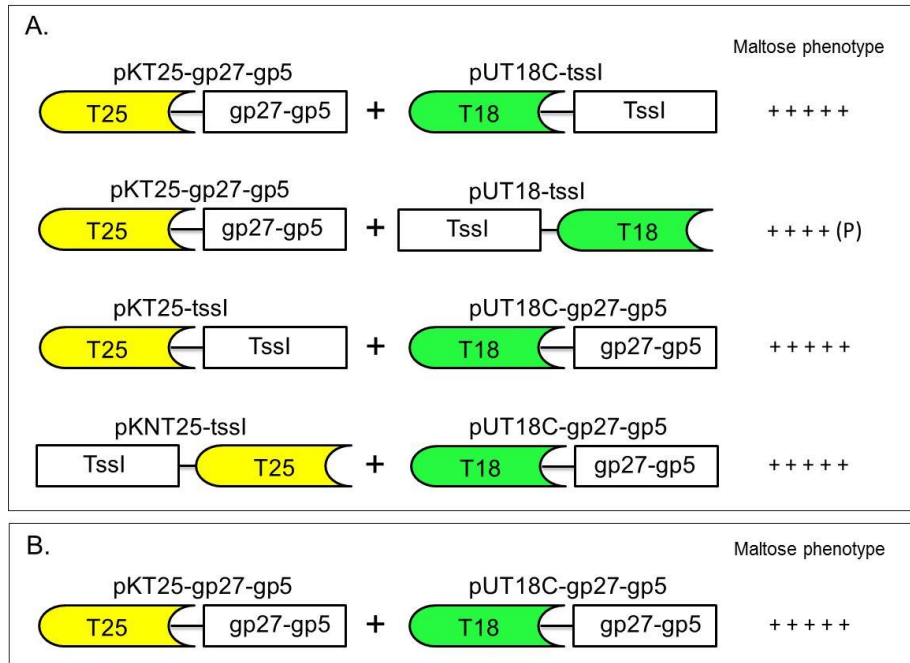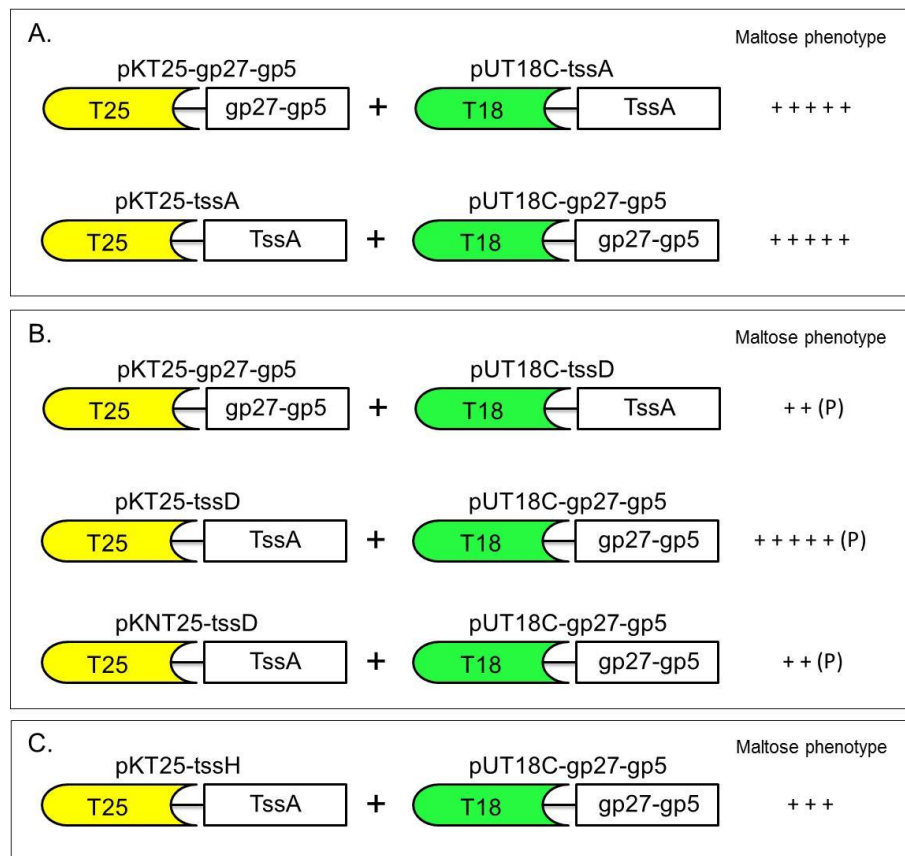
Figure 3.24. Analysis of interaction of the gp27.gp5-like region of TssI with TssA, TssD and TssH using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of the gp27.gp5-like region of TssI and TssA, TssD and TssH to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combination of the gp27.gp5like- region of TssI and TssA fusion proteins that gave rise to a maltose-positive phenotype. B. Combination of gp27.gp5-like region and TssD fusion proteins that yielded a maltose-positive phenotype. C. Combination of gp27.gp5-like region and TssH fusion proteins that yielded a maltose-positive phenotype.
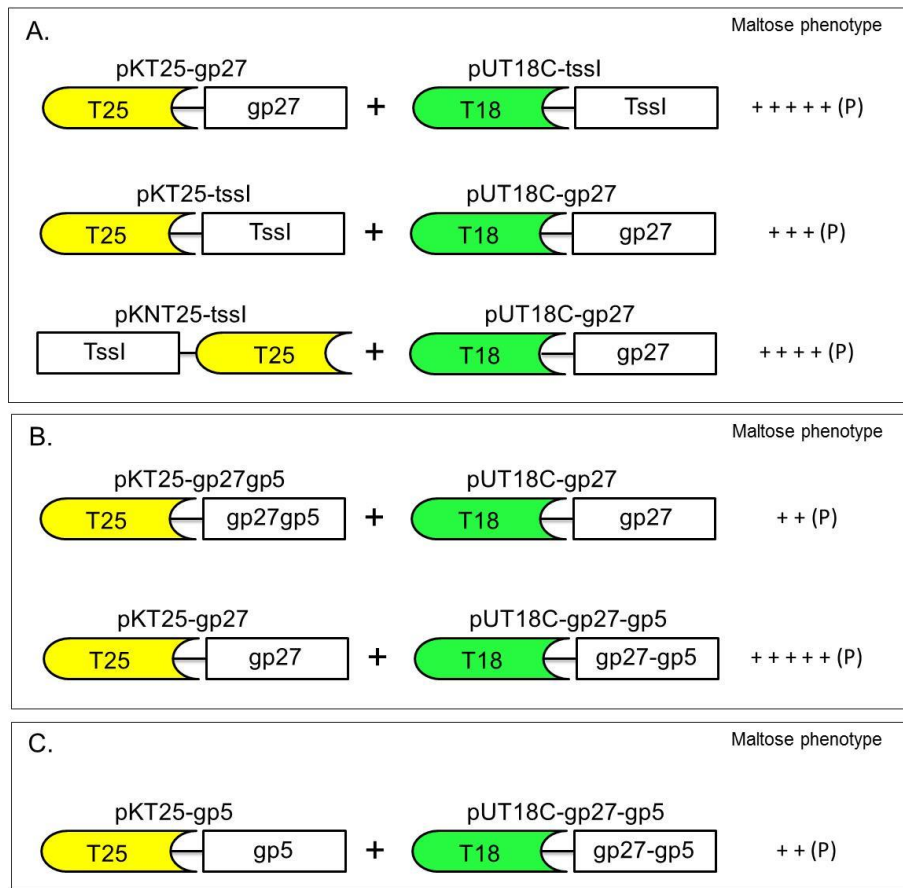
Figure 3.25. Analysis of interaction between the gp27- and gp5–like regions of TssI and full-length TssI and its gp27- and gp5-like regions using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of the gp27- and gp5-like region with gp27.gp5 or TssI to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combination of the gp27-like region of TssI and TssI fusion proteins that gave rise to a maltose-positive phenotype. B. Combination of fusion protein containing the gp27-like region of TssI and the gp27.gp5-like region of TssI fusion proteins that gave rise to a maltose-positive phenotype. C. Combination of fusion protein of the gp5-like region and the gp27.gp5-like region that yielded a maltose-positive phenotype.

3.8. Discussion

A large scale experiment was conducted to screen for protein-protein interactions between the 13 Tss proteins components of the T6SS of *B. cenocepacia.* Results from this comprehensive analysis provide useful information regarding potential protein-protein interactions within the T6SS and will contribute towards building a model of the T6SS. Figure 3.26 presents the results of the BACTH assay for all 13 subunits of the T6SS by combining the results of Shastri (2011) with those obtained in this work. Figure 3.27 presents a summary of the protein-protein interactions shown in Figure 3.26 along with those reported by other groups.

It is worth mentioning that the Mal$^+$ phenotype was observed more frequently when pKT25-Tss was combined with pUT18C-Tss i.e. the N-terminal cya domains and the C-terminal insert. Also, as the BACTH system is not quantitative, β-galactosidase assays are recommended to obtain quantitative data, particularly for weak Mal$^+$ phenotype, where a relevant interaction has been inferred.

The BACTH system is a powerful tool to screen for protein-protein interactions *in vivo*. The limitations of the BACTH system include the possibility of false positive interactions, as some proteins may have an intrinsic tendency to interact with other proteins i.e. they are "sticky proteins". False negative results may also occur, due to the fact that the fusion of the protein to the T18 or T25 domains may cause the protein to misfold, to be unstable or simply prevent the interaction with its partners. Also, the fact that the two hybrid proteins are expressed from two compatible plasmids; low copy number plasmid (pKT25 or pKNT25) and high copy number plasmid (pUT18C or pUT18); may lead to different relative amounts of the two hybrid proteins being produced. This may cause further problems and therefore the interaction has to be tested in the two possible combinations of T18 and T25 fusions (Battesti and Bouveret, 2012).

3.8.1. Self-association of T6SS subunits

The two-hybrid analysis of Shastri (2011) suggested that TssA, TssB, TssD, TssE, TssF and TssG self-associated (i.e. they form dimers or oligomers). The results for TssA, TssD and TssE were confirmed in this work. In addition, BACTH assay results presented here also suggest that TssH, TssI, TssK, TssL and TssM$_{CTD}$ self-associate. These results suggest that most of Tss proteins assemble as multimers in the T6SS

apparatus. While evidence for the oligomerisation of TssD has been provided by several groups, including three separate structural determinations (Section 1.5.1.1), no evidence for TssA or TssE oligomerisation has been provided a part from the analysis of Shastri (2011) that was also confirmed here.

3.8.1.1. Self-interaction of TssH

TssH (ClpV) is a member of the ring-forming Clp/Hsp100 proteins, which belong to an AAA$^+$ ATPase superfamily (*A*TPase *a*ssociated with a variety of cellular *a*ctivities). Like ClpB, TssH consists from four domains but with the addition of an N-terminal extension known as the N-domain (Figure 1.12). Screening of full-length TssH in the two-hybrid assay indicated that TssH can self-interact. This result agrees with the current model of TssH based on the structure of the closely related ClpB protein in which ClpB monomers are arranged in a hexameric ring.

To determine which region of TssH is able to self-interact, pairwise combinations of individual TssH domains were screened in the two-hybrid system. Interestingly, the NBD2 domain (the second ATPase domain, located at the C-terminus of TssH) was found to interact with full-length TssH protein and with NBD2 (Jones.R.A, 2013). This suggests that the oligomersation of TssH into hexamers is mediated through the NBD2 domain. This result is in agreement with the observation of Mogk and colleagues who showed that the C-terminus of NBD2 is essential for the formation of the ClpB hexamer (Mogk *et al*., 2003). For TssH domain organisations see Figure 1.12.

3.8.1.2. Self-interaction of TssI

TssI (VgrG) is structurally homologous to the bacteriophage T4 tail spike. The tail spike is composed of trimer of gp27.gp5 heterodimers (Pukatzki *et al*., 2007). Consistent with the oligomeric status of the T4 tail spike, when pairwise combinations of plasmids encoding TssI fusion proteins were co-expressed, a Mal$^+$ phenotype was observed in all four possible two-hybrid combinations.

Bioinformatics and biochemical analysis of VgrG-related proteins indicate that the N-terminal and central regions of TssI share structural homology with the gp44 protein of bacteriophage Mu, which is a structural orthologue of the bacteriophage T4 tail-spike proteins gp27 and gp5. The crystal structure of the N-terminal region of c3393, a TssI

from uropathogenic *E.coli* strain CFT073, is strikingly similar to the trimeric structure of the (gp27.gp5)$_3$ complex (Leiman *et al.*, 2009). The results of the TssI BACTH assays showed that the self-interaction of TssI can be detected by this type of assay and suggests that the *B. cenocepacia* TssI encoded by BCAM0148 may also form trimeric complex analogous to the gp27.gp5 trimeric complex.

TssI orthologues share the gp27.gp5 homologous region at their N-termini but diverge C-terminally to this region due to the presence of different C-terminal effector domains in evolved TssIs (Section 1.5.1.2). To confirm that the self-interaction of TssI was occurring by means of the gp27.gp5-like region, the gp27 and gp5 corresponding regions of *B. cenocepacia* TssI were identified based on multi-sequence alignment of a variety of TssI's from different Gram-negative bacteria, including BCAM0148 of *B. cenocepacia* (Figure 3.22). The involvement of the gp27.gp5, gp27 and gp5 regions of BCAM0148 TssI in TssI oligomerisation was analysed using the two-hybrid system. The results were consistent with the previously observed trimerisation of the gp27.gp5-like region of TssI from uropathogenic *E. coli* (Leiman *et al.*, 2009).

 Intriguingly, when individual gp27 and gp5 regions were tested with the fused gp27.gp5 region and with full-length TssI, five combinations involving gp27, two with gp27.gp5 and three with full-length TssI, gave rise to a maltose-positive phenotype. However, only one combination of gp5 and gp27.gp5 showed Mal$^+$ phenotype. This suggests that the gp27 moiety may be more important for the trimerisation of TssI than the gp5-like component. Furthermore, when gp27 and gp5 were tested for their ability to self-associate or cross-interact, a maltose-positive phenotype was not observed.

3.8.1.3. Self-interaction of TssJ

TssJ is an outer membrane-anchored periplasmic lipoprotein that forms one of the 13 core components of the T6SS. The crystal structures of TssJ from one of the T6SSs of the enteroaggregative *E. coli* (EAEC) strain (where it is known as SciN) and *Serratia marcescens* strain Db10 (where it is known as Lip) were recently determined (Rao *et al.*, 2011, Felisberto-Rodrigues *et al.*, 2011). Screening of TssJ-TssJ protein interaction via BACTH assays did not provide evidence for self-interaction. This result agrees with the recent data published on the TssJ of EAEC crystal structure which showed that TssJ is a monomer in solution (Felisberto-Rodrigues *et al.*, 2011). On the other hand, the crystal

structure of the TssJ of *S. marcescens* shows a tetrameric assembly probably consist from a dimer. However, if the N-terminus domain of the TssJ is anchored into the outer membrane, the tetramer structure will be incompatible with dimeric or tetrameric structure. Therefore, the authors proposed that the tetramer is likely to be a crystallographic artefact (Rao *et al.*, 2011).

3.8.1.4. Self-interaction of TssK

TssK is cytosolic protein and an essential component of the T6SS (Zoued *et al.*, 2013). The results of the BACTH assays revealed that TssK self-associates (Section 3.3) and therefore forms a dimer or oligomer. Indeed, this conclusion has been confirmed by SEC-MALLS (Size Exclusion Chromatography Multi-Angle Laser Light Scattering) analysis of TssK which showed that TssK is trimer (Sun and Thomas, unpublished data). Up to date study confirm the TssK trimerisation by MALS/QELS/UV/RI (multi-angle laser light scattering/quasi-elastic light scattering/absorbance/refractive index detectors) and SAXS (Small-angle X-Ray scattering) analysis (Zoued *et al.*, 2013).

3.8.1.5. Self-interaction of TssL

TssL shares amino acid sequence homology and membrane topology with the IcmH/DotU component of the T4SSb (Ma *et al.*, 2009b). Also the TssL and IcmH/DotU have similar secondary structure prediction and belong to the DUF2077 family (Pfam PF09850) (Durand *et al.*, 2012). Recently, Durand *et al* reported the crystal structure of TssL from EAEC and showed that it forms a dimer (Durand *et al.*, 2012). Consistent with these data, our two-hybrid analysis yielded a strong Mal$^+$ phenotype only when T25-TssL was co-expressed with T18-TssL. Therefore, the C-terminal region of TssL in both fusion proteins is free. Analysis of the crystal structure of the cytoplasmic domain of the TssL from EAEC for possible packing interactions showed that the largest contact area of the TssL dimer involves the helix 8 residues and the amino acids 148-177 of the linker region (the TssL cytoplasmic domain encode amino acids 1-183) (Durand *et al.*, 2012). This contact area is located at the C-terminal region of the TssL cytoplasmic domain, and therefore corresponds with the free C-terminal end of the TssL fusion protein which is involved in the TssL self-interaction observed in the BACTH assay.

3.8.1.6. Self-interaction of TssM

TssM shares amino acid sequence similarity and some of the characteristics of the IcmF component of the T4SSb. It was shown that TssM is essential for the secretion of TssD and it also interacts with TssL and TssJ to form the transmembrane sub-complex of the T6SS (Zheng and Leung, 2007, Ma *et al.*, 2009b, Felisberto-Rodrigues *et al.*, 2011). For the BACH analysis, TssM was divided into 2 fragments: TssM$_{NTD}$ (cytoplasmic domain, amino acids 58-445) and TssM$_{CTD}$ (periplasmic domain, amino acids 466-1314) to avoid inclusion of the TMDs. The result of the two-hybrid analysis suggests that TssM might self-associate through its periplasmic domain and thereby provide the periplasmic outer shell of T6SS complex. However, a recent study by Feliberto *et al.* (2011) showed that TssM$_{CTD}$ of EAEC T6SS-1 is a monomer in solution and the authors suggested that the TssM self-interaction might take place through their transmembrane helices (Felisberto-Rodrigues *et al.*, 2011).

3.8.2. Interactions between different T6SS subunits

The 13 core components of T6SS can be divided into three groups based on the bioinformatics analysis. The first group includes genes encoding the membrane-associated proteins (TssL-TssM-TssJ) that anchor the injection machinery to the cell envelope. The second group include genes encoding bacteriophage-like components (TssB-TssC-TssD-TssE-TssI) that constitute the injection machinary. TssH may also be considered as a component of the injection machinery as it is required to the contracted sheath disassembles. The third group include genes encoding proteins with unknown function (TssA, TssF, TssG, TssK) (Silverman *et al.*, 2012). This group probably comprises elements of the baseplate. The discussion of the interactions between different Tss proteins will use this classification.

3.8.2.1. Interactions between TssH and other T6SS subunits
3.8.2.1.1. Interactions between TssH and the membrane-associated proteins

Screening of TssH interactions with TssL in the two-hybrid assay indicated that the two proteins may interact (Section 3.4.1.11). It is noteworthy that in both positive combinations the C-terminal end of both proteins was free which might suggest that the C-terminal region of TssH and TssL are involved in the observed interaction between the two proteins. Interaction between TssH and TssJ were not detected when screened by two-hybrid assays. It is not surprising that no evidence was obtained for a TssH-TssJ

interaction, as TssJ is located in the periplasmic space whereas TssH is a cytoplasmic protein. Although TssH and TssM$_{NTD}$ are both present in the same cellular compartment there does not appear to be an interaction between these two proteins.

3.8.2.1.2. Interactions between TssH and the bacteriophage-like proteins

TssH was previously shown to be required for disassembley of the T6SS tail sheath-like tubule composed of polymerised TssB and TssC subunits (Kapitein *et al.*, 2013). Accordingly, a maltose-positive phenotype was observed in three combinations of TssH and TssB fusion proteins in the two-hybrid assays (Section 3.4.1.2). In agreement with this result a recent study in *V. cholerae* O1 strain A1552 showed an interaction between the N-terminus of TssH and TssB using the yeast two-hybrid assay, but the bacterial two-hybrid assay in that study was unable to detect that interaction. In addition, the N-terminus of TssH from the same strain of *V. cholerae* was found to interact with two TssB homologues encoded by *P. aeruginosa* and *Y. pseudotuberculosis*. Interestingly, when full-length TssH was used the interaction did not occur, which may be due to the low level of expression of the latter construct (Broms *et al.*, 2013). Also, Pietrosiuk and colleagues were unable to detect the interaction between full-length TssH and TssB (Pietrosiuk *et al.*, 2011).

On the other hand, the BACTH assay did not provide any evidence for an interaction between TssH and TssC. In agreement with this result, Broms and colleagues were unable to detect interaction between TssH and TssC using bacterial-two hybrid or yeast-two hybrid assays (Broms *et al.*, 2013). However, an interaction between TssH and TssC was demonstrated in other study. They showed that the TssH$_{N-domain}$ specifically binds to the N-terminal region of TssC and the interaction requires that both proteins are in oligomeric assembly (Pietrosiuk *et al.*, 2011). This might explain our result as TssH was co-expressed with TssC without TssB, therefore TssC is a monomer under these conditions. Moreover, TssC was found to degrade rapidly in absence of TssB (Broms *et al.*, 2013). Implementation of a trihybrid assay, where unfused TssB is coexpressed with TssC and TssH two-hybrid fusion proteins may allow detection of such interactions.

The BACTH assay gives a positive result for combinations of TssH with the tail tube protein TssD where the C-terminus of both proteins is free (Section 3.4.1.4).

As far as the tail spike protein, TssI, is concerned, a maltose-positive phenotype was observed in three pairwise combinations with TssH (Section 3.4.1.8), but when TssH fusion proteins were tested with the gp27.gp5 region of TssI the positive maltose phenotype was observed in only one combination. Also, TssH was found to interact with TssI BCAS0667 and TssI BCAL1294 of *B. cenocepacia* (Jones, 2013). If TssH is required for disassembly of the contracted TssB-TssC tail sheath following firing of the T6SS, it is not clear why TssH contacts TssD and TssI. These results might suggest that the translocation of TssI driven by the ATPase activity of TssH.

A very weak maltose-positive phenotype was observed in one pairwise combination of TssE, the putative baseplate component, and TssH fusion proteins (Section 3.4.1.5).

3.8.2.1.3. Interactions between TssH and T6SS proteins of unknown function

Screening for possible interactions between TssH and this group of proteins revealed a couple of interesting interactions. Two pairwise combinations of TssH and TssA fusion proteins gave rise to a strong maltose-positive phenotype when the C-terminal end of both proteins is free, suggesting that the two proteins may interact through their C-terminal regions. Indeed, the BACTH assays of TssA domains with TssH showed that $TssA_{CTD}$ gave rise to strong Mal^+ phenotype. In addition, the BACTH assays of TssA with TssH domains suggested that $TssH_{NBD2}$ and $TssH_{Ndom}$ interact with full-length TssA, as they gave rise to a strong maltose-positive phenotype in BACTH assays. Testing TssA domains with TssH domains fusion proteins gave rise to a weak Mal^+ phenotype between $TssA_{CTD}$ or $TssA_{NTD}$ with $TssH_{NBD2}$ and a moderate Mal^+ phenotype between $TssA_{CTD}$ or $TssA_{NTD}$ with $TssH_{Ndom}$ which may indicate that the domains of TssA and TssH are less stable than full-length proteins. TssA is predicted to be a cytoplasmic protein and is likely to be a baseplate component (Section 6.17). It is possible that the disassembly of the contracted tail sheath by TssH occurs at the baseplate.

The BACTH assay showed a weak to moderate maltose-positive phenotype between TssH and TssF in three pairwise combinations. As the function of TssF is unknown the relevance of TssH interaction with TssF is uncertain and the patchy appearance may indicate transient interaction. TssF was demonstrated to interact with TssA, -D, -E and –I and so may also be a baseplate component.

Maltose-negative phenotypes were observed in BACTH assays of TssH and TssG. In fact, a maltose-negative phenotype was observed in all pairwise combinations of TssG and all other Tss fusion protein.

Lastly, TssH was found to interact with TssK in three pairwise combinations of BACTH fusion proteins, (Section 3.4.1.10). The observations that TssK interacts with TssA and TssL suggests that it may be a component of the baseplate or it may tether the baseplate to the membrane-associated complex. As $TssH_{NBD2}$ is involved in a bit of interactions, the results could be artifactual because NBD2 of ClpB interacts with a lot of proteins to disaggregate them.

3.8.2.2. Interactions between TssI and other T6SS subunits
3.8.2.2.1. Interactions between TssI and the membrane-associated proteins

A trimer of TssI subunits forms the tip of the T6SS puncturing device which shares structural similarity with the gp27.gp5 tail spike complex of the bacteriophage T4 (Section 1.5.1.2). A transient interaction of TssI with the membrane-associated proteins might be expected as the TssD-TssI needle complex is projected through the cell envelope on exit from the cell. Some interactions between TssI and membrane-associated T6SS subunits were observed in the BACTH assays. Firstly, one combination of TssI and TssJ fusion proteins gave rise to colonies exhibiting a moderate maltose-positive phenotype. It is possible that the TssJ interacts with TssI in the periplasm as TssJ is anchored in the outer membrane. A weak to moderate Mal$^+$ phenotype was observed in three pairwise combinations of TssI and the cytoplasmic domain of TssL. The TssI-TssL cytoplasmic interaction might help to stabilise the baseplate complex. However, a maltose-positive phenotype was not observed when TssI was tested with the two TssM domains in the BACTH assays.

3.8.2.2.2. Interactions between TssI and the bacteriophage-like proteins

As TssI corresponds to the bacteriophage T4 tail spike, the interaction of TssI with other bacteriophage tail tube proteins would be expected. Accordingly, the BACTH assays identified several interesting interactions between TssI and other bacteriophage-like proteins.

Based on the current model of T6SS, TssI and TssD form a dynamic bacteriophage-like structure, where TssI probably acts as the cell puncturing device located at the tip of the TssD tail tube (Silverman *et al*., 2012). TssI is predicted to interact with the TssD protein via its β sandwich domains, as the TssI trimeric complex would form a hexamer of β sandwich domains which could interact with the hexameric ring of TssD tail tube (Cascales and Cambillau, 2012). The BACTH assay of TssI and TssD fusion proteins showed a moderate to strong maltose-positive phenotype in two pairwise combinations. Moreover, testing the gp27.gp5-like region of TssI with TssD by BACTH assay also gave rise to a moderate to strong maltose-positive phenotype in three pairwise combinations, which are consistent with the predicted interaction of β sandwich domains of the TssI gp27-like region with TssD. In agreement with this result, TssI BCAL1294 and TssI BCAS0667 have been previously shown to interact with TssD in the BACTH assay (Jones.R.A, 2013). As discussed in Section 1.5.1.1, TssI and TssD showed mutual dependence in order for them to be secreted (Zheng and Leung, 2007). This may indicate that TssI cannot be injected without TssD as it sits at the top of the TssD tube and is pushed out by it; and TssD cannot exit the cell and puncture another cell without TssI.

TssE is shares significant amino acid sequence similarity (about 40%) with the bacteriophage T4 gp25 component (Leiman *et al*., 2009, Lossi *et al*., 2011). TssE is a structural component of the bacteriophage T4 baseplate (Kostyuchenko *et al*., 2003). In bacteriophage T4, gp25 was shown to interact with the $(gp27.gp5)_3$ complex at rest or gp19 (TssD) in the activated state (Kostyuchenko *et al*., 2003, Kostyuchenko *et al*., 2005, Leiman *et al*., 2010). If TssE functions as the baseplate protein like gp25, then would be expected to interact with TssI and TssD. Consistent with this, one pairwise combination of TssI and TssE fusion proteins gave rise to a moderate maltose-positive phenotype (Section 3.4.2.5). Therefore, these results support the theory of TssE being structure and functional analogue of bacteriophage T4 gp25 protein.

TssB and TssC are predicted to form the tail sheath components of the T6SS. In fact, TssB/TssC form tubules with a similar organization to the bacteriophage tail sheath (polymerised gp18), (Section 1.5.1.3). The BACTH assay showed a moderate maltose-positive phenotype in two pairwise combinations of TssI and TssB fusion proteins (Section 3.4.2.2). In both combinations, the C-terminal end of TssI was free, which

suggest that the C-terminal end of TssI may play an important role in the observed interaction with TssB. However, the BACTH assay of TssI and TssC fusion proteins gave rise to only a very weak Mal$^+$ phenotype in one pairwise combination, (Section 3.4.2.3). This might be due to the instability of TssC when expressed without TssB. Also, the interaction between TssB and TssC promotes conformational changes which might favour the interaction with TssI. The tri-hybrid assay can be used to test this hypothesis. It is possible that TssI is transiently interacting with TssBC complex or it triggers the interaction between TssB and TssC.

3.8.2.2.3. Interactions between TssI and the T6SS proteins with unknown function

TssI was found to interact with TssA, TssF, and TssK in the BACTH assays (Section 3.4.2.1, 3.4.2.6, 3.4.2.9, respectively). Three pairwise combinations of TssI and TssA fusion proteins gave rise to a strong maltose-positive phenotype. In all three positive combinations the C-terminal region of TssA was free. Indeed, the two-hybrid analysis of individual TssA domain with TssI showed that TssA interacts with TssI via the C-terminal domain (TssA$_{CTD}$). Also, the gp27.gp5 like domain was found to interact with TssA.

The BACTH assay of TssI and TssF fusion proteins gave rise to a moderate maltose-positive phenotype in three pairwise combinations. A similar result was obtained in two pairwise combinations of TssI and TssF$_{NTD}$ fusion protein indicating that TssF$_{NTD}$ is involved in the interaction with TssI. In addition, the BACTH assay of TssI and TssK fusion proteins gave rise to a weak maltose-positive phenotype in two pairwise combinations. As TssA also interacts with TssF and TssK, and TssF interacts with TssK, these results might indicate that TssA, TssF and TssK form a complex in the cytoplasm (as they are predicted to be cytoplasmic proteins) which interacts with TssI.

3.8.2.3. Interactions between TssJ and other T6SS subunits

Localisation experiments demonstrate that TssJ is outer membrane-anchored lipoprotein which is located on the inner (i.e. periplasmic) face of the membrane (Aschtgen *et al.*, 2008). In addition, TssJ was shown to interact with TssM$_{CTD}$ using the yeast two-hybrid system and by SPR (Zheng and Leung, 2007, Felisberto-Rodrigues *et al.*, 2011). In this study, the BACTH system failed to reveal the observed interaction between TssJ and TssM$_{CTD}$. In fact, the only pairwise combination involving TssJ which yielded a Mal$^+$

phenotype TssJ-T25 and T18-TssI combination (Section 3.4.2.8). As TssJ is a small protein, it is possible that fusion with T18 or T25 interferes with the interaction of TssJ with other Tss proteins due to steric hindrance.

3.8.2.4. Interactions between TssK and other T6SS subunits
3.8.2.4.1. Interactions between TssK and the membrane-associated proteins
A weak to moderate maltose-positive phenotype was observed in three pairwise combinations of TssK and TssL fusion proteins (Section 3.4.4.4). In agreement with this result, TssK was shown to interact with TssL by immunoprecipitation (Zoued *et al*., 2013). It is possible that TssK along with other Tss proteins forms a sub-complex which might constitute the baseplate of T6SS or an extension of the membrane-spanning unit of the T6SS into the cytoplasm which surrounds the baseplate. Another hypothesis suggested that TssK might function as a connector between the membrane and phage-like sub-complexes (Zoued *et al*., 2013).

3.8.2.4.2. Interactions between TssK and the bacteriophage-like proteins
In addition to the interaction of TssK and TssI proteins describe above, TssK was shown to interact with TssD (the tail tube protein) in the BACTH assays (Section 3.4.4.2). Three pairwise combinations of TssK and TssD fusion proteins gave rise to a weak to moderate maltose-positive phenotype. This result was confirmed in recent study on the T6SS-1 gene cluster in EAEC which showed that TssK interacts with TssD (Zoued *et al*., 2013). Therefore, this result is in agreement with the hypothesis of TssK being a part of a sub-complex in the cytoplasm; and the interaction of TssK and TssD might trigger the formation of TssD tube or TssD might pass through this structure.

3.8.2.4.3. Interactions between TssK and other T6SS subunits
In this study TssK was found to interact with TssA and TssF using BACTH assays. A strong maltose-positive phenotype was observed in two pairwise combinations of TssK and TssA fusion proteins (Section 3.4.4.1). The BACTH result of TssK with $TssA_{CTD}$ showed a moderate maltose-positive phenotype which might suggest that the interaction between TssK and TssA is mediated via $TssA_{CTD}$ see section 3.5.7. On other hand, a moderate to strong maltose-positive phenotype was observed in four pairwise combinations of TssK and TssF fusion proteins (Section 3.4.4.3). Therefore, as discussed above, TssA, TssF and TssK may form a cytoplasmic sub-complex structure

which acts as the T6SS baseplate which envelope the TssD and TssI "syringe needle". Zoued *et al* showed that TssK interacted with TssA.

3.8.2.5. Interactions between TssL and other T6SS subunits
3.8.2.5.1. Interactions between TssL and the membrane-associated proteins

The interaction of TssL and TssM was the first documented protein-protein interaction between different T6SS subunits (Zheng and Leung, 2007, Ma *et al*., 2009b). In agreement with the work carried out by Ma *et al*., the BACTH assay of TssL and TssM$_{NTD}$ fusion proteins gave rise to a strong Mal$^+$ phenotype in two pairwise combinations. In both of these, the TssL C-terminal region was free (unfused). The TssL, TssM and TssJ proteins are predicted to form a membrane-spanning complex which acts as the outer shell of T6SS, as shown in Figure 3.28.

3.8.2.5.2. Interactions between TssL and the bacteriophage-like proteins

TssL was shown to interact with TssB (a component of the tail sheath) and TssD (the tail tube protein) using the BACTH system. The two positive pairwise combinations of TssL and TssB fusion proteins gave rise to a weak maltose-positive phenotype. A very weak Mal$^+$ phenotype was observed in one pairwise combination of TssL and TssD fusion proteins (Sections 3.4.5.2 and 3.4.5.4, respectively). This result might suggest that the TssL protein could be involved in the recruitment of TssD to form the tail tube.

3.8.2.5.3. Interactions between TssL and other T6SS subunits

TssL was found to interact with two Tss proteins of unknown function, TssA and TssF. Two pairwise combinations of TssL and TssA fusion proteins gave rise to a strong Mal$^+$ phenotype (Section 3.4.5.1). The interaction of these two proteins was observed when the C-terminal end of both proteins is free, which might suggest that the interaction is mediated through the C-terminal domains of both proteins. However, TssA$_{CTD}$ did not produce a positive result in the two-hybrid assay with TssL. A TssL and TssF fusion protein interaction was observed in three pairwise combinations with moderate to strong maltose-positive phenotypes (Section 3.4.5.1). If TssA and TssF along with other Tss proteins form a cytoplasmic sub-complex, then this sub-complex might interact with TssL which is anchored in the inner membrane. As TssL interacts with -A, -F and –K, TssA interacts with –F, -K and –L and TssK interact with –A, -F and –L, these results suggest that they are in close proximity near the IM as TssL is IM-anchored (Figure 3.28).

Figure **3.26. T6SS subunit interaction matrix based on pairwise combinations of BACTH plasmids encoding full-length TssA to TssI and TssK, and the soluble components of TssJ, TssL and TssM.** The cytoplasmic N-terminal and periplasmic C-terminal regions of TssM are considered separately. Results shown in red font are from Shastri (2011). Results in black font are the result of combinations tested in this study. [symbol] pKT25-Tss, [symbol] pKNT25-Tss, [symbol] pUT18C-Tss, [symbol] pUT18-Tss, (+) Mal$^+$ phenotype, (+^{P}) Mal$^+$ phenotype with patched pink/red colonies, (-) Mal$^-$ phenotype, (ND) not done, MN= TssM$_{NTD}$, MC= TssM$_{CTD}$.
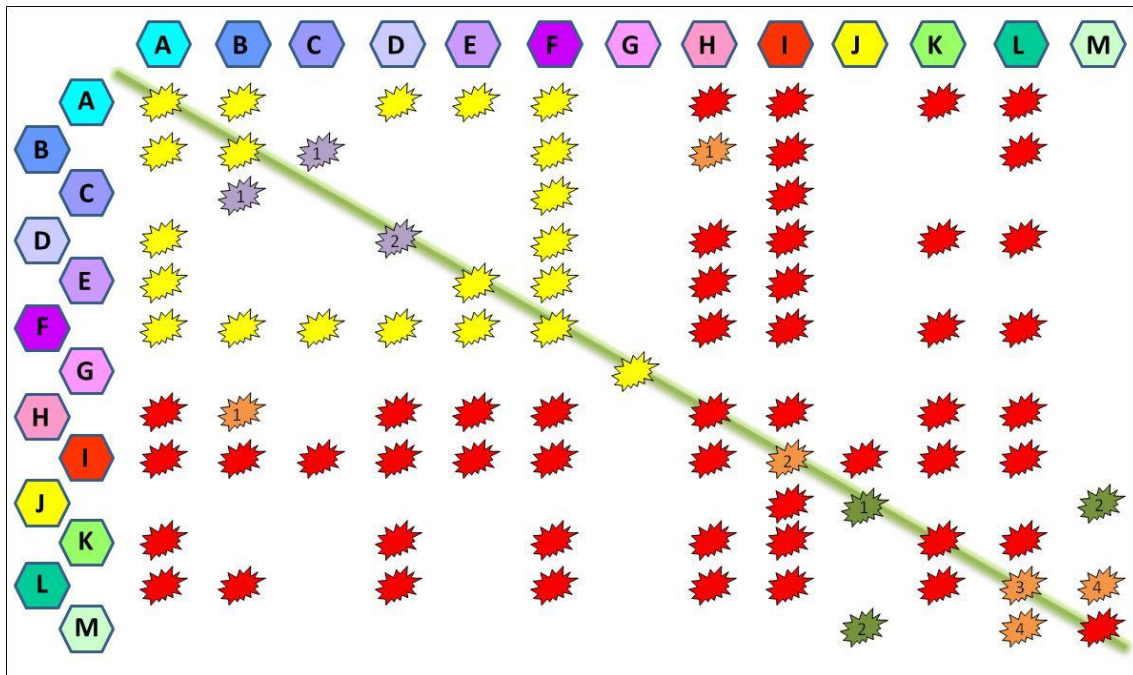
Figure **3.27. Potential protein-protein interactions within the T6SS.** Protein-protein interaction reported by Shastri (2011), obtained in this work, previously reported interactions confirmed in this work, previously reported interactions confirmed by Shastri (2011), previously reported interactions not detected in this work. 1. (Bonemann *et al.*, 2009). 2. (Mougous *et al.*, 2006). 1. (Broms *et al.*, 2013). 2. (Pukatzki *et al.*, 2007). 3. (Durand *et al.*, 2012) 4. (Ma *et al.*, 2009b). 1. (Rao *et al.*, 2011). 2. (Zheng and Leung, 2007, Felisberto-Rodrigues *et al.*, 2011).
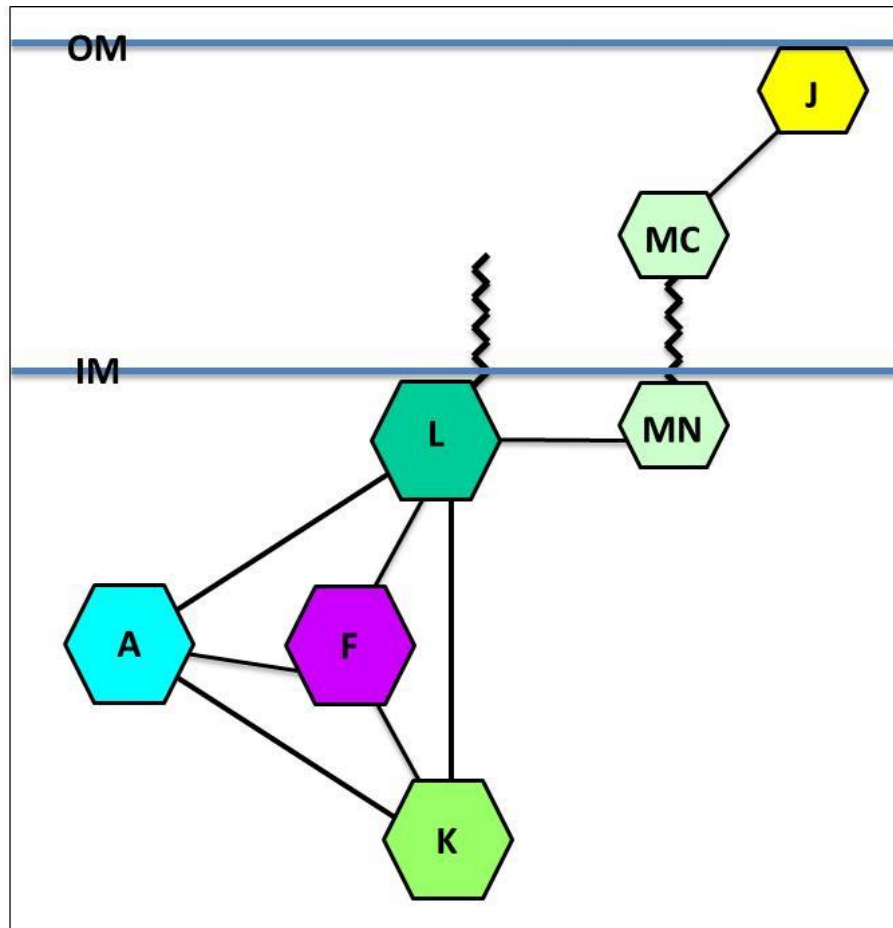
Figure 3.28. Schematic representation of interaction network between some Tss proteins.
The BACTH interaction between TssA, TssF, TssJ, TssK, TssL and TssM subunits of T6SS are
shown. The TssA, TssF, TssK and TssL proteins are predicted to form a complex close to the
inner membrane as TssL interact with the N-terminal domain of TssM (MN) which is located in
the cytoplasm. IM, inner membrane. OM, outer membrane. MC, C-terminal domain of TssM.
The zigzag motif corresponds to the single TMD at the C-terminus of TssL and the TMD
connecting the TssM.NTD to the CTD (the other two N-terminal TMDs in TssM are not
shown).

Chapter 4. Investigation of interactions of TssA^{E} and TalT with components of the T6SS using the bacterial adenylate cyclase two-hybrid system.

4.1. Introduction

To provide evidence that $TssA^E$ is a functional orthologue of $TssA^S$, it was decided to employ the two hybrid system to probe for interactions between $TssA^E$ and $TssA^S$-interacting T6SS subunits (including $TssA^S$ itself). For this analysis the *A. hydrophila* $TssA^E$ protein, AHA1844, was employed. Like many other bacteria that encode a $TssA^E$ orthologue, the *A. hydrophila* T6SS also encodes a homologue of the TssA-like protein, TaIT, which is not required for full T6SS function. Moreover, unlike $TssA^S$ and $TssA^E$, TaIT contains a TMD between the NTD and CTD (Section 1.5.4.2) (Figure 1.20, 4.14 and 4.15). To investigate the potential role of this protein, the *A. hydrophila* TaIT homologue, AHA1846, was also subjected to a two-hybrid analysis with $TssA^S$-interacting T6SS subunits.

4.2. Investigation into the oligomerisation of $TssA^E$

A. hydrophila AHA1844 was used as a representative of the $TssA^E$ class of TssA proteins for two-hybrid analysis. PCR amplification of the *AHA1844* gene was carried out from chromosomal DNA of *A. hydrophila* strain ATCC 7966 using pairs of primers designed specifically to amplify the gene. In each primer, a cleavage site for a restriction enzyme was included to allow cloning of the amplicon into the MCS of the BACTH plasmids (see appendix 1 for the primer sequences). The *AHA1844* gene was successfully cloned into all BACTH vectors generating both N-terminal and C-terminal fusions to T18 and T25 (pUT18-$TssA^E$, pUT18C-$TssA^E$, pKNT25-$TssA^E$ and pKT25-$TssA^E$). The cloning strategy was similar to the cloning of *tss* genes of *B. cenocepacia* into BACTH plasmids (Section 3.2).

To assess the possibility of $TssA^E$ self-interaction all possible pairwise combinations of plasmids encoding fusion of T18 and T25 to AHA1844 ($TssA^E$) were introduced into BTH101 and transformants were selected on MacConkey-maltose agar. It was observed that all four pairwise combinations of $TssA^E$ fusion proteins gave rise to deep red/purple colonies on MacConkey-maltose plates after 72 hours incubation at 30°C (Figure 4.1A). The very strong maltose positive phenotype of all four combinations suggests that $TssA^E$ self-interacts to form an oligomer.

In order to investigate which predicted domain of $TssA^E$ is responsible for the self-interaction, DNA encoding $TssA^E_{NTD}$ was amplified and cloned into all four two-hybrid

vectors, while that of $TssA^E_{MD-CTD}$ was cloned into pKT25 and pUT18C ($TssA^E_{NTD}$ fusion proteins consisted of amino acids 1-229 fused to T18 and T25, whereas $TssA^E_{MD-CTD}$ were comprised of amino acids 244-478). The combination of pKT25-$TssA^E_{MD-CTD}$ and pUT18C-$TssA^E_{MD-CTD}$ gave rise to a very strong Mal^+ phenotype with pink/red colonies on MacConkey maltose plates after 72 hours incubation at 30˚C, indicating that the MD and/ or the CTD of $TssA^E$ is important for oligomerisation (Figure 4.1B). Moreover, the four possible combinations of full-length $TssA^E$ and $TssA^E_{MD-CTD}$ fusion proteins also gave rise to a very strong Mal^+ phenotype after 72 hours incubation at 30˚C (Figure 4.1C). In contrast, three pairwise combinations of $TssA^E_{NTD}$ fusion proteins with itself gave rise only to a weak to moderate Mal^+ phenotype with patchy appearance after 120 hours incubation at 30˚C (Figure 4.2A). These results suggest that $TssA^E_{NTD}$ weakly or transiently self-interacts. Pairwise interaction analysis of $TssA^E_{NTD}$ and full length $TssA^E$ fusion proteins was carried out in all possible combinations to probe for $TssA^E_{NTD}$-$TssA^E$ interactions. Five out of eight combinations gave rise to a weak to moderate Mal^+ phenotype with patchy appearance after 120 hours incubation at 30˚C (Figure 4.2B).

As with *B. cenocepacia* $TssA^S$, $TssA^E$ of *A. hydrophila* oligomerises and the MD-CTD region appears to play an important role in this process. However, neither the MD nor the CTD of $TssA^E$ bears any obvious amino acid sequence similarity to the CTD of $TssA^S$ (Section 1.5.4.2). To determine which of these domains was responsible for the self-interaction of $TssA^E_{MD-CTD}$, DNA sequences encoding amino acids 224 to 386 (MD) and 381 to 478 (CTD) were amplified and cloned into pKT25 and pUT18C.

Interestingly, the BACTH assays showed that the combination of $TssA^E_{MD}$ with itself and $TssA^E_{CTD}$ with itself both gave rise to a very strong Mal^+ phenotype with pink/red appearance after 72 hours incubation at 30˚C (Figure 4.3A and B). Accordingly, $TssA^E_{CTD}$ fusion proteins also gave rise to a very strong Mal^+ phenotype in combination with $TssA^E_{MD-CTD}$ and $TssA^E$ fusion proteins (Figure 4.4A and B), and $TssA^E_{MD}$ fusion proteins in combination with $TssA^E_{MD-CTD}$ and $TssA^E$ fusion proteins gave rise to a very strong Mal^+ phenotype after 72 hours and a weak to moderate Mal^+ phenotype after 120 hours incubation at 30˚C, respectively (Figure 4.4C and D). However, the two pairwise combinations of $TssA^E_{MD}$ and $TssA^E_{CTD}$ fusion proteins gave rise to a maltose-negative phenotype in the BACTH assay (result not shown).

A maltose-negative phenotype resulted from all pairwise combinations between $TssA^E_{NTD}$ and $TssA^E_{MD}$ fusion proteins (result not shown). However, one plasmid combination encoding $TssA^E_{NTD}$ and $TssA^E_{CTD}$ fusion proteins (pKT25-$TssA^E_{CTD}$ and pUT18C-$TssA^E_{NTD}$) gave rise to a weak Mal^+ phenotype after 120 hours incubation at 30˚C (Figure 4.4E).

4.3. Investigation into the oligomerisation of TalT

To assess the possibility of TalT self-interaction, all possible pairwise combinations of plasmids encoding fusions of T18 and T25 to TalT were introduced into BTH101. Only one pairwise combination gave rise to a Mal^+ phenotype after 120 hours incubation at 30˚C, i.e. T25-TalT and T18-TalT (Figure 4.5A). However, the phenotype was less strong than observed for each of the full-length TssA and $TssA^E$ combinations (Figures 3.16A and 4.1A). To determine whether the putative C-terminal domain of TalT was responsible for the self-interaction, DNA encoding the C-terminal region of TalT was cloned into pKT25 and pUT18C. Two overlapping regions were cloned encoding amino acids 271-476 (CTD_L), which encodes the entire region C-terminal to the TMD, and amino acids 322-476 (CTD_S) which encodes the C-terminal region of TalT that is conserved in all TalT homologues (Figures 1.19 and 1.20C). The results show that as with the full-length TalT, the BACTH assay of pKT25-$TalT_{CTD.L}$ in combination with pUT18C-$TalT_{CTD.L}$ and pKT25-$TalT_{CTD.S}$ in combination with pUT18C-$TalT_{CTD.S}$ gave rise to a weak maltose-positive phenotype after 120 hours incubation at 30˚C (Figure 4.5B and C). Therefore, $TalT_{CTD}$ is involved in oligomerisation of TalT.

4.4. Investigation of potential interactions between $TssA^E$ and TalT

As the genes encoding $TssA^E$ and TalT are often present flanking *tssM*, it was decided to investigate whether these two proteins interacted with each other. However, Mal^- phenotypes were observed in all 8 pairwise combinations between plasmids encoding full-length $TssA^E$ and TalT fusion proteins (results not shown).
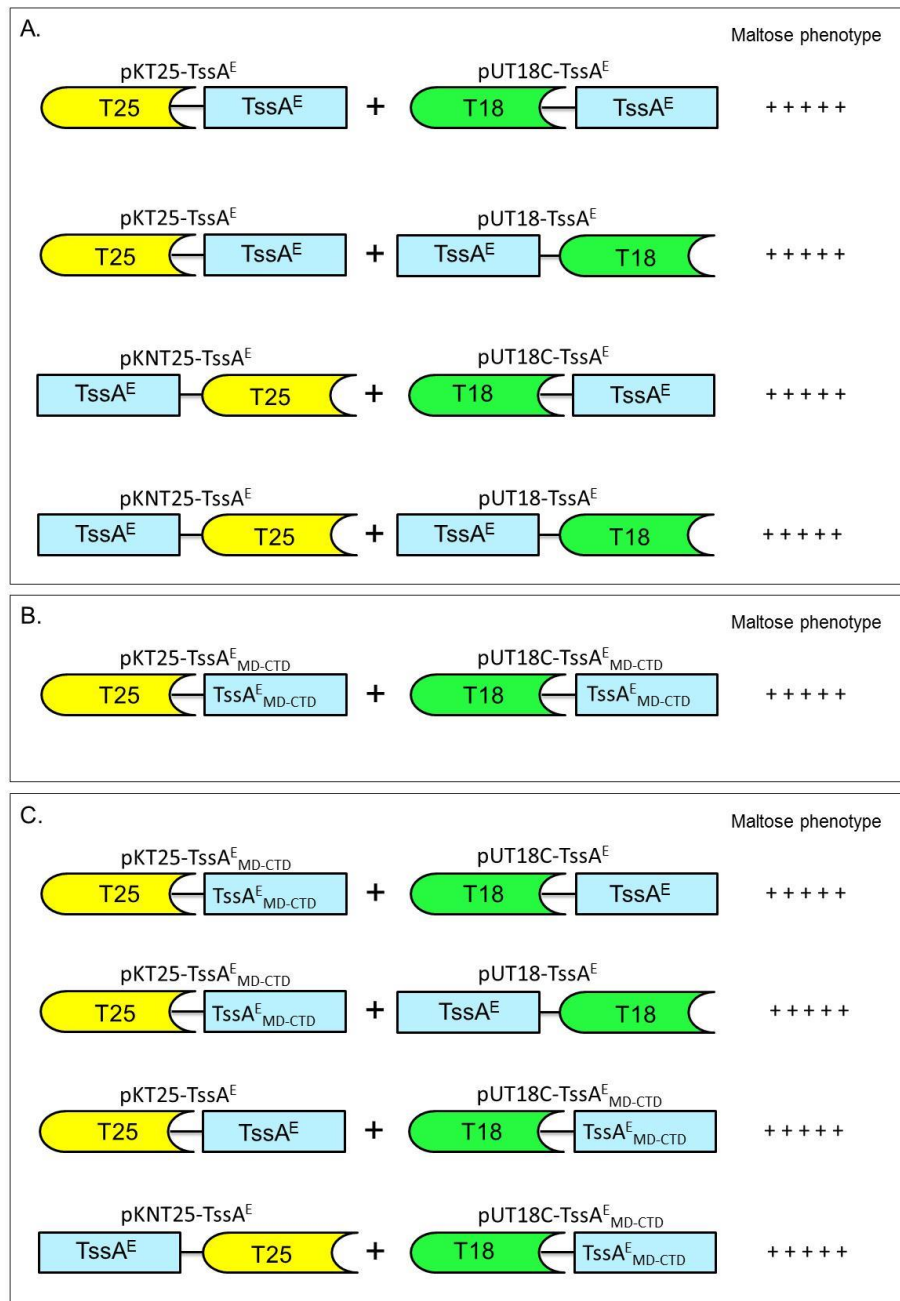
Figure 4.1. Analysis of interactions between TssA^E and TssA^E, TssA$^E_{MD\text{-}CTD}$ and TssA$^E_{MD\text{-}CTD}$, and TssA^E and TssA$^E_{MD\text{-}CTD}$ using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of TssA^E and TssA$^E_{MD\text{-}CTD}$ to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combinations of TssA^E and TssA^E fusion proteins that gave rise to a maltose-positive phenotype. B. Combinations of TssA$^E_{MD\text{-}CTD}$ and TssA$^E_{MD\text{-}CTD}$ fusion proteins that yielded a maltose-positive phenotype. C. Combinations of TssA^E and TssA$^E_{MD\text{-}CTD}$ fusion proteins that yielded a maltose-positive phenotype.

Figure 4.2. Analysis of interactions between TssA$^E_{NTD}$ and TssA$^E_{NTD}$, and TssA^E and TssA$^E_{NTD}$ using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of TssA^E and TssA$^E_{NTD}$ to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combinations of TssA$^E_{NTD}$ and TssA$^E_{NTD}$ fusion proteins that gave rise to a maltose-positive phenotype. B. Combinations of TssA^E and TssA$^E_{NTD}$ fusion proteins that gave rise to a maltose-positive phenotype.
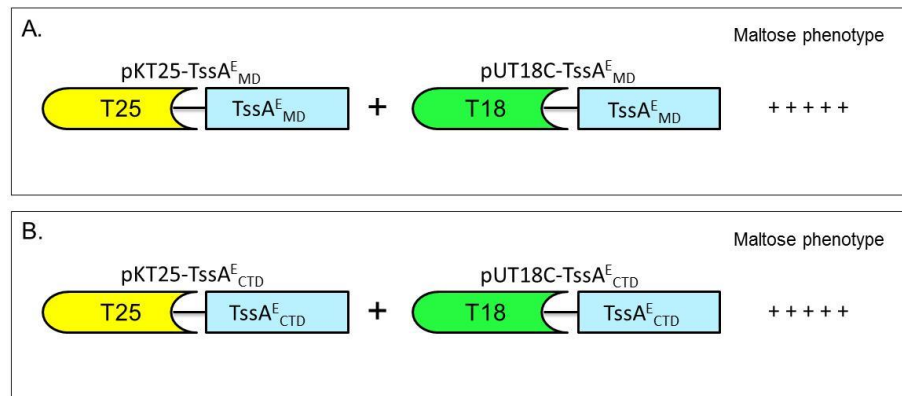
Figure 4.3. Analysis of TssA$^E_{MD}$ and TssA$^E_{CTD}$ self-interactions using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of TssA$^E_{MD}$ with TssA$^E_{MD}$ and TssA$^E_{CTD}$ with TssA$^E_{CTD}$ were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combinations of TssA$^E_{MD}$ and TssA$^E_{MD}$ fusion proteins that gave rise to a maltose-positive phenotype. B. Combinations of TssA$^E_{CTD}$ and TssA$^E_{CTD}$ fusion proteins that yielded a maltose-positive phenotype.
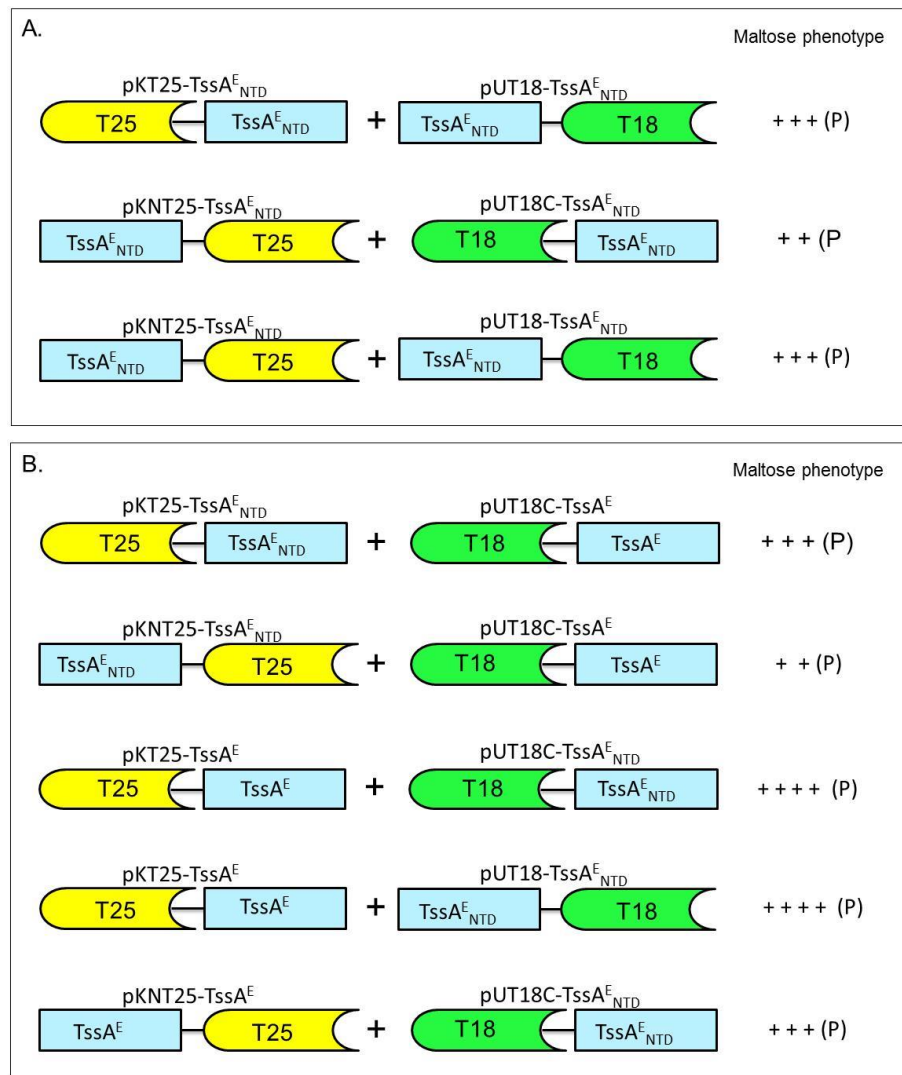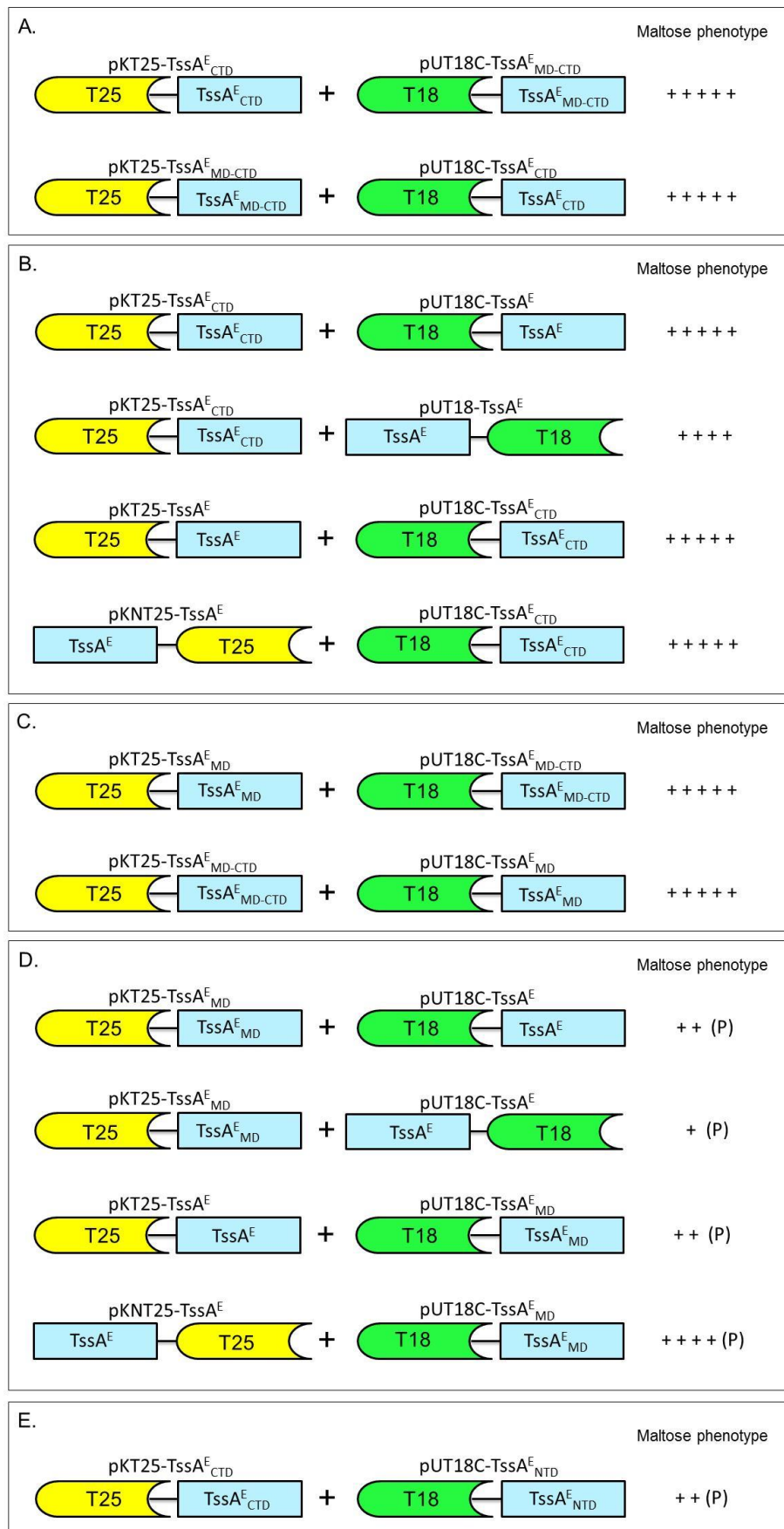
Figure 4.4. Analysis of interactions between TssA$^E_{CTD}$ and TssA$^E_{MD-CTD}$ or TssA^E, TssA$^E_{MD}$ and TssA$^E_{MD-CTD}$ or TssA^E, and TssA$^E_{CTD}$ interaction between with TssA$^E_{NTD}$ using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding T18 or T25 fusions of TssA$^E_{CTD}$ with TssA$^E_{MD-CTD}$ or TssA^E, TssA$^E_{MD}$ with TssA$^E_{MD-CTD}$ or TssA^E and TssA$^E_{CTD}$ with TssA$^E_{NTD}$ were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combinations of TssA$^E_{CTD}$ and TssA$^E_{MD-CTD}$ fusion proteins that gave rise to a maltose-positive phenotype. B. Combinations of TssA$^E_{CTD}$ and TssA$^E$ fusion proteins that yielded a maltose-positive phenotype. C. Combinations of TssA$^E_{MD}$ and TssA$^E_{MD-CTD}$ fusion proteins that yielded a maltose-positive phenotype. D. Combinations of TssA$^E_{MD}$ and TssA^E fusion proteins that yielded a maltose-positive phenotype. E. Combinations of TssA$^E_{CTD}$ and TssA$^E_{NTD}$ fusion proteins that yielded a maltose-positive phenotype.
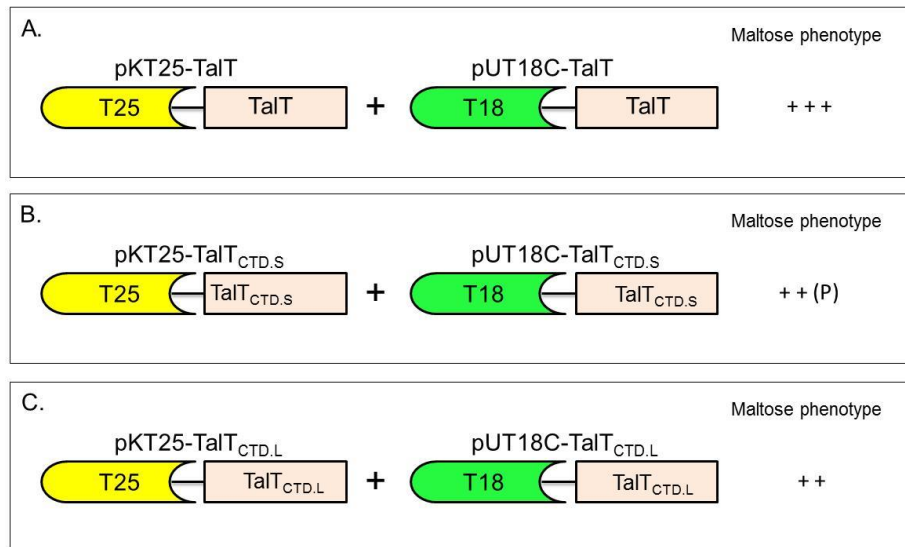
Figure 4.5. Analysis of the interaction between TalT and TalT, TalT_{CTD.S} and TalT_{CTD.S}, and TalT_{CTD.L} with TalT_{CTD.L} using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding T18 and T25 fusions of TalT with TalT, TalT_{CTD.S}, with TalT_{CTD.S} and TalT_{CTD.L} with TalT_{CTD.L} were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combinations of TalT and TalT fusion proteins that gave rise to a maltose-positive phenotype. B. Combinations of TalT_{CTD.S} and TalT_{CTD.S} fusion proteins that yielded a maltose-positive phenotype. C. Combinations of TalT_{CTD.L} and TalT_{CTD.L} fusion proteins that gave rise to a maltose-positive phenotype.

4.5. Interaction of TssA^E with TssA and TssA-interacting T6SS subunits

To determine whether TssA^E plays the same (or a very similar) functional role to TssA, the BACTH system was employed to probe for interactions between TssA^E and TssA, and between TssA^E and other Tss proteins which have been shown to interact with *B. cenocepacia* TssA, i.e. TssD, TssE, TssH, TssI, TssK and TssL. TssM was also tested because the genes encoding TssA^E and TalT flank *tssM* in most T6SS gene clusters where both *tssA^E* and *talT* are present (Figure 1.15). For convenience, plasmid constructs encoding T18 and T25 fusions to *B. cenocepacia* T6SS subunits, rather than *Aeromonas* T6SS subunits, were used in this assay.

4.5.1. Investigation into the cross-oligomerisation between TssA and TssA^E

Very strong Mal$^+$ phenotypes were observed when plasmids encoding full-length TssA fusion proteins were combined with those encoding full-length TssA^E fusion proteins for four out of the eight possible combinations tested. pKT25-TssA^E in combination with pUT18C-TssA and the analogous combination (pKT25-TssA and pUT18C-TssA^E) both yielded a very strong Mal$^+$ phenotype with deep purple colonies after 72 hours incubation. Whereas pKNT25-TssA^E in combination with pUT18C-TssA and the analogous combination of pKNT25-TssA with pUT18C-TssA^E both gave rise to a similarly strong phenotype but after 120 hours incubation, i.e. it took a longer incubation time to develop the strong maltose-positive phenotype (Figure 4.6A). These results suggest that despite the difference in the amino acid sequences of the C-terminal regions of TssA and TssA^E, they are still able to interact with each other.

Subsequently, heterologous interactions between TssA and TssA^E were probed by performing BACTH assays using the CTD of TssA and the MD-CTD region of TssA^E fused to the C-termini of T25 and T18. Interestingly, the two pairwise combinations of TssA$^E_{MD-CTD}$ and TssA$_{CTD}$ fusion proteins yielded a Mal$^+$ phenotype on MacConkey-maltose plates. The combination of pKT25-TssA$^E_{MD-CTD}$ and pUT18C-TssA$_{CTD}$ gave rise to a very strong maltose-positive phenotype with deep purple colonies after 72 hours incubation. The analogous combination (i.e. pKT25-TssA$_{CTD}$ and pUT18C-TssA$^E_{MD-CTD}$) also gave rise to a similar result but after 120 hours incubation and the colonies were pink/red coloured with a patchy appearance (Figure 4.6B). In addition, both possible combinations of TssA$^E_{MD-CTD}$ and full-length TssA fusion proteins (i.e. pKT25-TssA$^E_{MD-CTD}$ with pUT18C-TssA and pKT25-TssA with pUT18C-TssA$^E_{MD-CTD}$)

gave rise to a very strong maltose-positive phenotype after 72 hours incubation. Also, moderate to strong Mal$^+$ phenotypes were observed when plasmids encoding TssA$_{CTD}$ fusions were combined with those encoding full-length TssA^E fusions in four out of eight pairwise combinations (Figure 4.7A and B).

As TssA$_{CTD}$ interacts with TssA$^E_{MD-CTD}$, it was decided to investigate the contribution of the two putative domains of TssA$^E_{MD-CTD}$ to this interaction. Two of four possible combinations of TssA$^E_{CTD}$ and full-length TssA fusions gave rise to a strong maltose positive phenotype after 72 hours incubation (Figure 4.7C), whereas TssA$^E_{MD}$ fusions in combination with full-length TssA fusions gave rise to a Mal$^-$ phenotype in all four tested combinations. TssA$^E_{MD}$ in combination with TssA$_{CTD}$ also gave rise to maltose-negative phenotype (result not shown). However, one combination of TssA$^E_{CTD}$ and TssA$_{CTD}$ fusion proteins (pKNT25-TssA$_{CTD}$ in combination with pUT18C-TssA$^E_{CTD}$) gave rise to a weak maltose-positive phenotype after 120 hours (Figure 4.7D). Therefore, despite the difference in primary amino acid sequences TssA$_{CTD}$ and TssA$^E_{CTD}$ may perform a similar function.

4.5.2. Investigation into the interaction between TssA^E and TssD

A moderate to strong Mal$^+$ positive phenotype was observed after 120 hours incubation when TssA^E was fused to either the C- or N-terminal end of T25 and combined with TssD fused to the C-terminal end of T18 (Figure 4.8). The other combinations gave rise to a maltose-negative phenotype. These results suggest that TssA^E, like TssA, is able to interact with TssD.

4.5.3. Investigation into the interaction between TssA^E and TssE

Unlike the interaction observed between *B. cenocepacia* TssA and TssE, a maltose-negative phenotype was observed in all eight possible combinations of TssA^E and TssE fusion proteins (result not shown).
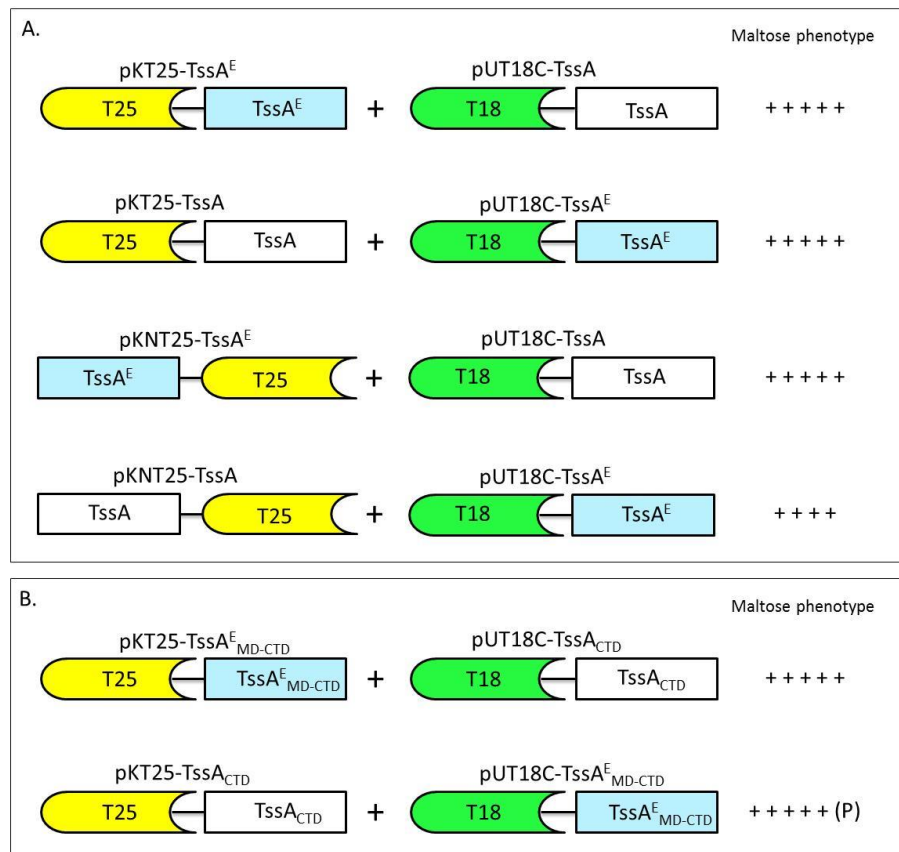
Figure 4.6. Analysis of interactions between TssA^E and TssA, and TssA$^E_{MD-CTD}$ and TssA$_{CTD}$ using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding T18 and T25 fusions of TssA^E with TssA and TssA$^E_{MD-CTD}$ with TssA$_{CTD}$ were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combinations of TssA^E and TssA fusion proteins that gave rise to a maltose-positive phenotype. B. Combinations of TssA$^E_{MD-CTD}$ and TssA$_{CTD}$ fusion proteins that yielded a maltose-positive phenotype.
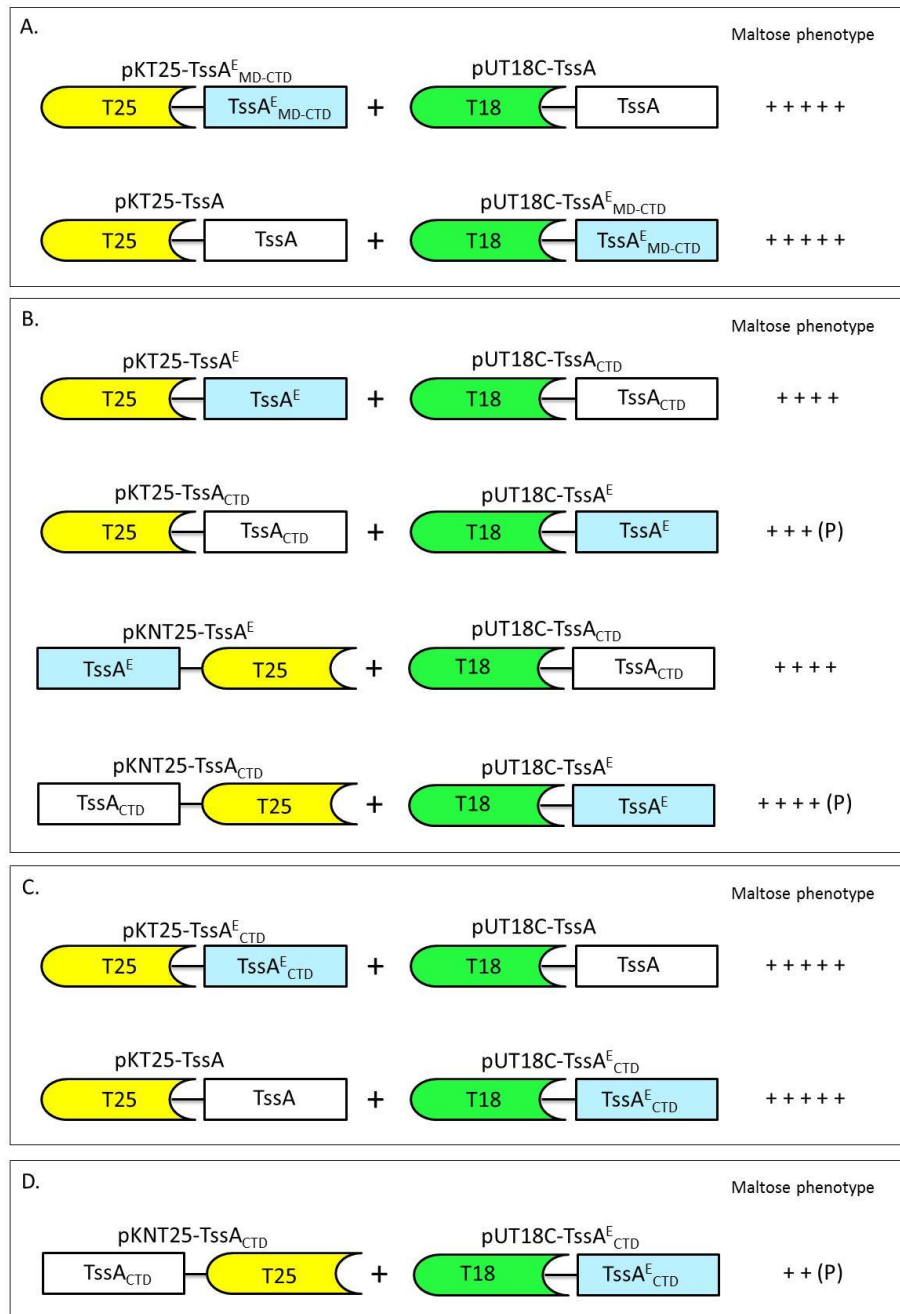
Figure 4.7. Analysis of interactions between TssA$^{E}_{MD-CTD}$ and TssA, TssA^{E} and TssA$_{CTD}$, TssA$^{E}_{CTD}$ and TssA, and TssA$^{E}_{CTD}$ with TssA$_{CTD}$ using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding T18 and T25 fusions of TssA$^{E}_{MD-CTD}$ with TssA, TssA^{E} with TssA$_{CTD}$, TssA$^{E}_{CTD}$ with TssA and TssA$^{E}_{CTD}$ with TssA$_{CTD}$ were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combinations of TssA$^{E}_{MD-CTD}$ and TssA fusion proteins that gave rise to a maltose-positive phenotype. B. Combinations of TssA^{E} and TssA$_{CTD}$ fusion proteins that yielded a maltose-positive phenotype. C. Combinations of TssA$^{E}_{CTD}$ and TssA fusion proteins that gave rise to a maltose-positive phenotype. D. Combinations of TssA$^{E}_{CTD}$ and TssA$_{CTD}$ fusion proteins that yielded a maltose-positive phenotype.

201
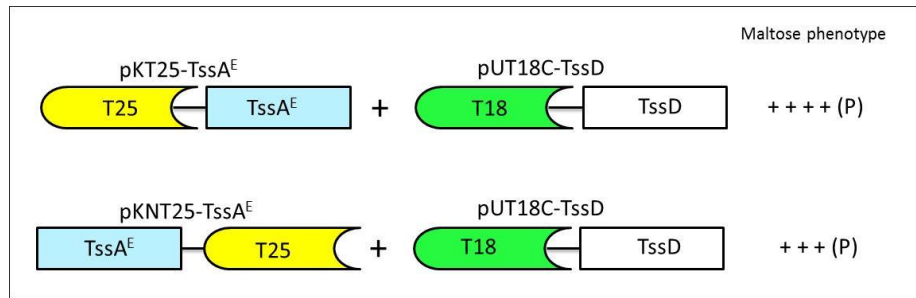
Figure 4.8. Analysis of ineractions between TssA^E and TssD interactions using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of TssA^E and TssD to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation.

4.5.4. Investigation into the interaction between TssA^E and TssH

Co-transformation of *E. coli* BTH101 strain with all possible plasmid combinations encoding TssA^E and TssH fusion proteins was carried out. Four out of six pairwise combinations of fusion proteins gave rise to a maltose-positive phenotype after 120 hours incubation, which was very strong in three combinations and moderate in one (Figure 4.9A). These results suggest that TssA^E, like TssA, can interact with TssH.

Three combinations of TssA$^E_{MD-CTD}$ and TssH fusion proteins yielded a maltose-positive phenotype (Figure 4.9B). Thus, T25-TssA$^E_{MD-CTD}$ with T18-TssH, and the analogous combination of T25-TssH with T18-TssA$^E_{MD-CTD}$, gave rise to a very strong Mal$^+$ phenotype after 72 hours incubation. While a very strong maltose-positive phenotype was observed for the TssH-T25 and T18-TssA$^E_{MD-CTD}$ fusion protein combination after 120 hours incubation.

Further analysis was carried out between the individual middle and C-terminal domains of TssA^E and TssH. T25-TssH in combination with T18-TssA$^E_{MD}$ gave rise to a weak maltose-positive phenotype after 120 hours incubation, but the analogous combination (T25-TssA$^E_{MD}$ in combination with T18-TssH) gave rise a maltose-negative phenotype (Figure 4.9C). A maltose-positive phenotype was observed in the two pairwise combinations of TssA$^E_{CTD}$ and TssH. Thus, T25-TssH in combination with T18-TssA$^E_{CTD}$ gave rise to a strong maltose-positive phenotype after 72 hours incubation, while the analogous combination of T25-TssA$^E_{CTD}$ with T18-TssH gave rise to a moderate maltose-positive phenotype after 120 hours incubation (Figure 4.9D).

4.5.5. Investigation into the interaction between TssA^E and TssI

Similarly to TssA, TssA^E was found to interact with the phage tail-spike homologue, TssI. Figure 4.10A shows that three combinations of TssA^E and TssI hybrid proteins yielded a maltose positive phenotype in the BACTH assay. In particular, the combinations of T25-TssA^E with T18-TssI and TssA^E-T25 with T18-TssI gave a strong maltose-positive phenotype. Further BACTH assays were carried out with plasmids encoding the fused MD-CTD domains of TssA^E (TssA$^E_{MD-CTD}$) to determine whether this region is involved in interactions with TssI. The results show that when pKT25-TssA$^E_{MD-CTD}$ was combined with pUT18C-TssI, colonies with a patchy Mal$^+$ phenotype were observed which required 120 hours incubation to become apparent (Figure 4.10B).

Other combinations of plasmids expressing $TssA^E_{MD-CTD}$ and TssI hybrid proteins gave rise to a Mal⁻ phenotype. Interestingly, both domains of $TssA^E_{MD-CTD}$ were observed to separately interact with TssI (Figures 4.10C and D). Therefore, like TssA, $TssA^E$ interacts with TssI using its CTD.

4.5.6. Investigation into the interaction between $TssA^E$ and TssK

Four out of eight plasmid combinations expressing $TssA^E$ and TssK hybrid proteins gave rise to a maltose-positive phenotype after 120 hours incubation (Figure 4.11A). However, only one combination of $TssA^E_{MD-CTD}$ with TssK (pKT25-$TssA^E_{MD-CTD}$ and pUT18C-TssK) gave rise to a maltose-positive phenotype (Figure 4.11B). Testing the individual NTD and CTD domains of $TssA^E$ with TssK showed that CTD interacts with TssK as judged from the weak maltose-positive phenotype observed for one combination of these fusion proteins (i.e. pKT25- $TssA^E_{CTD}$ with pUT18C-TssK) (Figure 4.11C).

4.5.7. Investigation into the interaction between $TssA^E$ and TssL

Results from the BACTH assays showed that when T25-$TssA^E$ or $TssA^E$-T25 was co-expressed with T18-TssL, a weak maltose-positive phenotype was observed after 120 hour incubation (Figure 4.12). When T25-$TssA^E_{MD-CTD}$ or $TssA^E_{MD-CTD}$-T25 was co-expressed with T18-TssL, a maltose-negative phenotype was observed. Therefore, the individual MD and CTD domains were not tested in combination with TssL.

4.5.8. Investigation into the interaction between $TssA^E$ and TssM

All combinations of $TssA^E$ and $TssA^E_{MD-CTD}$ with $TssM_{NTD}$ and $TssM_{CTD}$ fusion proteins gave rise to a maltose-negative phenotype (results not shown).
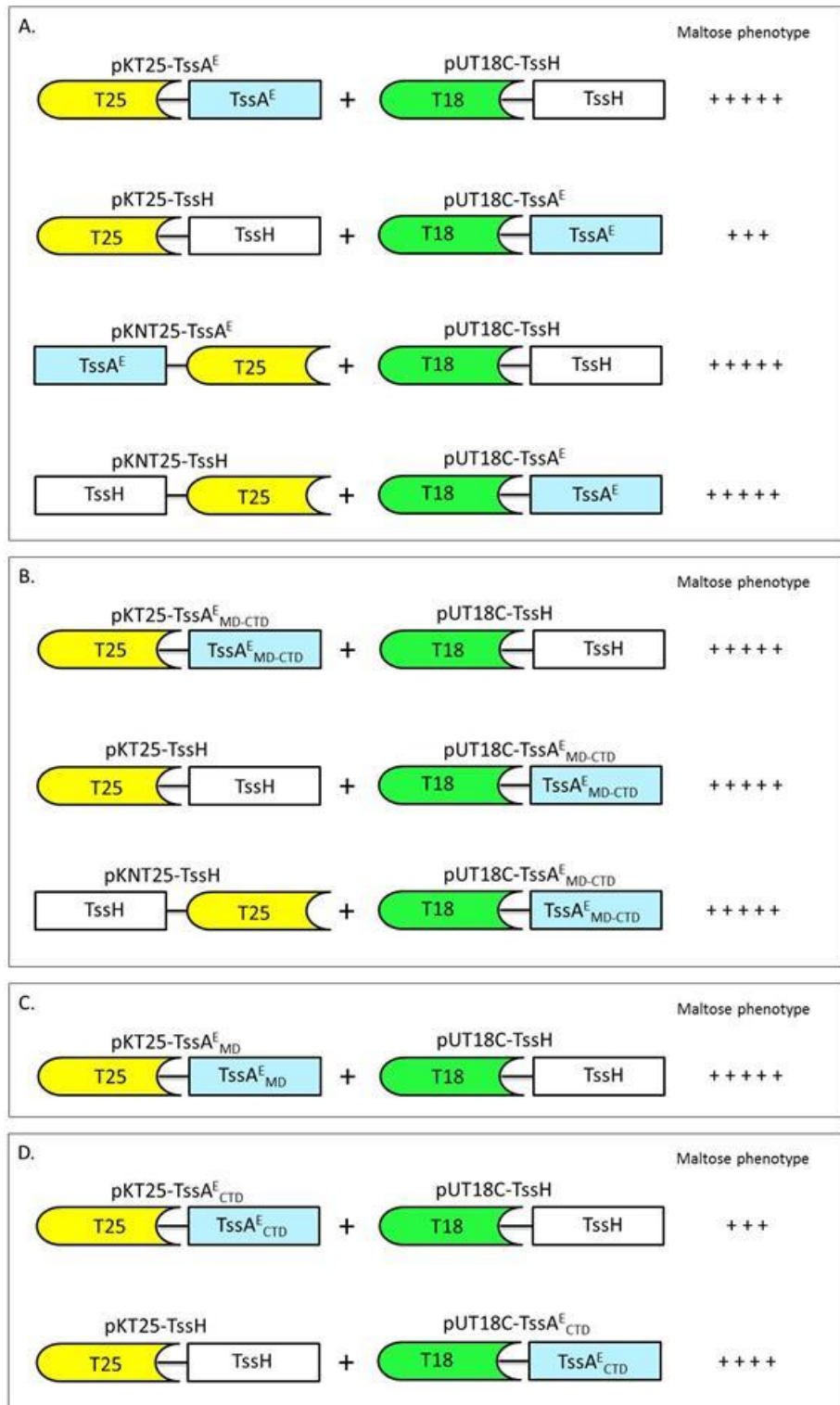
Figure 4.9. Analysis of interactions between TssA^E and its middle and C-terminal domains with TssH using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding T18 and T25 fusions of TssA^E with TssH, TssA$^E_{MD-CTD}$ with TssH, TssA$^E_{MD}$ with TssH and TssA$^E_{CTD}$ with TssH were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combinations of TssA^E and TssH fusion proteins that gave rise to a maltose-positive phenotype. B. Combinations of TssA$^E_{CTD}$ and TssH fusion proteins that yielded a maltose-positive phenotype. C. Combinations of TssA$^E_{MD}$ and TssH fusion proteins that gave rise to a maltose-positive phenotype. D. Combinations of TssA$^E_{CTD}$ and TssH fusion proteins that yielded a maltose-positive phenotype.

Figure 4.10. Analysis of interactions between TssA^E, TssA$^E_{MD-CTD}$, TssA$^E_{MD}$ and TssA$^E_{CTD}$ with TssI using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding T18 and T25 fusions of TssA^E with TssI, TssA$^E_{MD-CTD}$ with TssI, TssA$^E_{MD}$ with TssI and TssA$^E_{CTD}$ with TssI were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combinations of TssA^E and TssI fusion proteins that gave rise to a maltose-positive phenotype. B. Combinations of TssA$^E_{MD-CTD}$ and TssI fusion proteins that yielded a maltose-positive phenotype. C. Combinations of TssA$^E_{MD}$ and TssI fusion proteins that gave rise to a maltose-positive phenotype. D. Combinations of TssA$^E_{CTD}$ and TssI fusion proteins that yielded a maltose-positive phenotype.

207
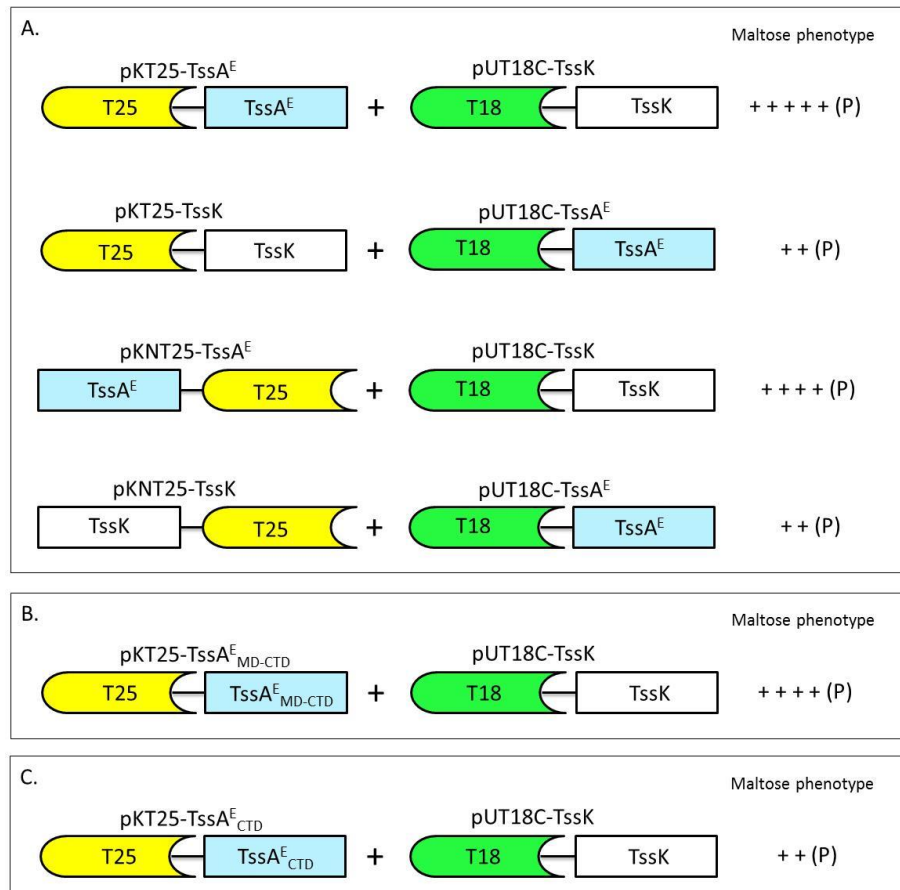
Figure 4.11. Analysis of interactions between TssA^E, TssA$^E_{MD-CTD}$ and TssA$^E_{CTD}$ with TssK using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding T18 and T25 fusions of TssA^E with TssK, TssA$^E_{MD-CTD}$ with TssK and TssA$^E_{CTD}$ with TssK were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combinations of TssA^E and TssK fusion proteins that gave rise to a maltose-positive phenotype. B. Combinations of TssA$^E_{MD-CTD}$ and TssK fusion proteins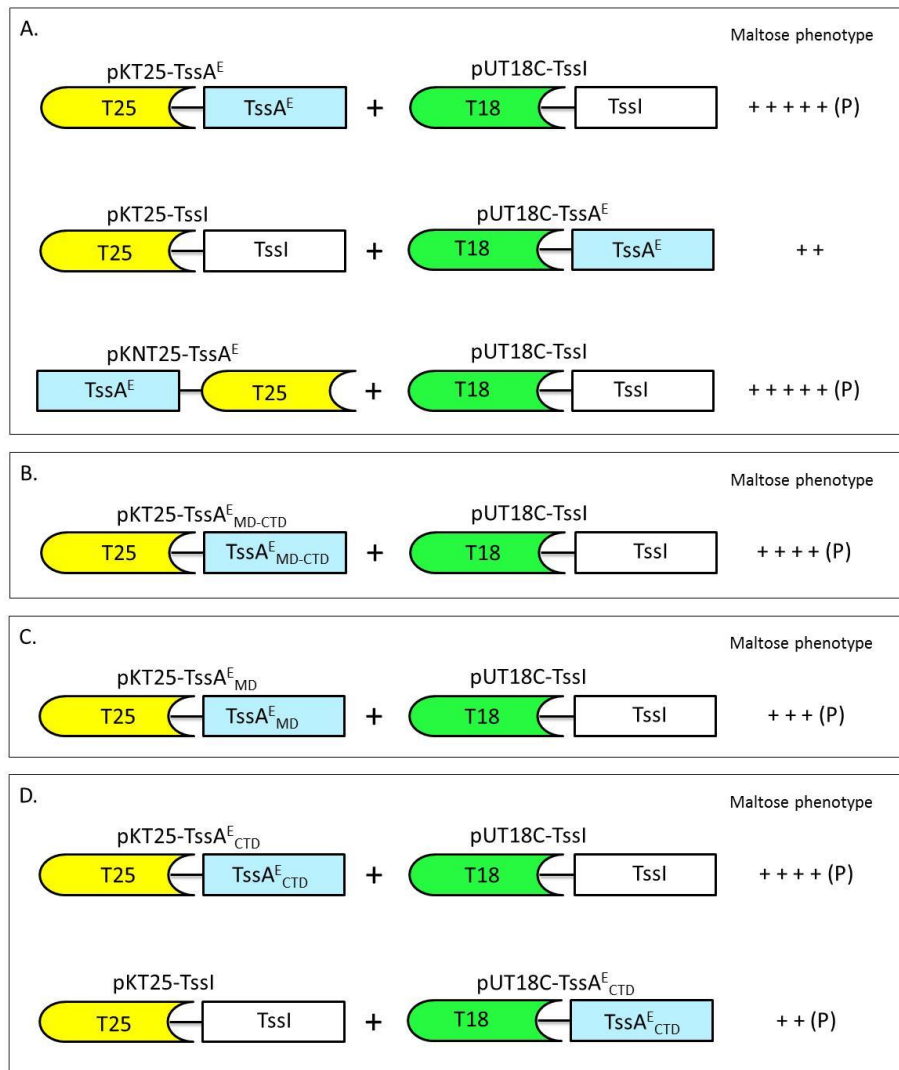 that yielded a maltose-positive phenotype. C. Combinations of TssA$^E_{CTD}$ and TssK fusion proteins that yielded a maltose-positive phenotype.
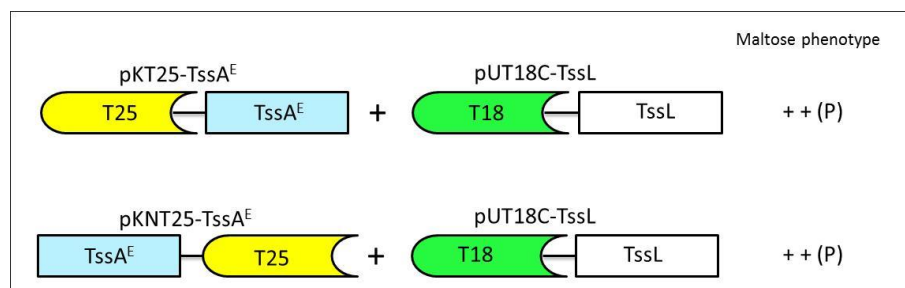
Figure 4.12. Analysis of interactions between TssA^E and TssL using the BACTH assay. Pairwise combinations of compatible BACTH plasmids encoding fusions of TssA^E and TssL to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation.

208

4.6. Investigation into potential interactions between TalT and TssA, and between TalT and TssA-interacting T6SS subunits

A maltose-negative phenotype was observed for all pairwise combinations of TalT with TssA, TssD, TssE, TssH, TssI, TssK, TssL and TssM fusion protein combinations (results not shown). It is possible that TalT is partitioning into the membrane (because it contains TMD), and the CTD is located in the periplasmic space (see Discussion). Therefore, interactions between $TalT_{CTD}$ and T6SS subunits that are normally periplasmically located, i.e. $TssM_{CTD}$, would not occur, as the BACTH fusion proteins are designed to be expressed in the cytoplasm were ATP is available. The BACTH assays were repeated using the $TalT_{CTD.S}$ (due to time constraints the BACTH assays were not carried out using the $TalT_{CTD.L}$). $TalT_{CTD.S}$ gave rise to a weak maltose-positive phenotype in pairwise combinations with three T6SS subunits following 120 hours incubation (Figure 4.15A, B and C). The pairwise combinations yielding a Mal^+ phenotype were pKT25-TssH with pUT18C-$TalT_{CTD.S}$ and pKT25-$TalT_{CTD.S}$ combined with either pUT18C-TssI or pUT18C-TssL.
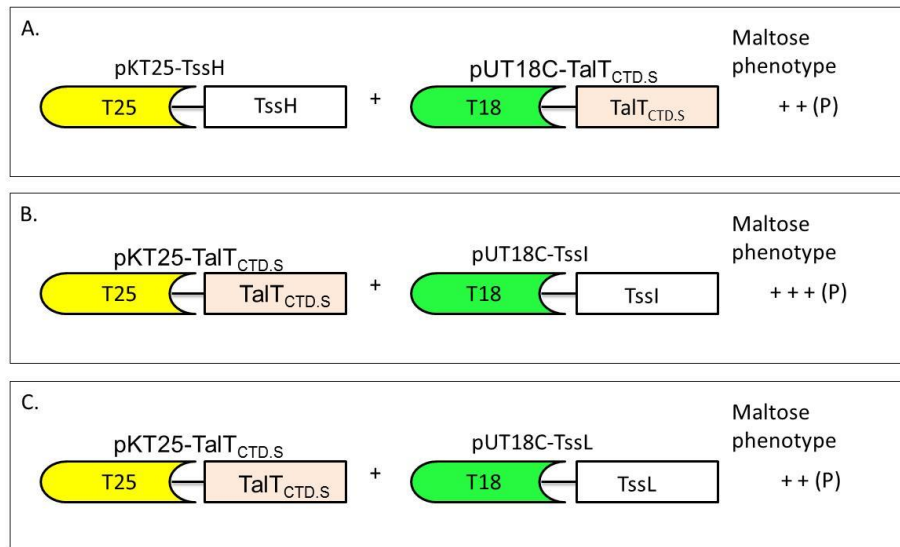
Figure 4.13. Analysis of interactions between TalT$_{CTD.S}$ and TssH, TssI and TssL using the BACTH assay. Pairwise combinations of compatible BACTH plasmids T18 and T25 encoding fusions of TalT$_{CTD.S}$ with TssH, TssI or TssL were introduced into *E. coli* strain BTH101. Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72 and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are shown. The strength of the maltose phenotype shown was scored after 120 h incubation. A. Combinations of TalT$_{CTD.S}$ and TssH fusion proteins that gave rise to a maltose-positive phenotype. B. Combinations of TalT$_{CTD.S}$ and TssI fusion proteins that yielded a maltose-positive phenotype. C. Combinations of TalT$_{CTD.S}$ and TssL fusion proteins that gave rise to a maltose-positive phenotype.

4.7. Discussion

Using the BACTH system, we demonstrated that the SciA class TssA oligomerises by means of its CTD, and that it interacts with other core components of the T6SS, i.e. TssB, TssD, TssE, TssF, TssH, TssI, TssK and TssL (see Chapter 3). Therefore, TssA was selected for further investigation. TssA^E was hypothesed to behave in a similar way despite the lack of amino acid sequence homology between C-terminal regions of the two TssA classes. Here it is demonstrated that full-length TssA^E self-interacts and does so through its C-terminal region. Furthermore, the predicted middle and C-terminal domains of TssA^E were both found to be important for this interaction, as each domain showed a strong self-interaction. Accordingly, BACTH assays showed that TssA$^E_{MD-CTD}$ interacted strongly with the full-length TssA^E and with the separate middle and C-terminal domains of TssA^E. The assay also suggested that TssA$^E_{NTD}$ may be involved in weaker self-interactions than that observed for the CTD.

Not only was TssA^E found to interact with itself, here it was also demonstrated to interact with a SciA class TssA and TssA$_{CTD}$, and this interaction involves TssA$^E_{MD-CTD}$. This is a remarkable observation given the absence of detectable amino acid sequence homology between the two CTD regions. Analysis of the individual MD and CTD domains of TssA^E with TssA and TssA$_{CTD}$ showed only a strong interaction between TssA$^E_{CTD}$ and full-length TssA, while it appeared to be involved in a weaker interaction with TssA$_{CTD}$. These results suggest that the TssA$^E_{CTD}$ plays a more important role than TssA$^E_{MD}$ in the cross-interaction of TssA^E with TssA (TssA^S).

Furthermore, TssA^E was shown to cross-interact with most of the T6SS core subunits that were previously demonstrated to interact with TssA, i.e. TssD, TssH, TssI, TssK and TssL (Figure 4.14). Of the TssA^S-interacting subunits tested, only TssE did not detectably interact with TssA^E. It worth to mention that *A. hydrophila* TssE was not used in these experiments and also TssA^E is quite different to TssA^S. Interestingly, TssA$^E_{MD-CTD}$ interacted with TssH, TssI and TssK. The analysis of TssA$^E_{MD}$ with these three T6SS subunits showed strong and moderate interactions with TssH and TssI, respectively. On the other hand, TssA$^E_{CTD}$ showed strong interactions with TssH and TssI and a weak interaction with TssK, supporting the observations made with TssA$^E_{MD-CTD}$. As TssA$^S_{CTD}$ also interacts with TssH, TssI and TssK, this may suggest some structural similarity between the TssA^S and TssA^E CTDs which cannot be inferred

from the amino acid sequences alone. However, unlike $TssA^S_{CTD}$, $TssA^E_{CTD}$ did not detectably interact with TssD.

TalT was shown to self-interact in one pairwise combination, i.e. T25-TalT with T18C-TalT. Also, $TalT_{CTD}$ was shown to self-interact, which suggests that the TalT self-interaction is mediated mainly through its CTD domain (similar to $TssA^S$ and $TssA^E$) (Figure 4.15). The TalT self-interaction gave rise to a weaker maltose-positive phenotype in the BACTH assay than $TssA^S$ and $TssA^E$ which both showed a strong maltose-positive phenotype in all four pairwise combinations. However, the presence of a predicted TMD in TalT, which would insert the protein into the inner membrane with the NTD located in the cytoplasm and the CTD in the periplasm, could explain why only one pairwise combination gave rise to a Mal^+ phenotype (Figure 4.16). The BACTH assay requires the two CyaA fragments (T18 and T25) to be located in the cytoplasm in order to interact and produce cAMP, as there is no ATP in the periplasmic space. Therefore, the self-interaction of TalT through its CTD would only give rise to a Mal^+ phenotype when T18 and T25 are both in the cytoplasm. Thus, the BACTH assay results for full-length TalT suggest the topology shown in Figure 4.16A. In addition, these results suggest that $TssA^S$ and $TssA^E$ are cytoplasmically located, as $TssA^S$, $TssA^E$ and TalT share a conserved N-terminal region and neither $TssA^S$ nor $TssA^E$ possess a predicted TMD.

A maltose-positive phenotype was not detected in the BACTH assays for any pairwise combination of TalT with TssD, TssE, TssH, TssI, TssK, TssL and TssM fusion proteins further reinforcing the idea that TalT is not orthologous to TssA. However, when the BACTH assay was carried out again with $TalT_{CTD.S}$ and the T6SS subunits, a weak maltose-positive phenotype was detected between $TalT_{CTD.S}$ and cytoplasmically located TssH, TssI and TssL. These results might indicate a weak or transient interaction between TalT and TssH, TssI and TssL during T6SS assembly. Perhaps surprisingly, $TalT_{CTD.S}$ did not interact with $TssM_{CTD}$. This is the only periplasmically located T6SS subunit domain that could interact with $TalT_{CTD.S}$ as this domain is not large enough to contact OM-anchored TssJ. The conserved genetic location of *talT* adjacent to *tssM* is also suggestive of such interaction. However, it is important to note that *A. hydrophila* TssM was not used in these experiments. A recent study on the *V. cholerae* O37 serogroup strain V52 showed that the TalT orthologue, VCA0121 (VasL)

is not an essential component of T6SS, as deletion of VasL did not inhibit secretion of Hcp (TssD) or bacterial virulence (Zheng *et al*., 2011). Therefore, the function of TalT remains unclear. One possibility is that TalT$_{CTD}$ is a sensor domain that transduces a periplasmic signal to the cytoplasmicaly located NTD which results in regulation of T6SS activity under certain conditions.
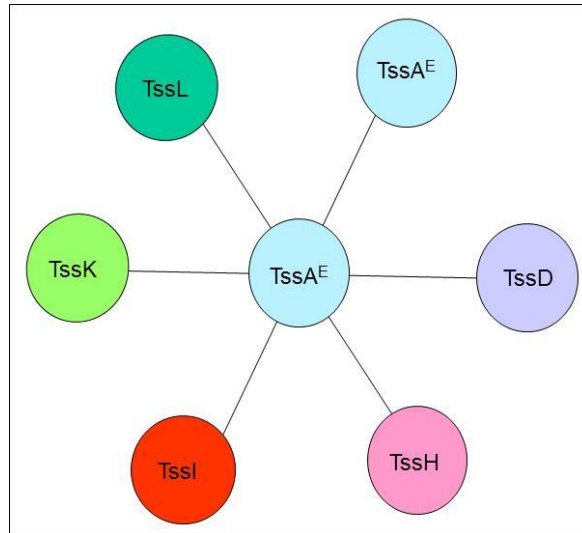
Figure 4.14. Summary of potential interactions of TssA^E in T6SS. Two-hybrid assay results suggest that TssA^E interacts with itself and with 6 other T6SS subunits i.e. TssA, TssD, TssH, TssI, TssK and TssL. Possible interactions between TssA^E and TssB and TssF were not tested.
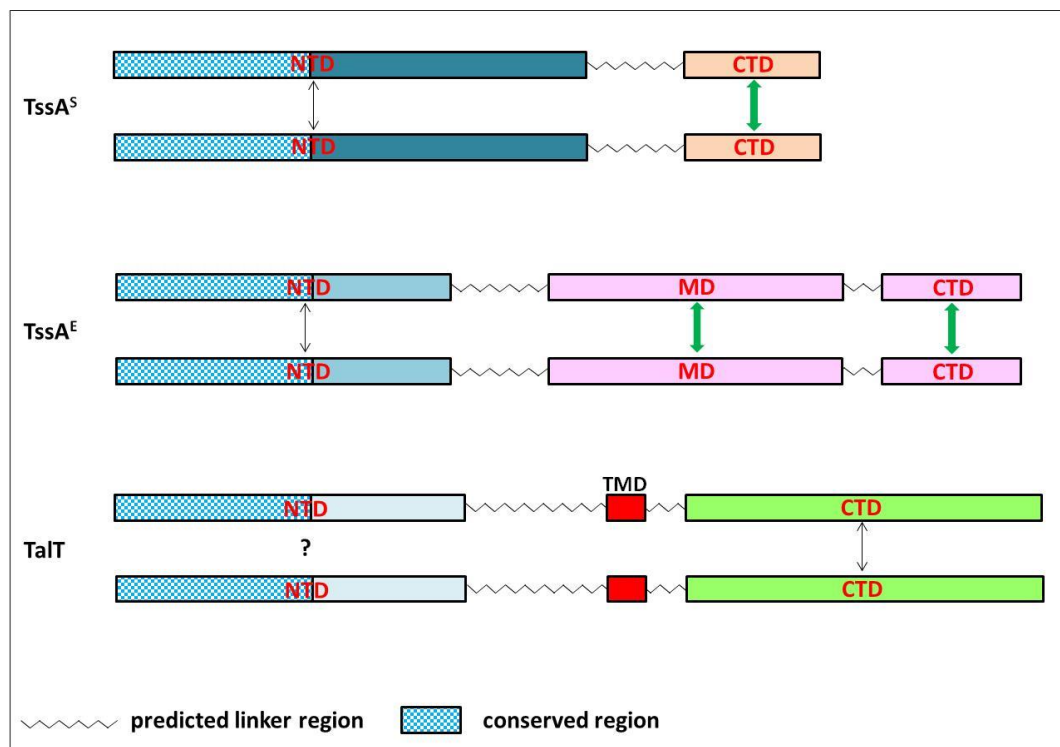
Figure 4.15. Summary of self-interactions of TssA^S, TssA^E and TalT domains. Self-interactions between domains of **TssA^S, TssA^E and TalT** as determined by the BACTH assay are shown. Domain combinations that gave rise to a strong maltose-positive phenotype after 120 h incubation are represented by green double-headed arrows, whereas combinations that gave rise to a moderate maltose-positive phenotype are represented by the black double-headed arrows. The interaction of the N-terminal region of TalT was not tested.
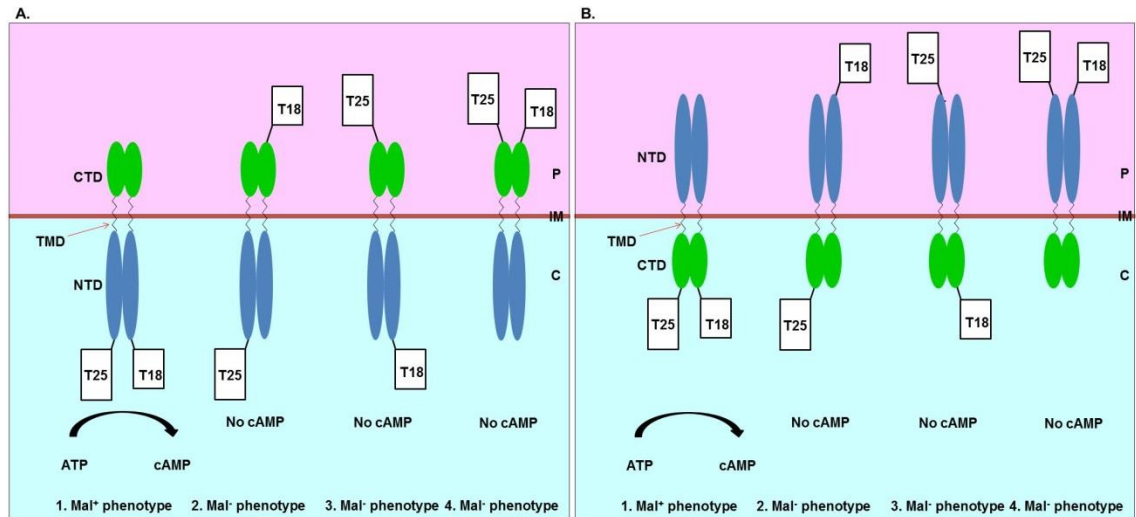
Figure 4.16. Schematic representations of pairwise combinations of TalT BACTH fusion proteins and the potential outcome in a BACTH assay depending on TalT orientation in the cytoplasmic membrane. TalT is predicted to contain two soluble domains, an N-terminal (NTD) and a C-terminal (CTD) domain, connected by a centrally located trans-membrane domain (TMD) which inserts the TalT protein into the inner membrane (IM). When T18 and T25 fragments fused to TalT become close to each other due to the self-interaction of $TalT_{CTD}$, cAMP will be produced in *E. coli* BTH101 cells only if they are both located in the cytoplasm as ATP is not available in the periplasmic space. cAMP will switch on the expression of the *mal* genes giving rise to Mal^+ colonies on MacConkey-maltose plates. A. Predicted BACTH assay phenotypes for a cytoplasmically located NTD. 1) pKT25-TalT co-expressed with pUT18C-TalT. 2) pKT25-TalT co-expressed with pUT18-TalT. 3) pKNT25-TalT co-expressed with pUT18C-TalT. 4) pKNT25-TalT co-expressed with pUT18-TalT. B. Predicted BACTH assay phenotypes for a cytoplasmically located CTD. 1) pKT25-TalT co-expressed with pUT18C-TalT. 2) pKT25-TalT co-expressed with pUT18-TalT. 3) pKNT25-TalT co-expressed with pUT18C-TalT. 4) pKNT25-TalT co-expressed with pUT18-TalT. The observed phenotypes are consistent with the topology shown in A.

Chapter 5. Biochemical analysis of TssA interactions with other T6SS subunits

5.1. Introduction

The BACTH assay results suggest that TssA plays an important role in the assembly of the T6SS as it interacts with many other subunits (TssB, TssD, TssE, TssF, TssH, TssI, TssK and TssL) (chapter 3). In order to provide biochemical evidence to support the two-hybrid assay results of TssA with the other T6SS subunits, pull down and co-immunoprecipitation assays were employed. Due to time constraints, only TssA interactions with TssD and TssE were investigated using these approaches.

5.2. Cloning and overexpression of individual *tss* genes

The basis of the pull down experiments was to co-express a His-tagged T6SS subunit in the same cell as an untagged subunit and then perform IMAC on the soluble fraction of the cell lysate. Proteins that interact with the tagged subunit will be retained on the column and co-eluted with the tagged protein by high concentrations of imidazole. The interacting proteins can be detected by western blotting using an antibody specific for the protein or an epitope-specific antibody.

5.2.1. Overproduction of His$_6$-TssE

In order to express TssA and TssE in the same cell, therefore allowing the two proteins to interact *in vivo*, the *tssE* gene was cloned into the p15A derivatives, pACYCDuet-1, which is compatible with the pET14b plasmid (pET14b-His$_6$.TssA was previously constructed by Shastri, 2011). Another useful feature of pACYCDuet-1 is that it allows the co-expression of two genes on the same plasmid which might be useful for pull down experiments of TssA with other Tss proteins. pACYCDuet-1 can also be used in combination with pETDuet-1, thereby allowing four proteins to be overexpressed in the same cell. pACYCDuet-1 contains the chloramphenicol resistance gene and two multiple cloning sites (MCS), each one is preceded by a T7 promoter/*lac* operator and ribosome binding site (rbs) (Figure 5.1).

The *tssE* gene was PCR amplified from pUT18C-TssE with KOD polymerase and cloned into the first MCS of pACYCDuet-1 using *Bam*HI and *Hin*dIII by including their sites in the forward (pACYCtssEforHis) and reverse (pACYCtssErev) primer, respectively (see appendix 1 for the primer sequences). This placed the *tssE* gene in frame with the vector His-tag coding sequence, resulting in an ORF encoding TssE with
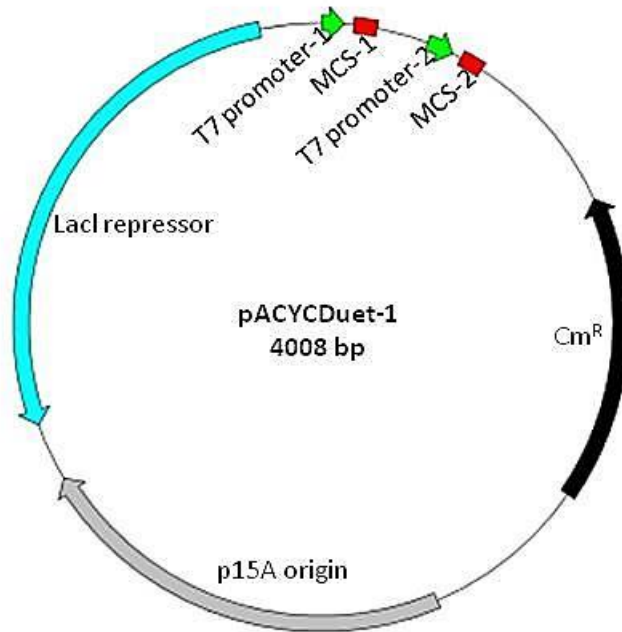
```
                                              (pACYC-T7-1for)
                                          ACYCDuetUP1 primer →              EcoNI
GCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATT
        T7 promoter-1                      lac operator
AGGAAAT TAATACGACTCACTATA GGGGAATTGTGAGCGGATAACAATTCC CCTGTAGAAATAATTTTGTTTAACT
                                                                    Ecl136I
        rbs        NcoI                   His•Tag         BamHI EcoRI SacI
TTAATA AGGAG ATATACC ATG GGCAGCAGC CATCACCATCATCACCAC AGCCA GGATCCGAATTCGAGCTC GGCGC
           MetGlySerSerHisHisHisHisHisHisSerGlnAspProAsnSerSerSerAlaA
                                                                    BsrGI
  BspMI                                                   (pACYC-T7-2for)
  PstI SalI HindIII NotI                                  DuetUP2 Primer →
GCCTGCAGGTCGACAAGCTTGCGGCCGCA TAA TGCTTAAGTCGAACAGAAAGTAATCGTA TTGTACACGGCCGCAT
rgLeuGlnValAspLysLeuAlaAlaAlaEnd                       ← DuetDOWN1 Primer
                                                        (pACYC-T7-1rev)
        T7 promoter-2                      lac operator
AATCGAAAT TAATACGACTCACTATA GGGGAATTGTGAGCGGATAACAATTCC CCATCTTAGTATATTAGTTAAGT
                                                FseI
        rbs      NdeI       BglII MunI   EcoRV NgoAIV    PvuI    AatII KpnI
ATAAGA AGGAG ATATA CATATG GCAGATCTCAATTGGATATCGGCCGGCC ACGCGATCGCT GACGTCGGTACCCTC
                 MetAlaAspLeuAsnTrpIleSerAlaGlyHisAlaIleAlaAspValGlyThrLeu
XhoI                         S-Tag                                   PacI
GAG TCTGGTAAAGAAACCGCTGCTGCGAAATTTGAACGCCAGCACATGGACTCGTCTACTAGCGCAGC TTAATTAA
GluSerGlyLysGluThrAlaAlaAlaLysPheGluArgGlnHisMetAspSerSerThrSerAlaAlaEnd

CCTAGGCTGCTGCC ACCGCTGAGCAATAACTAGC ATAACCCCTTGGGGCCTCTAAACGGGTCTTG
          ← T7 Terminator Primer
                 (T7rev)
```

Figure 5.1. The T7 expression vector pACYCDuet-1. A. Diagrammatic illustration of the overexpression vector pACYCDuet-1. p15A origin, origin of p15A replication. B. Nucleotide sequence of the multiple cloning site region of pACYCDuet-1. The T7 promoter-1 and the T7 promoter-2 sequences are highlighted in green. The *lac* operators are highlighted in grey. The hexahistidine tag coding sequence is highlighted in yellow. The start and stop codons are shown in pink font. The restriction sites are shown in red font. The S-Tag sequence is shown in purple font. The location of pACYC-T7-1for, pACYC-T7-1rev, pACYC-T7-2for and T7 terminator primer sequences are underlined in black.

an N-terminal 14 amino acid tag that includes six consecutive histidine residues. The hexa-histidine tag allows the binding of target proteins to a nickel resin. The integrity of the cloned DNA was confirmed by nucleotide sequence determination.

The above plasmid, pACYCDuet-His$_6$.TssE, was introduced into *E. coli* BL21(λDE3) and overexpression of His$_6$.TssE was carried out at 37°C in BHI and induced with 1 mM IPTG as described in Section 2.5.1.1. This resulted in production of a large amount of a protein corresponding to the predicted molecular weight of TssE (19.7 kDa including the tag) (Figure 5.2A). The protein was shown to be present in the insoluble fraction of *E. coli* cells (Figure 5.2B). Attempts were made to solubilise His$_6$.TssE by growing the cells at a lower temperature (20ºC) and inducing with lower concentrations of IPTG (0.01 and 0.1). However, there was no improvement in the amount of soluble protein obtained (results not shown). Untagged TssE was cloned into pET14b previously and the protein was overproduced in *E. coli* BL21(λDE3) but it was also found to be insoluble under the standard induction conditions (S.Shastri, 2011).

5.2.2. Overproduction of the TssI (VgrG) core region

DNA encoding the gp27.gp5 homologous region of TssI BCAM0148 (one of the 10 VgrG proteins encoded within the *B. cenocepacia* genome), which corresponds to the N-terminal and central regions of TssI (Figure 1.9 and Section 3.7), was amplified from pUT18C-TssI with primers TssIforpET and gp5CtermHisTagrev and blunt end ligated (i.e. without prior restriction digestion) into the *Sma*I site of pUC19 to give pUC19-gp27.gp5.His$_{10}$. These primers resulted in amplification of the first 526 codons of *tssI* and the reverse primer added a deca-histidine tag to the C-terminus of the encoded product. This region of *tssI* was transferred to pACYCDuet-1 by releasing the fragment from pUC19-gp27.gp5.His$_{10}$ using *Nco*I and *Bgl*II, and ligating it into pACYCDuet-1 digested with *Nco*I and *Bam*HI. This placed the core TssI coding sequence into the MCS of the upstream promoter of pACYCDuet-1. The cloning of the gp27.gp5-like region into pUC19 and pACYCDuet-1 vector was confirmed by DNA sequence determination.

Overproduction of core TssI from pACYCDuet-gp27.gp5.His$_{10}$ in *E. coli* BL21(λDE3) was carried out as described in Section 2.5.1.1. SDS-PAGE analysis showed that, after IPTG induction, core TssI was overproduced as judged by the appearance of a high

abundance protein of the expected molecular weight (60.7 kDa including the tag). After subjecting the crude extract (which contained the total cell protein) to sonication and centrifugation, most of the expressed target protein was released into the soluble fraction (Figure 5.3).

5.2.3. Overproduction of His$_{10}$.TssA

tssA was cloned into the MCS located at the upstream promoter of pACYCDuet-1 by releasing it from pET14b-His$_{10}$.TssA (Section 6.3) using *Nco*I and *Bam*HI and transferring it into pACYCDuet-1 digested with the same enzymes. The successful transfer of His$_{10}$.TssA to pACYCDuet-1 was confirmed by DNA sequence determination.

Overproduction of the N-terminally deca-histidine-tagged TssA polypeptide from pACYCDuet-His$_{10}$.TssA in *E. coli* BL21(λDE3) was performed using the standard condition (Section 2.5.1.1). Large amounts of a polypeptide of ~ 45 kDa were observed, which corresponds to the expected size of 44.5 kDa for His$_{10}$-TssA (Figure 5.4). A solubility test showed that TssA protein was present in the soluble fraction (result not shown).

5.3. Co-expression of two *tss* genes in BL21(λDE3) cells using compatible plasmids

The strategy used in this study to facilitate the investigation of protein-protein interactions *in vivo* was to co-overproduce the two proteins of interest in the same cells using a pair of compatible plasmids in which one of each pair of overproduced proteins was His-tagged. Therefore, BL21(λDE3) was transformed with three pairs of compatible plasmids expressing TssA in combination with TssD, TssE or core TssI (i.e. pACYCDuet-His$_6$.TssE with pET14b-TssA, pACYCDuet-gp27.gp5.His$_{10}$ with pET14b-TssA and pACYCDuet-His$_{10}$-TssA with pET14b-TssE) and the overexpression was carried out as described in Section 2.5.1.1. Construction of three of the required plasmids (pACYCDuet-His$_6$.TssE, pACYCDuet-gp27.gp5.His$_{10}$ and pACYCDuet-His$_{10}$-TssA) was described in Section 5.2. Construction of pET14b-TssA and pET14b-TssE was described in Section 6.8.2 and Shastri (2011), respectively. The results showed that, the pET14b derivatives gave good expression of the cloned gene, whereas the pACYCDuet derivatives did not express any detectable levels of protein when present in the same cells as a pET14b derivative (Figure 5.5).
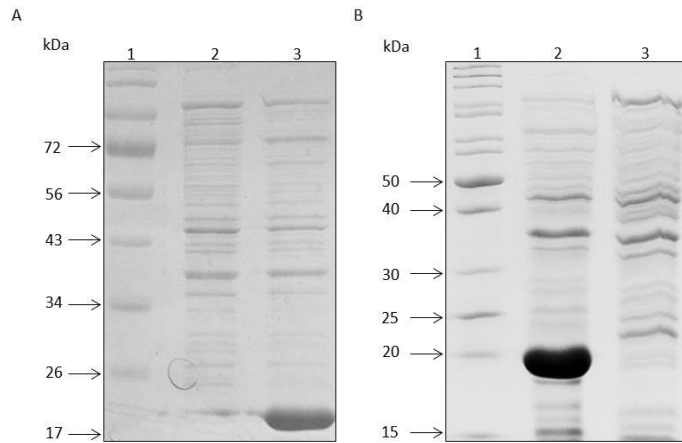
Figure 5.2. Analysis of His₆.TssE protein expression and solubility. *E. coli* strain BL21(λDE3) containing pACYCDuet-His$_6$.TssE was cultured in BHI medium supplemented with chloramphenicol to OD$_{600}$ 0.5 at 37°C whereupon a pre-induction sample was taken and 1 mM IPTG was added to the remainder of the culture to induce expression of the *tssE* gene for a further 3 hours. 15 µl of cleared cell lysate was electrophoresed in a 15 % SDS polyacrylamide gel. A. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, sample of uninduced BL21(λDE3) cells containing pACYCDuet-His$_6$.TssE; lane 3, sample of induced BL21(λDE3) cells containing pACYCDuet-His$_6$.TssE. B. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, crude extract from induced BL21(λDE3) cells containing pACYCDuet-His$_6$.TssE following cell lysis; lane 3, soluble fraction from induced BL21(λDE3) cells containing pACYCDuet-His$_6$.TssE.

Figure 5.3. Analysis of core TssI protein overproduction and solubility. *E. coli* strain BL21(λDE3) containing pACYCDuet-gp27.gp5.His$_{10}$ was cultured in BHI medium supplemented with chloramphenicol to OD$_{600}$ 0.5 at 37°C whereupon a pre-induction sample was taken and 1 mM IPTG was added to the remainder of the culture to induce expression of the core *tssI* gene for a further 3 hours. 15 µl of cleared cell lysate was electrophoresed in a 12 % and 10% SDS polyacrylamide gel, respectively. A. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, sample of uninduced BL21(λDE3) cells containing pACYCDuet-gp27.gp5.His$_{10}$; lane 3, sample of induced BL21(λDE3) cells containing pACYCDuet-gp27.gp5.His$_{10}$. B. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, crude extract from induced BL21(λDE3) cells containing pACYCDuet-gp27.gp5.His$_{10}$ after cell lysis; lane 3, soluble fraction from induced BL21(λDE3) cells containing pACYCDuet-gp27.gp5.His$_{10}$.
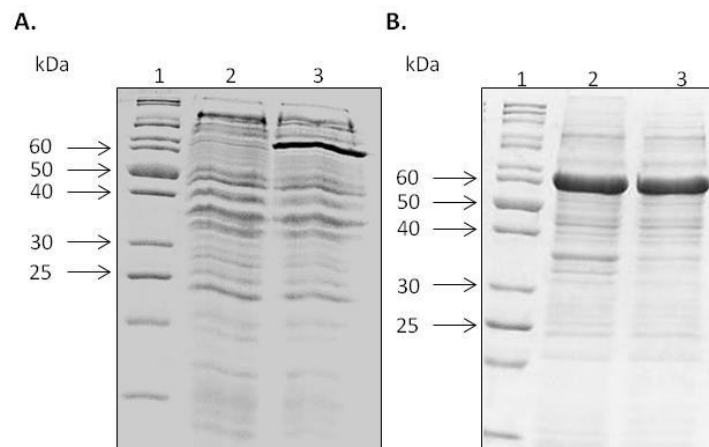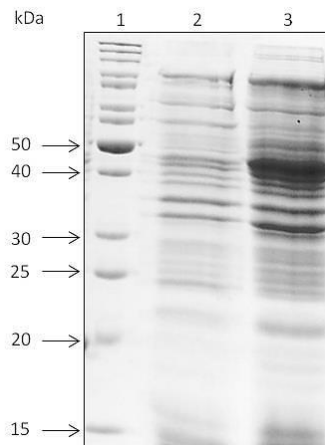
Figure 5.4. Analysis of His$_{10}$.TssA protein overproduction. *E. coli* strain BL21(λDE3) containing pACYCDuet-His$_{10}$.TssA was cultured in BHI medium supplemented with chloramphenicol to OD$_{600}$ 0.5 at 37°C whereupon a pre-induction sample was taken and 1 mM IPTG was added to the remainder of the culture to induce expression of the *tssA* gene for a further 3 hours. 15 µl of cleared cell lysate was electrophoresed in a 12 % SDS polyacrylamide gel. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, sample of uninduced BL21(λDE3) cells containing pACYCDuet-His$_{10}$.TssA; lane 3, sample of induced BL21(λDE3) cells containing pACYCDuet- His$_{10}$.TssA.
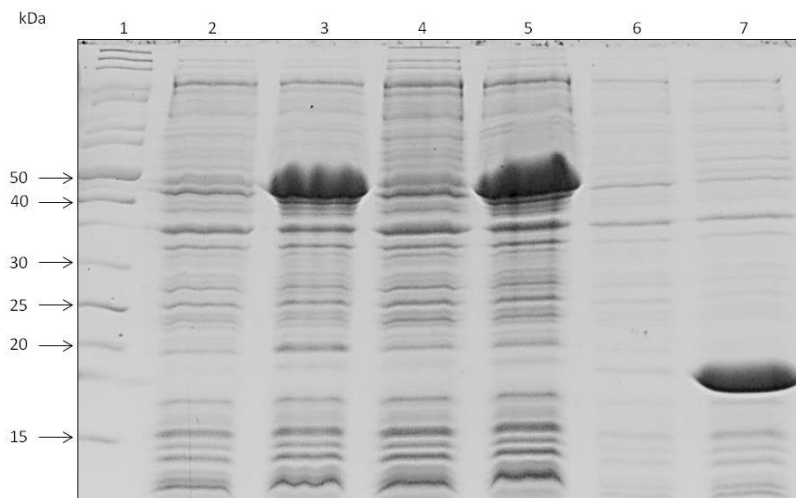
Figure 5.5. Analysis of co-expression of His$_6$.TssE with TssA, core TssI.His$_{10}$ with TssA and His$_{10}$-TssA with TssE. *E. coli* strain BL21(λDE3) containing combinations of pACYCDuet-His$_6$.TssE with pET14b-TssA, pACYCDuet-gp27.gp5.His$_{10}$ with pET14b-TssA and pACYCDuet-His$_{10}$.TssA with pET14b-TssE were cultured in BHI medium supplemented with chloramphenicol and ampicillin to OD$_{600}$ 0.5 at 37°C whereupon pre-induction samples were taken and 1 mM IPTG was added to the remainder of the cultures to induce expression of the genes for a further 3 hours. 15 µl of cleared cell lysate was electrophoresed in a 15 % SDS polyacrylamide gel. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, sample of uninduced BL21(λDE3) cells containing pACYCDuet-His$_6$.TssE with pET14b-TssA; lane 3, sample of induced BL21(λDE3) cells containing pACYCDuet-His$_6$.TssE with pET14b-TssA; lane 4, sample of uninduced BL21(λDE3) cells containing pACYCDuet-gp27.gp5.His$_{10}$ with pET14b-TssA; lane 5, sample of induced BL21(λDE3) cells containing pACYCDuet-gp27.gp5.His$_{10}$ with pET14b-TssA; lane 6, sample of uninduced BL21(λDE3) cells containing pACYCDuet-His$_{10}$.TssA with pET14b-TssE; lane 7, sample of induced BL21(λDE3) cells containing pACYCDuet-His$_{10}$-TssA with pET14b-TssE.

To account for this observation, it was hypothesised that the T7 promoters contained on pACYCDuet-1 are more tightly repressed than the T7 promoter on pET14b as each includes an overlapping *lac* operator and the pACYCDuet-1 vector also encodes the Lac repressor (Figure 5.1). Moreover, it is possible that the Lac repressor tetramer binds the two *lac* operators on the pACYCDuet-1 vector at the same time and thereby generates a 'repression loop' which effects even stronger repression of the T7 promoters in the pACYCDuet derivatives. The native *E. coli lacZ* promoter region contains three *lac* operators and pairs of these operators are bound by the LacI tetramer simultaneously (Gilbert and Muller-Hill, 1966).

To test the 'repression loop' hypothesis, the second T7lac promoter of pACYCDuet-His$_6$.TssE was deleted by digesting the plasmid with *Bsr*GI and *Acc*65I (*Acc*65I recognises the MCS *Kpn*I site) and then self-ligating the digested plasmid. The resulting clone, pACYCDuet-His$_6$.TssE-ΔT7.2, was identified by a diagnostic digestion with *Nde*I which is unable to cut the P$_{T7-2}$-deleted plasmid. Figure 5.6 shows that one of two clones screened was resistant to *Nde*I (clone 2) and its integrity was confirmed by DNA sequencing. After constructing pACYCDuet-His$_6$.TssE-ΔT7.2, the efficiency of expression of TssE was tested in the absence of another plasmid and the result showed that His$_6$.TssE was overproduced in a large amount (Figure 5.7). However, when the co-expression experiment with pET14b-TssA was repeated only His$_6$.TssA was overproduced (Figure 5.7).

It is possible that the "plain" T7 promoter of pET14b plasmid over-dominates the single T7lac promoter of pACYCDuet-His$_6$.TssE-ΔT7.2. In order to confirm that the T7 promoter on pET14b is causing the problem, it was removed from pET14b by cutting the plasmid with *Bam*HI and *Bgl*II followed by self-ligation. The resulting clone, pET14bΔT7, was confirmed by a diagnostic digest with both of the enzymes which were unable to cut the plasmid.

After constructing pET14bΔT7, the expression of pACYCDuet-His$_6$.TssE was tested in combination with pET14b-TssA, pET14b, pET14bΔT7 and pGS301 (the latter is derived from pTrc99A and employs the *E. coli* RNA polymerase-dependent *trc* promoter rather than a T7 promoter to express the *B. pseudomallei fur* gene). In addition,
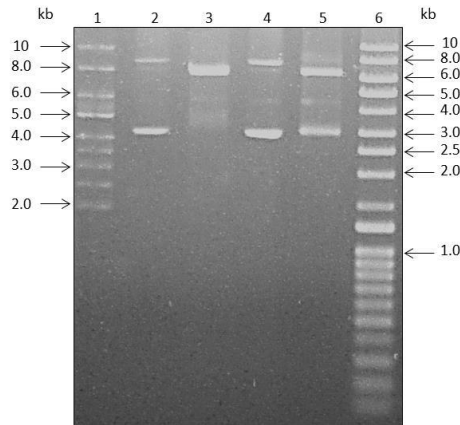
Figure 5.6. Screening for pACYCDuet-His$_6$.TssE-ΔT7.2. Plasmid DNA was prepared from two transformants potentially containing pACYCDuet-His$_6$.TssE-ΔT7.2 and digested by *Nde*I. DNA samples were analysed by electrophosis in a 0.8% agarose gel. Lane 1, Supercoiled ladder; lane 2, plasmid DNA from pACYCDuet-His$_6$.TssE-ΔT7.2 clone 1; lane 3, plasmid DNA from pACYCDuet-His$_6$.TssE-ΔT7.2 clone 1 digested by *Nde*I; lane 4, plasmid DNA from pACYCDuet-His$_6$.TssE-ΔT7.2 clone 2; lane 5, plasmid DNA from pACYCDuet-His$_6$.TssE-ΔT7.2 clone 2 digested by *Nde*I; lane 6, Q-step IV ladder.
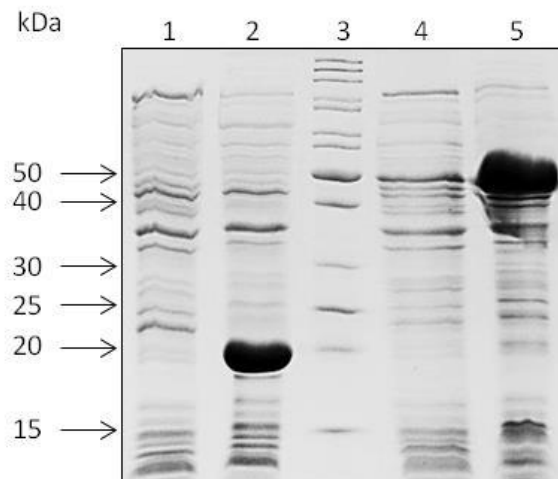
Figure 5.7. Effect of removal of the downstream T7lac promoter of pACYCDuet-1 on expression from the upstream T7lac promoter in the presence of a compatible plasmid containing the T7 promoter. *E. coli* strain BL21(λDE3) containing pACYCDuet-His$_6$.TssE-ΔT7.2 alone or in combination with pET14b-TssA were cultured in BHI medium supplemented with appropriate antibiotics to OD$_{600}$ 0.5 at 37°C whereupon pre-induction samples were taken and 1 mM IPTG was added to the remainder of the cultures to induce expression of the genes for a further 3 hours. 15 µl of cleared cell lysate was electrophoresed in a 15 % SDS polyacrylamide gel. Lane 1, sample of uninduced BL21(λDE3) cells containing pACYCDuet-His$_6$.TssE-ΔT7.2; lane 2, sample of induced BL21(λDE3) cells containing pACYCDuet-His$_6$.TssE-ΔT7.2; lane 3, EZ-Run Rec unstained protein ladder; lane 4, sample of uninduced BL21(λDE3) cells containing pACYCDuet-His$_6$.TssE-ΔT7.2 with pET14b-TssA; lane 5, sample of induced BL21(λDE3) cells containing pACYCDuet-His$_6$.TssE-ΔT7.2 with pET14b-TssA.

pACYCDuet-His$_{10}$.TssA was tested in combination with pUT18C-tssK and pUT18C-tssH. pUT18C utilises the *E. coli* RNA polymerase-dependent *lac* promoter to drive expression of T18-Tss protein fusions. SDS-PAGE analysis showed that His$_6$-TssE was overproduced when expressed on its own and in combination with pET14bΔT7, but not with pET14b-TssA or pET14b. Also, both proteins encoded by pACYCDuet-His$_6$.TssE and pGS301 were successfully overproduced when they were co-expressed in the same cells (*B. pseudomallei* Fur is 16.3 kDa). Similarly, when BL21(λDE3) was co-transformed with pACYCDuet-His$_{10}$.TssA and pUT18C-tssK or pUT18C-tssH, both pairs of encoded proteins were overproduced (T18-TssK is 70.8 kDa and T18-TssH is 117.3 kDa) (Figure 5.8). Therefore, it was concluded that pET14b and pACYCDuet-1 vectors cannot be used together to co-express two genes as the T7 promoter on the former over-dominates that of the latter. Given that this occurred under conditions of derepression of the T7 promoter, when no *Lac* repressor should be bound to the Lac operator, the mechanism of promoter dominance is not clear.

5.4. Co-overproduction of two T6SS subunits from the same expression vector

To solve the problem of pET14b being dominant over pACYCDuet-1, it was decided to express two T6SS subunit genes from the same plasmid. Therefore, the *tssD*, *tssH* and *tssI* (BCAM0148) genes were cloned separately into pACYCDuet-linkerHis1.TssA (Section 6.5) to generate plasmids expressing the following combinations: TssD with linkerHis$_6$.TssA, TssH with linkerHis$_6$.TssA and TssI with linkerHis$_6$.TssA. Also, these genes were cloned separately into pACYCDuet-1. For co-expression of TssA and TssE, *tssA* was cloned into pACYCDuet-His$_6$.TssE and a PCR product encoding TssE.His$_6$ was cloned into pACYCDuet-TssA. The details of these cloning experiments are included in the appropriate sections below.

5.4.1. Co-overproduction of His$_6$.TssE and TssA for pull downs

pACYCDuet-His$_6$.TssE-TssA was constructed by R. Sun and M. Thomas (unpublished results). The *tssA* gene was cloned into the downstream MCS of pACYCDuet-His$_6$.TssE by releasing it from pET14b-His$_6$.TssA (Section 6.2) using *Nde*I and *Bam*HI and transferring it into pACYCDuet-His$_6$.TssE digested with *Nde*I and *Bgl*II. Therefore, pACYCDuet-His$_6$.TssE-TssA expresses N-terminal His-tagged TssE and untagged TssA.
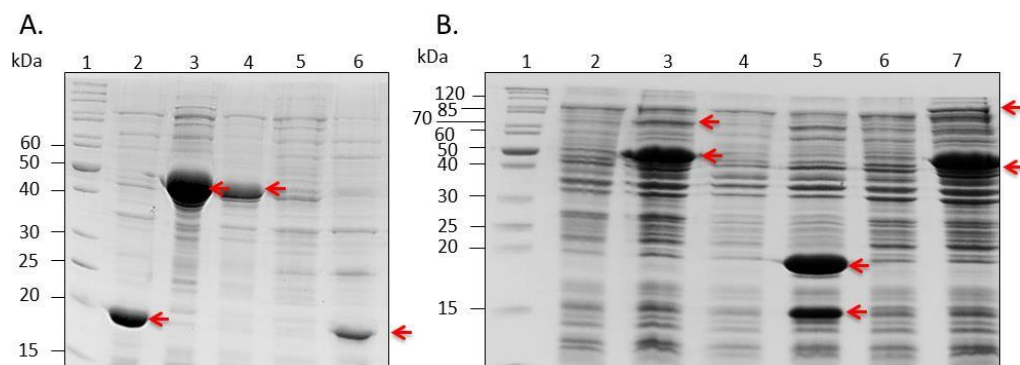
Figure 5.8. Effect of plasmids containing the T7, *trc* and *lac* promoters on the activity of the T7 *lac* promoters contained on pACYCDuet-1. *E. coli* strain BL21(λDE3) containing compatible pairs of expression plasmids were cultured in BHI medium supplemented with appropriate antibiotics to OD_{600} 0.5 at 37°C whereupon pre-induction samples were taken and 1 mM IPTG was added to the remainder of the cultures to induce expression of the genes for a further 3 hours. 15 µl of the cleared cell lysates were electrophoresed in 15 % SDS polyacrylamide gels. A. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, sample of induced BL21(λDE3) cells containing pACYCDuet-His$_6$.TssE; lane 3, sample of induced BL21(λDE3) cells containing pET14b-TssA; lane 4, sample of induced BL21(λDE3) cells containing pACYCDuet-His$_6$.TssE and pET14b-TssA; lane 5, sample of induced BL21(λDE3) cells containing pACYCDuet-His$_6$.TssE and pET14b; lane 6, sample of induced BL21(λDE3) cells containing pACYCDuet-His$_6$.TssE and pET14bΔT7. B. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, sample of uninduced BL21(λDE3) cells containing pACYCDuet-His$_{10}$.TssA and pUT18C-tssK; lane 3, sample of induced BL21(λDE3) cells containing pACYCDuet-His$_{10}$.TssA and pUT18C-tssK; lane 4, sample of uninduced BL21(λDE3) cells containing pACYCDuet-His$_6$.TssE with pGS301; lane 5, sample of induced BL21(λDE3) cells containing pACYCDuet-His$_6$.TssE with pGS301; lane 6, sample of uninduced BL21(λDE3) cells containing pACYCDuet-His$_{10}$.TssA and pUT18C-tssH; lane 7, sample of induced BL21(λDE3) cells containing pACYCDuet-His$_{10}$.TssA and pUT18C-tssH. Red arrows indicate the expected overproduced proteins.

An expression test of pACYCDuet-His$_6$.TssE-TssA was carried out in BL21(λDE3) growing at 37°C and the induction was performed with 0.5 mM IPTG. The result showed that both proteins were produced in high abundance. However, the solubility test showed that most of the His$_6$.TssE protein was present in the insoluble fraction, whereas TssA was soluble (Figure 5.9). Attempts were made to solubilise His$_6$-TssE by growing the cells at a lower temperature (20ºC) and inducing with a lower IPTG concentration (0.01 and 0.1 mM) for 16 hours. However, there was no improvement in the amount of soluble His$_6$-TssE protein obtained (result not shown). It was decided to proceed with the nickel resin pull down to determine whether there was a small amount His$_6$.TssE in the soluble fraction of the cell lysate of the culture grown at 20ºC and induced with 0.5 mM IPTG, which can bind to the nickel beads and pull down TssA.

The nickel resin pull down was carried out as described in Section 2.5.10. As controls, cleared cells lysates were prepared from BL21(λDE3) cells containing pACYCDuet-His$_6$.TssE and pACYCDuet-TssA following IPTG induction. The result showed that a small amount of a protein corresponding in size to His$_6$.TssE (19.7 kDa) was eluted from the nickel resin with 500 mM imidazole and was observed on a Coomassie blue-stained SDS-PA gel (Figure 5.10A). A protein of similar molecular weight was confirmed by immunodetection with Hisprobe-HRP (Figure 5.10B). Although a protein corresponding to TssA was not observed by SDS-PAGE, a weak signal corresponding to TssA could be detected by western blotting with anti-TssA.CTD polyclonal antibody. However, the TssA signal was also detected in the control where a cell lysate containing untagged TssA alone was applied to the resin. Therefore, the TssA that co-eluted with His$_6$.TssE might be due to non-specific binding of TssA to the nickel resin.

The soluble cell lysates in the experiment described above were prepared in a lysis buffer that contained 200 mM NaCl. To investigate the possibility that the high concentration of NaCl in the buffer interfered with the interaction of His$_6$.TssE with TssA. The cell lysates were prepared using a lysis buffer with different NaCl concentrations (0-500 mM NaCl). The results showed that no TssA could be detected by SDS-PAGE analysis using the alternative conditions (Figure 5.11). However, the lysis buffer with 50 mM NaCl was used in all subsequent pull down experiments unless stated otherwise.
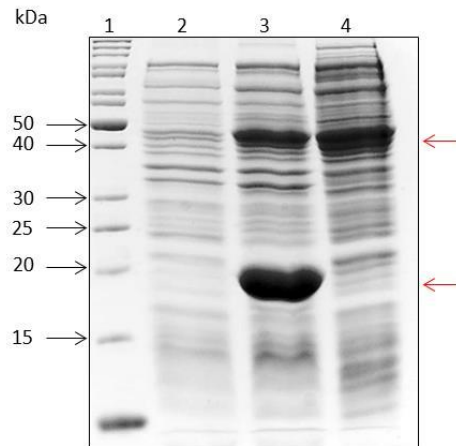
Figure 5.9. Analysis of His₆.TssE and TssA protein expression and solubility following expression from pACYCDuet-His₆.TssE-TssA. *E. coli* strain BL21(λDE3) containing pACYCDuet-His₆.TssE-TssA was cultured in BHI medium supplemented with chloramphenicol to OD_{600} 0.5 at 37°C whereupon a pre-induction sample was taken and 1 mM IPTG was added to the remainder of the culture to induce expression of the *tssE* and *tssA* genes for a further 3 hours. 15 µl of cleared cell lysate was electrophoresed in a 15 % SDS polyacrylamide gel. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, sample of uninduced BL21(λDE3) cells containing pACYCDuet-His₆.TssE-TssA; lane 3, sample of induced BL21(λDE3) cells containing pACYCDuet-His₆.TssE-TssA; lane 4, soluble fraction from induced BL21(λDE3) cells containing pACYCDuet-His₆.TssE-TssA. The location of TssA and His₆.TssE is indicated by red arrows.
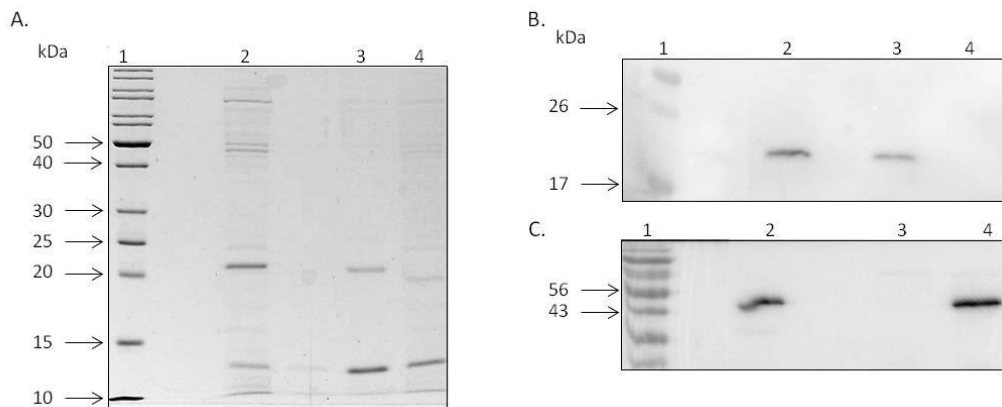
Figure 5.10. Attempted pull down of TssA by His₆.TssE. A. Co-purification of His₆.TssE with TssA. Soluble fraction from induced BL21(λDE3) cells containing co-overproduced His₆.TssE and TssA, His₆.TssE alone or TssA alone was incubated with nickel affinity beads. Eluted samples following washing were analysed for the presence of both His₆.TssE and TssA by SDS-PAGE followed by Coomassie staining. Lane 1, EZ-Run Rec Protein ladder; lane 2, TssA and His₆.TssE lysate eluted from nickel column; lane 3, His₆.TssE lysate eluted from nickel column; lane 4, TssA lysate eluted from nickel column. B. Immunoblotting of eluted proteins from nickel resin using Hisprobe-HRP (to detect His₆.TssE). C. Immunoblotting of eluted proteins from nickel resin using anti-TssA.CTD antibody (to detect TssA).
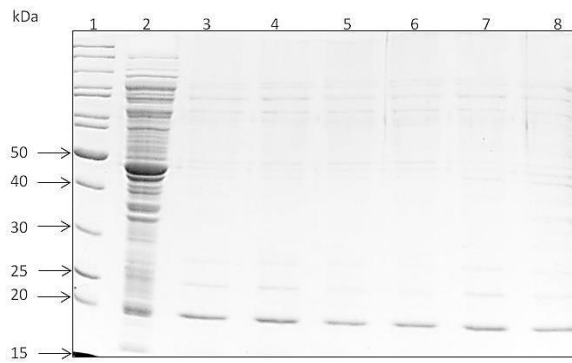
Figure 5.11. Effect of NaCl concentration on ability of His$_6$.TssE to pull down TssA. Clarified cell lysates were prepared from BL21(λDE3) cells containing pACYCDuet-His$_6$.TssE-TssA using a lysis buffer with different NaCl concentrations (0-500 mM NaCl). The soluble fraction from induced BL21(λDE3) containing His$_6$.TssE and TssA was incubated with nickel affinity beads. After washing the resin in a buffer containing the same concentration of NaCl and 10 mM imidazole, bound proteins were eluted in a buffer containing the same concentration of NaCl and 500 mM imidazole and analysed for the presence of both His$_6$.TssE and TssA by 15% SDS-PAGE followed by Coomassie blue staining. Lane 1, EZ-Run Rec Protein ladder; lane 2, soluble fraction prepared in lysis buffer containing 50 mM NaCl; lane 3-8, eluted proteins from the nickel beads pull down performed in the presence of 0, 50, 100, 150, 200 and 500 mM NaCl.

5.4.2. Co-overproduction of TssE.His$_6$ and TssA for pull downs

The BACTH assays showed that TssE interacts with TssA when the N-terminal region of TssE is free (Section 3.5.3). Therefore, it is possible that the His-tag at the N-terminus of TssE interfered with its proposed interaction with TssA. To avoid this potential problem, a new plasmid expressing TssA and TssE was constructed in which the His-tag was placed at the C-terminal end of TssE (i.e. pACYCDuet-TssE.His$_6$-TssA). To construct this plasmid, *tssE* was amplified from pUT18C-TssE with primers pACYCDuettssEfor and pACYCDuettssErevHis. The forward primer contained the first 20 base pairs of the *tssE* coding sequence preceded by two nucleotides (TC) to create a *Bsp*HI site with the first 4 nucleotides of *tssE* (ATGA). The reverse primer (pACYCtssErevHis) contained a *Bam*HI site, stop codon, the 18 base pairs of the His-tag coding sequence and the final 20 base pairs of the *tssE* coding sequence (excluding the stop codon). The PCR product was then ligated between the compatible *Nco*I and *Bam*HI sites of pACYCDuet-TssA giving rise to pACYCDuet-TssE.His$_6$-TssA.

The expression test of pACYCDuet-TssE.His$_6$-TssA was carried out at 20°C in BL21(λDE3) to increase the chance of obtaining soluble TssE.His$_6$ and the induction was performed with 0.5 mM IPTG. The result showed TssE.His$_6$ was produced in high abundance whereas expression of TssA was decreased relative to that observed in cells harbouring pACYCDuet-His$_6$.TssE-TssA (compare Figure 5.12 and 5.9). Again, the solubility test showed that TssE was mainly insoluble whereas the small amount of TssA was soluble (Figure 5.12). The result of nickel beads pull down of TssE.His$_6$ with TssA was similar to that for co-expressed His$_6$.TssE and TssA (result not shown).

5.4.3. Co-overproduction of VSVg.TssD and linkerHis1.TssA for pull down and co-immunoprecipitation

The *tssD* gene was PCR amplified from pUT18C-TssD and cloned into pACYCDuet-linkerHis1.TssA and pACYCDuet-1 using *Nco*I and *Bgl*II by including their sites in the forward and reverse primers, respectively (see appendix 1 for the primer sequences). The forward primer (TssDforpACYC.NtermVSVGTag) also encode the VSVg epitope tag sequence and the first 20 base pairs of the *tssD* coding sequence, whereas the reverse primer (TssDRevpACYC.NtermVSVGTag) also contained the final 20 base pairs of *tssD* coding sequence including the stop codon. The VSVg tag was included for convenient identification of TssD using western blotting, as it is relatively short tag and

VSVg tag antibody is available commercially. The PCR product was then ligated between the corresponding sites of pACYCDuet-linkerHis1.TssA and pACYCDuet-1 giving rise to pACYCDuet-VSVg.TssD-linkerHis1.TssA and pACYCDuet-VSVg.TssD, respectively.

The expression and solubility test of cells containing VSVg.TssD-linkerHis1.TssA and pACYCD-VSVg.TssD showed that TssD was produced in high abundance and was soluble when it is expressed separately and also in combination with TssA (the mol. wt of VSVg.-tagged TssD is 20 kDa). In contrast, although TssA was efficiently co-expressed with TssD, not much of it was present in the soluble fraction (the mol. wt of linkerHis-tagged TssA is 42.6 kDa) (Figure 5.13). For expression and solubility test of linkerHis1.TssA see Section 6.5.

Following the preparation of the soluble fractions of the cell lysates containing VSVg.TssD and linkerHis1.TssA, separately and together, a pull down was performed as described in Section 2.5.10. The cell lysate containing only VSVg.TssD was used as a control to ensure that retention of TssD on the nickel column in the presence of TssA was caused only by the interaction of VSVg.TssD with TssA, rather than interaction by VSVg.TssD alone with the resin. The result showed that VSVg-TssD bound to the resin in the presence and absence of TssA (Figure 5.14A).

In an attempt to find improved conditions which minimise the non-specific binding of TssD to the resin, the experiment was repeated using three different buffer conditions. The SDS-PAGE analysis showed that lowering the pH from 8.0 to 7.0, substituting KCl for NaCl or adding a non-ionic detergent gave similar results (Figure 5.14B, C and D).

The TssD amino acid sequence contains two histidine residues at the N-terminus (positions 3 and 5), and it is known that TssD forms hexamers. Therefore it was hypothesised that the TssD hexamer may have a natural histidine tag, as some of the histidine residues may be in close proximity. Moreover, it has been shown that three consecutive histidine residues are sufficient for protein purification by IMAC under non-denaturing conditions (Terpe, 2003). Also, it is not required to have adjacent histidines, they can be separated by one amino acid such as occurs with His3 and His5
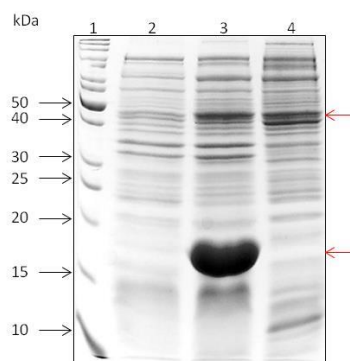
Figure 5.12. Analysis of TssE.His$_6$ and TssA protein expression and solubility following expression from pACYCDuet-TssE.His$_6$-TssA. *E. coli* strain BL21(λDE3) containing pACYCDuet-TssE.His$_6$-TssA was cultured in BHI medium supplemented with chloramphenicol to OD$_{600}$ 0.5 at 37°C whereupon a pre-induction sample was taken and 1 mM IPTG was added to the remainder of the culture to induce expression of the *tssE* and *tssA* genes for a further 3 hours. 15 µl of cleared cell lysate was electrophoresed in a 15 % SDS polyacrylamide gel. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, sample of uninduced BL21(λDE3) cells containing pACYCDuet-TssE.His$_6$-TssA; lane 3, sample of induced BL21(λDE3) cells containing pACYCDuet-TssE.His$_6$-TssA; lane 4, soluble fraction from induced BL21(λDE3) cells containing pACYCDuet-TssE.His$_6$-TssA. The location of TssA and TssE.His$_6$ is indicated by red arrows.
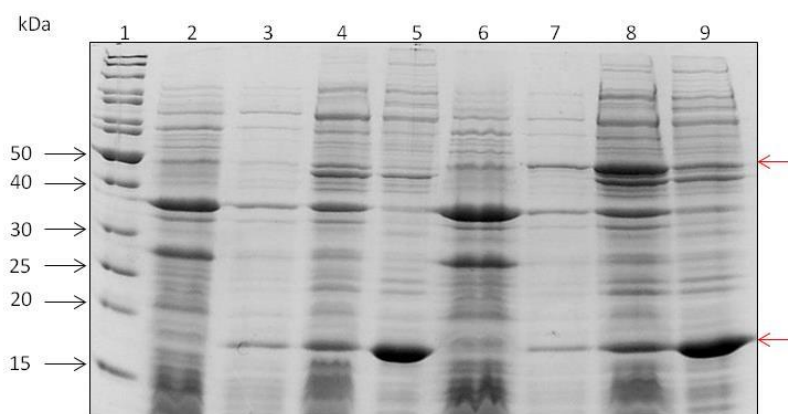
Figure 5.13. Analysis of VSVg.TssD expression and solubility in the absence and presence of linker His-tagged TssA. *E. coli* strain BL21(λDE3) containing pACYCDuet-VSVg.TssD-linkerHis1.TssA and pACYCDuet-VSVg.TssD were cultured in BHI medium supplemented with chloramphenicol to OD$_{600}$ 0.5 at 37°C whereupon pre-induction samples were taken and 1 mM IPTG were added to the remainder of the cultures to induce expression of the *tssA* and *tssD* genes for a further 3 hours. 15 µl of cleared cell lysate was electrophoresed in a 15 % SDS polyacrylamide gel. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, sample of uninduced BL21(λDE3) cells containing pACYCDuet-VSVg.TssD; lane 3, sample of induced BL21(λDE3) cells containing pACYCDuet-VSVg.TssD; lane 4, crude extract from induced BL21(λDE3) cells containing pACYCDuet-VSVg.TssD after cell lysis; lane 5, soluble fraction from induced BL21(λDE3) cells containing pACYCDuet-VSVg.TssD; lane 6, sample of uninduced BL21(λDE3) cells containing pACYCDuet-VSVg.TssD-linkerHis1.TssA; lane 7, sample of induced BL21(λDE3) cells containing pACYCDuet-VSVg.TssD-linker His1.TssA; lane 8, crude extract from induced BL21(λDE3) cells containing pACYCDuet-VSVg.TssD-linkerHis1.TssA after cell lysis; lane 9, soluble fraction from induced BL21(λDE3) cells containing pACYCDuet-VSVg.TssD-linkerHis1.TssA. Red arrows indicate the location of linker His-tagged TssA and VSVg-tagged TssD.
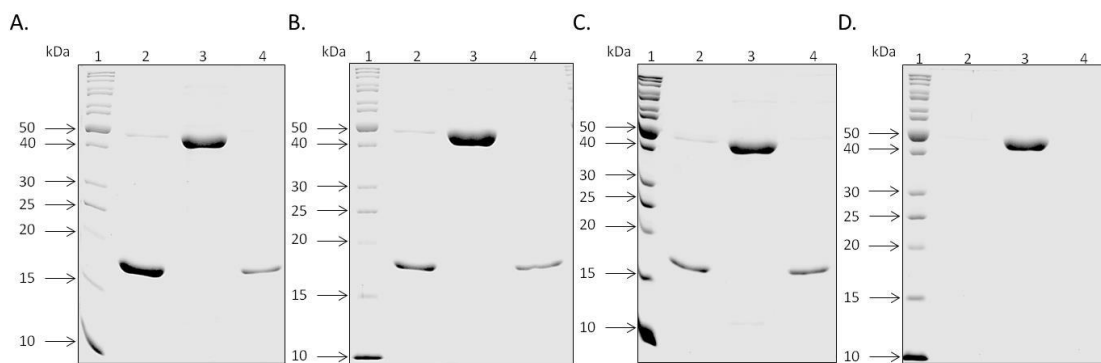
Figure 5.14. Attempted pull down of TssD by His-tagged TssA. The soluble fraction from induced BL21(λDE3) containing linkerHis1.TssA and VSVg-tagged TssD, VSVg-tagged TssD alone or linkerHis1.TssA alone was incubated with nickel affinity beads. After washing the resin, bound proteins were eluted in 500 mM imidazole and analysed for the presence of both linkerHis1.TssA and VSVg.TssD by 15% SDS-PAGE followed by Coomassie blue staining. Lane 1, EZ-Run Rec Protein ladder; lane 2, cell lysate containing linkerHis1.TssA and VSVg.TssD; lane 3, cell lysate containing linkerHis1.TssA; lane 4, cell lysate containing VSVg.TssD. The soluble cell lysate and the pull down was prepared and carried out in buffer containing A. 50 mM Tris-HCl pH 8.0, 50 mM NaCl, 10% glycerol and 10 mM imidazole. B. 50 mM Tris-HCl pH 7.0, 50 mM NaCl, 10% glycerol and 10 mM imidazole. C. 50 mM Tris-HCl pH 8.0, 50 mM KCl, 10% glycerol and 10 mM imidazole. D. 50 mM Tris-HCl pH 8.0, 50 mM NaCl, 10% glycerol, 10 mM imidazole and 0.1% triton.

in TssD (Block *et al.*, 2009). To overcome this potential problem, it was decided to modify TssD by replacing both histidine residues with amino acids that are more strongly conserved at these positions. In order to achieve this, the forward primer (TssDforpACYC.NtermVSVGTag.Mod) was designed to contain an *Nco*I site, VSVg tag sequence and the first 20 base pairs of *tssD* coding sequence with mutations to change the first histidine codon into an asparagine (AAT) and the second histidine to tyrosine (TAC). The mutation was based on alignment with other TssD sequences. The reverse primer was the same as that used for construction of pACYCDuet-VSVg.TssD (TssDRevpACYC.NtermVSVGTag). The *tssD* gene was PCR amplified from pUT18C-TssD using the above primers and the product was then ligated between the corresponding sites of pACYCDuet-linkerHis1.TssA and pACYCDuet-1 giving rise to pACYCDuet-VSVg.TssD$_m$-linkerHis1.TssA and pACYCDuet-VSVg.TssD$_m$.

An expression and solubility test performed on cells containing pACYCDuet-VSVg.TssD$_m$-linkerHis1.TssA and pACYCDuet-VSVg.TssD$_m$ showed that TssD$_m$ was produced in high abundance and was soluble when expressed separately and in combination with TssA. Linker1HisTssA co-expressed with TssD$_m$ was produced in a smaller amount compared to when it was coexpressed with TssD and again, little of it was present in the soluble fraction (Figure 5.15).

A pull down with the modified version of TssD was carried out as described in Section 2.5.10. SDS-PAGE analysis showed that some VSVg.TssD$_m$ with linkerHis1.TssA was retained on the column, but unfortunately the VSVg.TssD$_m$ was found to bind to the nickel resin in the absence of TssA (Figure 5.16). Therefore, the removal of the N-terminal histidines of TssD did not solve the problem of the non-specific binding of TssD to nickel resin.

Co-immunoprecipitation (Co-IP) was then employed as an alternative approach to investigate the interaction of TssA with TssD. Co-IP is a well-known technique to study protein-protein interactions and is based on immunoprecipitating one protein with a specific antibody and analysing the resulting immune complexes by SDS-PAGE. Western blotting can then be used to identify interacting partners.
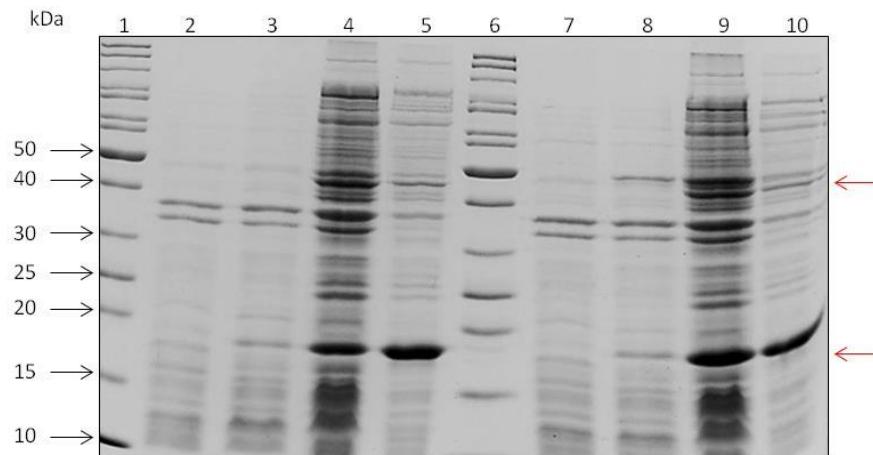
Figure 5.15. Analysis of VSVg.TssD$_m$ expression and solubility in the absence and presence of linker His-tagged TssA. *E. coli* strain BL21(λDE3) containing pACYCDuet-VSVg.TssD$_m$-linkerHis1.TssA and pACYCDuet-VSVg.TssD$_m$ were cultured in BHI medium supplemented with chloramphenicol to OD$_{600}$ 0.5 at 37°C whereupon pre-induction samples were taken and 1 mM IPTG were added to the remainder of the cultures to induce expression of the *tssA* and *tssD$_m$* genes for a further 3 hours. 15 μl of cleared cell lysate was electrophoresed in a 15 % SDS polyacrylamide gel. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, sample of uninduced BL21(λDE3) cells containing pACYCDuet-VSVg.TssD$_m$; lane 3, sample of induced BL21(λDE3) cells containing pACYCDuet-VSVg.TssD$_m$; lane 4, crude extract from induced BL21(λDE3) cells containing pACYCDuet-VSVg.TssD$_m$ after cell lysis; lane 5, soluble fraction from induced BL21(λDE3) cells containing pACYCDuet-VSVg.TssD$_m$; lane 6, sample of uninduced BL21(λDE3) cells containing pACYCDuet-VSVg.TssD$_m$-linkerHis1.TssA; lane 7, sample of induced BL21(λDE3) cells containing pACYCDuet-VSVg.TssD$_m$-linkerHis1.TssA; lane 8, crude extract from induced BL21(λDE3) cells containing pACYCDuet-VSVg.TssD$_m$-linkerHis1.TssA after cell lysis; lane 9, soluble fraction from induced BL21(λDE3) cells containing pACYCDuet-VSVg.TssD$_m$-linker His1.TssA. Red arrows indicate the location of linker His-tagged TssA and VSVg-tagged TssD$_m$.
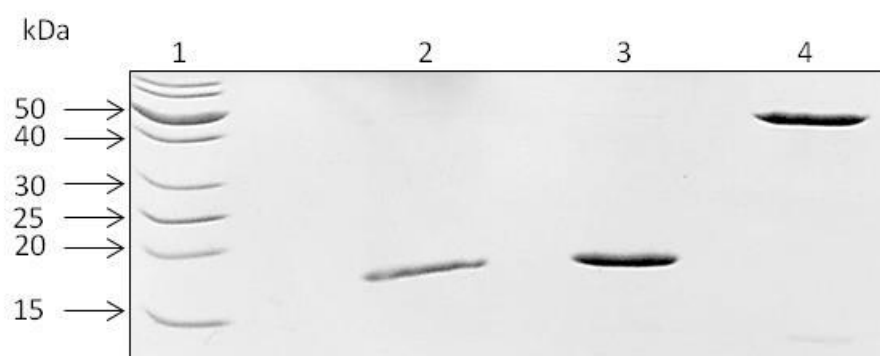
Figure 5.16. Attempted pull down of TssD$_m$ by His-tagged TssA. The soluble fraction from induced BL21(λDE3) containing linkerHis1.TssA and VSVg-tagged TssD$_m$, VSVg-tagged TssD$_m$ alone or linkerHis1.TssA alone was incubated with nickel affinity beads. After washing the resin, bound proteins were eluted in 500 mM imidazole and analysed for the presence of both linkerHis1.TssA and VSVg.TssD$_m$ by 15% SDS-PAGE followed by Coomassie blue staining. Lane 1, EZ-Run Rec Protein ladder; lane 2, cell lysate containing linkerHis1.TssA and VSVg.TssD$_m$; lane 3, cell lysate containing VSVg.TssD$_m$; lane 4, cell lysate containing linkerHis1.TssA.

Following the preparation of the soluble fraction of the cell lysate containing VSVg.TssD and linkerHis1.TssA, separately and together, a Co-IP was performed as described in Section 2.5.11. VSVg.TssD was specifically immunoprecipitated by anti-VSVg antibody and incubated with agarose beads coated with protein A. Material retained on the beads was separated by SDS-PAGE and immunodetected with anti-TssA polyclonal antibody to detect linkerHis1.TssA and anti-TssD polyclonal antibody to detect VSVg.TssD (Figure 5.17A and B). The result showed that although linkerHis1.TssA was present in the anti-VSVg precipitated material when coexpressed with VSVg.TssD, it could also be detected when expressed on its own. Subsequent experiments showed that linkerHis1.TssA could non-specifically bind to the protein A-agarose beads (result not shown). The experiment was repeated by precipitating VSVg.TssD with linkerHis1.TssA by anti-TssD antibody. The immunoprecipitated material was separated by SDS-PAGE and immunodetected with anti-TssA polyclonal antibody to detect linkerHis1.TssA and anti-VSVg antibody to detect VSVg.TssD. A similar result was obtained as when anti-VSVg was used for the immunoprecipitation (result not shown). To investigate whether the observed immunoprecipitation of VSVg.TssD with linkerHis1.TssA was due to specific antibody precipitation or not, the assay was repeated by incubating the protein mixture with the agarose beads in the absence of antibody. The result of the western blotting showed that both proteins bind non-specifically to the agarose beads (result not shown).

To overcome the problem of non-specific binding of the proteins of interest to protein A-agarose, the Co-IP experiment was repeated with some modification using anti-VSVg antibody. The NaCl concentration in the lysis and wash buffer was increased from 50 mM to 150 mM. Also the protein samples were pre-cleared by incubation with the protein A-agarose beads. Finally the wash step was repeated five times (instead of 3 times as in the standard protocol) and the duration of wash was increased to 15 minutes (instead of 5 minutes). Unfortunately, the result showed that linkerHis1.TssA was present in the anti-VSVg precipitated material when coexpressed with VSVg.TssD, but it could also be detected when expressed on its own (result not shown). This experiment was repeated without using an antibody to precipitate the protein of interest, i.e. just an incubation of cleared cell lysates with protein A-agarose, again the result showed that linkerHis1.TssA and VSVg.TssD were found to bind non-specifically to protein A-agarose beads.

A study by Williams *et al* in 2006 demonstrated the advantage of pre-treatment of protein A-sepharose (PAS) with 0.2 M glycine at pH 10.6 overnight at room temperature as it reduces the non-specific binding of proteins by 84% (Williams *et al*., 2006). Therefore, further attempts at Co-IP assays were carried out using glycine blocked EZview Red protein A affinity gel beads, and also BSA blocked EZview Red protein A affinity gel beads (Section 2.5.11). The result showed that the two target proteins still bind non-specifically to the beads (result not shown). Due to time constraints no further experiments were performed.
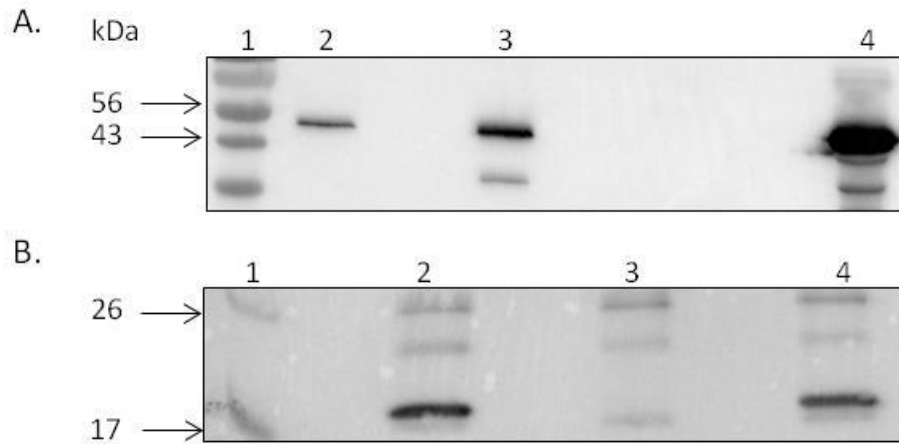
Figure 5.17. Identification of eluted proteins from the co-immunoprecipitation experiment of VSVg.TssD and linkerHis1.TssA by western blotting. The eluted protein samples from the EZview Red protein A affinity gel beads were further identified by western blotting. A. Samples were identified by immunoblotting using anti-TssA antibody. Lane 1, EZ-Run Rec Protein ladder; lane 2, cell lysate containing VSVg.TssD and linkerHis1.TssA; lane 3, cell lysate containing linkerHis1.TssA; lane 4, cell lysate containing linkerHis1.TssA not incubated with the beads as positive control B. Samples were identified by immunoblotting using anti-TssD antibody. Lane 1, EZ-Run Rec Protein ladder; lane 2, cell lysate containing VSVg.TssD and linkerHis1.TssA; lane 3, cell lysate containing linkerHis1.TssA; lane 4, cell lysate containing VSVg.TssD not incubated with the beads as positive control.

5.5. Discussion

To confirm the potential TssA interactions with some Tss proteins, it was planned to perform pull down experiments with TssD, TssE, TssH and TssI. In order to test *in vivo* interaction of the two proteins of interest, initially it was decided to co-express these proteins from two compatible plasmids, as several versions of *tssA* were already cloned into pET14b and there is only a single promoter in this plasmid. Therefore, the other *tss* genes (*tssD*, *tssE*, *tssH* and *tssI*) were cloned into pACYCDuet vector which is compatible with pET14b. Surprisingly, the results of the co-expression showed that only proteins that were encoded by pET14b were strongly expressed whereas expression of the protein encoded by the pACYCDuet plasmid was not detectable.

As pACYCDuet contains two T7lac promoters, it was speculated that a repression loop was created in the Duet vectors by the binding of a Lac repressor tetramer to both operators and this completely shut down expression of *tss* genes in the Duet clone (Gilbert and Muller-Hill, 1966). If this is true then deleting one of the operators (or maybe cloning a very long gene downstream of the first promoter to push the two operators far apart) would relieve the strong repression. However, deletion of the second T7lac promoter of pACYCDuet-His$_6$.TssE resulted in no improvement in the expression of TssE when combined with pET14b derivatives. Therefore, the loop suppression theory seems to be unlikely.

According to the commercial supplier of the Duet plasmids, Novagen, the combination of a "plain" T7 promoter pET plasmid (i.e. pET-3a-d, pET14b, pET-20b(+), etc.) with a T7lac promoter plasmid is not recommended. When T7 and T7lac promoters are used together, the T7 promoter will be preferentially targeted as it readily available and therefore lower expression is observed from the vector possessing a T7lac promoter. Surprisingly, in the work described in this chapter, the genes which were cloned into plasmids under the control of T7lac promoter (the pACYCDuet derivatives) yielded no detectable level of protein overexpression. It was predicted that the normal T7 promoter on pET14b somehow prevents expression from the T7lac promoters on pACYCDuet. Therefore, pET14bΔT7 was constructed and tested in combination with pACYCDuet-His$_6$.TssE which showed overproduction of TssE in this situation, indicating that promoter competition is the likely cause of the inhibition. To further substantiate the hypothesis that the problem is due to competition between T7 promoters on different

vectors, the vector containing the normal T7 promoter could be replaced by a plasmid that does not use such a promoter. In pGS301 the strong *trc* promoter on this plasmid uses the *E. coli* host RNA polymerase, not T7 RNA polymerase. As there will be no competition between vectors for T7 RNAP, all the T7 RNA polymerase will only bind to the pACYCDuet plasmid. pGS301 is in the same plasmid incompatibility group as pET, therefore it is compatible with pACYCDuet and encodes ampicillin resistance. Similarly, pUT18C-TssH and pUT18C-TssK utilise the *lac* promoter which also depends on *E. coli* RNA polymerase for expression of the T18-TssH and T18-TssK fusion proteins, respectively. The results of these experiments indicate that T7 and T7lac promoters are unsuitable for co-expression of two proteins, although it is still not clear why under inducing conditions no protein expression is detectable from the pACYCDuet derivatives.

The second approach was to clone the two genes of the interest on a plasmid containing two strong promoters, i.e. pACYCDuet. This method was often complicated by the unequal production of the two proteins. It was observed that the product of the gene cloned at the second MCS was expressed at a considerably lower level than the gene product cloned at the first MCS and this could interfere with demonstrating the full extent of the interaction between the two proteins of interest. In the absence of a gene inserted at the upstream promoter, expression of the protein from the downstream promoter was enhanced.

The unequal expression of the two proteins was not the only problem for the co-expression of His_6.TssE and TssA. Although His_6.TssE was produced in a large amount, it was insoluble even when the host strain was cultured at 20˚C and induced with lower concentrations of IPTG. Lossi *et al* were shown that GST-tagged TssE-3, one of the three TssE proteins found in *P. aeruginosa*, which contain three T6SS gene clusters, was soluble when the cells were grown at 20ºC and induced with 1 mM IPTG for 16 hours. Whereas the other two TssE were insoluble (Lossi *et al*., 2011).

Although the insolubility of His_6.TssE was not ideal for a pull down, an attempt was made using nickel beads to immobilise TssE. Unexpectedly, a weak protein band which corresponds in size to His_6.TssE was detected on a 15% SDS-PA gel and the identity of the protein was confirmed by immunoblotting. No detectable amounts of TssA protein

were seen in association with the eluted His_6.TssE except by western blotting. It was thought that the high concentration of NaCl (200 mM) in the lysis and wash buffers may have interfered with the interaction of TssE with TssA. This possibility was ruled out by testing different buffer conditions. Another possible reason for the inability of His_6.TssE to pull down TssA was the position of the His-tag. However, a pull down assay using TssE.His_6 gave rise to a similar result to the His_6.TssE experiment. Another alarming problem was the detection of TssA protein by immunoblotting in the absence of TssE which indicates that TssA binds non-specifically to the nickel beads.

The pull down assay of TssD and TssA was also not straight forward. The problem of non-specific interaction of a protein with resin also occured with VSVg.TssD. We speculated that the two histidine residues at the N-terminus of TssD might function as native hexa-histidine tag in the TssD hexamer. Mutation of these two histidine residues did not solve the problem. Therefore, alternative approach should be sought.

The Co-IP approach also had limitations when applied to study the possible interaction of TssD with TssA. It was observed that VSVg.TssD and linkerHis1.TssA either together or separately could bind to the protein A. Several attempts were carried out to overcome the problem of the non-specific binding. Unfortunately, all attempts were fruitless. Due to time constraints, no more experiments were performed.

Other methods could be employed to study protein-protein interaction such as surface plasmon resonance (SPR) and chemical crosslinking. SPR is a powerful tool to verify and study protein-protein interaction. The technique involves immobilisation of one of the interaction proteins on a sensor chip and the other protein is injected over the surface via a microfluidic system. This allows the monitoring in real-time of the interactions between the immobilized protein and the injected protein. One advantage of The SPR is that it requires only very small amounts of protein (Berggard *et al*., 2007). Proteins which are very close to each other inside a cell could be chemically crosslinked to probe for possible protein-protein interaction. Several crosslinkers are available with different properties. This technique helps to identify transient and/or weak interactions by capturing and stabilizing these interactions (Melcher, 2004).

Chapter 6. Characterisation of TssA

6.1. Introduction

The two-hybrid analysis suggested that TssA is of central importance to the T6SS as it interacts with so many other subunits (TssB, TssD, TssE, TssF, TssH, TssI, TssK and TssL). Evidence provided by Shastri (2011) indicated that it oligomerised into a large complex and the CTD was important for self-interaction. Moreover, no studies regarding the structure and function of TssA have been documented in literature. For these reasons it was decided to subject TssA to further structure and function studies.

## 6.2. Overproduction and purification of His$_6$.TssA

To facilitate the purification of large amounts of TssA, it was decided to clone the *tssA* gene into pET14b under control of the phage T7 promoter (Figure 6.1). T7 RNA polymerase (RNAP) is a highly active enzyme and more efficient in transcription of DNA into RNA than *E.coli* RNAP (Golomb and Chamberlin, 1974). Another advantage of using pET14b is that T7 RNAP is highly selective for its own promoter (Studier and Moffatt, 1986).

The *B. cenocepacia tssA* gene (BCAL0348) was previously amplified and cloned into pET14b using *Nde*I and *Bam*HI by including their sites in the forward and reverse primer, respectively. This placed the *tssA* gene in frame with the vector His-tag coding sequence, resulting in production of a N-terminal hexa-histidine-tagged TssA polypeptide, His$_6$.TssA (S.Shastri, 2011). The hexa-histidine tag allows the purification of target protein using a nickel or cobalt chelate column. However, when the plasmid DNA prepared from *E. coli* MC1061 cells containing pET14b-His$_6$.TssA was analysed by electrophoresis in a 0.8% agarose gel, an additional fragment was observed. Therefore, the gene was recloned.

Overexpression of the gene encoding His$_6$.TssA in *E. coli* BL21(λDE3) was carried out as described in Section 2.5.1.1. The *tssA* gene was successfully overexpressed resulting in production of large amounts of the protein. A solubility test was carried out as described in Section 2.5.2 and sample of the crude extract (total cell protein following sonication) and the supernatant (containing the soluble protein) following centrifugation of the crude extract was analysed by 12% SDS-PAGE. This showed that TssA was present in the soluble fraction (Figure 6.2). To purify TssA protein, the procedure was scaled up using 1 litre culture of BL21(λDE3)/pET14b-His$_6$.TssA.

Immobilized metal ion affinity chromatography (IMAC) was used for purification of His$_6$.TssA from the cleared lysate. Gravity column chromatography with a step elution technique was performed with 2 ml HIS-Buster Nickel Affinity resin (Amocol) in 0.8 x 4 cm polypropylene columns (Bio-rad). 5 ml of the cleared lysate from 1 litre culture of *E. coli* BL21(λDE3) which over-expressed His$_6$.TssA was loaded onto the column. The flow-through was collected and recycled several times (when this protein was purified previously it was found to be poorly bound to the nickel affinity resin (S.Shastri, 2011). After washing the column with 20 volumes of wash buffer, a step elution was carried out using increasing concentrations of imidazole (Section 2.5.3) and 2 ml fractions were collected and analysed by SDS-PAGE (Figure 6.3). This showed that most of the protein applied to the column passed straight through the column and appeared in the flow-through even after recycling it several times and less than 5% of the available soluble protein was purified by IMAC. Therefore, several experiments were carried out to investigate the possible reason (or reasons) for the poor binding of His$_6$.TssA to the nickel column and try improving retention of the protein by nickel sepharose.

Disulfide bond formation between cysteine thiol (sulfhydryl) groups of a protein generally occurs in an oxidative environment. Therefore disulfide bonds are mostly formed in the extracellular compartment of bacteria, although they can be formed in the cytoplasmic compartment under conditions of oxidative stress. However, during protein purification, thiol groups of cysteine residues that do not normally participate in the disulphide bond formation, may form intra- or intermolecular disulphide bonds. The TssA polypeptide contains one cysteine residue at position 104. Therefore, it is possible for intermolecular disulphide bonds to form between two TssA molecules under oxidative condition. To exclude this possibility, a sample of purified His$_6$.TssA was boiled with an equal volume of 2X Laemmli buffer in the presence and absence of β-mercaptoethanol and analysed by electrophoresis in a 12% SDS polyacrylamide gel. No dimerization or multimerisation of His$_6$.TssA molecules was detected in the sample that had not been exposed to β-mercaptoethanol (Figure 6.4). This indicates that the poor binding of TssA to nickel-sepharose is not due to masking of the His tag through disulphide-induced multimerisation of TssA oligomers. Here it was observed that the TssA protein migrated with a size corresponding to 50 kDa (Figure 6.4).

A

BamHI
XhoI
NdeI
NcoI
XbaI

T7 Promoter

AmpR

pET14b
4671 bp

pBR322 origin

rop
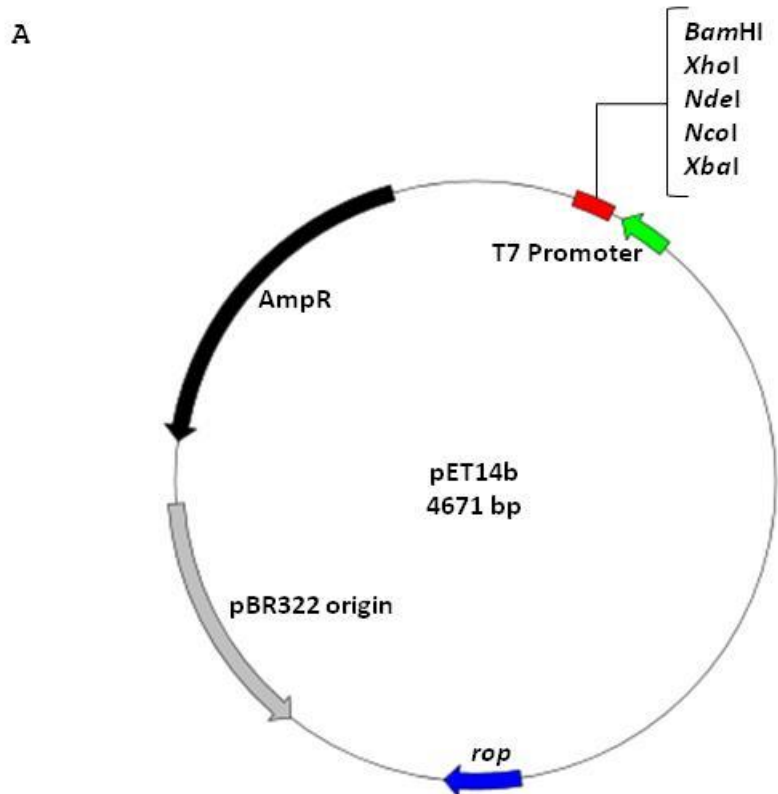
B

pET14b cloning region:

```
                            T7 promoter primer
BglII                          T7 promoter                              XbaI
AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTA
    S.D.        NcoI              hexahistidine tag              TCS          NdeI   XhoI
AGAAGGAGATATACCATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCTCGAG
BamHI                                          T7 terminator primer
GATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCCTTGGGGC
 T7 terminator
CTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATATC
```

Figure 6.1. Overexpression vector pET14b. A. Diagrammatic illustration of the overexpression vector pET14b. *rop*, repressor of primer; pBBR322 origin, origin of pBR322 replication. B. Nucleotide sequence of the multiple cloning site region of pET14b. TCS, thrombin cleavage site is highlighted in green. The T7 promoter and T7 transcription terminator regions are underlined. The T7 promoter and terminator primers are highlighted in yellow. The hexahistidine tag is highlighted in teal. The start codon is highlighted in blue and the stop codon in pink. The restriction sites are shown in red font.
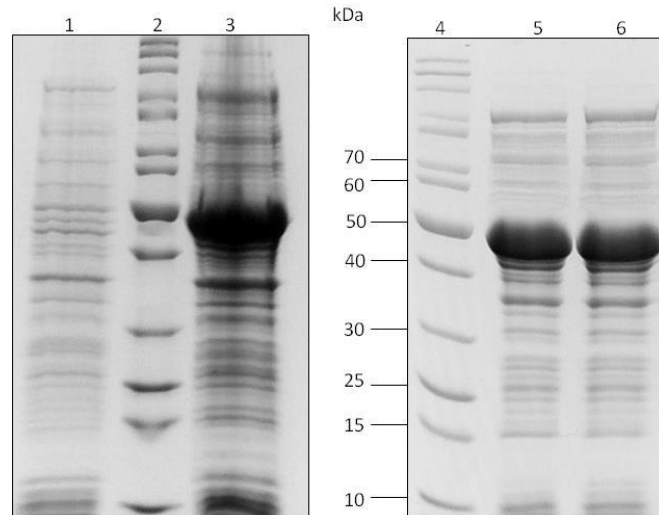
Figure 6.2. Analysis of His$_6$-tagged TssA protein expression and solubility. *E. coli* strain BL21(λDE3) cells containing pET14b-His$_6$.TssA was cultured in BHI medium supplemented with ampicillin to OD$_{600}$ 0.5 at 37°C whereupon a pre-induction sample was taken and 1 mM IPTG was added to the remainder of the culture to induce expression of the *tssA* gene for a further 3 hours. 15 µl of cleared cell lysate was electrophoresed in a 12 % SDS-PAGE. Lane 1, sample of uninduced BL21(λDE3) cells containing pET14b-His$_6$.TssA; lane 2 and 4, EZ-Run Rec unstained protein ladder; lane 3, sample of induced BL21(λDE3) cells containing pET14b-His$_6$.TssA; lane 5, crude extract from induced BL21(λDE3) cells containing pET14b-His$_6$.TssA after cell lysis; lane 6, soluble fraction from induced BL21(λDE3) cells containing pET14b-His$_6$.TssA.
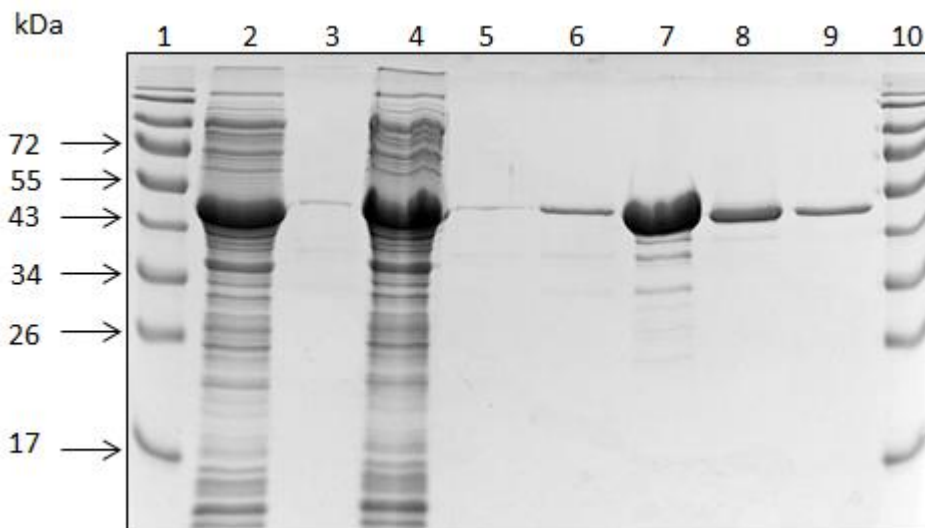
Figure 6.3. Purification of His$_6$.TssA by IMAC. The soluble fraction from lysed induced *E. coli* BL21(λDE3) cells containing pET14b-His$_6$.TssA was prepared and loaded onto a nickel affinity column. Samples were analysed by electrophoresis in a 12% SDS-PAGE. Lanes 1 and 10, Fermentas Prestained Protein Ladder; lane 2, soluble fraction applied to the nickel-sepharose column; lane 3, wash with 25 mM imidazole; lane 4, sample of the recycled flow-through of the His$_6$.TssA load; lane 5, wash with 80 mM imidazole; lanes 6-9, 1 ml fractions eluted with 250 mM imidazole.

250

The possibility that the His tag was masked by TssA oligomerisation was addressed by performing IMAC in the presence of a high concentration of urea which will denature the folded protein and make the His-tag more available. In the first experiment, urea was added to the cleared cell lysate containing His_6.TssA to a final concentration of 5 M followed by IMAC as described in Section 2.5.3. In the second experiment 5 M urea was included in all subsequent IMAC steps (load, wash, and elution).

These experiments showed that inclusion of 5 M urea in the purification of TssA did not result in an increase in the amount of His_6.TssA which bound to the nickel column (Figure 6.5). As 5 M urea may not have resulted in sufficient unfolding of TssA to unmask the His tag, the experiment was repeated in the presence of 8 M urea. However, even at such a high concentration of urea there was no increase in His_6.TssA retention on the nickel column (data not shown).

In order to check for the presence of the His-tag on His_6.TssA, the flow-through was re-applied to a fresh column. This test assesses the ability of the remaining TssA in the flow-through to bind to nickel. For example, it is possible that the hexa-histidine tag is cleaved off from most of the His_6.TssA molecules by cellular protease after expression or during purification. The result showed that the flow-through contains a constant fraction of His_6.TssA molecules that are able to bind nickel-sepharose (result not shown). This suggested that the binding capacity of the column for His_6.TssA was being exceeded.
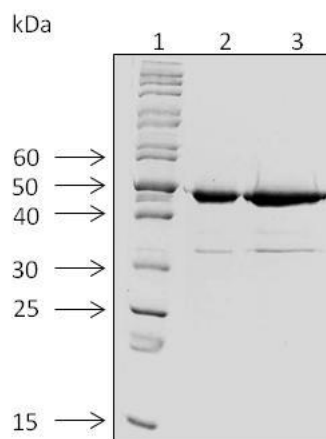
Figure 6.4. Testing His₆.TssA protein for intermolecular disulfide bond formation. Purified His₆.TssA samples were boiled with an equal volume of 2X Laemmli buffer in the presence and absence of β-mercaptoethanol. Samples were electrophoresed in a 12 % SDS-PAGE. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, His₆.TssA sample without β-mercaptoethanol; lane 3, His₆.TssA sample with β-mercaptoethanol.
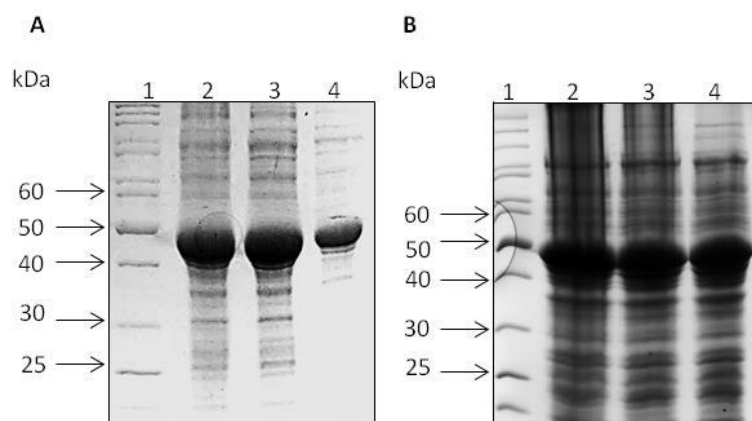
Figure 6.5. Polyacrylamide gel electrophoretic analysis of His₆.TssA purification by IMAC in presence of 5 M urea. Urea was added to the soluble fraction of lysed E. coli BL21(λDE3) cells containing pET14b-His₆.tssA to a final concentration of 5 M and applied to two 1 ml GE Healthcare HisTrap HP columns. The flow-through was collected and the first column was washed with wash buffer without urea. Whereas the second column was washed with wash buffer containing 5 M urea. A) Lane 1, EZ-Run Rec unstained protein ladder; lane 2, sample of the cell extract with 5 M urea loaded on the nickel column; lane 3, flow-through; lane 4, wash with wash buffer without urea. B) Lane 1, EZ-Run Rec unstained protein ladder; lane 2, sample of the cell extract with 5 M urea loaded on the nickel column; lane 3, flow-through; lane 4, wash with wash buffer containing 5 M urea.

6.3. Overproduction and purification of His_{10}.TssA

Many researchers have used Deca-His-tags on their target proteins to enhance binding to nickel-sepharose. For example, Lee and Kim demonstrated a clear correlation between the His-tag length and binding affinity by comparing variable His-tag length in a bacterial T7 expression system (Lee and Kim, 2009). Also, Mohanty and Wiener showed that the tag length has a great effect upon the protein yield (Mohanty and Wiener, 2004). Therefore, in an attempt to improve the yield of TssA, the N-terminal histidine tag was extended from His_6 to His_{10}. A complementary pair of oligonucleotides (DecaHisFor) was designed containing an *Nco*I cohesive end at the 5' end followed by sequences encoding 10 histidines and a thrombin cleavage site, and ending with an *Nde*I cohesive end (see appendix 1 for the primer sequences). The pair of oligonucleotides were annealed and ligated with *Nco*I/*Nde*I-digested pET14b-His_6.TssA. Desired recombinants were identified by PCR screening and nucleotide sequencing.

Overexpression and solubility testing of the gene encoding His_{10}.TssA in *E. coli* BL21(λDE3) was carried out as described in Section 2.5.2. His_{10}.TssA was produced in high abundance and found to be soluble (Figure 6.6). After preparing the soluble fraction from 1 litre culture of BL21(λDE3)/pET14b-His_{10}.TssA (Section 2.5.2), 5 ml of the soluble fraction was used in purification of His_{10}.TssA using IMAC as described in Section 2.5.3. The results showed that the percentage of His_{10}.TssA bound to the nickel column was similar to that for His_6.TssA (i.e. less than 5%), (Figure 6.7). However, His_{10}.TssA was eluted from the column at approximately 300 mM imidazole, whereas His_6.TssA was previously eluted at about 90 mM imidazole. Therefore, the longer tag increased the affinity of TssA for nickel but not the percentage of molecules bound. This is consistent with the idea that the capacity of the nickel-sepharose column for TssA is relatively low.
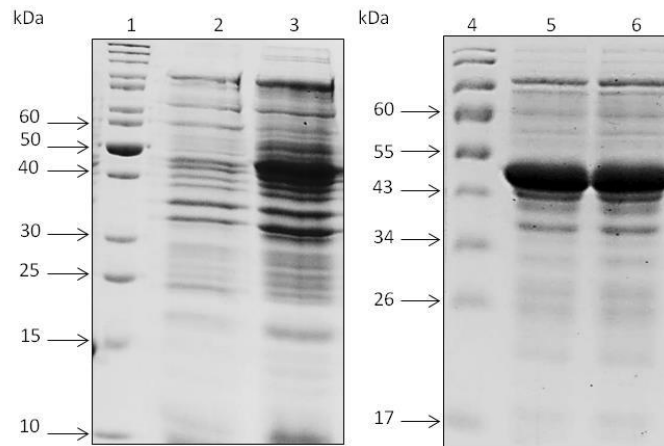
Figure 6.6. Analysis of His$_{10}$.TssA protein expression and solubility. *E. coli* BL21(λDE3) containing pET14b-His$_{10}$.*tssA* was cultured at 37°C in BHI supplemented with ampicillin to OD$_{600}$ 0.5, whereupon a pre-induction sample was taken and 1 mM IPTG was added to the remainder of the culture to induce expression of *tssA* for a further 3 hours. A sample of cleared cell lysate was electrophoresed in a 15 % SDS-PAGE. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, sample of uninduced BL21(λDE3) cells containing pET14b-His$_{10}$.*tssA*; lane 3, sample of induced BL21(λDE3) cells containing pET14b-His$_{10}$.*tssA*; lane 4, EZ-Run Rec pre-stained protein ladder; lane 5, IPTG-induced His$_{10}$.TssA in total cell extract, after cell lysis; lane 6, soluble fraction of cell extract.
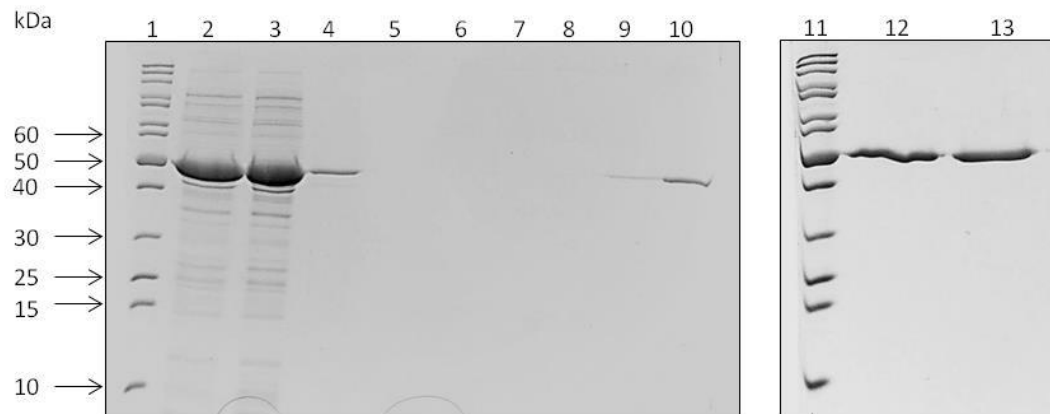
Figure 6.7. Purification of His10.TssA by IMAC. The soluble fraction from lysed *E. coli* BL21(λDE3) containing pET14b-His$_{10}$.*tssA* was applied to a 1 ml GE Healthcare HisTrap HP column in lysis buffer. His$_{10}$.TssA was eluted using a linear gradient of 10-500 mM imidazole. Lanes 1 and 11, EZ-Run Rec unstained protein ladder; lane 2, soluble fraction from cell extract; lane 3, flow-through from the nickel affinity column; lane 4, eluate following 10 mM imidazole wash of the nickel affinity column; lanes 5-10 and 12-13, gradient elution of His$_{10}$.TssA from nickel affinity column (fractions 1-8).

6.4. Overproduction and purification of TssA.His$_{10}$

Extension of the His-tag from six residues to ten did not increase the proportion of soluble TssA that bound to the nickel resin. As the location of the His tag has been shown to affect it's availability for binding to metal ions (Block *et al*., 2009) it was decided to change the location of the His$_{10}$-tag to the C-terminus of TssA. To do this, *tssA* was amplified from pET14b-His$_6$.TssA using primers tssAnonHisFor and tssACtermHisTag (see appendix 1 for the primer sequences).

The forward primer contained an *Xba*I site, Shine-Dalgarno sequence and the first 20 base pairs of *tssA* coding sequence, whereas the reverse primer contained a *Bam*HI site, stop codon, 10 histidine codons and the final 20 base pairs of *tssA* coding sequence (a thrombin cleavage site was not introduced into the construct). The PCR product was then ligated between the corresponding sites of pET14b giving rise to pET14b.TssA.His$_{10}$.

Expression and solubility test of TssA.His$_{10}$ showed that this variant of TssA was produced in high abundance and was soluble (Figure 6.8). The purification of TssA.His$_{10}$ was carried out using IMAC. The results showed that the percentage of TssA.His$_{10}$ bound to the nickel column was similar to that observed for His$_6$.TssA (less than 5%), (Figure 6.9). However, TssA.His$_{10}$ was eluted from the column at approximately 160 mM imidazole, whereas His$_6$.TssA was previously eluted at about 90 mM. Therefore, the longer tag at the C-terminal also increased the affinity of TssA for nickel but not the percentage of molecules bound.
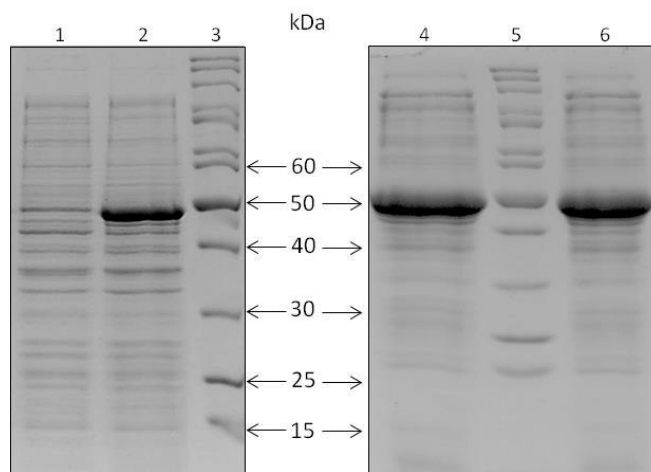
Figure 6.8. Analysis of TssA.His$_{10}$ protein expression and solubility. *E. coli* BL21(λDE3) containing pET14b.TssA.His$_{10}$ was cultured at 37°C in BHI medium supplemented with ampicillin to OD$_{600}$ 0.5, whereupon a pre-induction sample was taken and 1 mM IPTG was added to the remainder of the culture to induce expression of *tssA* for a further 3 hours. A sample of cleared cell lysate was electrophoresed in a 15 % SDS-PAGE. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, sample of uninduced BL21(λDE3) cells containing pET14b.TssA.His$_{10}$; lane 3, sample of induced BL21(λDE3) cells containing pET14b.TssA.His$_{10}$; lane 4, EZ-Run Rec pre-stained protein ladder; lane 5, IPTG-induced TssA-His$_{10}$ in total cell extract, after cell lysis; lane 6, soluble fraction of cell extract.
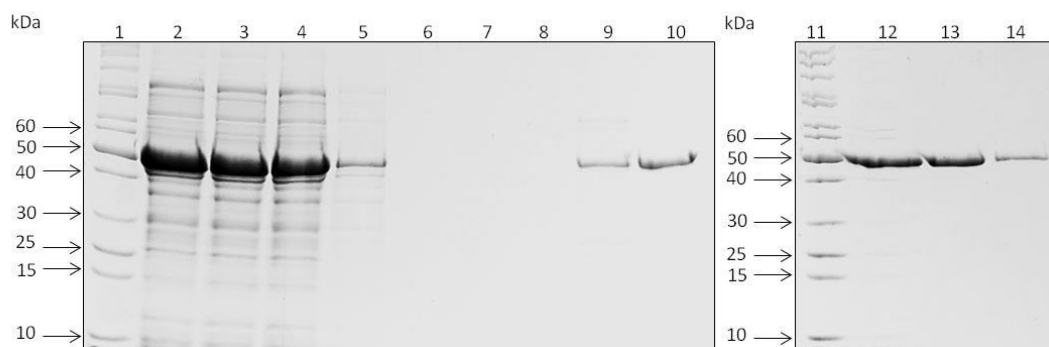
Figure 6.9. Purification of TssA.His10 by IMAC. The soluble fraction from lysed *E. coli* BL21(λDE3) containing pET14b.TssA.His$_{10}$ was applied to a 1 ml GE Healthcare HisTrap HP column in the lysis buffer. TssA.His$_{10}$ was eluted using a linear gradient of 10-500 mM imidazole. Lanes 1 and 11, EZ-Run Rec unstained protein ladder; lane 2, soluble fraction from cell extract; lane 3, flow-through from the nickel affinity column; lane 4, eluate following 10 mM imidazole wash of the nickel affinity column; lanes 5-10 and 12-14, gradient elution of TssA.His$_{10}$ from nickel affinity column (fractions 1-9).

6.5. Overproduction and purification of linker His-tagged TssA derivatives

Based on the two domain model of TssA in which the NTD is separated from the CTD by a long interdomain linker (Figure 3.16), and the observation that the predicted interdomain linker is susceptible to proteolysis, see below (Section 6.6.4), which suggested that this region is exposed, it was decided to locate the His tag in the interdomain linker. As the TssA linker region varies in length and sequence in different bacterial species, it may tolerate the insertion of additional amino acids without loss of function. To increase the chances of placing the His tag in an accessible location, two forms of linker His-tagged TssA were constructed where the tag was inserted at a different location in the linker. These two sites were chosen based on observed proteolytic cleavage sites in the interdomain linker (Section 6.6.4). To introduce the six histidine codons, SOE-PCR (Splicing by Overlap Extension PCR) was used. The PCR involved the amplification of two segments of *tssA* coding sequence either side of the insertion site in the linker region. pET14b.TssA encoding untagged TssA (Section 6.8), was used as a template for both fragments. To place the His tag near the N-terminus of the linker, the first fragment (encoding the N-terminal domain) was amplified using pET14biotAfor and TssAlinkerHis1rev primers, and the second fragment (encoding the C-terminal domain) using pET14biotArev and TssAlinkerHis1for primers. The resulting PCR products were electrophoresed in a 0.8% agarose gel and extracted. In order to generate the full length *tssA* gene containing the linker His-tag, the two first round PCR products were mixed together and used as template for a second PCR step with primers pET14biotAfor and pET14biotArev. The product of the second PCR step was cut with *Nde*I and *Bam*HI and ligated into the *Nde*I and *Bgl*II sites of pACYCDuet-1 to give rise to pACYCDuet-linkerHis1.TssA. pACYCDuet-linkerHis2.TssA was generated by the same process but with TssAlinkerHis2for substituted for TssAlinkerHis1for and TssAlinkerHis2rev substituted for TssAlinkerHis1rev. pACYCDuet-1 is a p15A-derived T7 expression vector (Section 5.2.1).

BL21(λDE3) cells were transformed with both pACYCDuet.TssA derivatives in order to check the expression and solubility of linkerHis1.TssA and linkerHis2.TssA (Section 2.5.2). Both versions of linker His-tagged TssA were expressed very well and were present in the soluble fraction (Figure 6.10). Purification of linkerHis1.TssA and linkerHis2.TssA was carried out using nickel affinity chromatography as described in section 2.5.3. It was observed that a large fraction of linkerHis1.TssA and

linkerHis2.TssA did not bind to the nickel resin and passed straight through the column (Figure 6.11) although it appeared that a larger fraction of the protein was retained on the column than is the case for the other His-tagged TssA derivatives.

To assess the possibility of the TssA which bound to the resin being a small sub-population of deformed proteins which were in some way folded incorrectly, a batch of washed nickel beads slurry in microcentrifuge tubes was used to assess the binding capacity of unbound protein. Cell lysate containing linkerHis1.TssA was applied to fresh nickel-sepharose and incubated for 30 minutes. Then the beads were spun down at low speed and the supernatant was transferred into a second microcentrifuge tube containing fresh washed nickel beads and incubated for another 30 minutes. This process was repeated 5 times. The bound protein in each tube was washed and eluted with 255 and 500 mM imidazole and all samples were analysed on 12% SDS-PA gels. The result showed that a small proportion of unbound TssA from each flow-through is able to bind to fresh nickel beads (Figure 6.12), consistent with the hypothesis that a relatively small amount of the TssA protein saturates the nickel beads (Section 6.2).

To confirm that the introduced mutation in the non-conserved linker region does not interfere with the previously observed TssA self-interaction a BACTH assay was conducted (Section 3.5.1). To do this, linkerHis1.TssA was cloned into pKT25 and pUT18C. Using pACYCDuet-TssAlinkerHis1 as a template, two separate PCR reactions were performed; the first with the primers CtermiotAfor and CtermiotArev to introduce an upstream *Pst*I restriction site and a downstream *Xba*I site for ligation into pKT25. The second PCR contained the primers pUT18CiotAfor and CtermiotArev to introduce the same restriction sites for ligation into pUT18C. The BACTH assay of T25-linkerHis1.TssA with T18-linkerHis1.TssA gave rise to a strong Mal$^+$ phenotype similar to the Zip control (Figure 6.13). This indicates that insertion of six histidines into the interdomain linker is unlikely to adversely affect the self-association of TssA.
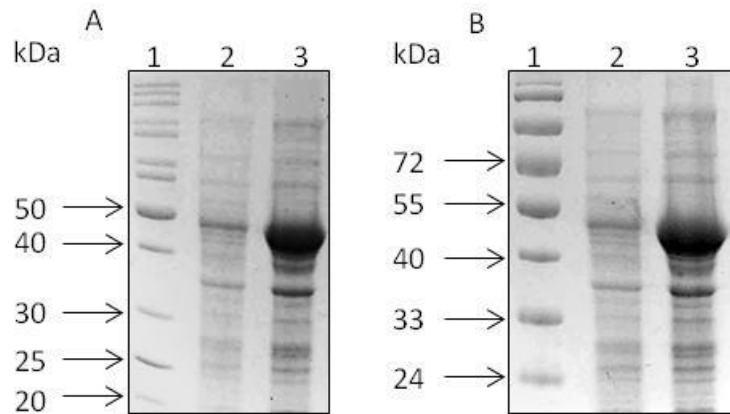
Figure 6.10. Analysis of linkerHis1.TssA and linkerHis2.TssA protein expression.
Coomassie blue stained 12% SDS polyacrylamide gel showing the overproduction of linkerHis1.TssA and linkerHis2.TssA. (A) Linker His1.TssA. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, sample taken prior to induction; lane 3, sample taken after 3 hours of induction with 1 mM IPTG. (B) Linker His2.TssA. Lane 1, EZ-Run Rec pre-stained protein ladder; lane 2, sample taken prior to induction; lane 3, sample taken after 3 hours of induction with 1 mM IPTG.
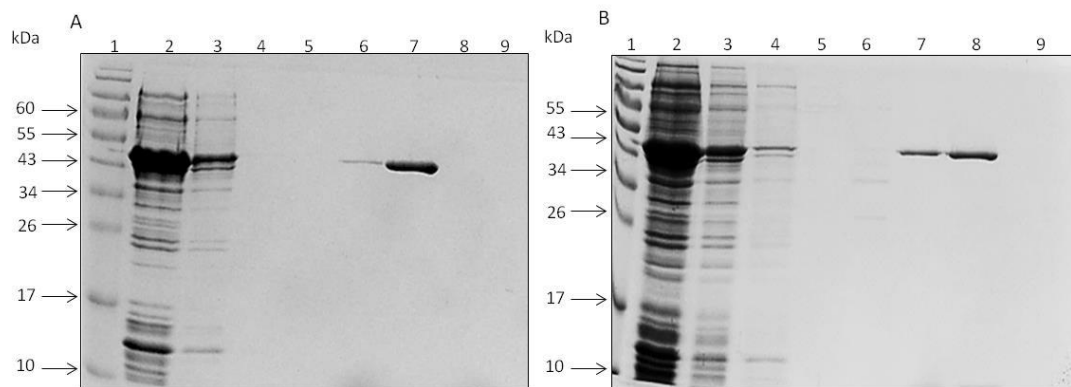
Figure 6.11. Purification of linker-His-tagged TssA by IMAC. The soluble fraction from lysed *E. coli* BL21(λDE3) cells containing pACYCDuet-linkerHis1.TssA or pACYCDuet-linkerHis2.TssA was applied separately to a 1 ml GE Healthcare HisTrap HP column in lysis buffer. The linkerHis1.TssA or linkerHis2.TssA was eluted using a linear gradient of 10-500 mM imidazole. Samples were analysed on a 12% SDS-PAGE. (A) LinkerHis1.TssA. Lane 1, EZ-Run Rec pre-stained protein ladder; lane 2, soluble fraction from cell lysate; lane 3, flow-through from the column; lane 4, eluate following 10 mM imidazole wash of the nickel affinity column; lanes 5-9, 5 ml fractions from the gradient elution with 10-500 mM imidazole (fractions 1-5). (B) LinkerHis2.TssA. Lane 1, EZ-Run Rec pre-stained protein ladder; lane 2, soluble fraction from cell lysate; lane 3, flow-through from the column; lane 4, eluate following 10 mM imidazole wash of the nickel affinity column; lanes 5-9, 5 ml fractions from the gradient elution with 10-500 mM imidazole (fractions 1-5).
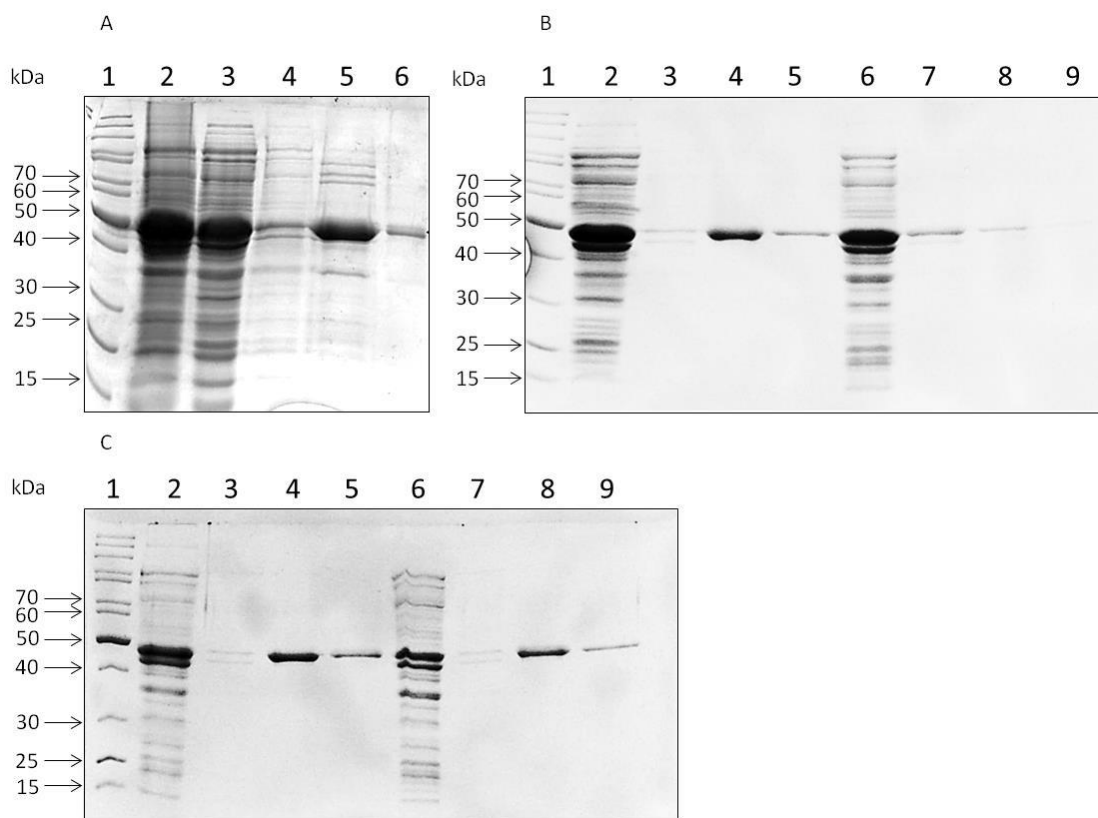
Figure 6.12. Serial batch IMAC for linkerHis1.TssA shows nickel sepharose has a low capacity for the TssA protein. LinkerHis1.TssA was applied sequentially to five batches of nickel-sepharose beads and each time incubated for 30 minutes. Then the beads were spun down and the supernatant was transferred into a clean microcentrifuge tube containing fresh washed nickel beads and incubated for another 30 minutes. The bound protein in each tube was washed and eluted with 255 and 500 mM imidazole and samples of eluted protein were analysed on 12% SDS-PAGE. (A) Batch 1. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, cell lysate load; lane 3, flow-through; lane 4, wash; lane 5, 255 mM imidazole elution; lane 6, 500 mM imidazole elution. (B) Batches 2 and 3. Lane 1, EZ-Run Rec unstained protein ladder; lanes 2 and 6, unbound protein from batches 1 and 2, respectively; lanes 3 and 7, flow-through from batches 2 and 3, respectively; lanes 4 and 8, 255 mM imidazole elution from batches 2 and 3, respectively; lanes 5 and 9, 500 mM imidazole elution from batches 2 and 3, respectively. (C) Batches 4 and 5. Lane 1, EZ-Run Rec unstained protein ladder; lanes 2 and 6, unbound protein from batches 3 and 4, respectively; lanes 3 and 7, flow-through from batches 4 and 5, respectively; lanes 4 and 8, 255 mM imidazole elution from batches 4 and 5, respectively; lanes 5 and 9, 500 mM imidazole elution from batches 4 and 5, respectively.
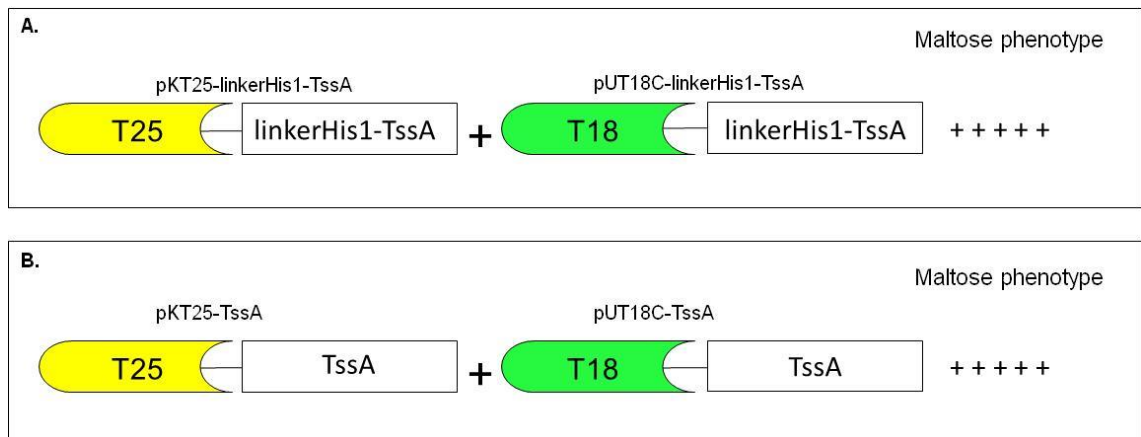
Figure 6.13. Analysis of linkerHis1.TssA or TssA self-interaction using the BACTH assay.
Pairwise combinations of compatible BACTH plasmids encoding fusions of linkerHis1.TssA or
TssA to the T18 and T25 components of CyaA were introduced into *E. coli* strain BTH101.
Transformants were scored for their maltose phenotype on MacConkey-maltose agar after 72
and 120 h incubation at 30°C. Combinations that gave rise to a maltose-positive phenotype are
shown. The strength of the maltose phenotype shown was scored after 72 h incubation; A.
Combination of linkerHis1.TssA fusion proteins that gave rise to a maltose-positive phenotype.
B. Combination of TssA fusion proteins that yielded a maltose-positive phenotype.

6.6. Analysis of His$_6$.TssA

6.6.1. Purification of His$_6$.TssA

In order to increase the purity of His$_6$.TssA, IMAC was used as a first step (Figure 6.14), followed by anion exchange chromatography (Section 2.5.4). To do this, the partially purified TssA following IMAC was dialysed against two changes of 20 mM Tris-HCl (pH 8.0) (pI of TssA is 4.95). The dialysed protein was cleared by centrifugation at 13,200 rpm for 30 minutes and the supernatant was passed through a 0.45 μm syringe filter. The dialysed and cleared protein was applied to a Q sepharose column (GE Health care), and a gradient elution with 0-1.0 M NaCl was performed. Fractions collected were then analysed by electrophoresis in a 12% SDS polyacrylamide gel (Figure 6.15). Under the conditions of this experiment, His$_6$.TssA protein did not bind to Q sepharose as most of the protein was present in the flow-through. Moreover, more than 50% of the protein precipitated during the dialysis. To increase the solubility of His$_6$.TssA present in the dialysis buffer, the experiment was repeated but with 50 mM NaCl included in the buffer. However, a similar result was obtained (data not shown). The salt concentration in the TssA sample was then diluted to 25 mM and the sample re-applied to the Q sepharose column, but again the result showed that most of the protein was present in the flow-through (data not shown). The experiment was repeated once again using dialysis buffer at pH 9.0 and the elution buffer also at pH 9.0, but the protein still passed straight through the column.

As purification of His$_6$.TssA by IMAC followed by anion exchange chromatography was unsuccessful, and as it is known to form a large complex, it was decided to attempt to purify His$_6$.TssA by IMAC followed by size exclusion chromatography (SEC). Partially purified His$_6$.TssA protein (5 ml at 0.7 mg ml^{-1}) was prepared by nickel affinity chromatography and then concentrated 25 fold using a 30 kDa Amicon *Ultra*-4 centrifugal filter unit (i.e. final volume approximately 200 μl). Size exclusion chromatography was performed on a Superose 6 column (GE Healthcare) (Section 2.5.6) following injection of 100 μl of the concentrated protein using a 100 μl loop (Figure 6.16A). Figure 6.16B shows the applied and eluted samples from the Superose 6 column. The elution time of His$_6$.TssA was 14.52 minutes. The experiment was repeated with known molecular weight standards to allow a calibration curve to be generated as shown in (Figure 6.17) this allowed an estimation of the His$_6$.TssA complex as 1.9 MDa.
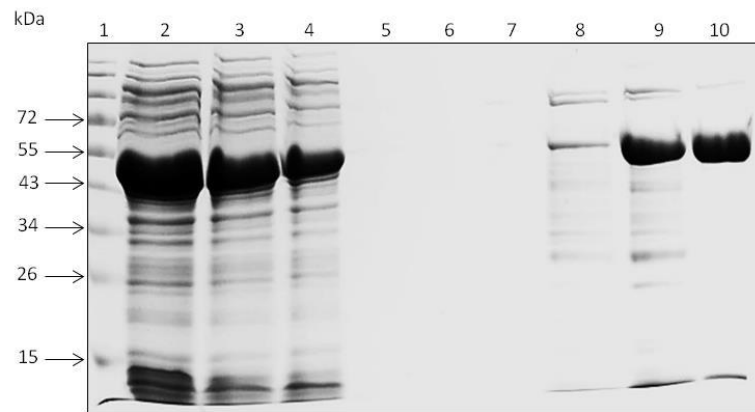
Figure 6.14. Purification of His₆.TssA by IMAC. The soluble fraction from lysed *E. coli* BL21(λDE3) cells containing pET14b-His$_6$.TssA was applied to a 1 ml GE Healthcare HisTrap HP column in lysis buffer. His$_6$.TssA was eluted using a linear gradient of 10-500 mM imidazole. Samples were analysed on a 12% SDS-PAGE. Lane 1, EZ-Run Rec pre-stained protein ladder; lane 2, soluble fraction from cell lysate; lane 3, flow-through from the column; lane 4, eluate following 10 mM imidazole wash of the nickel affinity column; lanes 5-10, 5 ml fractions from the gradient elution with 10-500 mM imidazole.
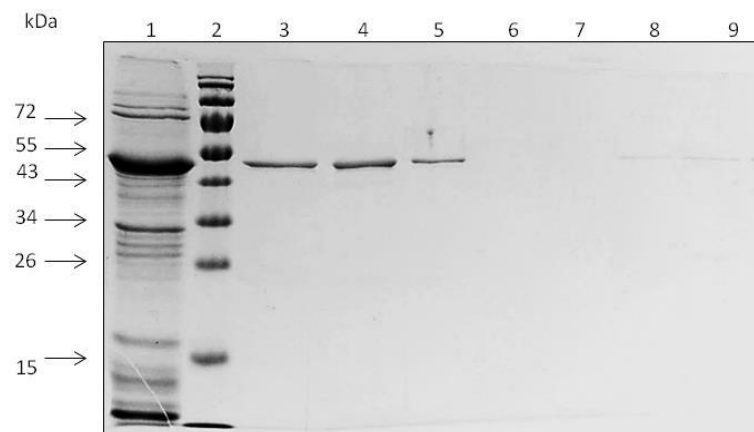
Figure 6.15. Purification of His$_6$.TssA by anion exchange chromatography. Pooled fractions 9 and 10 from the IMAC column (Figure 6.14) were dialysed overnight against 20 mM Tris-HCl (pH 8.0) buffer. Dialysed partially purified His$_6$.TssA protein was loaded on a 1 ml GE Healthcare HiTrap Q HP column and gradient elution was carried out from 0-1000 mM NaCl in 20 mM Tris-HCl buffer (pH 8.0). Lane 1, insoluble His$_6$.TssA following dialysis; lane 2, EZ-Run Rec pre-stained protein markers; lane 3, sample of protein loaded on the Q column; lane 4, flow-through from the column; lane 5, wash; lanes 6-9, 5 ml fractions from the gradient elution with 0-1000 mM NaCl.
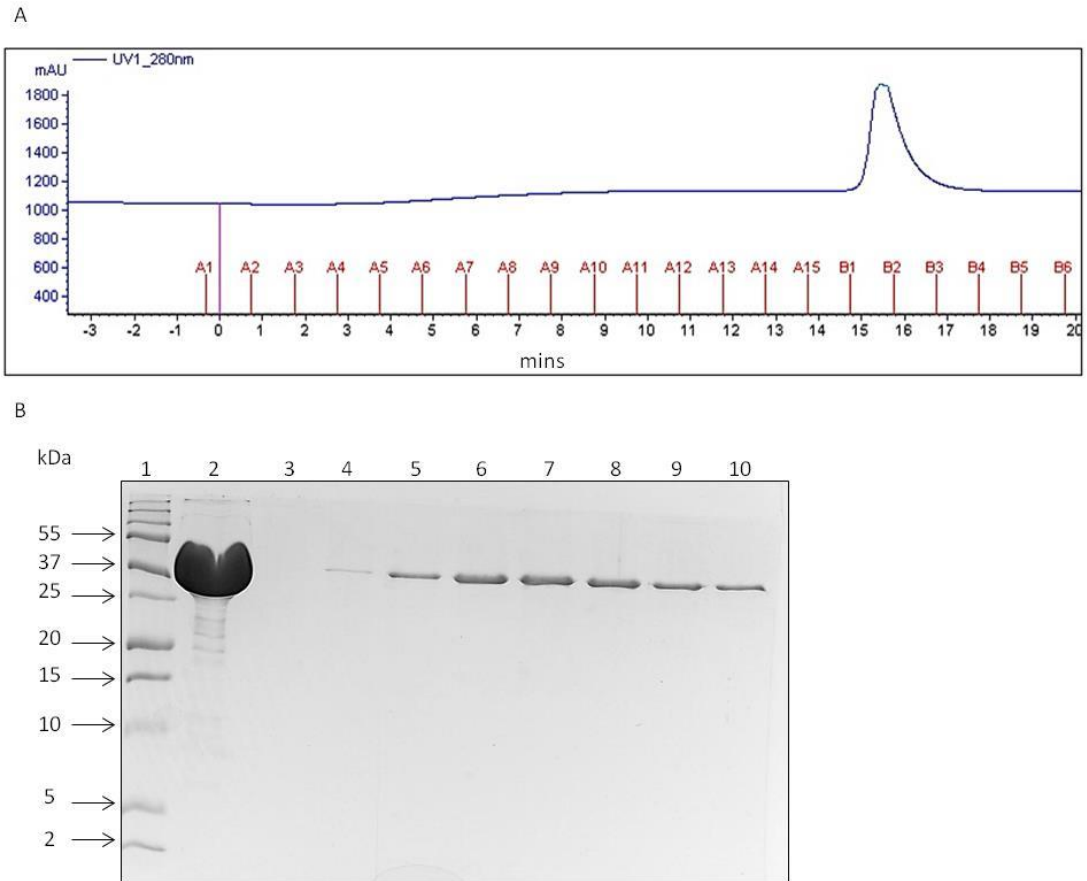
Figure 6.16. Purification of His$_6$.TssA by size exclusion chromatography. His$_6$.TssA protein eluted from a HisTrap column was concentrated into 0.2 ml of 50 mM Tris-HCl buffer, 200 mM NaCl, 10% glycerol (pH 8.0) and 0.1 ml was applied to a Superose 6 gel filtration column (GE Healthcare). A. Elution profile of His$_6$.TssA from gel filtration column monitored at 280nm. B. SDS PAGE analysis of His$_6$.TssA eluted from gel filtration column. Lane 1, EZ-Run Rec prestained protein ladder; lane 2, His$_6$.TssA protein applied to gel filtration column; lanes 3-10, fractions A13-A15 and B1-B6 in trace (A).
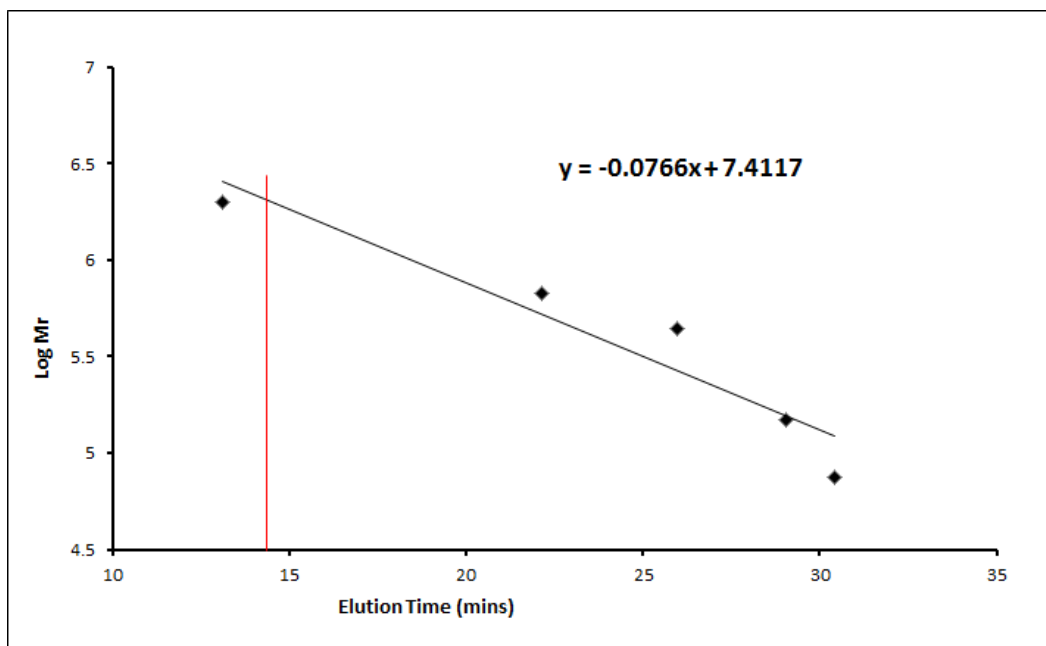
Figure 6.17. Estimation of the molecular weight of His$_6$.TssA by SEC. Calibration curve showing the elution times of a set of known protein standards (Section 2.5.6) plotted against the log$_{10}$ of their molecular weight. A line of best fit between the points was generated using Microsoft Excel and the equation of the line is shown. The peak elution time of His$_6$.TssA (14.52 minutes) is shown by the vertical line.

6.6.2. Analysis of the oligomerisation status of His$_6$.TssA by sedimentation velocity analytical ultra-centrifugation ($_{SV}$AUC)

The estimated size of His$_6$.TssA by SEC is over one million Daltons, which suggests that His$_6$.TssA is forming a very large complex. This is in agreement with the result of the BACTH assay which indicated that TssA self-associates, and preliminary $_{SV}$AUC data suggesting that TssA forms a large complex (S.Shastri, 2011). In order to confirm that TssA oligomerises and the extent of oligomerisation, a more highly purified sample of His$_6$.TssA protein was subjected to $_{SV}$AUC at the Astbury Centre for Structural Molecular Biology at the University of Leeds. Analytical ultracentrifugation (AUC) is a well-known physical biochemistry technique used to study protein interactions and protein self-association (Lebowitz *et al*., 2002). In this technique the sedimentation of macromolecules in a centrifugal field is monitored to determine their hydrodynamic and thermodynamic characteristics in solution (Schuck, 2003). There are two types of AUC which can both be used for the characterisation of protein; sedimentation velocity (SV) and sedimentation equilibrium (SE). SV is used for the determination of the oligomeric state and the stoichiometry of interactions by assigning sedimentation coefficients which allows for modelling of the hydrodynamic shape of proteins and protein complexes. SE is used for determination of the protein molar mass and the study of protein self-association and heterogeneous interactions (Lebowitz *et al*., 2002).

Results were analysed by whole-boundary profile analysis using the program Sedfit version 11.9 (July 2010) (Schuck, 2000). The partial specific volume of the sample and buffer density and viscosity were calculated using Sednterp version 1.09 (March 2006) (Laue, 1992). Radial absorbance scans at 280 nm were recorded for each cell over a period of 10.5 hours. The scans are shown in (Figure 6.18) for each cell, with the blue trace showing the earliest scan and red the latest. Using the raw data files the samples were subjected to continuous c(s) distribution analysis. The signal at 20 S (Figure 6.19) represents higher molecular weight material. Therefore, the calculated molecular weight (mol. wt) and composition of 280 nm signal for the two concentrations of TssA showed similar sedimentation properties, the majority of the sample being a higher order multimer of ~ 850 kDa. Considering the molecular weight of a single TssA molecule, which is ~ 43 kDa including the His-tag, this large multimer of TssA protein corresponds to a complex containing approximately 20 monomers of TssA.

Another sample of purified His$_6$.TssA protein was subjected to $_{SE}$AUC analysis. The $_{SV}$AUC experiment was carried out first to assess the possibility of $_{SE}$AUC analysis with this protein prep. The sample was run at 3 different concentrations (17.9, 4.5, 1.9 µM). The concentration range chosen showed a broad distribution of molecules in solution (Figure 6.20). This experiment was carried out at 4˚C so the measured sedimentation coefficients vary significantly to the previous value collected for His$_6$.TssA at 20˚C. To compare data from different experiments, the data from $_{SE}$AUC experiment was extrapolated to represent values at 20˚C in water (S$_{20,W}$). The results indicated that at 1.9 µM, TssA existed predominantly as a monomer (S$_{20,w}$ 3.1 ±0.2) and at 17.9 µM it gave rise to a major oligomeric species (S$_{20,w}$ 18.9 ±0.3). However, it was not possible to estimate the size of the oligomeric species.

It was decided not to proceed with the $_{SE}$AUC experiment, because even at the highest concentration of TssA available, less than 50% of the protein formed the higher oligomeric species. One reservation is that the higher concentration created at the bottom of the cell might push more protein into the higher oligomeric species, creating problems for accurate data analysis (Amy Barker, personal communication).
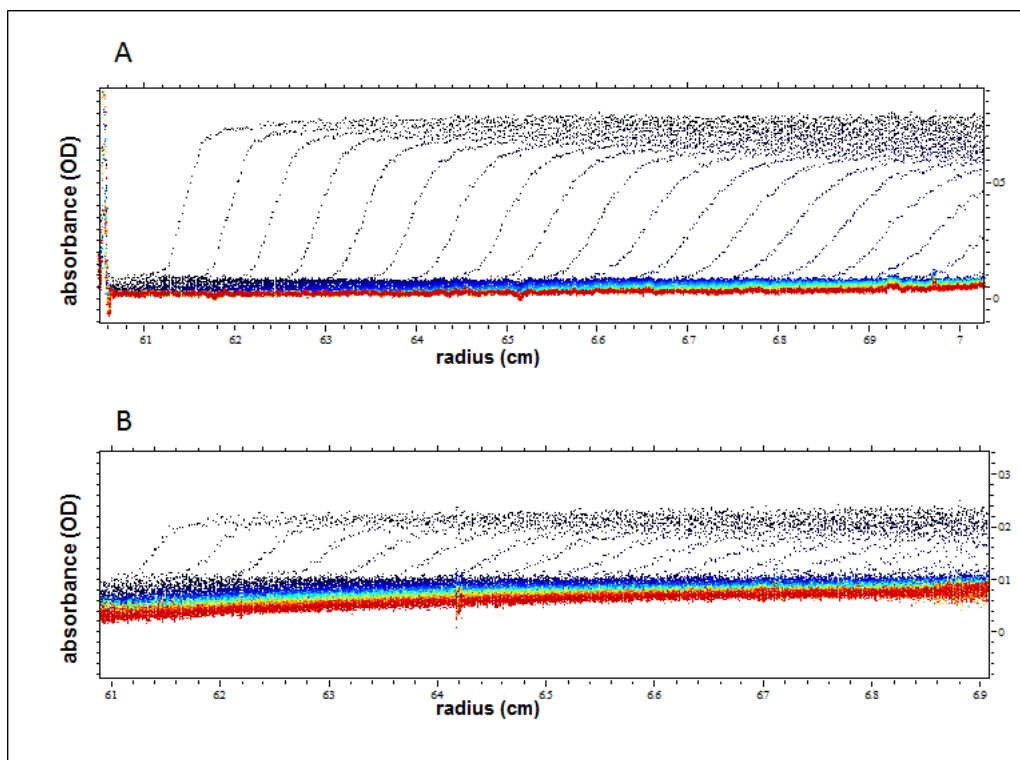
Figure 6.18. Analytical ultra-centrifugation radial absorbance scans of His$_6$.TssA at 280 nm. For each cell the blue trace shows the earliest scan and red the latest. (A) Raw data scans for cell 1 and (B) raw data scans for cell 2.
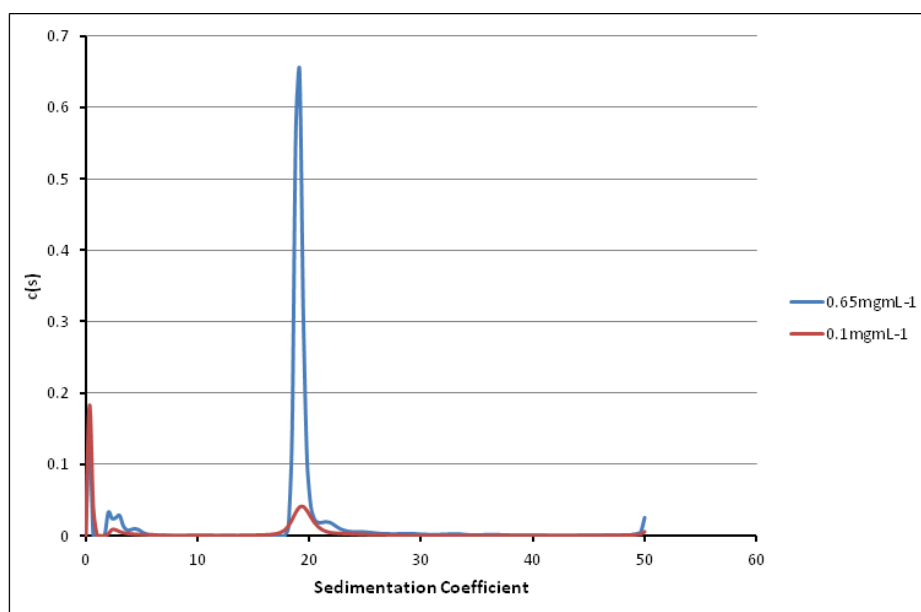
Figure 6.19. Sedimentation velocity analytical ultra-centrifugation ($_{SV}$AUC) of His$_6$.TssA. Sedimentation coefficient c(s) distributions up to 50 S. The signal at 20 S represents higher molecular weight material.
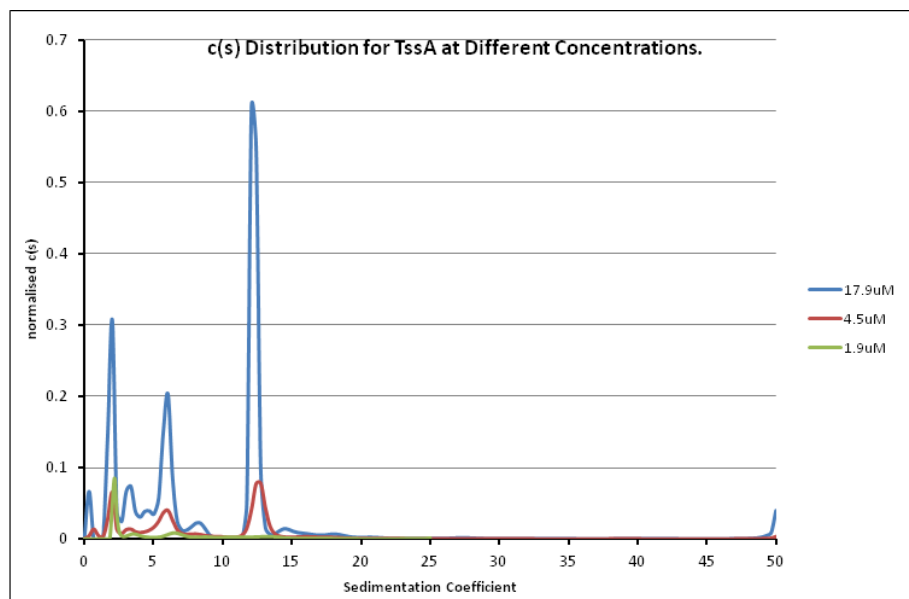
Figure 6.20. Sedimentation velocity analytical ultra-centrifugation ($_{sv}$AUC) of different concentrations of His$_6$.TssA. Sedimentation Coefficient c(s) distributions up to 50 S for TssA at different concentration (17.9, 4.5 and 1.9 μM).

6.6.3. Transmission electron microscopy of (TEM) of His$_6$.TssA

The data obtained from SEC and $_{sv}$AUC suggest that His$_6$.TssA is forming a large complex, although the approximate sizes obtained from these two techniques differ markedly. As the His$_6$.TssA oligomers may be large enough to be visualised by electron microscopy after negative staining, purified His$_6$.TssA was subjected to EM analysis in buffer containing 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, 10% glycerol and 80 mM imidazole (Department of Molecular Biology and Biotechnology, University of Sheffield). The results of negative stain EM showed that His$_6$.TssA formed rings with a variable number of outer projections (Figure 6.21). The rings showed a variation in size and some were distorted or not well formed. The internal diameter of the rings ranged from 20 to 25 nm and the external diameter (including the projections) varied from 40 to 45 nm. In order to find the optimal conditions for visualising the His$_6$.TssA rings, different buffer conditions were tested (Figure 6.21 A-D). This showed that buffers A and C were optimum for ring formation.

Also, the His$_6$.TssA protein which was purified in the presence of 8 M urea and renatured by sequential dialysis against buffers containing 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, 10% glycerol and 6-4-2-0 M urea (Section 6.2) was analysed by negative stain EM. However, this sample showed some poorly formed rings in addition to protein aggregates (result not shown).
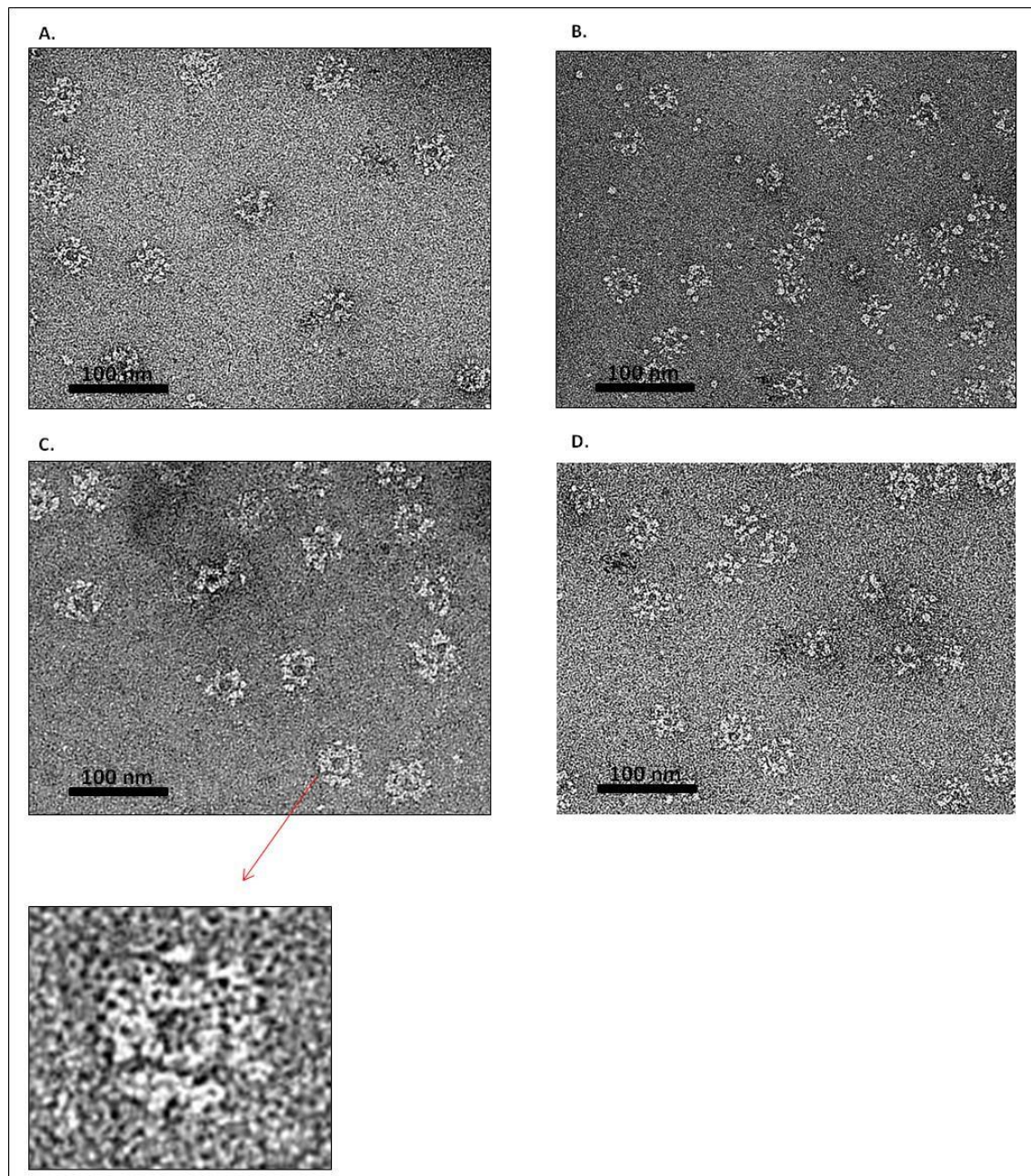
Figure 6.21. Negative stain electron micrograph of His$_6$.TssA in different buffers. The EM images of His$_6$.TssA under all four conditions show large ring-like structures with a central pore and projections around the outside. Negative staining was carried out using 0.75% uranyl formate. His$_6$.TssA was analysed in buffer containing A. 25 mM Tris-HCl (pH 7.4), 150 mM NaCl and 20% glycerol; B. 25 mM Tris-HCl pH (6.8), 25 mM NaCl; C. 25 mM Tris-HCl (pH 6.8), 500 mM NaCl; D. 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, 10% glycerol and 80 mM imidazole.

6.6.4. Analysis of His$_6$.TssA proteolytic cleavage products

During one period, purification of His$_6$.TssA was hampered by the presence of a contaminating protease that resulted in degradation of TssA following IMAC. After applying the dialysed protein to a Q sepharose column, the eluted protein was observed to be degraded into 2 major species of approximately 35 and 40 kDa (Figure 6.22). This could occur if hydrolysis of the TssA polypeptide chain occurred at different positions within the interdomain linker. To verify this, the degradation products were analysed by N-terminal sequencing and mass spectrometry.

6.7.4.1. N-terminal sequence analysis of His$_6$.TssA proteolytic cleavage products

A sample of each of the two fragments was subjected to N-terminal sequence analysis. To do this, the degraded His$_6$.TssA protein eluted from the Q sepharose column (at a concentration 0.4 mg ml^{-1}) in 300 mM NaCl was loaded into 3 lanes of a 12% SDS-PA gel. Following electrophoresis the gel was blotted onto a ProBlot^{TM} membrane and N-terminal sequence analysis was carried out in the Department of Molecular Biology and Biotechnology (MBB), University of Sheffield.

The N-terminal sequence analysis results confirmed that the observed two major species are degradation products of TssA, as both have the N-terminal sequence G-S-S-H-H-H-H-H-H. As expected, the N-terminal formyl methionine residue had been removed, as the side-chain of the second amino acid (glycine) is very small (Dalboge *et al*., 1990, Hirel *et al*., 1989, Solbiati *et al*., 1999). The size of these N-terminal fragments supported the hypothesis that proteolytic cleavage had occurred within or close to the interdomain linker. As cleavage here would release a small C-terminal fragment of the protein corresponding to the CTD predicted from multiple sequence alignment and secondary structure prediction (Section 1.5.4.2). The purification of His$_6$.TssA was repeated using the same conditions, but a lower molecular weight cut off (CO) dialysis bag (3500 Da) was used rather than the 8000 Da CO dialysis membrane used in the previous experiment. This was employed to ensure that small degradation fragments of His$_6$.TssA will remain inside the dialysis tube.

The proteolytic products were fractionated on a 15% SDS-PA gel which revealed several products of <15 kDa and small amounts of two products at ~ 20 kDa (Figure 6.23). The fractionated polypeptides were blotted onto a ProBlot^{TM} membrane and those

of ~ 8, 10, 11, 15 and 20 kDa were subjected to N-terminal amino acid sequence analysis. Only the 15 kDa polypeptide fragment was analysed from the flow-through of the Q column, whereas the remaining of the degradation products were analysed from the dialysed partially purified His$_6$.TssA sample. N-terminal analysis revealed that in some cases there was heterogeneity of N-termini within each product. For example, the N-terminal sequence analysis of the 15 kDa product revealed two closely spaced digestion sites: ↓AAQ↓QAQPE, whereas analysis of the 10 kDa product revealed three closely spaced proteolysis sites: ↓QQ↓TA↓SRPPVT (Figure 6.23 and Table 6.1). The cleanest sequence of all was obtained from the high abundance 8 kDa fragment which had a single N-terminus close to that of the 10 kDa fragment: ↓TIAGIQN. All these cleavage sites occur in the predicted interdomain linker. However, the N-terminus of the 11 kDa fragment and the N-terminus of one of the polypeptides running as ~ 20 kDa had N-termini in a similar location within the predicted NTD: ↓RDPE↓HADD. Thus, the 11 kDa species is likely to have a truncated C-terminus, probably through a second cleavage within the interdomain linker. The other major species of ~ 20 kDa had the same N-terminus as full-length His$_6$.TssA. The N-terminal sequences of the analysed His$_6$.TssA degradation products are shown in (Figure 6.23 and Table 6.1).
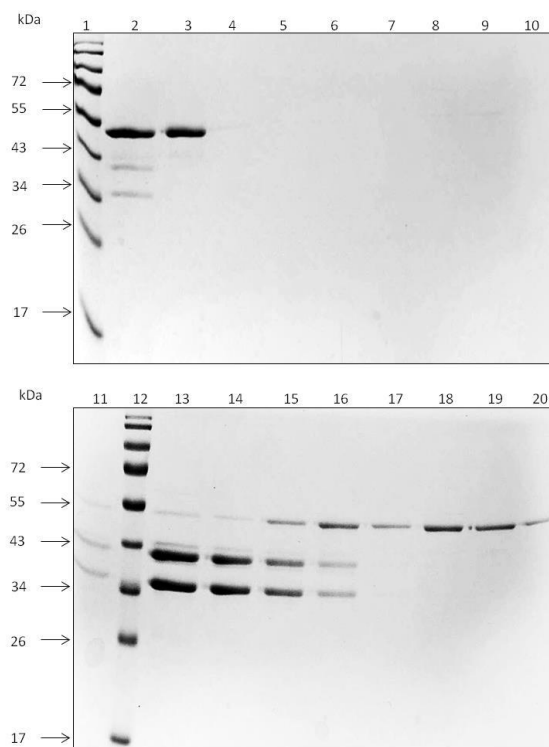
Figure 6.22. Analysis of proteolytically cleaved His₆.TssA. Dialysed partially purified His₆.TssA protein present in 20 mM Tris-HCl (pH 8.0) was loaded on a 1 ml GE Healthcare HiTrap Q HP column and isocratic gradient elution was carried out from 0-500 mM NaCl in 20 mM Tris-HCl (pH 8.0) buffer. Samples were analysed by 12% SDS-PAGE. Lanes 1 and 12, EZ-Run Rec pre-stained protein markers; lane 2, sample of protein loaded on the HiTrap Q HP anion exchange column; lane 3, flow-through of the column, lane 4, column wash; lanes 5-7, samples from 1 ml fractions eluted in 100 mM NaCl; lanes 8-10, samples from 1 ml fractions eluted in 200 mM NaCl; lanes 11, 13 and 14, samples from 1 ml fractions eluted in 300 mM NaCl; lanes 15-17, samples from 1 ml fractions eluted in 400 mM NaCl; lanes 18-20, samples from 1 ml fractions eluted in 500 mM NaCl.
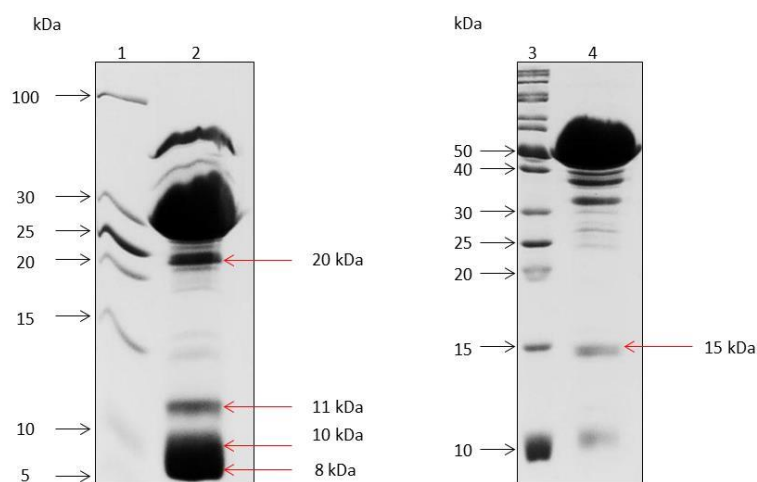
Figure 6.23. Polyacrylamide gel electrophoretic analysis of degraded His₆.TssA protein.
Concentrated samples of degraded His₆.TssA protein were fractionated by electrophoresis in a 15% SDS-PAGE. The red arrows indicate His₆.TssA degradation products which were analysed by N-terminal sequence analysis. Lane 1, low range protein ladder; lane 2, concentrated sample of dialysed partially purified His₆.TssA protein present in 20 mM Tris-HCl (pH 8.0); lane 3, EZ-Run Rec unstained protein marker; lane 4, concentrated sample of the flow-through of the Q column.

Table 6.1. N-terminal amino acid sequences of His₆.TssA degradation products.

Peptide[a]	N-terminal sequence
20 kDa	G-S-S-H-H-H-H- H-A-D-X-X-X-
15 kDa	A-A-Q-Q-A-Q-P-E Q-A-Q-P-E-X-I-E
11 kDa	R-D-P-E-H-A-D-D-
10 kDa	Q-Q-T-A-S-R-P T-A-S-R-P-P-V S-R-P-P-V-T-
8 kDa	T-I-A-G-I-Q-N

[a] sizes were estimated based on migration in SDS-PA gels.

6.6.4.2. Mass spectrometry analysis of His$_6$.TssA degradation product

In order to identify the C-terminal amino acid sequence (site of cleavage) of the two large proteolytic fragments of His$_6$.TssA (Figure 6.22), the sample was analysed by mass spectrometry at the Astbury centre for Structural Molecular Biology, University of Leeds following dialysis against water (button dialysis).

Four species of His$_6$.TssA fragment were identified using the mass spectrometer that had the same N-terminus as the full length His$_6$.TssA (Figure 6.24). Their calculated molecular masses suggest that they would run as two pairs of co-migrating polypeptides in SDS polyacrylamide gels, consistent with what was observed (Figure 6.22). In each case, the C-terminus is located within the predicted interdomain linker.

A summary of all the cleavage sites identified by N-terminal sequence analysis and mass spectrometry is presented in Figure 6.25.

Average Mass = 43811.5756, Monoisotopic Mass = 43784.9146
N-Terminus = H, C-Terminus = OH

```
1    GSSHH HHHHS SGLVP RGSHM PINLP ELLTP ISEAS PSGDD LLFSN EFDAI QDARR YDDPT LDQGE VVTEI KEADW GFVVD HAGEL
86   LRTRT KDLRL AVWLT EALAL EDGIT GLTEG YALLE GLCRE FWDTF HPLPE DDDIE HRLGN VAWLS GRTAE LLRAV PLTDG ASNAF
171  STLDW EVAQH VAQSI KRDPE HADDI ARGKP SIEQI DASRR VTSIA FYTAL LANLK AFEFA LDAFE ERLVE RAGDS APSFR QARDA
256  FETVY RLAER FAREQ GYTGS APHTQ AAQQA QPERI EPVFG QPIQT EETHV QQQTA SRPPV TQTIA GIQNR AQAVD QLRAV ARYFR
341  QTEPH SPVAY LADKA AEWAD MPLHK WLESV VKDDG SLSHI RELLG VRPDE QS
```

Average Mass = 34438.0278, Monoisotopic Mass = 34417.0903   **Measured masses**
N-Terminus = H, C-Terminus = OH

```
1    GSSHH HHHHS SGLVP RGSHM PINLP ELLTP ISEAS PSGDD LLFSN EFDAI QDARR YDDPT LDQGE VVTEI KEADW GFVVD HAGEL
86   LRTRT KDLRL AVWLT EALAL EDGIT GLTEG YALLE GLCRE FWDTF HPLPE DDDIE HRLGN VAWLS GRTAE LLRAV PLTDG ASNAF       34440 Da
171  STLDW EVAQH VAQSI KRDPE HADDI ARGKP SIEQI DASRR VTSIA FYTAL LANLK AFEFA LDAFE ERLVE RAGDS APSFR QARDA
256  FETVY RLAER FAREQ GYTGS APHTQ AAQQA QPERI EPVFG QPIQT EETHV QQQ
```

Average Mass = 34181.7662, Monoisotopic Mass = 34160.9731
N-Terminus = H, C-Terminus = OH

```
1    GSSHH HHHHS SGLVP RGSHM PINLP ELLTP ISEAS PSGDD LLFSN EFDAI QDARR YDDPT LDQGE VVTEI KEADW GFVVD HAGEL       34181 Da
86   LRTRT KDLRL AVWLT EALAL EDGIT GLTEG YALLE GLCRE FWDTF HPLPE DDDIE HRLGN VAWLS GRTAE LLRAV PLTDG ASNAF
171  STLDW EVAQH VAQSI KRDPE HADDI ARGKP SIEQI DASRR VTSIA FYTAL LANLK AFEFA LDAFE ERLVE RAGDS APSFR QARDA
256  FETVY RLAER FAREQ GYTGS APHTQ AAQQA QPERI EPVFG QPIQT EETHV Q
```

Average Mass = 31537.8693, Monoisotopic Mass = 31518.6643
N-Terminus = H, C-Terminus = OH

```
                                                                                                             31539 Da
1    GSSHH HHHHS SGLVP RGSHM PINLP ELLTP ISEAS PSGDD LLFSN EFDAI QDARR YDDPT LDQGE VVTEI KEADW GFVVD HAGEL
86   LRTRT KDLRL AVWLT EALAL EDGIT GLTEG YALLE GLCRE FWDTF HPLPE DDDIE HRLGN VAWLS GRTAE LLRAV PLTDG ASNAF
171  STLDW EVAQH VAQSI KRDPE HADDI ARGKP SIEQI DASRR VTSIA FYTAL LANLK AFEFA LDAFE ERLVE RAGDS APSFR QARDA
256  FETVY RLAER FAREQ GYTGS APHTQ AAQ
```

Average Mass = 31267.5808, Monoisotopic Mass = 31248.5315
N-Terminus = H, C-Terminus = OH

```
                                                                                                             31267 Da
1    GSSHH HHHHS SGLVP RGSHM PINLP ELLTP ISEAS PSGDD LLFSN EFDAI QDARR YDDPT LDQGE VVTEI KEADW GFVVD HAGEL
86   LRTRT KDLRL AVWLT EALAL EDGIT GLTEG YALLE GLCRE FWDTF HPLPE DDDIE HRLGN VAWLS GRTAE LLRAV PLTDG ASNAF
171  STLDW EVAQH VAQSI KRDPE HADDI ARGKP SIEQI DASRR VTSIA FYTAL LANLK AFEFA LDAFE ERLVE RAGDS APSFR QARDA
256  FETVY RLAER FAREQ GYTGS APHTQ
```

Figure 6.24. Average molecular masses of the largest His$_6$.TssA degradation products. The four species of N-terminal His$_6$.TssA fragment that were identified and their masses measured by mass spectrometry are shown.

277
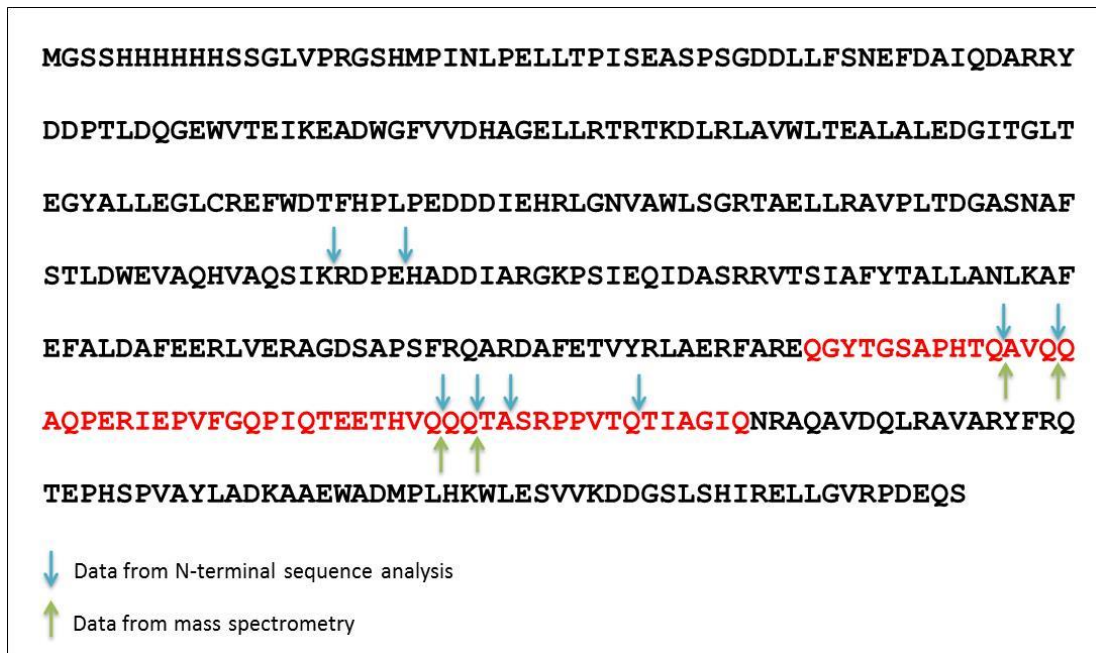
Figure 6.25. Summary of the cleavage sites of His₆.TssA protein. The amino acid sequence of the His₆.TssA protein is shown. The predicted linker region is in red font, the green up arrows indicate the site of cleavages identified by mass spectrometry and the blue down arrows indicate the sites of cleavage identified by N-terminal amino acid sequence analysis.

6.7. Analysis of other His-tagged TssA derivatives

The purification of N- and C-terminal Deca-His-tagged TssA and linkerHis1.TssA and linkerHis2.TssA was carried out using the same protocol employed for the purification of His_6.TssA (IMAC followed by Superose 6 SEC). SEC was used to determine the size of all four proteins. The elution time of His_{10}.TssA, TssA.His_{10}, linkerHis1.TssA and linkerHis2.TssA were 16.17, 16.06, 15.68 and 16.51 minutes, respectively, giving estimated sizes of 1.49, 1.50, 1.62 and 1.40 MDa, respectively. The purified proteins were then analysed by negative stain EM, and the results showed all four proteins form a ring structure similar to those formed by His_6.TssA (Figure 6.26A and B).
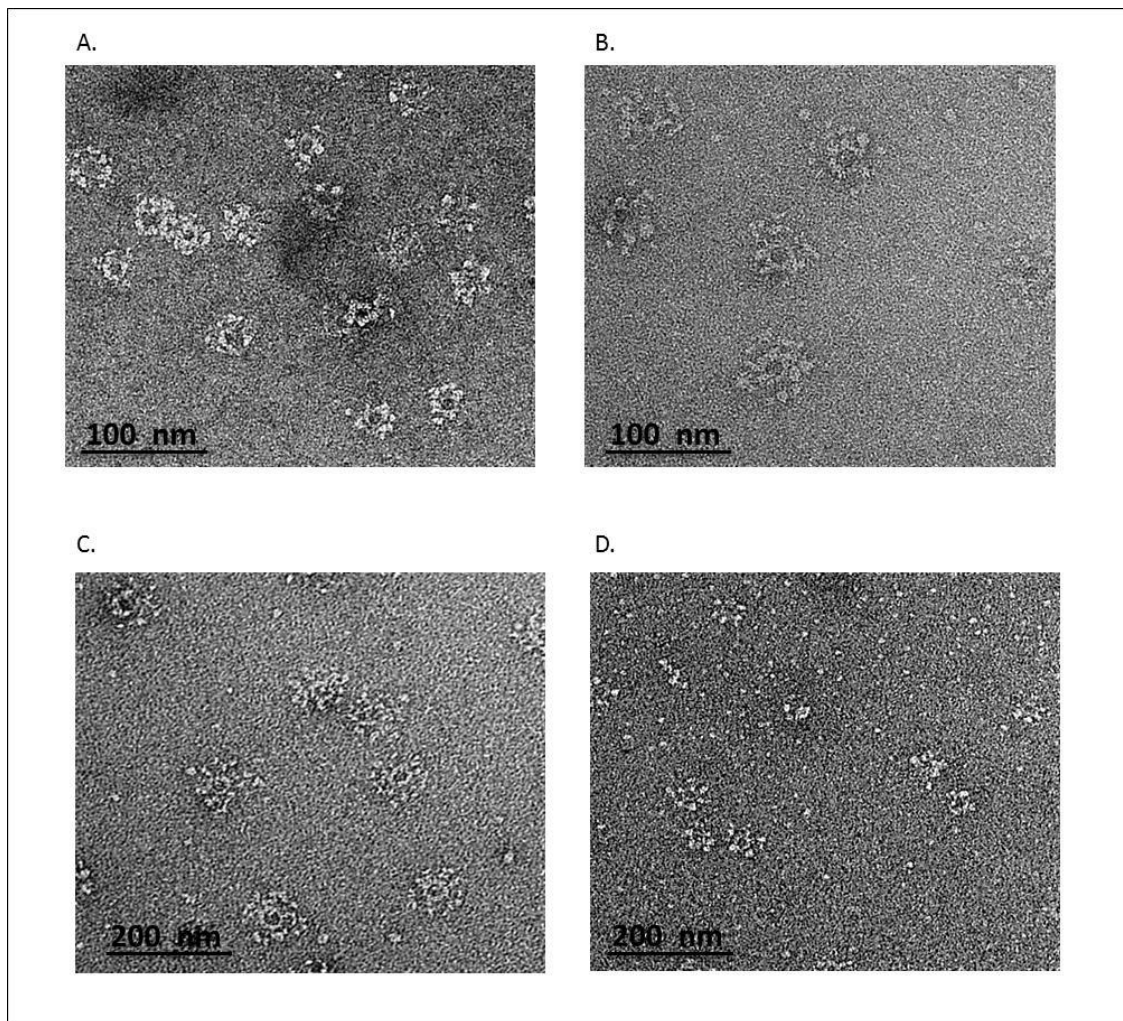
Figure 6.26. Negative stain electron micrographs of His$_{10}$.TssA, TssA-His$_{10}$, linkerHis1.TssA and linkerHis2.TssA. Negative staining was carried out using 0.75% uranyl formate. The protein was analysed in buffer containing 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, 10% glycerol and 10 mM imidazole. A. EM image of His$_{10}$.TssA complexes. B. EM image of TssA.His$_{10}$ complexes. C. EM image of linkerHis1.TssA complexes. D. EM image of linkerHis2.TssA complexes.

.

6.8. Analysis of native (untagged) TssA

6.8.1. Cloning and overproduction of untagged TssA

tssA was cloned into pET14b without fusion to the His tag coding sequence. The forward primer (tssAnonHisFor) was designed to include an *Xba*I site at the 5' end, followed by the pET14b derived Shine-Dalgarno sequence and the first 20 bp of the *tssA* coding sequence (see appendix 1 for the primer sequences). The reverse primer (pET14b-iotArev) contained a *Bam*HI site, stop codon and the last 20 bp of the *tssA* coding sequence. The *tssA* gene was PCR amplified from *B. cenocepacia* strain H111 genomic DNA, digested with *Xba*I and *Bam*HI and cloned into pET14b cut with same restriction enzymes. This resulted in removal of the histidine tag and thrombin cleavage site coding sequences from the plasmid.

Overexpression of the gene encoding the untagged TssA protein in *E. coli* BL21(λDE3) was carried out as described in Section 2.5.1.1. This resulted in production of a large amount of a protein migrating with a size corresponding to 50 kDa as observed for the other TssA derivatives (Figure 6.27). The protein was shown to be present in the soluble fraction of *E. cole* cells (result not shown).
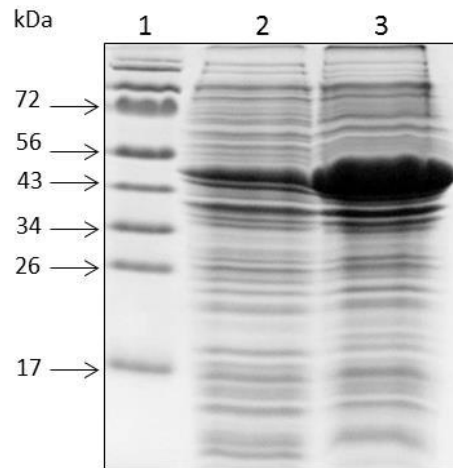
Figure 6.27. Overexpression of native TssA. Coomassie brilliant blue stained 12% SDS polyacrylamide gel showing the overproduction of untagged TssA in BL21(λDE3) containing pET14b.TssA. Lane 1, EZ-Run Rec prestained protein ladder; lane 2, total cell protein prior to induction; lane 3, total cell protein following 3 hours of induction with 1 mM IPTG.

6.8.2. Purification of native TssA

6.8.2.1. Polyethyleneimine precipitation

Polyethyleneimine (PEI) was used to precipitate nucleic acids from the soluble fraction of the crude cell lysate. To determine the optimum concentration of PEI to use, titration of PEI precipitation was carried out on 1 ml of the soluble fraction following induction of TssA synthesis. 10 µl of 5% PEI in water (pH 7.0) was added to the supernatant. After mixing, the sample was centrifuged at 13,200 rpm for 5 minutes. The supernatant was transferred to clean micro-centrifuge tube and treated with another 10 µl of 5% PEI as in the previous step. This process was repeated until no pellet was formed (i.e. five times). Each supernatant was analysed by electrophoresis in a 12% SDS PA gel (Figure 6.28). The result showed that > 90% of TssA was precipitated at a final concentration of 0.15% PEI.

Although TssA is not known to be a DNA binding protein, it is precipitated by PEI. In order to remove the nucleic acids from the lysate without precipitating TssA, precipitation was carried out in the presence of 0.2% PEI over a range of medium to high ionic strength lysis buffers containing 0.6-1.0 M NaCl. The results showed that in the presence of buffer containing 0.9-1.0 M NaCl > 90% of TssA remained soluble at a PEI concentration that precipitated nucleic acids (Figure 6.29). Therefore, in the subsequent purification procedures of untagged TssA, PEI co-precipitation of nucleic acids and TssA was carried out using 0.2% PEI in water. The precipitate was resuspended in the same original volume of lysis buffer containing 1.0 M NaCl and a clearing spin carried out to remove the insoluble nucleic acids.
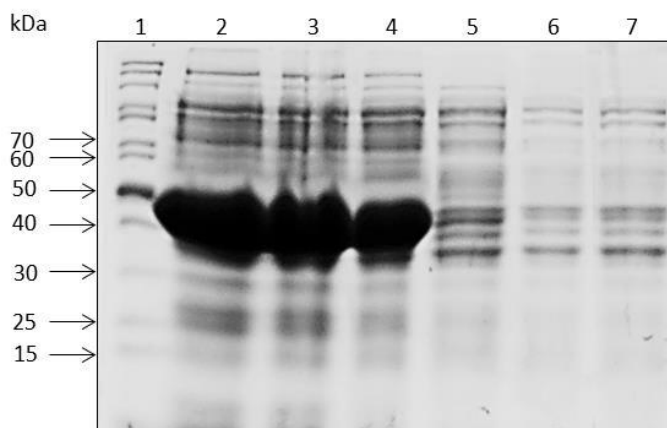
Figure 6.28. Titration of PEI precipitation of untagged TssA. 10 µl of 5% PEI in water (pH 7.0) was added to 1 ml of soluble fraction from lysed *E. coli* BL21(λDE3) cells containing pET14b.TssA. The mixture was centrifuged and the supernatant transferred to a clean microcentrifuge tube to which another 10 µl of PEI was added. Sequential transfer was carried out until no pellet was formed. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, cell lysate (containing untagged TssA); lanes 3-7, supernatant after sequential addition of 10 µl of 5% PEI.
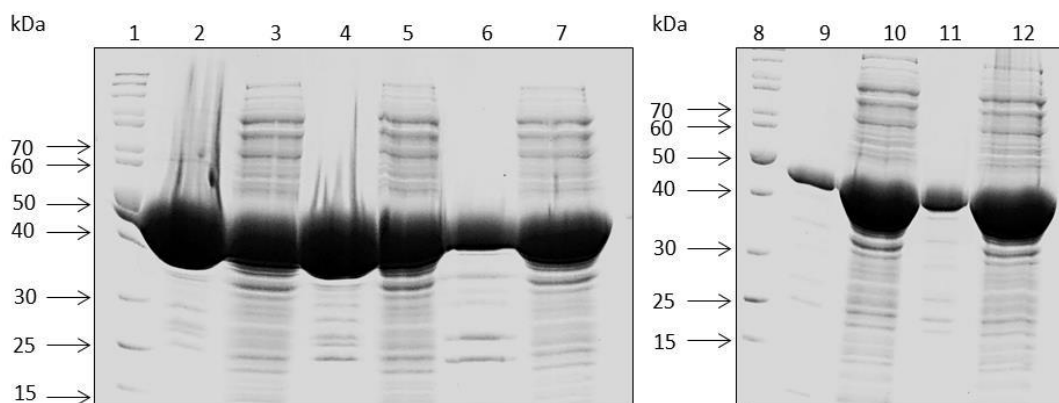
Figure 6.29. Determination of buffer ionic strength to be used to selectively precipitate nucleic acids with PEI from a lysate containing untagged TssA. Cell lysates which contained soluble untagged TssA were prepared in lysis buffer (50 mM Tris-HCl (pH 8.0), 10% glycerol) containing 0.6-1.0 M NaCl. 1 ml of the soluble fraction containing 0.6, 0.7, 0.8, 0.9 and 1.0 M NaCl was transferred to a microcentrifuge tube and 40 µl of 5% PEI was added to each tube and mixed. The samples were centrifuged at 13,200 rpm for 5 minutes and the supernatant transferred to clean tubes. Samples of the supernatant and pellet were analysed by 12% SDS-PAGE. Lanes 1 and 8, EZ-Run Rec unstained protein ladder; lane 2, sample from the pellet generated in the presence of in 0.6 M NaCl; lane 3, sample from the supernatant 0.6 M NaCl; lane 4, sample from the pellet generated in the presence of 0.7 M NaCl; lane 5, sample from the supernatant containing 0.7 M NaCl; lane 6, sample from the pellet generated in the presence of 0.8 M NaCl; lane 7, sample from the supernatant containing 0.8 M NaCl; lane 9, sample from the pellet generated in the presence of 0.9 M NaCl; lane 10, sample from the supernatant containing 0.9 M NaCl; lane 11, sample from the pellet generated in the presence of 1 M NaCl; lane 12, sample from the supernatant containing 1 M NaCl.

6.8.2.2. Ammonium sulfate precipitation

The remaining PEI in solution can be a problem for subsequent steps of protein purification, therefore it should be eliminated. PEI can be removed from the protein of interest by ammonium sulfate precipitation (ASP). In addition, ASP can be used as a purification step for the target protein. To determine the salting out point of TssA, increasing concentrations of ammonium sulfate (5%, 10%, 15%, 20% and 30%) were tested. The results showed that more than 95% of TssA protein was precipitated by 30% ammonium sulfate (Figure 6.30). Therefore, a 30% ammonium sulfate precipitation step was included in the purification procedure.

6.10.2.3. Purification of untagged TssA by size exclusion chromatography

Further purification of untagged TssA was carried out using SEC. 10 ml of soluble fraction of a cell lysate containing overproduced TssA was sequentially precipitated with 0.2 % PEI and 30% ammonium sulfate, as described above (Figure 6.31A). After ASP the pellet was re-suspended in 5 ml lysis buffer and dialysed overnight against 1 litre of the same buffer with two changes to remove the ammonium sulfate. In order to use the purified untagged TssA protein for AUC, the buffer was then changed to 10 mM sodium phosphate (pH 7.4), 50 mM NaCl using an Amicon Ultra-4 centrifugal filter with a 100 kDa cut off. After buffer exchange, the sample was concentrated to 200µl (TssA concentration 20 mg ml^{-1}). A 100 µl loading loop used to apply 100µl of the concentrated sample onto a Superose 6 gel filtration column. 0.5 ml fractions were collected and the elution of TssA was monitored at 280 nm (Figure 6.31B and C). The elution time of untagged TssA was 16.56 minutes giving an estimated size for untagged TssA as 1.39 MDa.
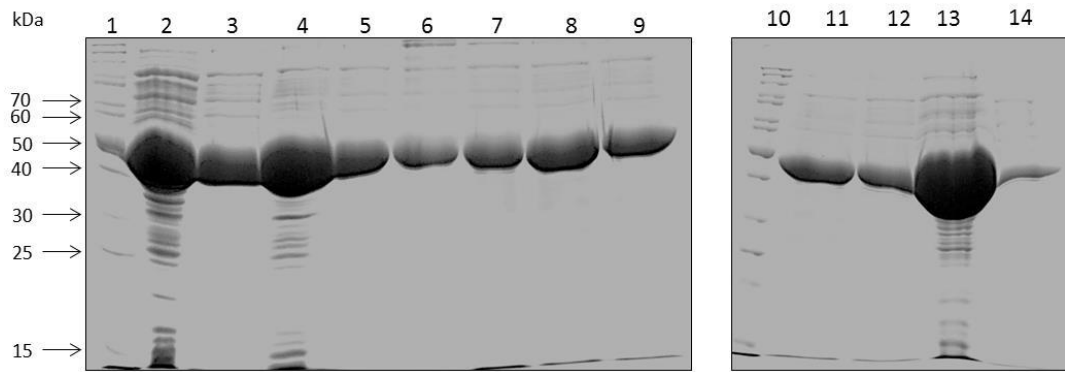
Figure 6.30. Titration of ammonium sulfate precipitation of untagged TssA protein. 1 ml of untagged TssA treated with 40 µl of 5% (w/v) PEI and the recovered precipitate was resuspended in 1 ml of a buffer containing 50 mM Tris-HCl (pH 8.0), 1 M NaCl, 10 % glicerol and 10 mM imidazole. Ammonium sulfate was slowly added to a final concentration of 5, 10, 15, 20 and 30%. Following addition of ammonium sulfate, the precipitate was recovered by centrifugation at 13,000 rpm for 5 minutes. Samples of the pellet and the supernatant were analysed by 12% SDS-PAGE. Lanes 1 and 10, EZ-Run Rec unstained protein ladder; lane 2, soluble fraction of cell lysate containing untagged TssA; lane 3, resuspended untagged TssA in 1 M NaCl buffer after precipitation with 0.2 % PEI; lanes 4, 6, 8, 11 and 13, insoluble material following precipitation with 5, 10, 15, 20 and 30% ammonium sulphate, respectively; lanes 5, 7, 9, 12 and 14, supernatant following precipitation with 5, 10, 15, 20 and 30% ammonium sulphate, respectively.
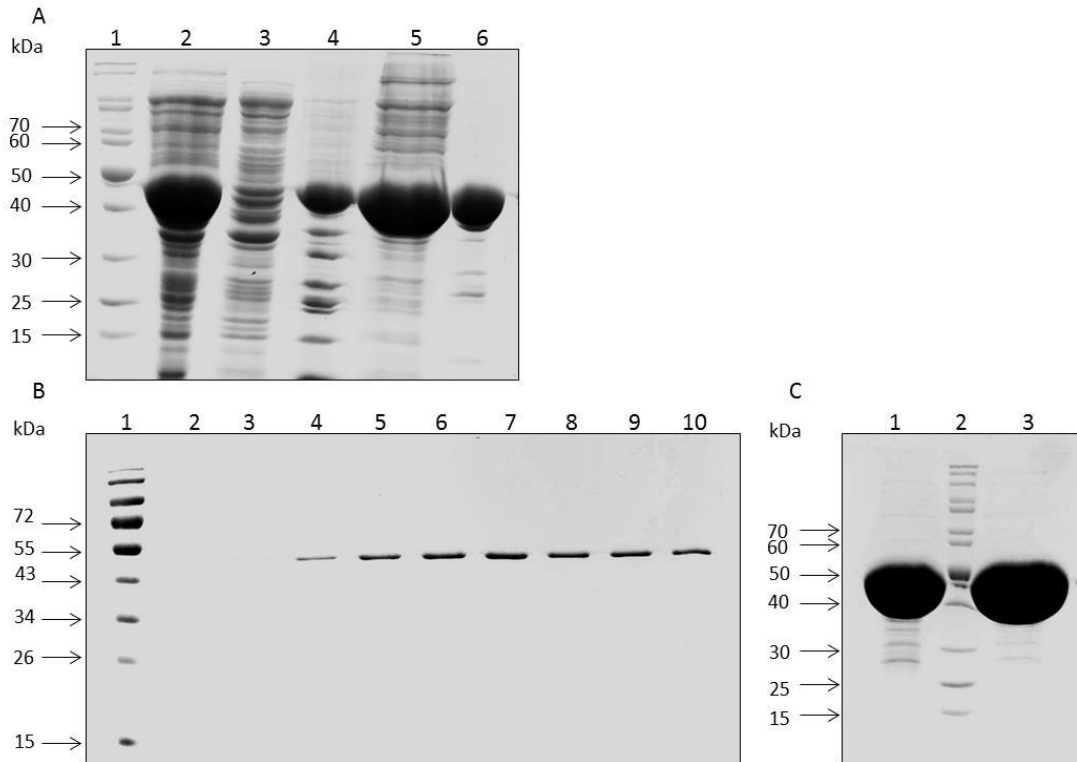
Figure 6.31. Purification of untagged TssA. The untagged TssA protein was partially purified by PEI and ammonium sulfate precipitation (A) then concentrated into 0.1 ml of 50 mM Tris-HCl buffer, 200 mM NaCl and 10% glycerol (pH 8.0) and applied to a Superose 6 gel filtration column (B and C). A. Sequential PEI and ammonium sulfate precipitation of untagged TssA. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, cell lysate supernatant containing untagged TssA; lane 3, the supernatant following PEI precipitation; lane 4, remaining insoluble material after re-suspension of the PEI pellet in 1.0 M NaCl; lane 5, soluble material following resuspension of the PEI precipitate in 1.0 M NaCl; lane 6, TssA pellet resuspended in lysis buffer following 30% ammonium sulfate precipitation. B. SDS PAGE analysis of untagged TssA eluted from gel filtration column. Lane 1, EZ-Run Rec pre-stained protein ladder; lanes 2-3, eluted fractions before the TssA peak detected on the UV trace; lanes 4-10, fractions corresponding to the TssA peak on the UV trace. C. Lane 1, concentrated sample of untagged TssA prior to SEC; lane 2, EZ-Run Rec unstained protein ladder; lane 3, pooled concentrated untagged TssA (fractions 4-10) following elution from gel filtration column.

6.8.3. Analysis of the oligomerisation status of untagged TssA by analytical ultra-centrifugation

6.8.3.1. Analysis of untagged TssA by sedimentation velocity analytical ultra-centrifugation ($_{SV}$AUC)

$_{SV}$AUC analysis was carried out using 3 concentrations of TssA (26, 19 and 1 µM). Radial absorbance scans at 220 and 250 nm as well as interference data were recorded for each cell over 10.5 hours (Figures 6.32 and 6.33). Using the raw data files shown in Figures 6.32 and 6.33, the samples were subjected to continuous c(s) distribution analysis. For the calculation of molecular weight at the 250 nm absorbance signal, the result showed that the two highest concentrations of untagged TssA gave similar sedimentation properties (the concentration of the lowest (1 µM) was too low for detection using this optical system). Both concentrations showed species sedimenting at approximately 2, 4 and 10 S; with estimated mol. wt ranging from 130 kDa, 257-358 kDa and 1.0-1.8 MDa respectively (Figure 6.34).

An estimate for the molar mass is calculated based on the best-fit frictional ratio information on each sedimenting species. The asymmetry of the molecules in solution is measured by the frictional coefficient (f/fo); where 1.0 denotes a perfect sphere whereas 1.2-1.8 is usual for a folded protein. This asymmetry factor is measured as an average for all species in solution. The data show that the two concentrations of untagged TssA analysed are highly asymmetric or contain unfolded regions.

The result of interference data analysis for the 26 and 19 µM concentrations were comparable to absorbance data collected for those concentrations. Analysis detailed levels of highly asymmetry or unfolded protein and generated similar mol. wt values. The interference system measures differences in refractive index between sample and buffer solutions, and can be used to analyse the lowest concentration of untagged TssA. The result showed that the major species was spherical and sedimented at approximately 2 S, with a calculated mol. wt of 12.2 kDa. The two concentrations were also analysed to double-check that both optical systems showed similar results, which they did.
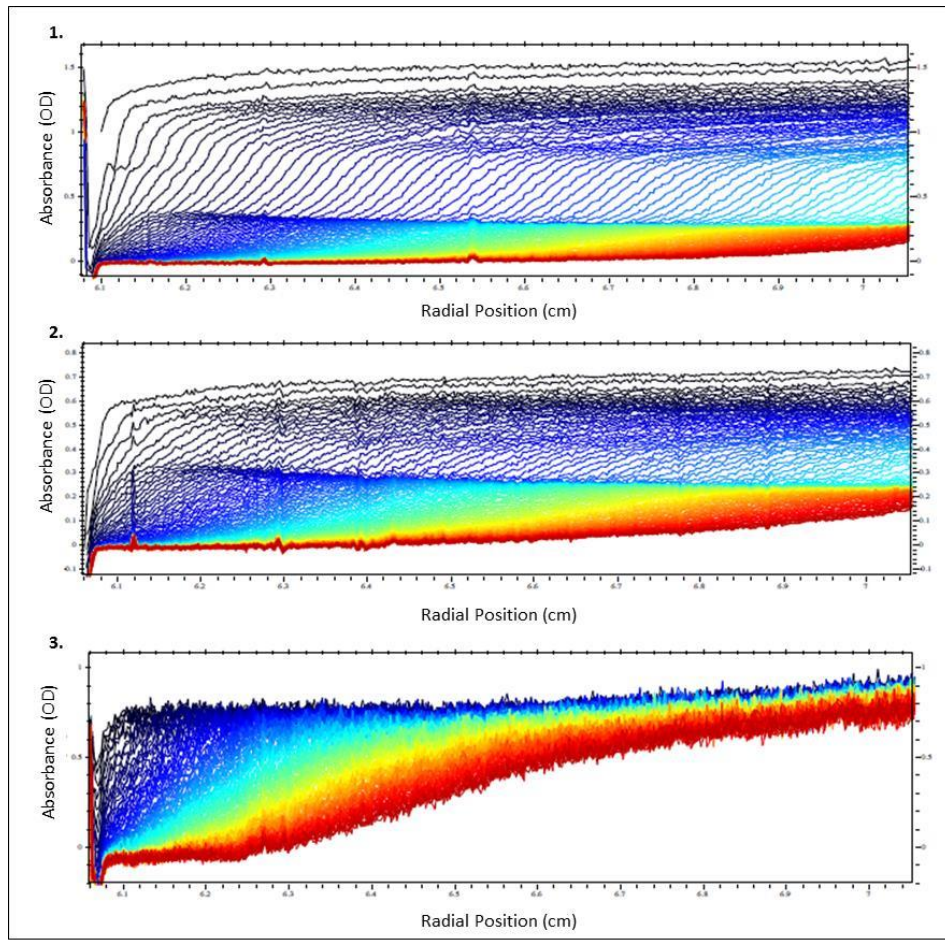
Figure 6.32. Analytical ultracentrifugation radial absorbance scans of untagged TssA at 220 and 250 nm. The data are shown for each cell with blue showing the earliest scan and red the latest. (1) Absorbance at 250 nm raw data scans for cell 1, (2) Absorbance at 250 nm raw data scans for cell 2 and (3) Absorbance at 220 nm raw data scans for cell 3.
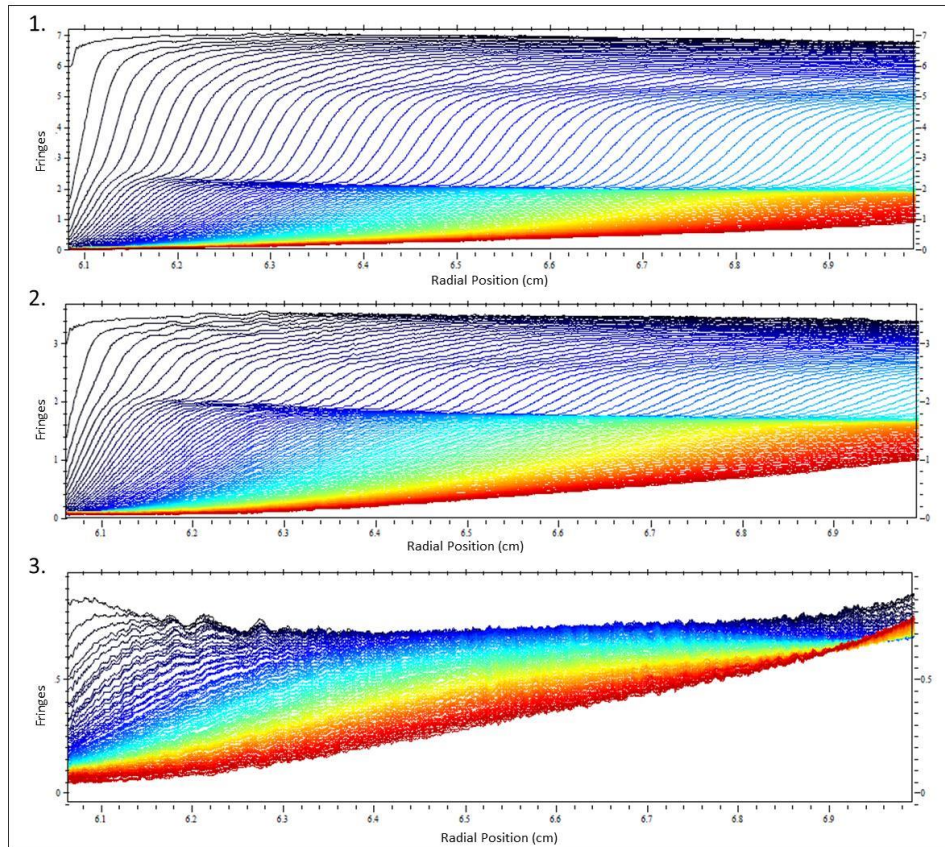
Figure 6.33. Analytical ultracentrifugation interference scans of untagged TssA. The data are shown for each cell with blue showing the earliest scan and red the latest. (1) Interference raw data scans for cell 1, (2) interference raw data scans for cell 2 and (3) interference raw data scans for cell 3.
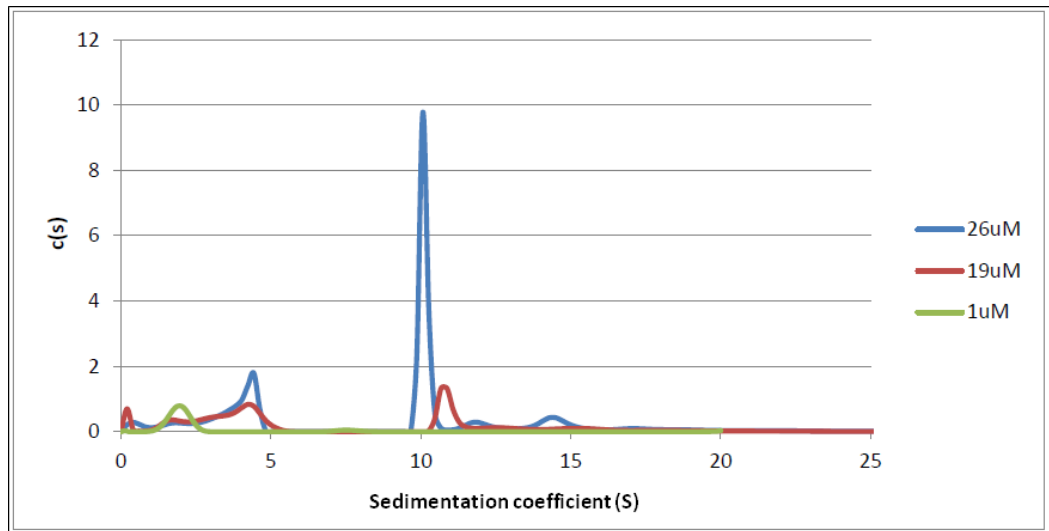
Figure 6.34. Sedimentation velocity analytical ultra-centrifugation ($_{sv}$AUC) of untagged TssA. c(s) distributions up to 25S are shown.

6.8.3.2. Analysis of untagged TssA by sedimentation equilibrium analytical ultra-centrifugation ($_{SE}$AUC)

Following $_{SV}$AUC analysis of untagged TssA, $_{SE}$AUC analysis was carried out at 3 concentrations of TssA (26, 19 and 3 μM). Samples (0.125 ml) were centrifuged in 1.2 cm pathlength 6-sector centrepiece cells (samples in RH sectors, reference buffer in LH sectors) with sapphire windows in a 4-place An-60 Ti analytical rotor running in an Optima XL-I analytical ultracentrifuge (Beckman Instruments) at a temperature of 4˚C. Running speeds were 5,500, 18,000 and 22,000 rpm and were calculated assuming a monomer molecular weight of 41.8 kDa. Radial absorbance scans at wavelengths 250 and 280 nm with 20 replicates at a radial step size of 0.001 cm were performed and interference scans were also collected. Scans were judged to be at equilibrium by using Sedfit v11.9 by looking at the difference between successive scans (Schuck, 2000). Due to the range of concentrations present, two wavelengths were used to determine concentration gradients during the run time; 250 and 280 nm. The total run time was just over 65 hours. Tables 6.2-6.5 show analyses of the data obtained from $_{SE}$AUC. It is possible that the sample may contain some contaminants which explain the low molecular weight of species 3 in table 6.2 and 6.4.

All concentrations were analysed individually and globally to compare similarities and to draw accurate conclusions regarding the stoichiometry and molecular weights of species in solution. Shown below in the tables 6.2, 6.3, 6.4 and 6.5 are the best fits given for the data ranges analysed, modelled using SedPhat Species Analysis. None of the data could be modelled accurately using monomer: n-mer, where n=2-20 (and monomer is 41.8 kDa). It resulted in poor fits and software crashing.

Table 6.2 $_{SE}$AUC analysis of untagged TssA at 26 μM**.**

Untagged TssA (26 μM)	Determined molecular weight (Da)		
	Species 1	Species 2	Species 3
5,500 rpm	54, 063	849,070	119
18,000 + 22,000 rpm	54, 063	803,267	1,192
All speeds	52, 456	848,363	1,169

Table 6.3 $_{SE}$AUC analysis of untagged TssA at 19 μM**.**

Untagged TssA (19 μM)	Determined molecular weight (Da)		
	Species 1	Species 2	Species 3
5,500 rpm	53,569	870,568	1,119
18,000 + 22,000 rpm	34,392	533,213	1,119
All speeds	61,484	814,788	1,120

Table 6.4 $_{SE}$AUC analysis of untagged TssA at 3 μM**.**

Untagged TssA (3 μM)	Determined molecular weight (Da)		
	Species 1	Species 2	Species 3
5,500 rpm	0	1,027,929	573
18,000 + 22,000 rpm	103,717	792,342	2043
All speeds	*	*	*

* Software would not allow for global analysis of the two data sets at differing rotor speeds.

Table 6.5 Global analysis of untagged TssA**.**

Untagged TssA	Determined molecular weight (Da)		
	Species 1	Species 2	Species 3
All conc, all speeds	60,125	909,444	1,329

6.8.4. Transmission electron microscopy of untagged TssA

A sample of purified untagged TssA was analysed by transmission EM following negative stain with uranyl formate (0.75%). The result showed that the untagged TssA protein forms a ring structure similar to the His-tagged TssA, although the details of the external ring of projections are less clear (Figure 6.35).

6.9. Overexpression and purification of His$_6$.TssA.NTD

To further characterise TssA, the individual domains were analysed separately. To purify large amounts of TssA.NTD, the coding sequence was fused in frame with the His tag coding sequence of pET14b. To do this, the TssA.NTD coding sequence was PCR amplified with a forward primer (pET14b-iotAfor) containing an *Nde*I site and a reverse primer (pET14b-iotA-NTDrev) containing a *Bam*HI site (see appendix 1 for the primer sequences). Therefore, the plasmid-encoded TssA.NTD contains the Hexa-histidine tag at the N-terminal end.

Overexpression of the *tssA.NTD* gene in *E. coli* BL21(λDE3) was performed as described before (Section 2.5.1.1). Large amounts of a polypeptide of ~30 kDa were observed, which corresponded to the expected size of 30.8 kDa for His$_6$.TssA.NTD. A solubility test showed that TssA.NTD protein was present in the soluble fraction (Figure 6.36).

The purification of His$_6$.TssA.NTD was carried out using IMAC as described in Section 2.5.3 (Figure 6.37). Unlike full-length His-tagged TssA, almost all of the His$_6$.TssA.NTD bound to the nickel column in the presence of 10 mM imidazole. His$_6$.TssA.NTD was eluted from the nickel column at ~ 75 mM imidazole and 4 ml of approximately 1 mg ml^{-1} His$_6$.TssA.NTD was obtained. For further purification of TssA.NTD, an ion exchange column (Q sepharose) was used (Section 2.5.4). TssA.NTD bound to the Q column and was eluted with 1.0 M NaCl buffer (Figure 6.38). Following this, the purified protein was dialysed against low salt buffer (20 mM Tris-HCl (pH 8.0), 25 mM NaCl) using an Amicon Ultra-4 centrifugal filter with a 10 kDa cut off.
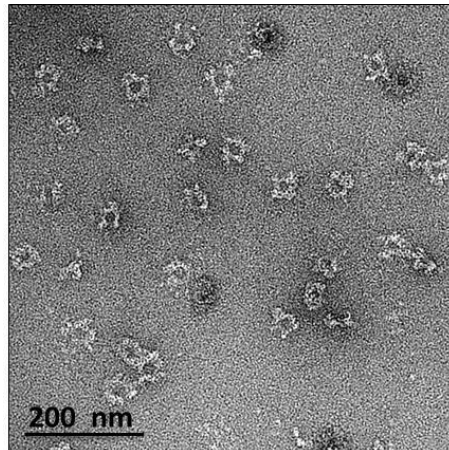
Figure 6.35. Negative stained EM images of purified untagged TssA. TssA was analysed in buffer containing 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, 10% glycerol.
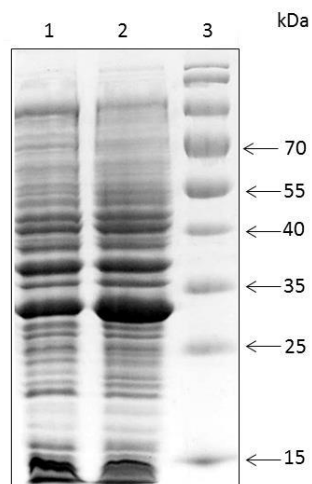
Figure 6.36. Analysis of His$_6$.TssA.NTD protein expression. *E. coli* strain BL21(λDE3) containing pET14b-His$_6$.TssA.NTD was grown to OD$_{600}$ 0.5 in BHI medium supplemented with ampicillin at 37°C whereupon a pre-induction sample was taken and IPTG was added to a final concentration of 1 mM to induce expression of the *tssA.NTD* gene for 3 hours. 15 µl of cleared cell lysate was electrophoresed in a 12 % SDS-PAGE. Lane 1, total cell protein prior to induction; lane 2, total cell protein following 3 hours of induction with 1 mM IPTG; lane 3, EZ-Run Rec prestained protein ladder.
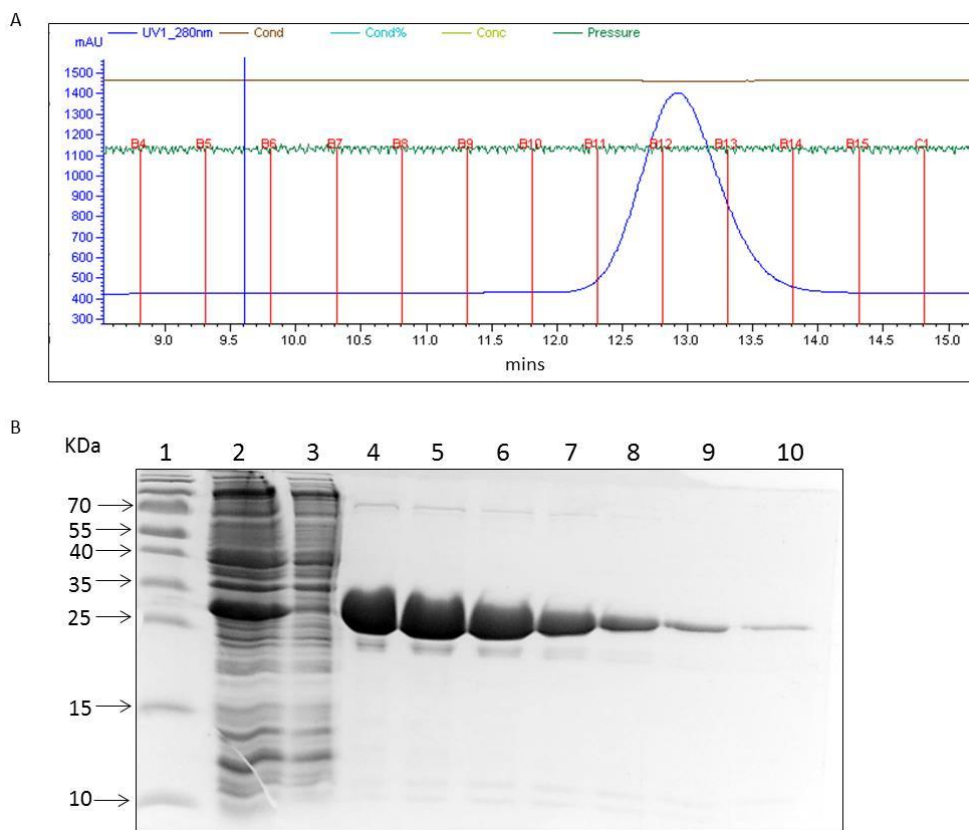
Figure 6.37. Purification of His6.TssA.NTD by IMAC. The soluble fraction from lysed *E. coli* BL21(λDE3) cells containing pET14b-His$_6$.TssA.NTD was applied to a 1 ml GE Healthcare HisTrap HP column in lysis buffer. His$_6$.TssA.NTD was eluted using a linear gradient of 10-500 mM imidazole. A. Elution profile of His$_6$.TssA.NTD from the IMAC column monitored at 280 nm. B. Lanes 1, EZ-Run Rec prestained protein ladder; lane 2, soluble fraction of cell lysate containing His$_6$.TssA.NTD; lane 3, flow-through from the nickel affinity column; lanes 4-10, gradient elution of His$_6$.TssA.NTD from the nickel affinity column.
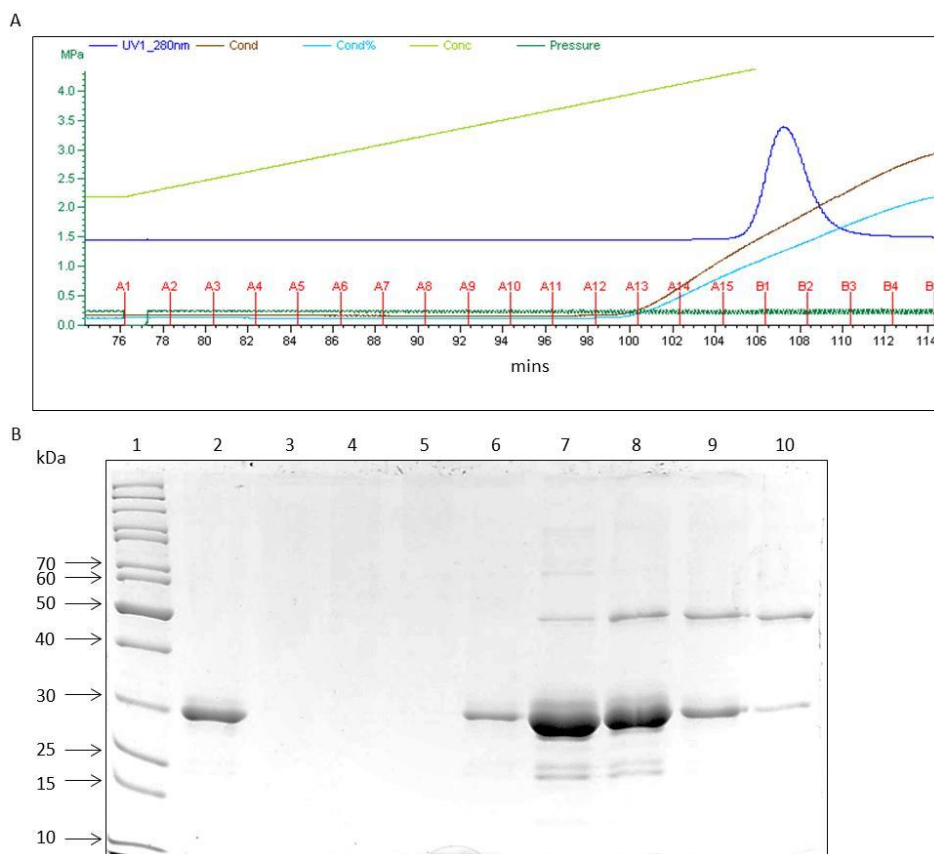
Figure 6.38. Purification of His₆.TssA.NTD by anion exchange. Peak fractions from the IMAC column were pooled and dialysed overnight into 20 mM Tris-HCl (pH 8.0) buffer. Dialysed partially purified His₆.TssA.NTD protein was loaded on a 1 ml GE healthcare HiTrap Q HP column and gradient elution was carried out from 0-1000 mM NaCl in 20 mM Tris-HCl buffer (pH 8.0). A. Elution profile of His₆.TssA.NTD from anion exchange column monitored at 280nm. B. Lane 1, EZ-Run Rec unstained protein marker; Lane 2, His₆.TssA.NTD after dialysis (load of the Q column); lane 3, flow-through from the column; lane 4, wash; lanes 5-10, fractions A14-A15 and B1-B4 in trace (A).

6.10. Analysis of His$_6$.TssA.NTD

6.10.1. Estimation of the oligomerisation status of His$_6$.TssA.NTD using size exclusion chromatography (SEC)

In order to assess the oligomerisation state of His$_6$.TssA.NTD, the purified protein was concentrated to 10 mg ml^{-1} using a 10 kDa cut off Amicon *Ultra*-4 centrifugal filter unit. Size exclusion chromatography on a Superose 12 column was performed as described in Section 2.5.6. 100 µl of the concentrated protein was applied to the column and elution was monitored at 280 nm (Figure 6.39). The experiment was repeated three times and the column was calibrated with protein standards of known size (Figure 6.40). Using the average elution time for His$_6$.TssA.NTD gave an estimated molecular weight of 42.95 kDa which was higher than expected for a monomer of His$_6$.TssA.NTD (30.8 kDa) but far too small to be a dimer. This suggests that His$_6$.TssA.NTD is a monomer under the conditions present in the buffer used, but may exhibit asymmetry.
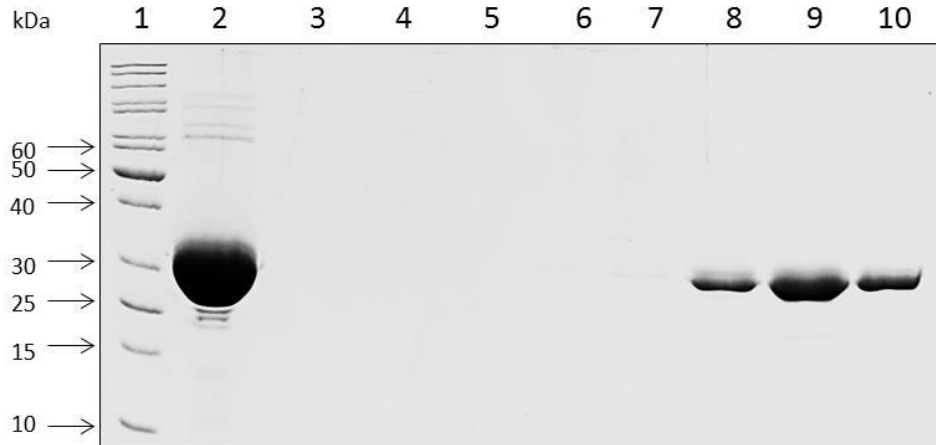
Figure 6.39. Purification of His6.TssA.NTD by gel filtration chromatography. The His$_6$.TssA.NTD protein eluted from the anion exchange column was concentrated to 10 mg ml^{-1} in 50 mM Tris-HCl (pH 8.0) buffer containing 200 mM NaCl and 10% glycerol and applied to a Superose 12 gel filtration column (GE Healthcare). Fractions eluted from the column were analysed by SDS-PAGE. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, His$_6$.TssA.NTD applied to gel filtration column; lanes 3-10, fractions eluted from the gel filtration column.
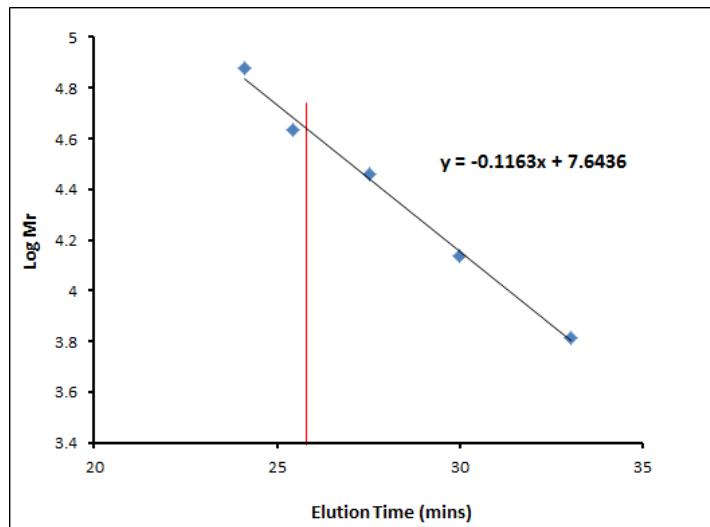
Figure 6.40. Determination of size of His$_6$.TssA.NTD by SEC. The graph shows the elution times of a set of known standarads plotted against the $\log_{10}$ of their molecular weight. A line of the best fit between the points was generated using Microsoft Excel and the equation of the line is shown. Substituting x for the peak elution time of His$_6$.TssA.NTD (25.89 minutes, red line) in the equation gave an estimated molecular weight of 42.95 kDa.

6.10.2. Analysis of His$_6$.TssA.NTD by circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy is a useful technique for rapid determination of the secondary structure content of a protein. Also, it can be used to determine whether a purified protein is folded and to study the effect of a mutation on its conformation or stability. CD is defined as unequal absorption of left- and right-circularly polarised beams of UV light (Greenfield, 2006). UV light can be divided into two regions according to the wavelengths, "near-UV" and "far-UV" regions. The secondary structure composition of any protein can be determined by the "far-UV" spectral region (190-250 nm). The peptide bond can absorb these wavelengths when it is located in a regular, folded protein. Zero CD intensity means the absence of regular structure, while spectra containing both negative and positive signals mean an ordered structure. The "near-UV" spectral region (250-300 nm) is used in CD spectroscopy to determine the tertiary structure of the protein. The aromatic amino acids (phenylalanine, tyrosine and tryptophan) and disulfide bonds are the chromophores that absorb these wavelengths. If the analysed protein produces significant near-UV signals, this indicates that the protein is folded into a well-ordered structure (www.ap-lab.com/circular_dichroism.htm access date 12/02/2013). The CD spectra of poly-L-lysine is used as a standard to analyse the spectra of other proteins and polypeptides, and to estimate the percentage of their α-helices, β-sheet and random coil (Greenfield and Fasman, 1969).

To investigate the conformation of TssA.NTD, 0.1 ml of purified His$_6$.TssA.NTD at a concentration of 1 mg ml^{-1} (32.5 μM) was loaded into a 0.02 cm cell length and subjected to CD analysis in the 190-300 nm wavelength range. The analysis was repeated using 100 μl of the protein at 5 μM (Figure 6.41). Calculation of the mean residue ellipticity of His$_6$.TssA.NTD indicated that the protein has 47.6% α-helices, 27.9% β-strand, 0.0% β-turn and 24.5% random coil.
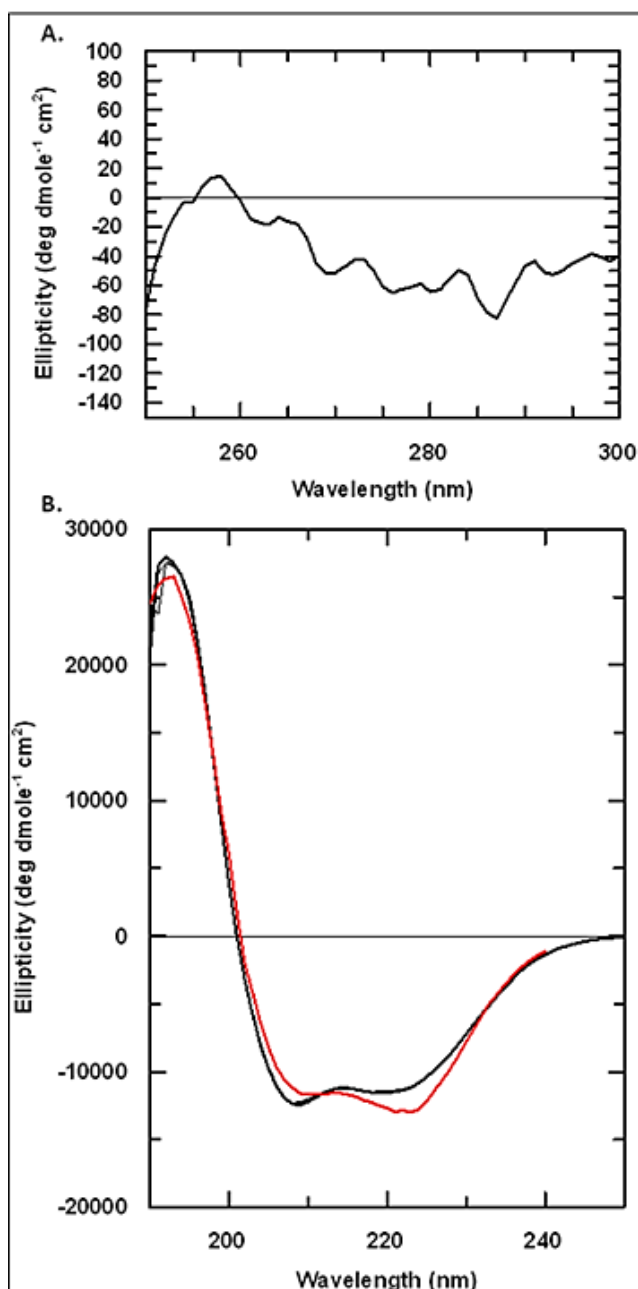
Figure 6.41. Circular dichroism spectra of His$_6$.TssA.NTD. A. Near UV spectrum (aromatic) for tertiary structure analysis. B. Far UV spectrum (amide region) for secondary structure analysis (shown in red).

6.10.3. Purification of untagged TssA.NTD

TssA.NTD was used to generate anti.TssA antibodies. In order to remove the His-tag from the protein, thereby avoiding the possibility of generating antibodies to the tag, which might interfere with subsequent pull down experiments, thrombin cleavage was carried out (Section 2.5.9). This was possible because pET14b encodes a cleavage site for the site-specific protease thrombin downstream of the hexahistidine tag coding sequence (Figure 6.1).

To determine the optimal conditions for digestion, a small scale optimisation experiment was carried out (Section 2.5.9). This showed that complete digestion of the sample was obtained with 0.02 U per µl of thrombin after 16 hours (result not shown). Therefore, the experiment was scaled up proportionately for 1 ml of His_6.TssA.NTD at 1 mg ml^{-1}. After scaling up the reaction, the digestion was still incomplete after 16 hours (Figure 6.42). The partially digested protein was applied to a nickel chelate column, and the flow through containing TssA.NTD without the His-tag was collected (subtractive IMAC) (Figure 6.42). The biotinylated thrombin was removed from the reaction mixture with streptavidin agarose. Untagged TssA.NTD was used for polyclonal antibody generation in a rabbit (Yorkshire Bioscience Ltd).
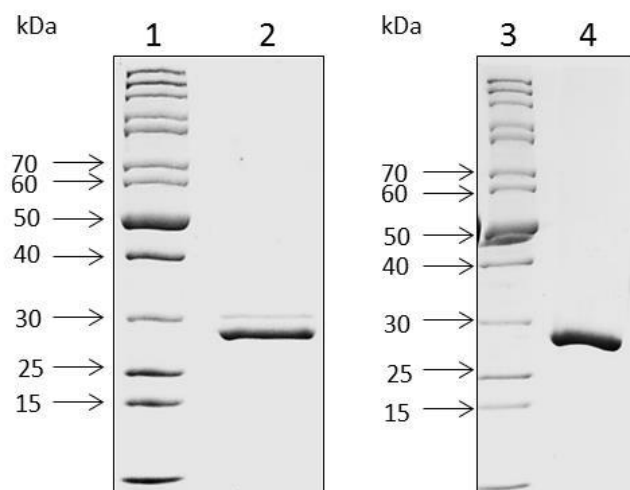
Figure 6.42. Thrombin digestion of His$_6$.TssA.NTD. His$_6$.TssA.NTD (1 ml of 1 mg ml^{-1}) was incubated for 16 hours at room temperature with 0.02 U µl^{-1} thrombin and then applied to a nickel chelate column in 50 mM Tris-HCl, 200 mM NaCl and 10% glycerol (pH 8.0) buffer. The flow-through containing TssA.NTD without His-tag was collected. Samples were analysed by 12% SDS-PAGE. Lanes 1 and 3, EZ-Run Rec unstained protein ladder; lane 2, His$_6$.TssA.NTD protein following thrombin digestion; lane 4, the flow through of the nickel chelate column.

6.11. Overexpression and purification of His$_6$.TssA.CTD

A segment of the *tssA* gene encoding TssA.CTD (70 amino acids) and 12 amino acids of the interdomain linker was previously cloned into pET14b in frame with the hexahistidine tag and thrombin cleavage site coding sequences (S.Shastri, 2011). Therefore, the resulting plasmid, pET14b.TssA.CTD, encodes N-terminally His-tagged TssA.CTD (His$_6$.TssA.CTD). Overexpression of His$_6$.TssA.CTD in *E. coli* BL21(λDE3) was carried out at 37°C in BHI as described in Section 2.5.1.1. However, production of the His$_6$.TssA.CTD protein was not detected under these conditions (predicted size is 11,270 Da), (result not shown). Attempts were made to overproduce His$_6$.TssA.CTD by growing the cells in alternative growth media such as LB and AIM and at a lower temperature (30ºC). However, there was no improvement in the amount of protein obtained (result not shown).

It was thought that TssA.CTD may be not stable in the absence of TssA.NTD. Therefore, the two domains were coexpressed. In order to achieve this, compatible pairwise combinations of the colEI-derived plasmids pET14b-His$_6$.TssA.CTD and pET14b-His$_6$.TssA.NTD, and the p15Aderived plasmids pKT25.TssA.CTD, pKNT25.TssA.CTD, pKT25.TssA.NTD and pKNT25.TssA.NTD were co-transformed into different types of *E. coli* competent cells BL21(λDE3), and its derivatives C41(λDE3) and C43(λDE3). The expression test for all combinations was carried out as described in Section 2.5.1.1. The result showed that overproduction of a protein corresponding in size to that of TssA.CTD was not observed in any combinations expressed in BL21(λDE3), whereas using C41(λDE3) and C43(λDE3) as host strains revealed overproduction of TssA.CTD protein for some plasmid combinations. These combinations included three involving His$_6$.TssA.CTD (i.e. pKNT25.TssA.NTD with pET14b-His$_6$.TssA.CTD in C41(λDE3), the same combination in C43(λDE3), and pKT25.TssA.NTD with pET14b-His$_6$.TssA.CTD in C43(λDE3)) (Figures 6.43 and 6.44).

The expression of *tssA.CTD* was then carried out in the absence of TssA.NTD using C41(λDE3) and C43(λDE3). The result showed that His$_6$.TssA.CTD was overproduced in both strains in the absence of TssA.NTD (Figure 6.45). Unfortunately, the solubility test showed that His$_6$.TssA.CTD protein was insoluble under the conditions employed. Further overexpression experiments were carried out using lower concentrations of IPTG (1.0, 0.5, 0.1 mM and 10 µM) at 30 and 37˚C in both cell types C41(λDE3) and

C43(λDE3). This showed that His$_6$.TssA.CTD was highly expressed in C41(λDE3) at 37˚C when induced by 1.0 and 0.5 Mm, and 10 μM IPTG, and at 30˚C when induced by 1.0 mM IPTG (Figure 6.46). Furthermore, His$_6$.TssA.CTD was highly expressed in C43(λDE3) at 37˚C when induced by 1.0 and 0.5 mM IPTG, and at 30˚C when induced by 1.0 mM IPTG (Figure 6.47). However, no soluble protein was obtained under these conditions (Figures 6.48, 6.49 and 6.50).

TssA.CTD was previously cloned into the BACTH vectors pUT18C and pUT18 (S.Shastri, 2011) and pKT25 (this study), which create fusion proteins with the T25 or T18 domains of cyaA (Section 3.5.1). To test if T18 or T25 is able to solubilise TssA.CTD, the fusion proteins encoded by pUT18C.TssA.CTD, pUT18.TssA.CTD and pKT25.TssA.CTD were overproduced in C41(λDE3) with 1 mM IPTG at 25 and 37˚C. The results showed that the TssA.CTD fusion proteins encoded by two of the three vectors (pUT18-TssA.CTD and pUT18C-TssA.CTD) were overproduced at 37˚C as judged by the appearance of polypeptides of the expected sizes (T18.TssA.CTD, 29 kDa; TssA.CTD-T18, 30 kDa) (Figures 6.51). However, in these two cases the fusion protein was found in the insoluble fraction. A protein corresponding in size to T25-TssA.CTD (33.4 kDa) was present in increased abundance in the crude cell lysate of cells containing pKT25-TssA.CTD following induction but it is not clear whether this corresponds to T25-TssA.CTD as a protein of similar size was observed in cell containing the vector pKT25 (Figure 6.52). In any case it was insoluble.
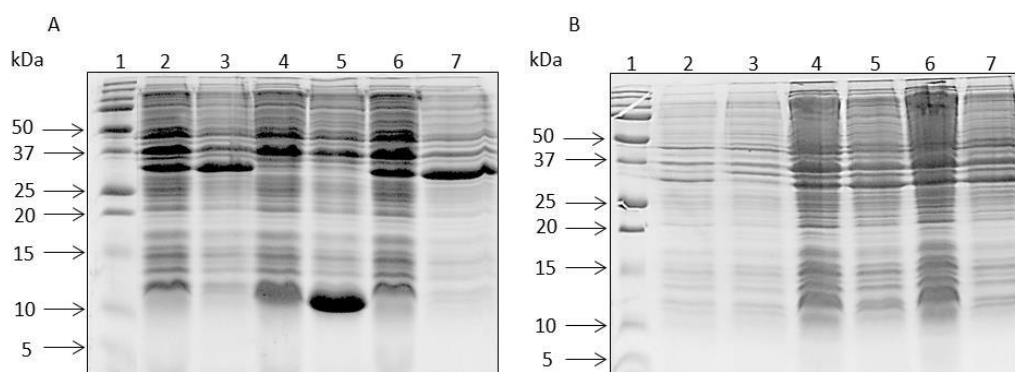
Figure 6.43. Co-expression of TssA.CTD and TssA.NTD proteins. Different plasmid combinations encoding TssA.CTD and TssA.NTD (or as T18- or T25 fusion proteins) were co-transformed into different *E. coli* expression strains and cultured in BHI media supplemented with ampicillin and kanamycin at 37°C and induced for 3 hours with 1mM IPTG. A. Lane 1, EZ-Run Rec prestained protein ladder; lane 2, uninduced TssA.NTD and T25.TssA.CTD in C41(λDE3); lane 3, IPTG induced TssA.NTD and T25.TssA.CTD; lane 4, uninduced TssA.NTD-T25 and TssA.CTD in C41(DλE3); lane 5, IPTG induced TssA.NTD-T25 and TssA.CTD; lane 6, uninduced TssA.NTD and T25.TssA.CTD in C41(λDE3); lane 7, IPTG induced TssA.NTD and T25.TssA.CTD. B. Lane 1, EZ-Run Rec prestained protein ladder; lane 2, uninduced TssA.NTD and T25.TssA.CTD proteins in BL21(λDE3); lane 3, IPTG induced T25.TssA.NTD and T25.TssA.CTD; lane 4, uninduced TssA.NTD and TssA.CTD-T25 proteins in C43(λDE3); lane 5, IPTG induced TssA.NTD and TssA.CTD-T25; lane 6, uninduced TssA.NTD and T25.TssA.CTD in C43(λDE3); lane 7, IPTG induced TssA.NTD and T25.TssA.CTD.
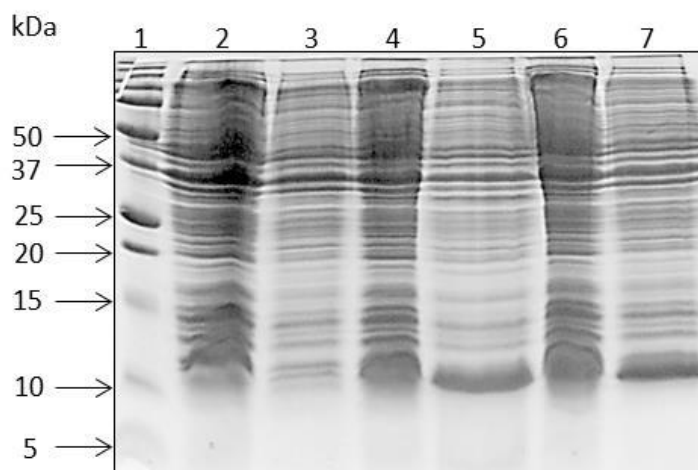
Figure 6.44. Co-expression of TssA.CTD and TssA.NTD proteins. Different plasmid combinations encoding TssA.CTD and TssA.NTD (or as T18- or T25 fusion proteins) were co-transformed into different *E. coli* expression strains and cultured in BHI media supplemented with ampicillin and kanamycin at 37°C and induced for 3 hours with 1mM IPTG. Lane 1, EZ-Run Rec prestained protein ladder; lane 2, uninduced TssA.NTD and TssA.CTD-T25 in BL21(λDE3); lane 3, IPTG induced T25.TssA.NTD and TssA.CTD-T25 in BL21(λDE3); lane 4, uninduced T25.TssA.NTD and TssA.CTD in C43(λDE3); lane 5, IPTG induced T25.TssA.NTD and TssA.CTD in C43(λDE3); lane 6, uninduced TssA.NTD-T25 and TssA.CTD in C43(λDE3); lane 7, IPTG induced TssA.NTD-T25 and TssA.CTD in C43(λDE3).
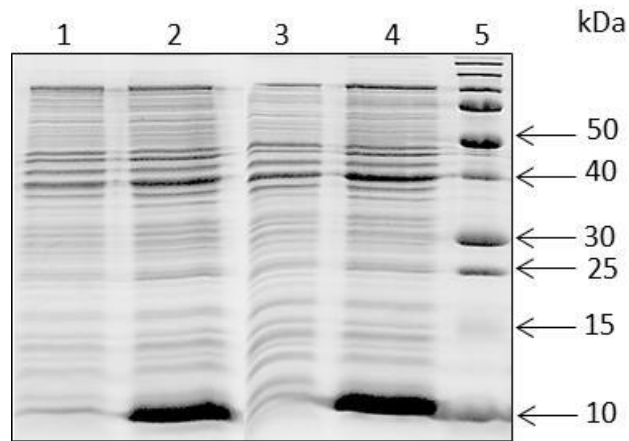
Figure 6.45. Optimization of His₆.TssA.CTD protein expression using different host strains.
E. coli strains C41(λDE3) and C43(λDE3) containing pET14b.TssA.CTD were cultured in BHI
broth at 37°C and induced for 3 hours with 1mM IPTG. Samples were electrophoresed in a 15 %
SDS-PAGE. Lane 1, uninduced pET14b.TssA.CTD in C41(λDE3); lane 2, IPTG induced
TssA.CTD in C41(λDE3); lane 3, uninduced pET14b.TssA.CTD in C43(λDE3); lane 4, IPTG
induced TssA.CTD in C43(λDE3); Lane 5, EZ-Run Rec prestained protein ladder.
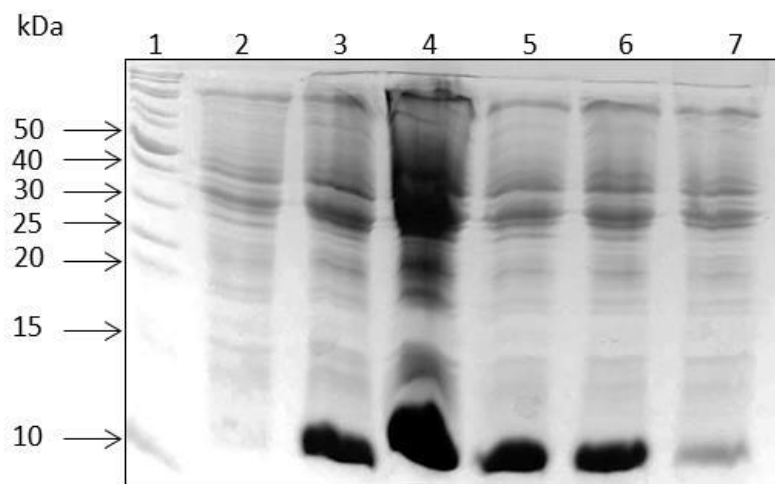
**Figure 6.46. Optimization of His₆.TssA.CTD protein expression in C41(λDE3) using
different IPTG concentrations and temperatures**. *E. coli* C41(λDE3) containing pET14b-
His.TssA.CTD was cultured in 250 ml of BHI media supplemented with ampicillin at 37°C. At
an OD$_{600}$ of 0.5 a preinduction sample was taken and the culture was distributed into 5 flasks
each containing 50 ml. Three cultures were incubated at 37°C and IPTG was added to a final
concentration of 1, 0.5 mM and 10 μM to induce expression of *tssA.CTD*. The other two
cultures were incubated at 25 and 30°C in the presence of 1 mM IPTG. All cultures were
incubated for 3 hours and following sonication samples of cleared cell lysate were
electrophoresed in a 15 % SDS polyacrylamide gel. Lane 1, EZ-Run Rec unstained protein
ladder; lane 2, uninduced C41(λDE3) cells containing pET14b-His6.TssA.CTD grown at 37°C;
lane 3, TssA.CTD protein from the culture incubated at 37°C and induced with 0.5 mM IPTG;
lane 4, TssA.CTD protein from the culture incubated at 37°C and induced with 1 mM IPTG;
lane 5, TssA.CTD protein from the culture incubated at 37°C and induced with 10 μM IPTG;
lane 6, TssA.CTD protein from the culture incubated at 30°C and induced with 1 mM IPTG;
lane 7, TssA.CTD protein from the culture incubated at 25°C and induced with 1 mM IPTG.
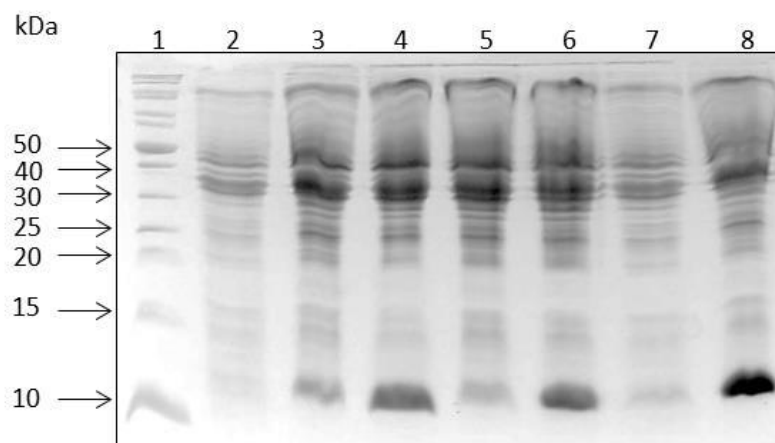
Figure 6.47. Optimization of His$_6$.TssA.CTD protein expression in C43(λDE3) using different IPTG concentrations and temperatures. *E. coli* C43(λDE3) containing pET14b-His.TssA.*CTD* was cultured in 300 ml BHI media supplemented with ampicillin at 37°C. At an OD$_{600}$ 0.5 a preinduction sample was taken and the culture was distributed into 6 flasks each containing 50 ml. Four cultures were incubated at 37°C and IPTG was added to a final concentration of 1.0, 0.5 mM and 0.1 mM, and 10 μM to induce expression of *tssA.CTD*. The other two cultures were incubated at 25 and 30°C in the presence of 1 mM IPTG. All cultures were incubated for 3 hours and following sonication samples of cleared cell lysate were electrophoresed in a 15 % SDS polyacrylamide gel. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, uninduced C43(λDE3) cells containing pET14-His$_6$.*tssA.CTD* grown at 37°C; lane 3, TssA.CTD protein from the culture incubated at 37°C and induced with 0.1 mM IPTG; lane 4, TssA.CTD protein from the culture incubated at 37°C and induced with 0.5 mM IPTG; lane 5, TssA.CTD protein from the culture incubated at 37°C and induced with 10 μM IPTG; lane 6, TssA.CTD protein from the culture incubated at 30°C and induced with 1 mM IPTG; lane 7, TssA.CTD protein from the culture incubated at 25°C and induced with 1 mM IPTG; lane 8, TssA.CTD protein from the culture incubated at 37°C and induced with 1 mM IPTG.
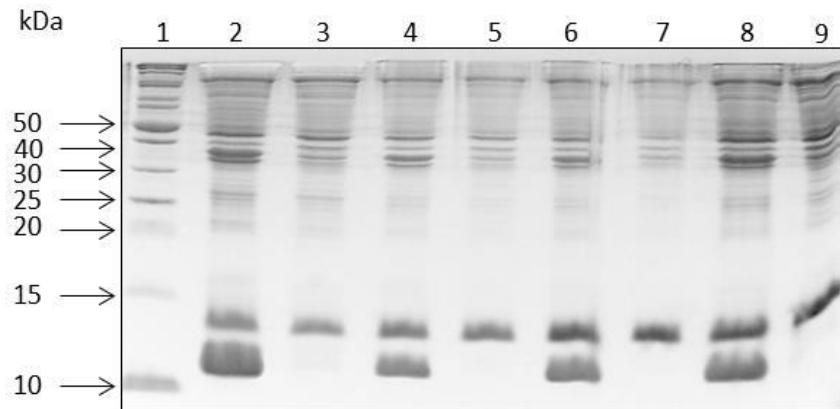
Figure 6.48. Solubility test of His6.TssA.CTD protein in C41(λDE3) at 37°C.
His$_6$.TssA.CTD protein was expressed from pET14b plasmid in *E. coli* strain C41(λDE3). A culture was also included that contained pKNT25.TssA.NTD in addition to pET14b-His$_6$.TssA.CTD to provide a marker for His$_6$.TssA.CTD. Cells were cultured in BHI media at 37°C until OD$_{600}$ 0.5 and induced for 3 hours with different concentrations of IPTG. 15 μl of the protein samples were electrophoresed in a 15% SDS polyacrylamide gel. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, crude cell lysate from cells containing pKNT25.TssA.*NTD* and pET14b-His$_6$.TssA.CTD induced with 1 mM IPTG; lane 3, cleared lysate (soluble fraction) following induction of TssA.NTD-T25 and His$_6$.TssA.CTD production; lane 4, crude cell lysate containing His$_6$.TssA.CTD following induction with 0.1 mM IPTG; lane 5, cleared lysate following induction of His$_6$.TssA.CTD; lane 6, crude cell lysate from cells containing His$_6$.TssA.CTD following induction with 0.5 mM IPTG; lane 7, cleared lysate following induction of His$_6$.TssA.CTD production; lane 8, crude cell lysate from cells containing His$_6$.TssA.CTD following induction with 10 μM IPTG; lane 9, cleared lysate following induction of His$_6$.TssA.CTD.
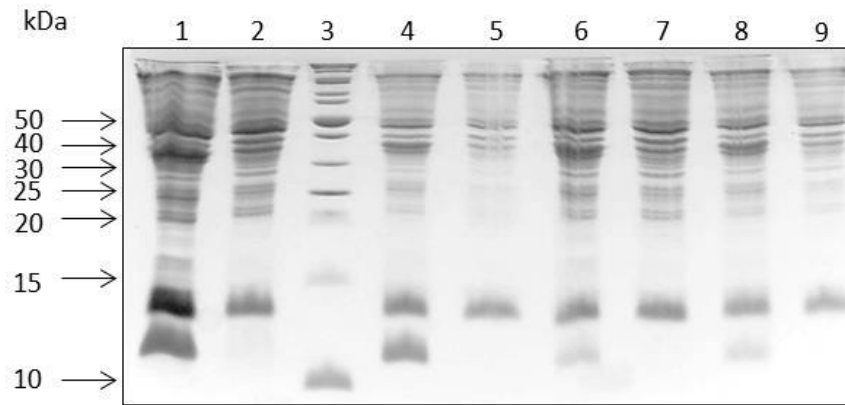
Figure 6.49. Solubility test of His6.TssA.CTD protein in C43(λDE3) at 37°C.
His$_6$.TssA.CTD protein was expressed from pET14b plasmid in *E. coli* strain C43(λDE3). Cells were cultured in BHI media at 37°C until OD$_{600}$ 0.5 and induced for 3 hours with different concentrations of IPTG. 15 μl of the protein samples were electrophoresed in a 15% SDS-PAGE. Lane 1, crude cell lysate from cells containing pET14b-His$_6$.TssA.CTD induced with 1 mM IPTG; lane 2, cleared lysate (soluble fraction) following induction of His$_6$.TssA.CTD production; lane 3, EZ-Run Rec unstained protein ladder; lane 4, crude cell lysate containing His$_6$.TssA.CTD following induction with 0.5 mM IPTG; lane 5, cleared lysate following induction of His$_6$.TssA.CTD; lane 6, crude cell lysate from cells containing His$_6$.TssA.CTD following induction with 0.1 mM IPTG; lane 7, cleared lysate following induction of His$_6$.TssA.CTD production; lane 8, crude cell lysate from cells containing His$_6$.TssA.CTD following induction with 10 μM IPTG; lane 9, cleared lysate following induction of His$_6$.TssA.CTD.
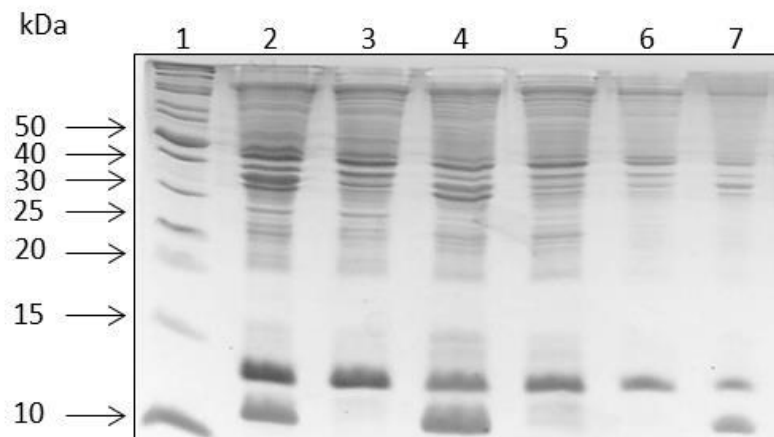
Figure 6.50. Solubility test of His6.TssA.CTD protein in C41(λDE3) and C43(λDE3) at 30°C. His$_6$.TssA.CTD protein was expressed from pET14b plasmid in *E. coli* strain C41(λDE3). A culture was also included that contained pKNT25.TssA.NTD in addition to pET14b-His$_6$.TssA.CTD to provide a marker for His$_6$.TssA.CTD. Cells were cultured in BHI media at 30°C until OD$_{600}$ 0.5 and induced with 1 mM IPTG for 3 hours. 15 μl of the protein samples were electrophoresed in a 15% SDS-PAGE. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, crude cell lysate from C43(λDE3) cells containing pET14b-His$_6$.TssA.CTD; lane 3, cleared lysate (soluble fraction) following induction of His$_6$.TssA.CTD production; lane 4, crude cell lysate from C41(λDE3) cells containing His$_6$.TssA.CTD; lane 5, cleared lysate following induction of His$_6$.TssA.CTD; lane 6, crude cell lysate from C41(λDE3) cells containing pKNT25.TssA.*NTD* and pET14b-His$_6$.TssA.CTD; lane 7, cleared lysate following induction of TssA.NTD-T25 and His$_6$.TssA.CTD production.
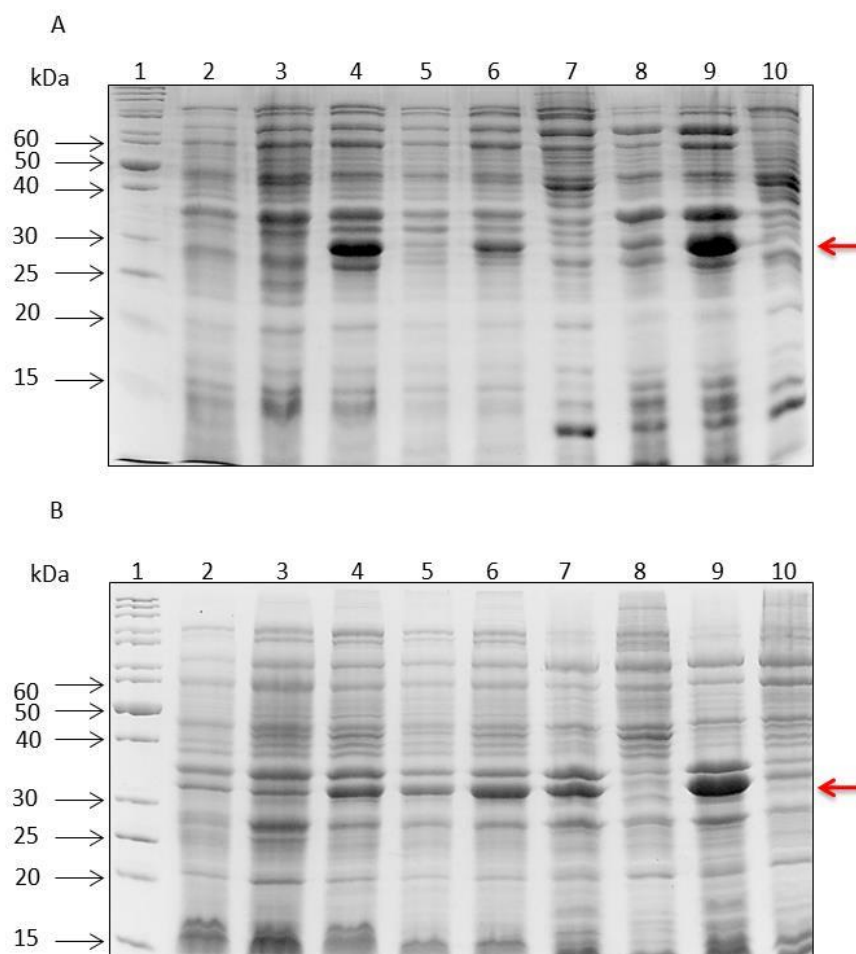
Figure 6.51. Analysis of T18.TssA.CTD and TssA.CTD-T18 fusion protein expression and solubility. *E. coli* C41(λDE3) containing pUT18C.TssA.CTD, pUT18C, pUT18.TssA.CTD or pUT18 were cultured at 37°C in BHI medium supplemented with ampicillin to OD_{600} 0.5. At the desired OD, preinduction samples were taken and 1 mM IPTG was added to the remainder of the culture. Each culture was divided into 2 flasks each containing 50 ml and one was incubated at 25°C and the other at 37°C. All cultures were incubated for 3 hours. Samples of total protein and the soluble fraction were analysed by 12% SDS-PAGE. A. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, uninduced cells containing pUT18C grown at 37°C; lane 3, IPTG induced T18 protein in cell growing at 37°C; lane 4, IPTG induced T18.TssA.CTD protein from the culture incubated at 37°C; lane 5, uninduced cells containing pUT18C.TssA.CTD grown at 37°C; lane 6, IPTG induced T18.TssA.CTD protein from the culture incubated at 25°C; lane 7, crude cell lysate following induction of T18.TssA.CTD at 25°C; lane 8, cleared lysate following induction of T18.TssA.CTD at 25°C; lane 9; crude cell lysate following induction of T18.TssA.CTD at 37°C; lane 10, cleared lysate following induction of T18.TssA.CTD at 37°C. Arrow indicates location of T18-TssA.CTD. B. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, uninduced cells containing pUT18 grown at 37°C; lane 3, IPTG induced T18 protein; lane 4, IPTG induced TssA.CTD-T18 protein from the culture incubated at 37°C; lane 5, uninduced cells containing pUT18.TssA.CTD grown at 37°C; lane 6, IPTG induced TssA.CTD-T18 protein from the culture incubated at 25°C; lane 7, crude cell lysate following induction of TssA.CTD-T18 at 25°C; lane 8, cleared lysate following induction of TssA.CTD-T18 at 25°C; lane 9, crude cell lysate following induction of TssA.CTD-T18 at 37°C; lane 10, cleared lysate following induction of TssA.CTD-T18 at 37°C. Arrow indicates location of TssA.CTD-T18.
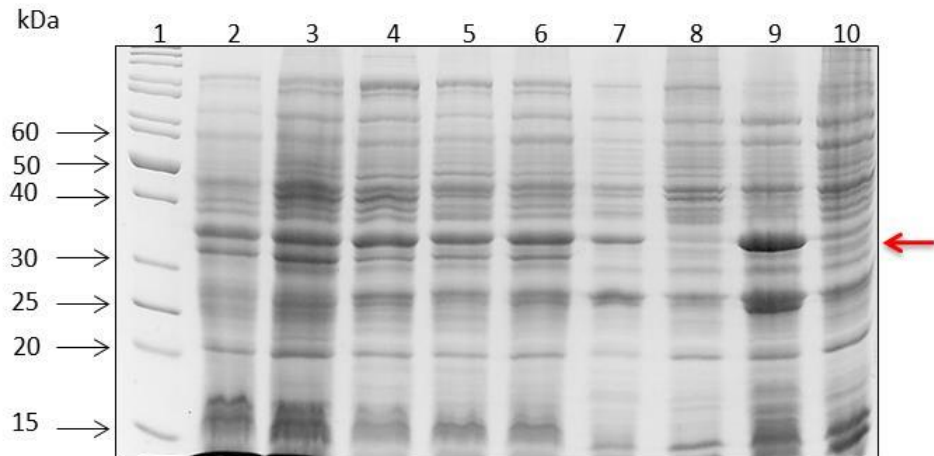
Figure 6.52. Analysis of T25.TssA.CTD fusion protein expression and solubility. *E. coli* BL21(λDE3) containing pKT25.TssA.CTD or pKT25 were cultured at 37°C in BHI medium supplemented with kanamycin to OD_{600} 0.5. At the desired OD, preinduction samples were taken and 1 mM IPTG was added. The culture of C41(λDE3) cells containing pKT25.TssA.CTD was divided into 2 flasks each containing 50 ml and one was incubated at 25°C and the other at 37°C. All cultures were incubated for 3 hours. Samples of total protein and the soluble fraction were analysed by 12% SDS-PAGE. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, uninduced cells containing pKT25 grown at 37°C; lane 3, IPTG induced T25 protein; lane 4, uninduced cells containing pKT25.TssA.CTD grown at 37°C; lane 5, IPTG induced T25.TssA.CTD protein from the culture incubated at 25°C; lane 6, IPTG induced T25.TssA.CTD protein from the culture incubated at 37˚C; lane 7, crude cell lysate from C41(λDE3) cells containing pKT25.TssA.CTD incubated at 25°C; lane 8, cleared lysate following induction of T25.TssA.CTD at 25°C; lane 9, crude cell lysate from C41(λDE3) cells containing pKT25.TssA.CTD incubated at 37°C; lane 10, cleared lysate following induction of T25.TssA.CTD at 37°C. Arrow indicates location of T25-TssA.CTD.

6.12. Purification of His$_6$.TssA.CTD from inclusion bodies

As all attempts to solubilise the His$_6$.TssA.CTD protein by varying the induction conditions were unsuccessful, it was decided to solubilise the protein inclusion bodies in the presence of 8 M urea as described in Section 2.5.14. The solubilised inclusion bodies were then purified by affinity chromatography on a Ni-chelated agarose column in the presence of urea, as described in Section 2.5.3. The result of SDS-PAGE analysis showed that most of the solubilised protein binds to the nickel affinity column and it binds with high affinity as it is not eluted until a concentration of 200 mM imidazole (Figure 6.53.A). The purified protein was subjected to a step wise dialysis into 20 mM Tris-HCl (pH 8.0) buffer containing descending concentrations of urea (6, 4, 2 and 0 M urea) in an attempt to allow the domain to refold. However, a white precipitate was observed in the dialysis bag when the protein was dialysed against the buffer containing 4 M urea. After overnight dialysis in the buffer with 0 M urea, 0.5 ml of the suspension of His$_6$.TssA.CTD was centrifuged for 30 minutes at 13.600 rpm to recover the precipitate. The supernatant was transferred into a clean micro-centrifuge tube and the remaining pellet was re-suspended in 50 μl 2x Laemmli buffer and boiled for 10 minutes. 50 μl of the protein supernatant was boiled for 10 minutes with 50 μl 2x Laemmli buffer. Both samples were analysed by 15% SDS-PAGE, which showed that all the protein had precipitated as no protein was detected in the supernatant (Figure 6.53.B). The suspension of purified His$_6$.TssA.CTD was sent to Yorkshire Bioscience for polyclonal antibody generation in a rabbit.
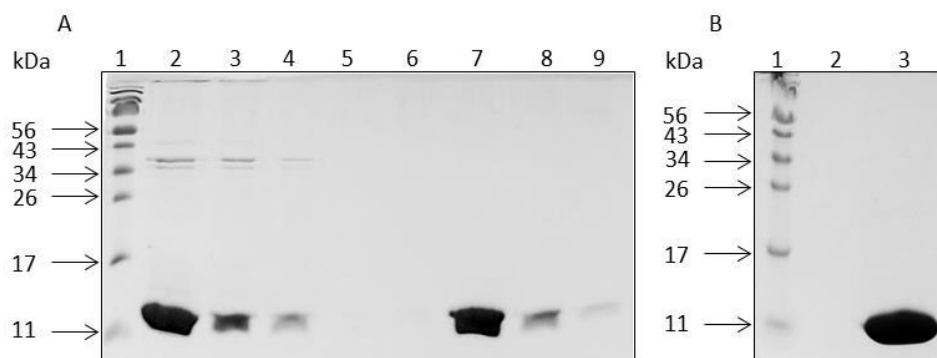
Figure 6.53. IMAC purification of His6.TssA.CTD from inclusion bodies. The insoluble material from *E. coli* C41(λDE3) cells expressing His$_6$.TssA.CTD following induction at 37˚C for 3 hours with 1 mM IPTG was resuspended in 2% sodium deoxycholate solution and after sonication and wash, the inclusion bodies were resuspended in buffer containing 50 mM Tris-HCl (pH 8.0) and 200 mM NaCl, 10 mM imidazole, 10% glycerol and 8 M urea and the cleared lysate was applied to a 1 ml HisTrap HP column (GE Healthcare). His$_6$.TssA.CTD was eluted using a linear gradient of 10-500 mM imidazole. A. IMAC purification of denatured His$_6$.TssA.CTD. Lane 1, EZ-Run Rec prestained protein ladder; lane 2, resuspended insoluble protein from cells overproducing His$_6$.TssA.CTD; lane 3, flow through from the column; lane 4, wash of the nickel affinity column; lanes 5-9 gradient elution of His$_6$.TssA.CTD from nickel affinity column, fractions 1-5. B. Dialysis of IMAC-purified denatured His$_6$.TssA.CTD. Lanes 1, EZ-Run Rec prestained protein ladder; lane 2, sample from the supernatant of dialysed His$_6$.TssA.CTD protein; lane 3, sample of the insoluble material after dialysis of His$_6$.TssA.CTD.

6.13. Cloning and expression of His$_6$.TssA.CTD-S.tag

Although TssA.CTD could be overproduced in *E. coli* as a His-tagged protein, obtaining sufficient soluble protein for analysis was unachievable so far. Fusing the polypeptide to some solubility tags might help to solubilise the protein. Some tags that could improve the solubility of proteins are the S-tag and MBP-tag (maltose binding protein).

Bovine pancreatic ribonuclease A (RNase A) is cleaved by the protease subtilisin into two fragments called RNase A S-peptide (residues 1-20) and S-protein (residues 21-124) (Richards, 1955, Richards and Vithayathil, 1959). The fragments together form an active ribonuclease, but when they are separated they have no enzymatic activity. (Wyckoff *et al*., 1967a, Wyckoff *et al*., 1967b). Only residues 1-15 of S-peptide (known as S-tag) are needed to form a complex with S-peptide with full activity (Potts *et al*., 1963). The amino acid composition of the S-tag (Lys-Glu-Thr-Ala-Ala-Ala-Lys-Phe-Glu-Arg-Gln-His-Met-Asp-Ser) makes this protein extremely soluble with no defined structure. Therefore, the function and the folding of the fusion protein are unlikely to be disturbed by the S-tag.

The pACYCDuet-1 vector encodes the S-tag sequence downstream from the second MCS (Figure 5.1). TssA.CTD was amplified without its stop codon using the primers pACYC-ACTDfor and pACYC-ACTDrev which contain *Bam*HI and *Kpn*I sites, respectively (see appendix 1 for the primer sequences), and ligated between the *Bam*HI and *Kpn*I sites of pACYCDuet-1, creating pACYCDuet-His$_6$.TssA.CTD-S.tag in which the encoded TssA.CTD polypeptide possessed an N-terminal His$_6$-tag and a C-terminal S-tag.

The induction of the T7 promoter on PACYCDuet-His$_6$.TssA.CTD-S.tag in BL21(λDE3) cells growing in BHI medium at 37˚C with 1 mM IPTG did not give rise to an overproduced protein band corresponding to the expected size of His$_6$.TssA.CTD-S.tag (12.92 kDa), (result not shown). Therefore, the expression experiment was repeated in different media (LB and AIM) at different temperatures (30 and 37˚C) in various host strains (BL21(λDE3), C41(λDE3) and C43(λDE3)) in order to optimise the expression level. Unfortunately, no improvement in the amount of the desired protein was obtained under these conditions (result not shown).

316

6.14. Expression and purification of an MBP.TssA.CTD fusion protein

In a further attempt to solubilise TssA.CTD it was fused to maltose-binding protein (MBP), as MBP has an exceptional ability to promote the solubility of proteins to which it is fused. In addition, MBP acts as a natural affinity tag which can be used for the purification of its fusion derivatives (Kapust and Waugh, 1999). It was shown that MBP is superior in enhancing the solubility of a diverse group of otherwise aggregation-prone proteins in comparison to glutathione S-transferase (GST) and thioredoxin (Kapust and Waugh, 1999). It is unclear why MBP is such remarkable solubility enhancer protein. It has been suggested that it might function as a general molecular chaperone for its fusion protein by transitory sequestration of aggregation-susceptible folding intermediates of the fused protein thereby interfering with their self-association (Richarme and Caldas, 1997).

The plasmid pMAL-c5X can be used to express and purify N-terminal MBP-tagged fusion proteins. The vector has a MCS located downstream from the *E. coli* MBP coding sequence (*malE* between the MBP coding sequence and the MCS is a sequence encoding a polyasparagine linker followed by four codons specifying a cleavage site for the protease, factor Xa (Figure 6.54).). Hence the plasmid can be used to express an MBP-fusion protein that can be subsequently proteolytically cleaved into MBP and the target protein (di Guan *et al*., 1988, Maina *et al*., 1988). The MBP coding sequence is under control of the strong "*tac*" promoter and *malE* translation initiation signals; therefore it can express the cloned gene very efficiently. MBP binds tightly to amylose and therefore allows the purification of the expressed MBP-fusion protein by affinity chromatography on amylose resin (Kellermann and Ferenci, 1982).

tssA.CTD was cloned into pMAL-c5X by releasing it from pET14b.TssA.CTD using *Nde*I and *Bam*HI and transferring it into pMAL-c5X digested with the same enzymes. The cloning of *tssA.CTD* into pMAL-c5X vector was confirmed by DNA sequence determination. pMAL-c5X.TssA.CTD and pMAL-c5X were transferred to NEB Express competent cells in order to test the expression and solubility of MBP.TssA.CTD (Section 2.5.2). NEB Express is an *E. coli* B strain similar to BL21(λDE3), except that it lacks the T7 RNA polymerase gene on the chromosome (the *tac* promoter is served by *E. coli* RNA polymerase). SDS-PAGE analysis showed that, after IPTG induction, both proteins (MBP.TssA.CTD fusion protein and MBP) were

overproduced as judged by the appearance of high abundance proteins at the expected molecular weight (51.5 and 42.5 kDa, respectively) (Figure 6.55). After subjecting the crude extract (which contains the total cell protein) to sonication and centrifugation, most of the expressed target proteins were released into the soluble fraction (Figure 6.55).

The soluble protein was then purified by affinity chromatography on an amylose resin column (Section 2.5.5). SDS-PAGE analysis showed that the target protein exhibited high-affinity for the amylose column, as MBP.TssA.CTD fusion protein and MBP were barely detectable in the flow-through and the wash of the amylose affinity column (Figure 6.56 and 6.57).

6.15. Analysis of MBP.TssA.CTD fusion protein
6.15.1. Estimation of MBP.TssA.CTD molecular weight by size exclusion chromatography

SEC was carried out as described before (Section 2.5.6) for further purification of MBP.TssA.CTD and for the estimation of its molecular weight (Figure 6.58 and 6.59). The SEC column was also calibrated with appropriate protein standards (Figure 6.60). As a comparison, MBP was also subjected to SEC. MBP and MBP.TssA.CTD eluted corresponding to proteins with molecular weights of 46.9 kDa and 1.54 MDa, respectively. Thus, whereas MBP exists as a monomer, the MBP.TssA.CTD fusion protein forms a large complex under the conditions used.

6.15.2. Negative stain electron microscopy of MBP.TssA.CTD fusion protein

The obtained result from SEC suggests that MBP.TssA.CTD fusion protein forms a large complex. Therefore, the purified MBP.TssA.CTD fusion protein was analysed by negative stain EM. Strikingly, the result showed that the fusion protein is able to form a ring structure with irregular external edge (Figure 6.61A). This demonstrated that the TssA.CTD was necessary and sufficient for ring formation.
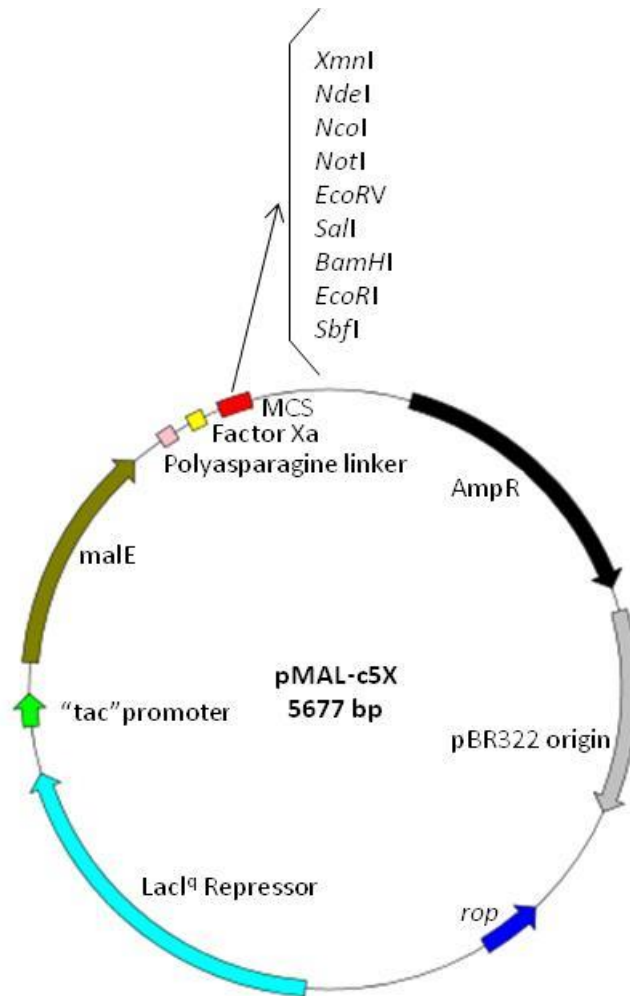
Figure 6.54. Diagrammatic representation of the overexpression vector pMAL-c5X. *rop*, repressor of primer; pBBR322 origin, origin of plasmid replication; AmpR, ampicillin resistance gene shown in black; MCS, multiple cloning site shown in red; factor Xa shown in yellow; polyasparagine linker shown in pink; *malE* gene encoding maltose-binding protein is shown in green; the "*tac*" promoter region is shown in light green; *LacI^q* repressor is shown in turquoise .
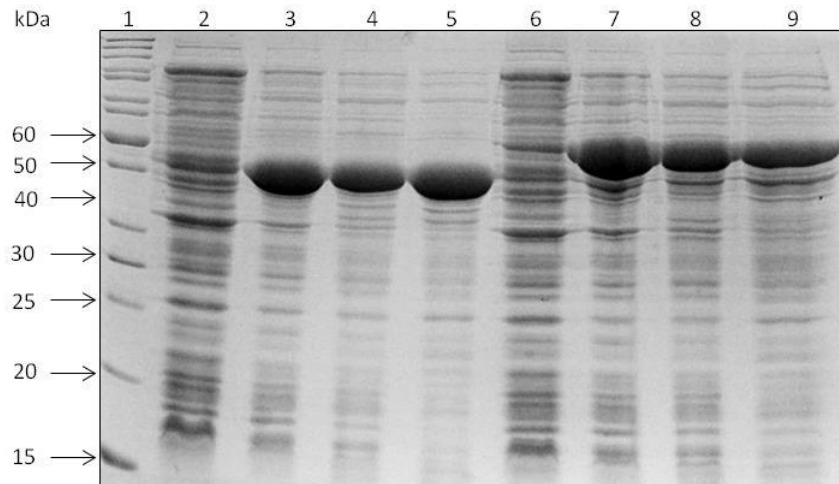
Figure 6.55. Analysis of MBP and MBP.TssA.CTD expression and solubility. *E. coli* NEB Express cells containing pMAL-c5X or pMAL-c5X.TssA.CTD were cultured in BHI medium supplemented with ampicillin to OD_{600} 0.5 at 37°C. When the cells reached the desired OD, a pre-induction sample was taken and 0.3 mM IPTG was added to the reminder of the cultures to induce overexpression of MBP and MBP-TssA.CTD, and the cultures were left to incubate for a further 3 hours. Cell lysates were electrophoresed in a 15 % SDS-PAGE. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, uninduced MBP; lane 3, IPTG induced MBP; lane 4, cell lysate total protein containing MBP; lane 5, soluble fraction containing MBP; lane 6, uninduced MBP.TssA.CTD; lane 7, IPTG induced MBP.TssA.CTD; lane 8, cell lysate total protein containing MBP.TssA.CTD; lane 9, soluble fraction of cell lysate containing soluble MBP.TssA.CTD.
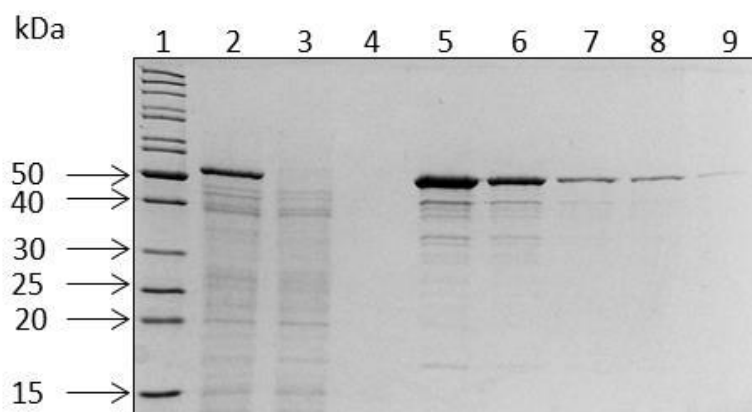
Figure 6.56. Purification of MBP.TssA.CTD fusion protein by amylose affinity chromatography. The soluble fraction from lysed *E. coli* ER2523 (NEB Express) cells containing pMAL-c5X.TssA.CTD was applied to a gravity column packed with 2 ml of amylose resin equilibrated with column buffer (20 mM Tris-HCl (pH 7.4), 200 mM NaCl, 1 mM EDTA, 10 mM β-ME and 1 mM azide). MBP.TssA.CTD fusion protein was eluted using column buffer containing 10 mM maltose. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, soluble fraction from cells containing MBP.TssA.CTD; lane 3, flow-through from the column; lane 4, wash of the column with column buffer; lanes 5-9, 3 ml fractions of MBP.TssA.CTD eluted from the column with column buffer containing 10 mM maltose .
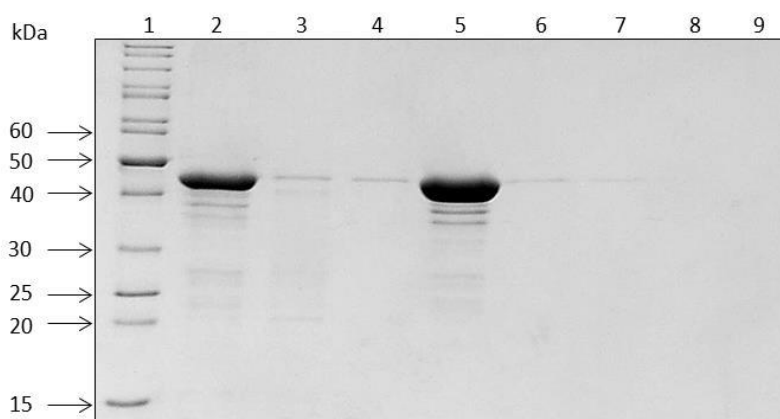
Figure 6.57. Purification of MBP by amylose affinity chromatography. The soluble fraction from lysed *E. coli* ER2523 (NEB Express) cells containing pMAL-c5X was applied to a gravity column packed with 2 ml of amylose resin and calibrated with column buffer. MBP was eluted using the column buffer containing 10 mM maltose. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, total cell protein supernatant containing MBP; lane 3, flow-through from the gravity column; lane 4, wash of the gravity column with column buffer; lanes 5-9, 3 ml fractions eluted from the amylose resin column with column buffer containing 10 mM maltose.
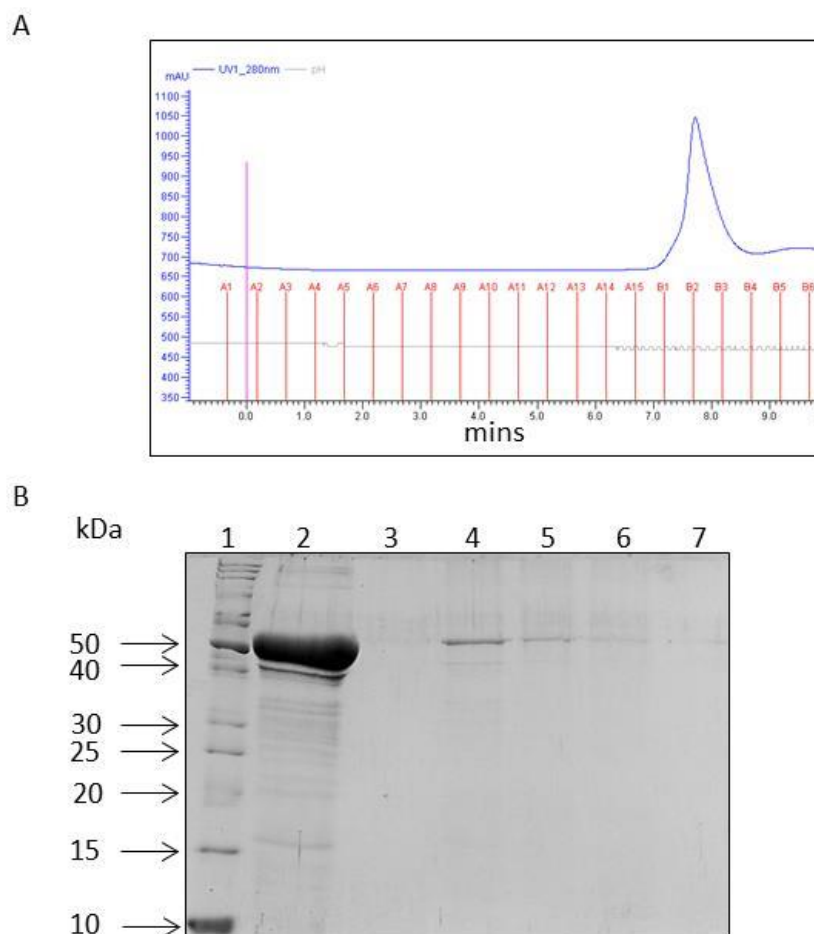
Figure 6.58. Purification of MBP.TssA.CTD fusion protein by gel filtration chromatography. The MBP.TssA.CTD fusion protein eluted from the amylose resin column was concentrated to 4 mg ml^{-1} in column buffer and 0.1 ml was applied to a Superose 12 gel filtration column (GE Healthcare). A. Elution profile of MBP.TssA.CTD fusion protein from gel filtration column monitored at 280 nm. B. SDS-PAGE analysis of MBP.TssA.CTD fusion protein eluted from the gel filtration column. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, concentrated MBP.TssA.CTD (load for the SEC); lanes 3-7, fractions A15 and B1-4 in trace (A).
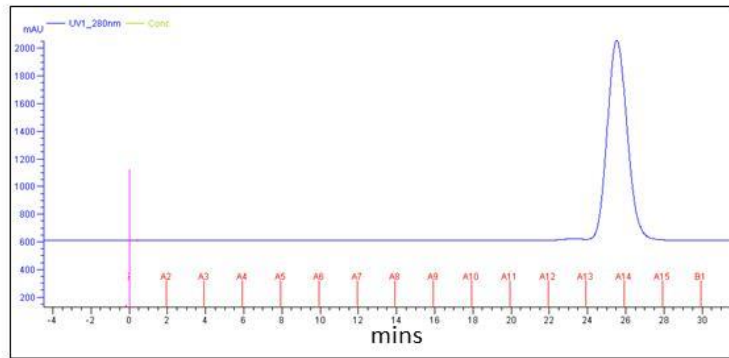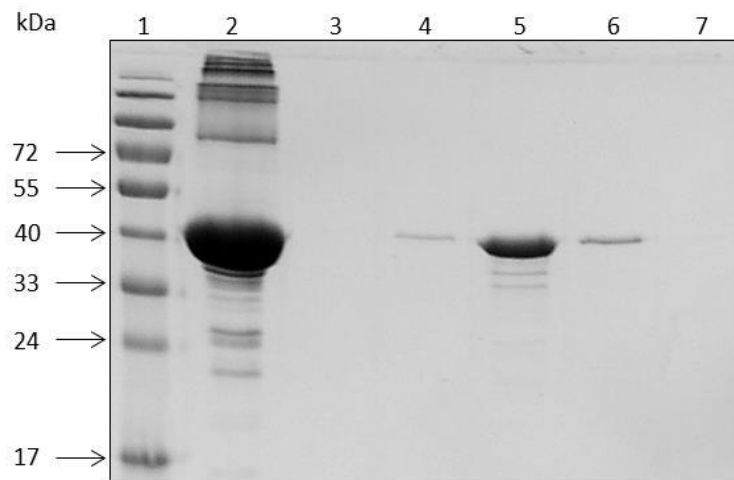
Figure 6.59. Purification of MBP protein by gel filtration chromatography. MBP eluted from the amylose resin column was concentrated to 4 mg ml^{-1} in column buffer and 0.1 ml was applied to a Superose 12 gel filtration column (GE Healthcare). A. Elution profile of MBP from gel filtration column monitored at 280 nm. B. SDS-PAGE analysis of MBP eluted from gel filtration column. Lane 1, EZ-Run Rec prestained protein ladder; lane 2, concentrated MBP (load for the SEC); lanes 3-7, fractions A13-15 in trace (A).
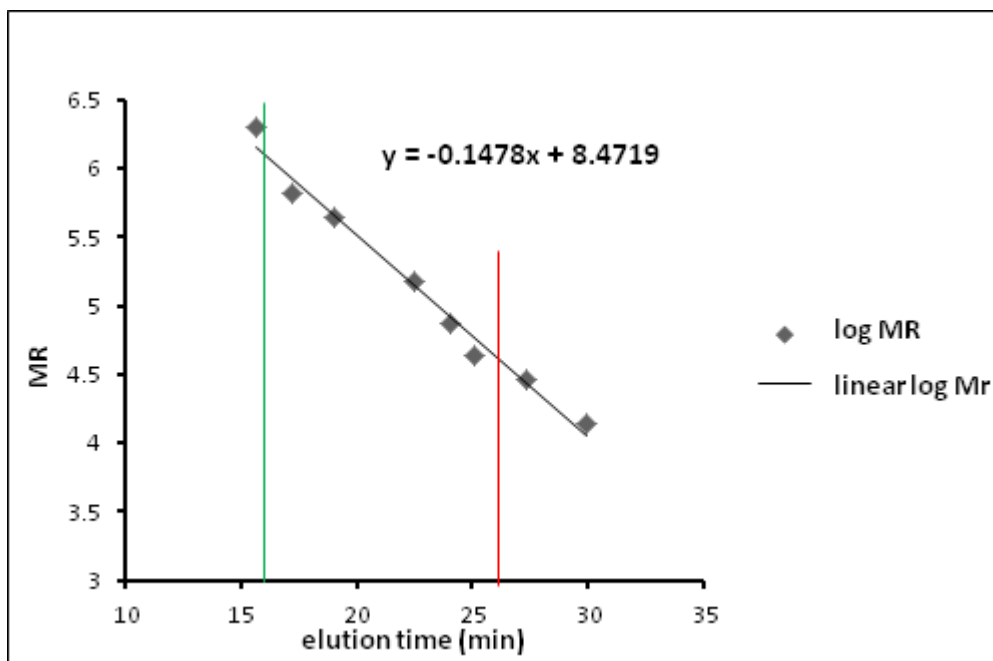
Figure 6.60. Estimation of the molecular weight of MBP and MBP.TssA by gel filtration. The graph showing the elution times of a set of known standarads plotted against the $\log_{10}$ of their molecular weight. A line of the best fit between the points was generated using Microsoft Excel and the equation of the line is shown. The peak elution time of MBP (25.53 minutes) and for MBP.TssA.CTD (15.45 minutes) are shown by the red and green vertical line, respectively. Substituting x for the peak elution time of MBP and MBP.TssA.CTD in the equation gave estimated molecular weights 46.9 kDa and 1.54 MDa, respectively.

6.15.3. Cleavage of MBP.TssA.CTD fusion protein by factor Xa

As TssA.CTD was shown to be capable of forming rings, it was decided to purify TssA.CTD following cleavage of MBP-TssA.CTD. In order to purify TssA.CTD without any fusion tag, factor Xa was used to digest the MBP.TssA.CTD fusion protein, as described in Section 2.5.8. The result of the pilot experiment showed that complete digestion of MBP.TssA.CTD at a concentration of 1 mg ml^{-1} with factor Xa at 200 μg ml^{-1} was obtained after 24 hours incubation at room temperature (result not shown). Therefore, 1.75 ml of MBP.TssA.CTD fusion protein (1 mg ml^{-1}) was mixed with 87.5 μl of factor Xa (200 μg ml^{-1}) and incubated at room temperature for 24 hours. Afterwards, the solution of digested protein was applied to a 4 ml Amicon concentrator tube with 100 kDa cut off and topped up with column buffer. The 4 ml of diluted solution of digested MBP.TssA.CTD fusion protein was concentrated down to 0.5 ml. Fresh 3.5 ml of column buffer was added and concentrated again. The process was repeated 5 times, whereupon the digested fusion protein was analysed by SDS-PAGE. The result showed that the cleavage of the target protein after 24 hours incubation was incomplete (Figure 6.62).

Therefore the partially digested MBP.TssA.CTD fusion protein (0.5 ml) was digested with factor Xa (87.5 μl at 200 μg ml^{-1}) for a further 12 hours incubation at room temperature. To separate factor Xa from the target protein, a 15 ml Amicon concentrator tube with 100 kDa cut off was used as described above. The buffer exchange was performed using 100 ml column buffer to ensure that all the released MBP passed through the filter. The SDS-PAGE analysis of the second digestion showed that the TssA.CTD protein was obtained without any residual MBP (Figure 6.62).

6.15.4. Negative stain electron microscopy of TssA.CTD

A sample from the obtained TssA.CTD protein after cleavage with factor Xa was analysed by negative stain EM. Interestingly, the result of the negative stain EM revealed a ring structure with smooth external edge (Figure 6.61 B).
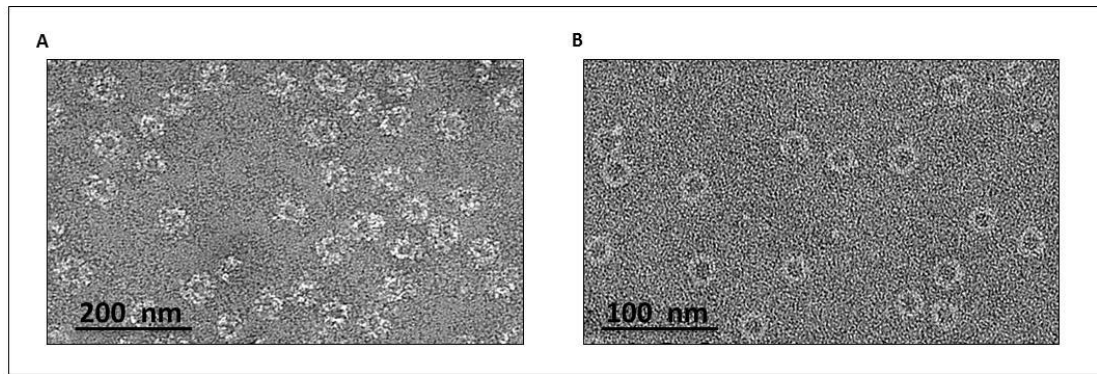
Figure 6.61. Negative stained EM images of MBP.TssA.CTD before and after cleavage of MBP by factor Xa. A. MBP.TssA.CTD showing a ring-like structure with a central pore and irregular projections around the outside. B. TssA.CTD after removal of the MBP by cleavage with factor Xa showing clear ring structure.
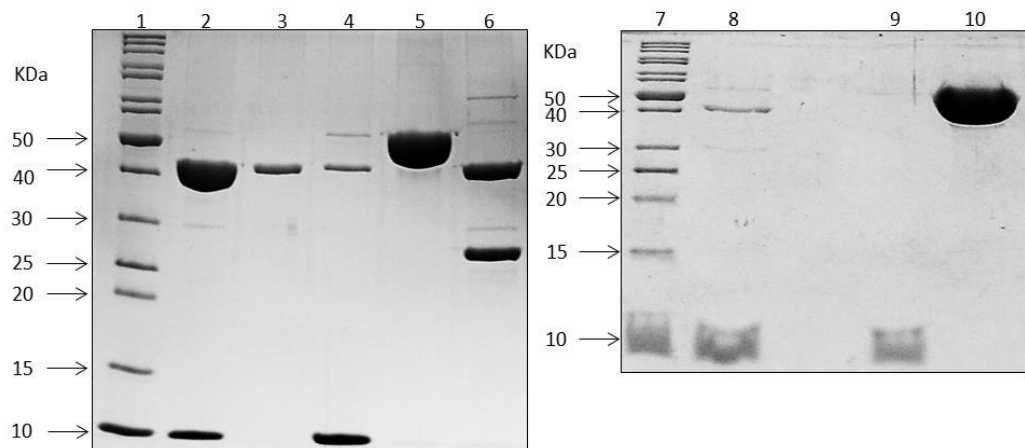
Figure 6.62. Cleavage of the MBP.TssA.CTD fusion protein by factor Xa. The MBP.TssA.CTD fusion protein at 1 mg ml^{-1} was incubated with 200 µg ml^{-1} factor Xa for 24 hours at room temperature, following which a 100 kDa Amicon Ultra-4 centrifugal filter unit was used to separate the digested MBP from TssA.CTD. Lanes 1 and 7, EZ-Run Rec unstained protein ladder; lane 2, digested MBP.TssA.CTD fusion protein after 24 hours; lane 3, flow-through of 100 kDa Amicon concentrator showing proteolytically released MBP; lanes 4 and 8, concentrated filtered digestion showing almost completely digested MBP-TssA.CTD; lane 5, undigested MBP.TssA.CTD fusion protein (negative control); lane 6, digested paramyosin (positive control); lane 9, concentrated filtered digestion showing TssA.CTD after further digestion with factor Xa; lane 10, flow-through of the concentrator showing MBP.

6.16. Expression and purification of TssA^E

The gene encoding TssA^E, the second class of TssA, from the *A. hydrophila* T6SS (Section 1.5.4.2) was amplified and cloned into pACYCDuet-1 using *Bam*HI and *Hin*dIII enzymes by including their sites in the forward and reverse primers, respectively (see appendix 1 for the primer sequences). This places the TssA^E coding sequence downstream of and in frame with the vector His tag coding sequence. Therefore, the target protein, His$_6$- TssA^E, can be purified by a nickel or cobalt chelate column.

Overexpression of the gene encoding the His$_6$.TssA^E protein in *E. coli* BL21(λDE3) was carried out as described in Section 2.5.1.1. TssA^E was successfully overproduced as concluded by the large amount of a protein migrating at the expected size of 52.6 kDa in SDS polyacrylamide gels. However, the solubility test showed that most of the protein was insoluble (Figure 6.63). Therefore, in an attempt to obtain more soluble protein the conditions were modified by growing the culture at a lower temperature (30°C) and induced with lower concentrations of IPTG (100, 10 and 1 μM). Also, the incubation time was extended to overnight with samples taken at 3 and 24 hours. An increase in the amount of His$_6$.TssA^E present in the soluble fraction was obtained with 100 μM IPTG after 3 hours incubation at 30°C (Figure 6.64). However, the protein was not expressed to detectable levels with 10 and 1 μM IPTG after 3 hours (result not shown). The yield of soluble protein from overnight incubation with 100 μM IPTG was lower than that observed for the induction with 100 μM IPTG after 3 hours (result not shown). This might suggest that the extended incubation time may lead to degradation of the desired protein in the soluble fraction or partitioning of the soluble protein into the insoluble fraction.

Purification of His$_6$.TssA^E was carried out using IMAC (Figure 6.65). SDS-PAGE analysis showed that no more than 50% of the soluble target protein bound to the column even after recycling it through the column. However, the His$_6$.TssA^E that did bind was eluted from the column at a relatively high concentration of imidazole (285 mM). As the protein obtained in this way contained many additional proteins (or degradation products), a second purification step was carried out using SEC. Fractions from the nickel column which contained partially purified His$_6$.TssA^E were pooled and concentrated using an Amicon Ultra-4 centrifugal filter unit with a 30 kDa cut off.

During the concentration of the protein sample, a yellowish precipitate began to form in the concentrator, but there was sufficient soluble protein remaining to permit SEC (result not shown). The soluble fraction was used to inject the gel filtration column. However, the elution time of $His_6.TssA^E$ was shorter than that of the largest calibration standard (~2MDa), meaning that molecular weight estimation was inaccurate. Nevertheless, it was concluded that $His_6.TssA^E$ forms a very large complex, as expected. SDS-PAGE analysis confirmed that most of the $His_6.TssA^E$ protein had precipitated during the concentration step.

6.16.1. Transmission electron microscopy of $His_6.TssA^E$

The $His_6.TssA^E$ protein purified by IMAC and SEC was submitted for analysis by EM. Unfortunately, the images showed aggregates indicating the presence of precipitated protein (data not shown). Therefore, a sample of more diluted partially purified protein from the nickel column was submitted for analysis. Again the EM images showed a lot of protein aggregation. However, amorphous structures with a distinct hole in the centre with an approximate diameter of 18 nm were also present (Figure 6.66). This suggests that under appropriate conditions it might be possible to obtain more uniform EM images. Due to time constraints no further protein purification experiments were performed with $His_6.TssA^E$.
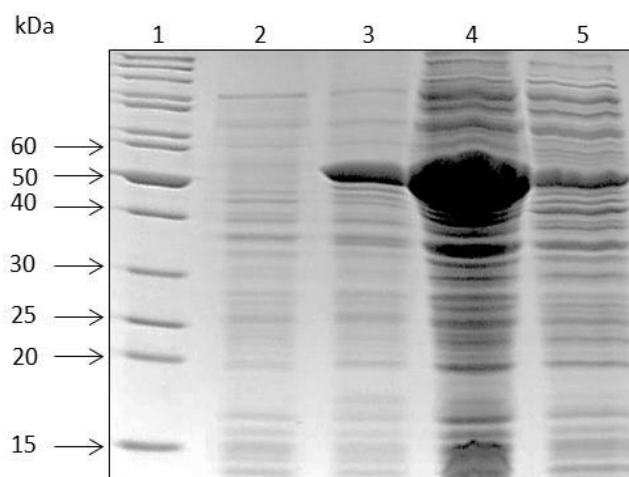
Figure 6.63. SDS-PAGE analysis of His$_6$.TssA^E overexpression at 37°C. *E. coli* BL21(λDE3) was transformed with pACYCDuet-His$_6$.TssA^E and cultured in BHI medium supplemented with chloramphenicol to OD$_{600}$ 0.5 at 37°C. When the cells reached the desired OD, a pre-induction sample was taken and 1 mM IPTG was added to the remainder of the culture which was incubated for a further 3 hours. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, total cell protein containing uninduced pACYCDuet-His$_6$.TssA^E; lane 3, total cell protein following IPTG induction of His$_6$.TssA^E synthesis; lane 4, crude cell extract following induction of His$_6$.TssA^E synthesis; lane 5, soluble fraction following induction of His$_6$.TssA^E synthesis.

Figure 6.64. SDS-PAGE analysis of His$_6$.TssA^E overexpression at 30°C. *E. coli* BL21(λDE3) was transformed with pACYCDuet-His$_6$.TssA^E and cultured in BHI medium supplemented with chloramphenicol to OD$_{600}$ 0.5 at 30°C. When the cells reached the desired OD, a pre-induction sample were taken and 100 μM IPTG was added to the remainder of the culture which was incubated for a further 3 hours. Lane 1, EZ-Run Rec unstained protein ladder; lane 2, total cell protein containing uninduced pACYCDuet-His$_6$.TssA^E; lane 3, total cell protein following IPTG induction of His$_6$.TssA^E synthesis; lane 4, crude cell extract following induction of His$_6$.TssA^E synthesis; lane 5, soluble fraction following induction of His$_6$.TssA^E synthesis.
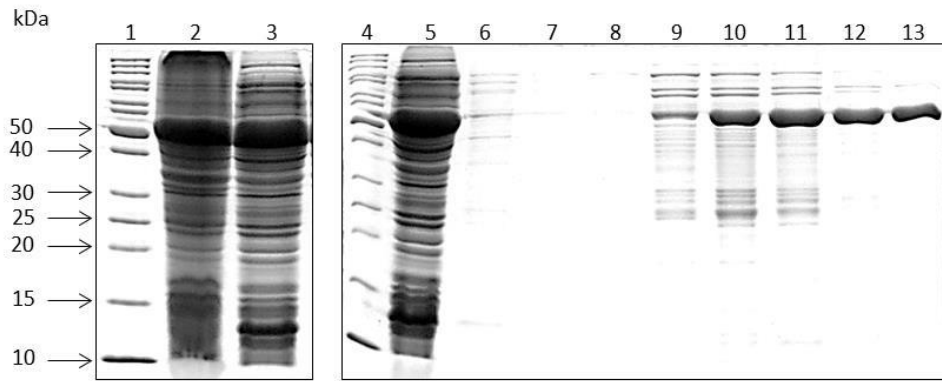
Figure 6.65. Purification of His₆.TssAᴱ by IMAC. The soluble fraction from lysed *E. coli* BL21(λDE3) cells containing IPTG induced pACYCDuet.His₆.TssAᴱ was applied to a 1 ml HisTrap HP column (GE Healthcare) in the lysis buffer. His₆.TssAᴱ was eluted using a linear gradient of 10-500 mM imidazole. Lanes 1 and 4, EZ-Run Rec unstained protein ladder; lane 2, insoluble fraction of total cell protein; lane 3, soluble fraction of cell extract (load for the nickel column); lane 5, flow-through from the nickel affinity column; lane 6, wash of the column; lanes 7-13, gradient elution of His₆.TssAᴱ from nickel column.
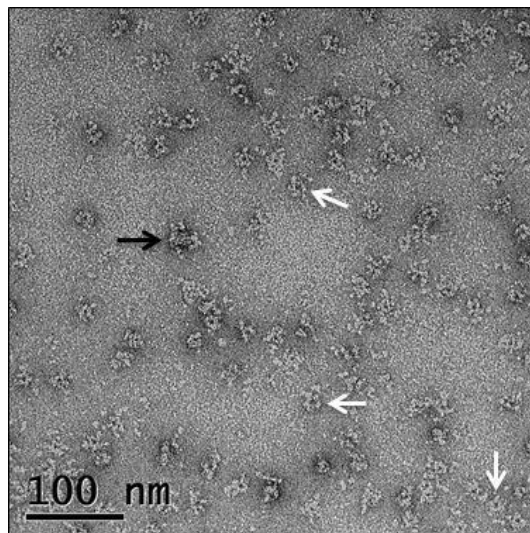
Figure 6.66. Negative stain EM of His₆.TssAᴱ. Amorphous structures with central holes are indicated by white arrows. Protein aggregates are indicated by black arrow.

6.17. Discussion

The T6SS core component, TssA, was successfully overproduced in *E. coli* and found to be highly soluble in an untagged and a variety of His-tagged forms. Figure 6.67 show diagrammatic representation of the different His-tagged forms of TssA constructed in this study. All forms of full-length TssA migrated in SDS-PA gels corresponding to a protein with a monomeric molecular weight of ~ 50 kDa, whereas the predicted mol. wt of untagged TssA is 41.8 kDa, His_6.TssA is 43.9 kDa, linkerHis1.TssA and linkerHis2.TssA are 42.6 kDa, His_{10}.TssA is 44.5 kDa and TssA-His_{10} is 43.2 kDa. The overproduced protein was confirmed as TssA by western blotting using HisProbe-HRP which detects a protein at the same position as the overproduced protein detected on Coomassie blue stained SDS-PA gels (Figure 6.68). Also, the mass spectrometry results of His_6.TssA degradation confirm that the protein was TssA. The apparent higher molecular mass of TssA may be due to the large number of proline residues in TssA (5% of total amino acids, 22 proline in 386 amino acids) which slows the migration of TssA in SDS-PA gels. This phenomenon has been observed previously in a proline-rich protein (Ziemer *et al*., 1982).

When the His_6-tag was fused at the N-terminal end of TssA, the recombinant protein showed a low efficiency of binding to nickel resin. This might suggest that the His-tag was being cleaved or steric hindrance was preventing interaction of the His-tag with the nickel resin. A few experiments were carried out in an attempt to establish the reason behind the reduced affinity of His_6.TssA for nickel resin. The possibility that the His-tag was being degraded or that the nickel-bound protein represented a sub-population of deformed TssA protein with an "unmasked" tag was excluded by the observation that a fraction of His_6.TssA in the flow through was retained on the column following sequential reapplication to fresh nickel beads. It is quite common for the His-tag to be buried inside a folded protein or to be masked by oligomerisation. Therefore, denaturing the protein with a chaotropic agent like urea, might improve the binding affinity of the protein to nickel resin by relaxing the folded protein and making the His-tag more accessible. Unfortunately, in case of His_6.TssA, using 5 and 8 M urea failed to improve retention of the protein on the column. In addition, extending the histidine tag and changing the position of the His-tag from the N-terminal end to the C-terminal end and to the non-conserved linker region of TssA did not improve the binding affinity of TssA protein to nickel resin. Subsequent analysis of TssA protein revealed a likely reason for

the reduced nickel affinity. This showed that TssA forms a very large complex (> 1 MDa). It is known that the capacity of chromatography columns for proteins is reduced as the size of the protein increase. Therefore, the large complex of TssA is likely blocking the binding sites on the metal resin and so only a small percentage of the TssA protein can bind to the resin. This explanation is supported by comparing the behaviour of His_6.TssA and His_{10}.TssA on nickel sepharose. Whereas His_6.TssA eluted from the nickel resin at a relatively low imidazole concentration (~ 90 mM), His_{10}.TssA eluted from the nickel column at a much higher imidazole concentration (~ 300 mM). Therefore, decaHis-tag facilitates tighter binding of TssA to the nickel resin but it does not increase the percentage of bound protein. Further support for this idea was obtained by the observation that His_6.TssA.NTD, which does not oligomerise, bound quantitatively and very tightly to nickel-sepharose.

SEC, which measures the hydrodynamic volume of a protein, gave an estimate of the molecular weight of TssA as > 1 MDa. In one experiment the estimated size of His_6.TssA was 1.9 MDa. However, the fact that the best fit line is skewed towards low molecular weight proteins may have caused an over estimation of the size of TssA. When the best fit line was redrawn through the three points with the highest molecular weights, His_6.TssA has an estimated size of 1.3 MDa (Figure 6.69). As TssA was subsequently shown to form a ring, it will not be possible to obtain an accurate estimation of the molecular weight by SEC.

In order to obtain a more accurate determination of the size of His_6.TssA, the highly purified protein was analysed by $_{sv}$AUC. This gave a value for the sedimentation coefficient of TssA as 19.1 S (Figure 6.19) with a corresponding mol. wt of ~ 850 kDa. The estimated mol. wt by $_{sv}$AUC were quite different to previous estimates (S.Shastri, 2011) and may contain a significant error, because there might be some dynamic dissociation, which may give rise to somewhat lower size estimation for the TssA complex. $_{sv}$AUC analysis of different dilutions of untagged TssA gave sedimentation coefficients of 2S, 4S and 10S which might indicate that TssA is assembled first into small subassembly complexes. A precise calculation of the His_6.TssA molecular weight was obtained from SEC-MALLS analysis, which gave a value of 1,238 kDa (R. Sun and M. Thomas unpublished data). This suggest that 30 monomers of TssA associate together to form the multimeric TssA complex.

The TssA complex was observed to form ring-like structures by EM. These rings have an average number of six projections and an internal diameter ranging from 20 to 25 nm and external diameter varying from 40 to 45 nm. Several protein subunits of the T6SS share structural and functional similarities with phage T4 proteins, (Section 1.5.1). A striking similarity between the EM images of the TssA ring structure and the bacteriophage T4 baseplate was observed (Figure 6.70). Therefore, we propose that TssA forms a major component of the baseplate of T6SS. The baseplate of phage T4 is formed from the central hub (the tail spike complex) surrounded by six wedges. The baseplate wedge consists of 7 different proteins (gp11-gp10-gp7-gp8-gp6-gp53-gp25), some of which are present in multiple copies. There is a strict order of assembly except for gp11, which can assemble at any stage since it is peripheral (it probably associates with gp10). The association of gp53 to the immediate precursor complex stimulates spontaneous association of the wedges to form a six-pointed star shaped structure without gp25 or the hub. Addition of gp25 further facilitates wedge ring formation (Yap *et al*., 2010). The TssA ring with the six projections might correspond to this T4 phage baseplate precursor. Supporting evidence for this hypothesis is the interaction between TssA and TssE demonstrated by the BACTH assay (Section 3.5.3). TssE is similar to gp25, and gp25 is a component of the phage T4 baseplate wedge (Yap *et al*., 2009). The interaction of TssA with the tail spike protein TssI is consistent with TssA being the major component of the baseplate and it can be predicted that TssI inserts into the TssA ring.

Another T6SS core component that shows an interaction with TssA in the BACTH assay is TssD (Hcp). TssD was shown to be structurally similar to the bacteriophage T4 tail tube protein gp19, and the crystal structure of TssD in *P. aeruginosa* revealed hexameric rings which stack on top of each other forming a tubular structure with an internal diameter of 40 Å (4 nm) and external diameter of 90 Å (9 nm) (Mougous *et al*., 2006). According to the current model of T6SS function, the baseplate complex formation triggers TssD tube polymerisation to begin and to pass through the baseplate before traversing the outer membrane of the cell, delivering TssI trimers or other T6SS effectors to the target cell or extracellular milieu (Basler *et al*., 2012). If TssA functions as the baseplate for the T6SS, the internal diameter of the TssA ring should be wide enough to accommodate the TssD hexameric ring. The internal diameter of TssA obtained from EM visualisation (20-25 nm) is more than wide enough to accommodate

the TssD hexameric ring. This suggests that other components of T6SS like TssE might interact with TssA and make the internal diameter smaller.

The data obtained from the analysis of TssA proteolysis support the prediction of the TssA secondary structure by PSIPRED (Figure 1.20A) which proposed that TssA consists of two domains connected by a non-conserved linker region of 55 amino acids.

Unlike full-length TssA, the TssA.NTD monomer migrates at the expected size (30.8 kDa) in SDS-PA gels. This suggests that the anomalous migration of the full-length protein is due to amino acid sequences in the linker and/or CTD. The CD spectroscopic analysis of TssA.NTD showed that 47.6% of the protein is organised in helices whereas the PSIPRED program predicts that 63.6% of TssA.NTD is organised into α-helices (159 amino acids from a total 250) (Figure 6.71). Random coil represents 24.5% of TssA.NTD according to CD spectroscopy and 36.4% according to PSIPRED. No β-strands are predicted by the PSIPRED analysis whereas CD analysis detects 27.9%. No beta-turn can be detected by either method.

Overexpression of TssA.CTD was attempted in *E. coli* BL21(λDE3), but no TssA.CTD was detectable by Coomassie blue staining. However, the recombinant protein was successfully overproduced in C41(λDE3) and C43(λDE3). Unfortunately, despite intensive efforts, it has not been possible to produce the recombinant TssA.CTD protein in a soluble form required to carry out biochemical analyses of structure and function.

TssA.CTD was successful overproduced and solubilised by fusion to MBP. Following purification, the size of the MBP.TssA.CTD fusion protein was estimated to be 1.29 MDa by SEC. Moreover, EM images showed that MBP.TssA.CTD forms a ring structure with irregular external protrusions which are likely to be the MBP component. EM images of TssA.CTD after proteolytic removal of MBP showed rings with smooth edges confirming the suggestion that MBP gives rise to the irregular protrusions. This result demonstrates that TssA.CTD is necessary and sufficient for TssA ring formation and is in agreement with the BACTH assay results which showed that TssA.CTD is responsible for the TssA self-interaction.

His$_6$.TssA^E protein was expressed to high levels in BL21(λDE3) at 37°C, but ~ 90% of the expressed protein was insoluble. The effect of lower temperature and IPTG concentrations were tested in order to obtain more soluble protein. ~ 50% of the expressed protein was soluble when the culture was grown at 30°C and induced with 100 μM IPTG. The reduced rate of protein expression that occurs under these conditions may facilitate correct folding of the protein. Similar to TssA, the capacity of the nickel resin for TssA^E was low. However, the fraction of protein that bound to the resin bound with higher affinity than TssA as it eluted from the column at 285 mM imidazole, suggesting that the His-tag is more exposed. Concentration of the partially purified protein led to protein aggregation. This may have been due to overconcentration of the protein or its limited solubility. Therefore, it is important to determine the limit of TssA^E protein solubility for future purification of this protein.

His$_6$.TssA^E protein was predicted to form a ring structure similar to TssA, as the BACTH assays showed that TssA^E is able to self-interact. Accordingly, visualising partially purified His$_6$.TssA^E protein by negative stain EM revealed an amorphous structure with a hole in the middle. Due to time constraints, no further analysis of TssA^E was performed.
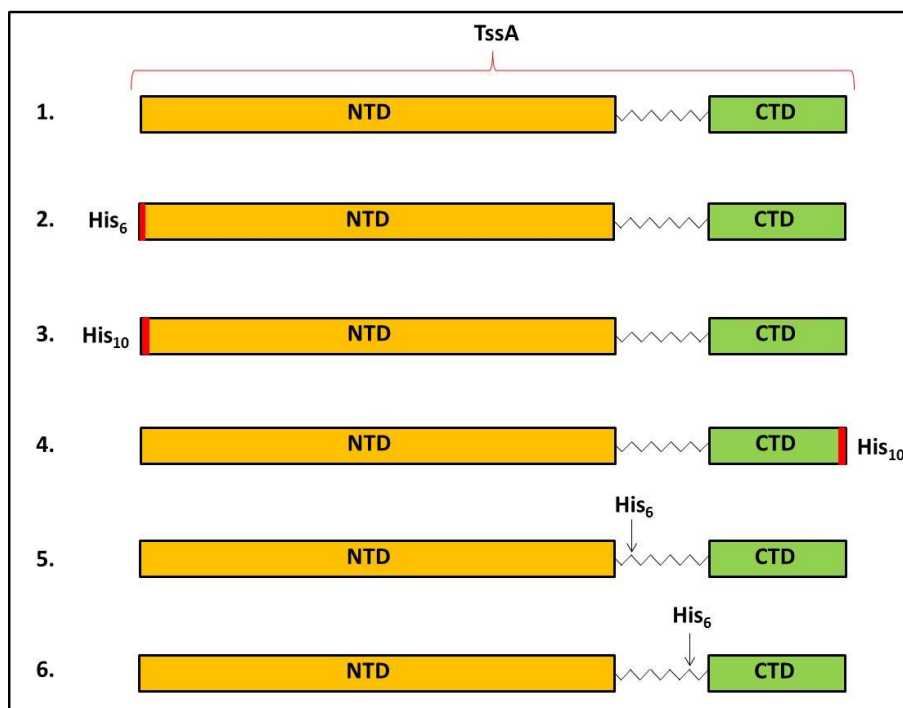
Figure 6.67. Schematic representation of the different forms of TssA used in this work. His$_6$, hexahistidine tag; His$_{10}$, decahistidine tag; NTD, N-terminal domain of TssA protein; CTD, C-terminal domain of TssA protein. 1. untagged TssA; 2. His$_6$.TssA; 3. His$_{10}$.TssA; 4. TssA.His$_{10}$; 5. linkerHis1.TssA; 6. linkerHis2.TssA.
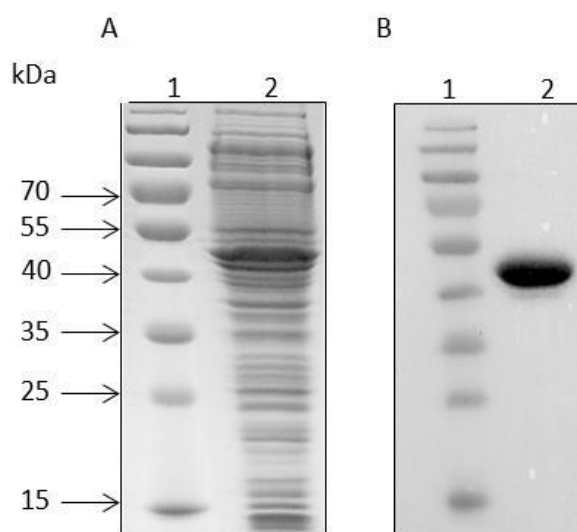
Figure 6.68. Detection of the overproduced His$_6$.TssA by western blot. The soluble fraction from lysed *E. coli* BL21(λDE3) cells containing pET14b-His$_6$.TssA was electrophoresed in a 12 % SDS-PAGE. Then the protein sample was blotted on transfer membrane to be probed by a HisProbe-HRP. A. SDS-PAGE analysis of His$_6$.TssA. Lane 1, EZ-Run Rec pre-stained protein ladder; lane 2, cell lysate following induction of His$_6$.TssA synthesis. B. Western blotting of cell lysate containing His$_6$.TssA and detection with HisProbe-HRP. Lane 1, EZ-Run Rec pre-stained protein ladder; lane 2, cell lysate following induction of His$_6$.TssA synthesis.
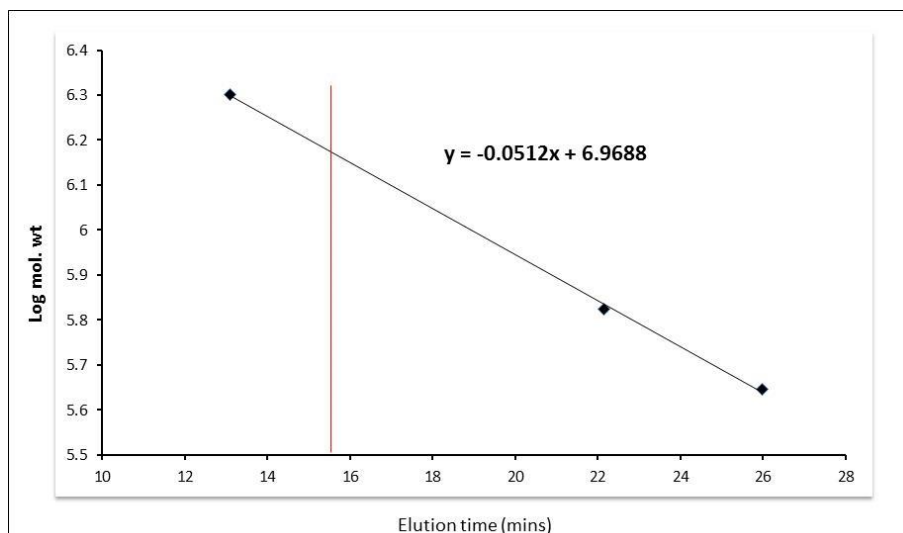
Figure 6.69. Estimation of the size of His₆.TssA. Figure 6.21 was redrawn using only the three proteins with the highest molecular weight. A line of the best fit between these points was generated using Microsoft Excel and the equation of the line is shown. The peak elution time of His₆.TssA (14.78 minutes, red line) was used to estimate the molecular weight of His₆.TssA as 1.3 MDa.
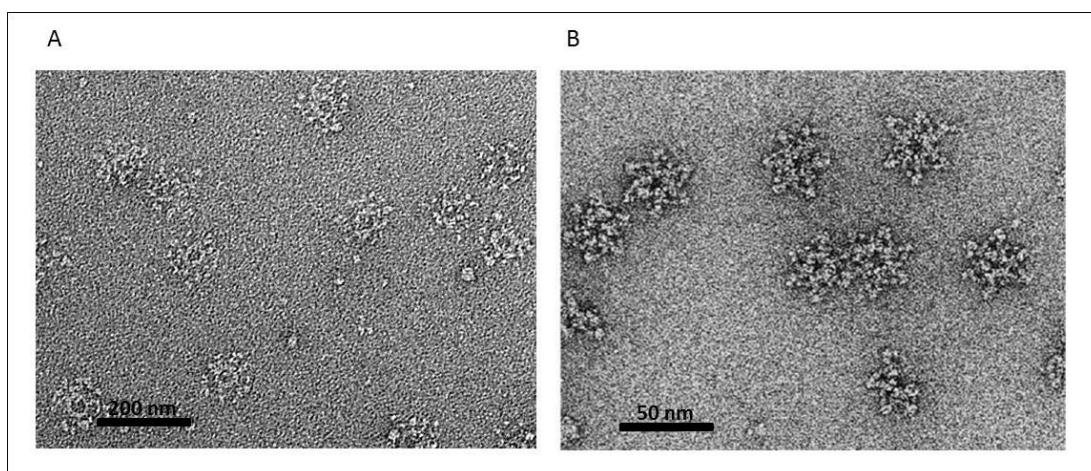
Figure 6.70. Negative stained EM images of TssA and the bacteriophage T4 baseplate. A. EM micrograph of His₆.TssA showing a ring structure with second external ring of projections around it. B. EM micrograph of phage T4 baseplate wedges (gp10-gp7-gp8-gp6-gp53) following hexamerisation into complexes adopting a six-fold star configuration (Yap *et al*., 2010).

```
Conf: ▐███████░█░░░█░██████████████████████████░░░██████▐
Pred:   ───────◖████───────────────────◖████████████───────
Pred: CCCCHHHHHCCCCCCCCCCCCCCCCCCHHHHHHHHHHHCCCCC
  AA: MPINLPELLTPISEASPSGDDLLFSNEFDAIQDARRYDDP
              10        20        30        40

Conf: ▐████████░█████░██████████████████░░░█░████████████▐
Pred:   ──────────────────◖██████████████████◖██████████
Pred: CCCCCCCCCCCCCCCHHHHHHHHHHHHHHHCCCHHHHHHHH
  AA: TLDQGEWVTEIKEADWGFVVDHAGELLRTRTKDLRLAVWL
              50        60        70        80

Conf: ▐████░░░░░░█░██████████████░░░██████████████░████▐
Pred:   ◖████────◖█████████████████───────────────────◖█
Pred: HHHHHCCCCHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCCCCH
  AA: TEALALEDGITGLTEGYALLEGLCREFWDTFHPLPEDDDI
              90        100       110       120

Conf: ▐███████████████████████░░░░░░░░██░████░████████▐
Pred:   ◖██████████████████████───────────◖█████████████
Pred: HHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCHHHHHHHH
  AA: EHRLGNVAWLSGRTAELLRAVPLTDGASNAFSTLDWEVAQ
              130       140       150       160

Conf: ▐████░░░░██░███░░░░░█░░░░█░████████████░░░█████████▐
Pred:   ◖████───────◖██──◖███───◖█████───◖████████────────
Pred: HHHHHHCCCCCCHHHHHCCCCCHHHHHHHHHHHCCCHHHHHH
  AA: HVAQSIKRDPEHADDIARGKPSIEQIDASRRVTSIAFYTA
              170       180       190       200

Conf: ▐████████████████████████████████░█████████████████▐
Pred:   ◖████████████████████████████───◖█████████████████
Pred: HHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCHHHHHHHHHHHH
  AA: LLANLKAFEFALDAFEERLVERAGDSAPSFRQARDAFETV
              210       220       230       240

Conf: ▐██████████░░░█▐
Pred:   ◖███████───────
Pred: HHHHHHHHCC
  AA: YRLAERFARE
              250
```

Legend:

◖██████ - Helix.

⟶ - Strand.

────── - Coil.

Conf: ▌▄▄▐██E Confidence of prediction.

Pred : predicted secondary structure.

AA : target sequence.

Figure 3.71. PSIPRED secondary structure prediction of *B. cenocepacia* TssA.NTD. The 250 amino acid sequence of *B. cenocepacia* TssA.NTD was analysed by PSIPRED v3.3 (Buchan *et al*., 2010). H, α-helix; C, random coil.

Chapter 7. Final discussion and future directions

7.1. Final discussion

This study was aimed at identifying the protein-protein interactions occurring between the T6SS core subunits. Many interactions between the 13 subunits of the *B. cenocepacia* T6SS were identified using the BACTH system (Figures 3.26 and 3.27). All Tss proteins showed self-interaction except TssC. Some of these self-interactions are documented in the literature which include self-association of TssD (Hcp) (Mougous *et al*., 2006), TssE (Leiman *et al*., 2009), TssI (VgrG) (Pukatzki *et al*., 2007), TssJ (Rao *et al*., 2011), TssK (Zoued *et al*., 2013), TssL (Durand *et al*., 2012) and TssM (Felisberto-Rodrigues *et al*., 2011). However, apart from these interactions a number of possible new self-interactions were revealed. TssA, TssB, TssF, TssG and TssH all showed evidence of self-association according to the BACTH results obtained in this study.

BACTH assay results suggested that *B. cenocepacia* TssA self-interacts mainly through its C-terminal domain and forms an oligomer. To determine whether other 'TssA^S class' proteins similarly self-interact through their CTDs, Shastri (2011) used the BACTH system to screen for interactions between the C-terminal region of *P. aeruginosa* TssA orthologue (encoded by PA0082) with itself and with *B. cenocepacia* TssA. The results showed that full length TssA or TssA$_{CTD}$ fusion proteins in combination with PA0082$_{CTD}$ fusion proteins gave rise to a strong Mal$^+$ phenotype. Combinations of PA0082$_{CTD}$ with itself also gave rise to strong Mal$^+$ phenotype suggesting that the protein may also oligomerise. These results indicate that all TssA^S proteins (i.e. SciA class) self-associate. It could also be speculated that the two groups of TssA proteins of the 'enteric class' TssA^E and TalT could also self-associate to form oligomers via their CTDs. In this study we showed that both TssA^E and TalT also self-associate and they do so mainly through their C-terminal regions. In addition to the Tss subunits self-interactions, a large number of known and novel hetero-oligomeric interactions were identified in this study. In particular, TssA was found to interact with a number of Tss proteins, i.e. TssB, TssD, TssE, TssF, TssH, TssI, TssK and TssL. This complex network of TssA interactions with other T6SS components led to the hypothesis that TssA has a key role in the assembly of the T6SS. Moreover, TssA^E interacts with TssA, TssD, TssH, TssI, TssK and TssL. Further analysis of TssA and TssA^E showed that both of them form rings. These results indicate that TssA and TssA^E are structural and functional orthologues.

TssA was demonstrated to be organised into two domains, $TssA_{NTD}$ and $TssA_{CTD}$, linked by non-conserved linker region. Interestingly, $TssA^E$ was predicted to have a similar domain organisation (Section 1.5.4.2). $TssA_{NTD}$ protein analysis suggests that it forms a monomer, whereas $TssA_{CTD}$ forms a large complex which can be visualised by EM as rings. SEC analysis of TssA suggests that the size of TssA complex is > 1 MDa.

The fact that many components of T6SS share structural similarity with bacteriophage T4 and TssA interacts with many bacteriophage-like proteins such as TssD (the tail tube), TssE (the baseplate protein gp25) and TssI (tail spike) led to the hypothesis that TssA might be a structural homologue to a bacteriophage T4 baseplate despite there being no shared amino acid similarity. A striking similarity between the EM images of the TssA ring and the bacteriophage T4 baseplate was observed (Figure 6.70). Therefore, TssA was hypothesed to be the major component of the baseplate of the T6SS. Also, TssA interacts with TssK and TssL which are connected to the IM. Therefore, TssK and TssL might suspend the T6SS baseplate to the IM. Unfortunately, the pull-down experiments to demonstrate these interactions with TssD and TssE were unsuccessful.

TssF is another interesting T6SS protein which showed, in addition to self-association, interactions with TssA, TssB, TssC, TssD, TssE, TssH, TssI, TssK and TssL. Therefore, TssF is predicted to play an essential role in T6SS assembly. TssF is predicted to stabilise the T6SS baseplate or form part of it. The *tssF* gene was found to be organised in tandem with *tssG* gene in almost all T6SS gene clusters which suggests an interaction between the two proteins or a functional relationship. In fact the *tssF* and *tssG* genes are fused in *E. coli* CFT073 such that the C-terminus of TssF is fused to the N-terminus of TssG (Mark Thomas unpublished data). Surprisingly, the BACTH assay did not detect an interaction between TssF and TssG (S.Shastri, 2011).

TssH (ClpV) was also found to interact with many Tss proteins, i.e. TssA, TssB, TssD, TssE, TssF, TssI, TssK and TssL. TssH is closely related to ClpB which forms a hexameric ring and functions as a chaperone that assists in protein folding and prevents aggregation (Lee *et al.*, 2003). As TssH showed evidence of self-association, it is predicted that also TssH form a hexameric ring like ClpB. Also, TssH might have a similar ability to ClpB to recognise many different proteins and therefore it is possible that a number of observed interactions involving TssH are not meaningful in term of

T6SS assembly. Figure 7.1 shows a model for T6SS, which is consistent with the data presented in this thesis and shows the possible role of TssA, TssE, TssF and TssG forming a baseplate on the cytoplasmic face of the inner membrane.

7.2. Future directions

Several unsuccessful attempts were made during this study to provide biochemical evidence for TssA interactions with other Tss proteins. The pull-downs may work better with other subunits that interact more strongly with TssA. Alternative approaches for demonstrating TssA interactions with other Tss subunits can be carried out such as surface plasmon resonance (SPR) which is more sensitive technique and also can measure the binding affinities and cross-linking which can stabilise the complex of interacting proteins. Moreover, it's possible to co-express three or four Tss proteins and trying the pull-downs or purifying subassembly complexes and then doing EM. Future work with TssA also may include β-galactosidase assays for the BACTH results which quantified the efficiency of complementation between the two hybrid proteins, averaging of the EM images which will provide more details of TssA ring and localisation studies which will provide valuable information about the TssA.

Similar analysis could be carried out on $TssA^E$ protein which could include protein purification, estimation of the protein size by SEC, EM and SEC-MALLS to determine the number of $TssA^E$ subunits. Further experiments would include limited proteolysis to confirm the predication of $TssA^E$ domains. Also, purification and analysis of $TssA^E_{CTD}$ would be useful. β-galactosidase assays and pull-down for the BACTH results of $TssA^E$. X-ray crystallography of TssA and $TssA^E$ or their domains would be valuable and possibly allowing the identification of areas of structural homology between the two classes of TssA.
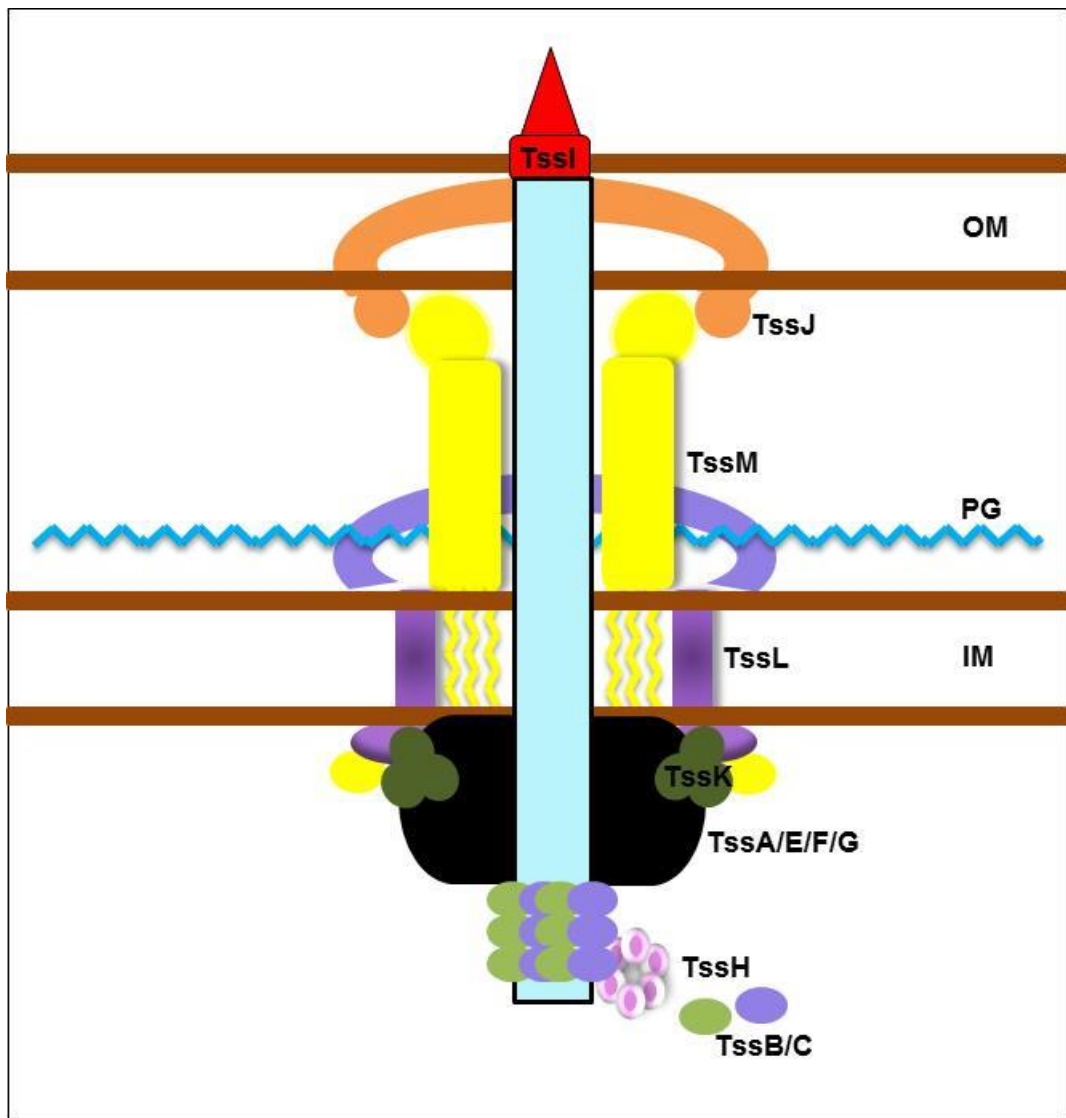
Figure 7.1. Schematic representation of the T6SS assembly. The diagram shows a model for the T6SS of *B. cenocepacia*, which includes TssL with a peptidoglycan-binding domain (PGB). The predicted T6SS baseplate contains TssA/E/F/G and forms a big cytoplasmic structure, which may stabilise T6SS.

8. Appendices

8.1. Appendix 1. Primers used in this study

Primers used to amplify and sequence *tss* genes for the BACTH assay		
Primer name	**Sequence (5´ to 3´)**	**Restriction site**
TssI for	GCGCTCTAGAGATGGATGCGCACAGCATGAT	*Xba*I
C-tssI rev	GCGCCCCGGGAGATCTAGTCTATGCTTCCGACGAAG	*Sma*I/*Bgl*II
N-tssI rev	GCGCCCCGGGAAAGATCTCCTGCTTCCGACGAAGGACTGA	*Sma*I/*Bgl*II
tssJ for(PKT25)	GCGCCTGCAGAACTTCTGGCCGGATGCGCGGC	*Pst*I
For2tssj	GCGCCTGCAGACTTCTGGCCGGATGCGCGGC	*Pst*I
C-tssJ rev	GCGCGGATCCTTAACCGCAGGAGACCGAAG	*Bam*HI
N-tssJ rev	GCGCGGATCCAAACCGCAGGAGACCGAAGACA	*Bam*HI
tssL for	CGCGTCTAGAAATGAGCTACGCGCCTTCCTT	*Xba*I
C-tssL rev	GCGCGGTACCTTACTTCAAGCGGTGCGCGATCT	*Kpn*I
N-tssL rev	GCGCGGTACCACTTCAAGCGGTGCGCGATCT	*Kpn*I
tssM1 for	GCGCTCTAGAAGTGAAGCGTGCGAACCAGCA	*Xba*I
C-tssM1 rev	GCGCGGTACCTTAGAAGAAGGTCGCGTAGCGCA	*Kpn*I
N-tssM1 rev	GCGCGGTACCAAAAGAAGGTCGCGTAGCGCAG	*Kpn*I
tssM2 for	GCGCGGATCCCACGATCGGCAACCAGCAGCT	*Bam*HI
C-tssM2 REV	CGCGGGTACCGTCATTGCGCCCTCTGTGCA	*Kpn*I
N-tssM2 rev	CGCGGGTACCAATTGCGCCCTCTGTGCATTCG	*Kpn*I
tssH for	GCGCTCTAGAA ATGAGCACGCCTCTGAAGAC	*Xba*I
C-tssH rev	GCGCGGTACCTCATTCGACGGTGTACGCGA	*Kpn*I
N-tssH rev	GCGCGGTACCAATTCGACGGTGTACGCGAATT	*Kpn*I
tssK for	GCGCGGATCCAATGAGTTATTCGGCCAAGGT	*Bam*HI
C-tssK rev	GCGCGGTACCTCATGATGTGACCGCGATCA	*Kpn*I
N-tssK rev	GGTACCAATGATGTGACCGCGATCAGTT	*Kpn*I
TSSH-IN For	CAGAACAACCCGATCATGAC	
TSSH-INRev	TTGATCGTGACCATCTTGCG	
TSSI-IN For	ACGCGAACGTTCGACTACAA	
TSSI-IN Rev	CGATATTGACGTGCTGATCC	
M2-IN For	ATCTCGTTCTACCTGTCCCG	
M2-IN Rev	GTCCCGACGAACTTCGCGAT	
pUT18Cfor	ATGTACTGGAAACGGTGCCG	
pUT18Crev	actatgcggcatcagagcag	
pKNT25rev	GTAACCAGCCTGATGCGATT	
M13rev	ggaaacagctATGACCATG	
pKT25for	ATTCGGTGACCGATTACCTG	
pKT25rev	cccagtcacgacgttgtaaa	
pUT18rev	CCACAACAAGTCGATGCGTT	
Rev.gp27	GCGCAGATCTTCACACGTTGTTCACGATGATCC	*Bgl*II

For.gp5	GCGCTCTAGAGTATCGAGTTGCCATCGAAGC	*Xba*I
Rev.gp5	GCGCAGATCTTTAAAACTCCTGGGACCGGTAG	*Bgl*II
CiotAshortCTDfor	GCGCCTGCAGGTCCTCCGGTGACGCAGACGAT	*Pst*I
pUT18CAshortCTDfor	GCGCCTGCAGTCCTCCGGTGACGCAGACGAT	*Pst*I

Primers used to amplify and sequence *AHA* genes for the BACTH assay

Primer name	Sequence (5´ to 3´)	Restriction site
AHA1844fullfor	GCGCTCTAGAGATGAGCTATCAACACCCCT	*Xba*I
AHA1844-Nrev	GCGCGGTACCCATTTCGACAACGGCGCCGC	*Kpn*I
AHA1844-Crev2	GCGCGGTACCGGCATTCATTTCGACAACGG	*Kpn*I
AHA1844for	GCGCTCTAGACGTCGACGTCGACAGTTCCA	*Xba*I
AHA1844NTDrev	GCGCGGTACCTTAGCTGGAACTGTCGACGTCGACG	*Kpn*I
AHA1844NTDrev2	GCGCGGTACCGTGGAACTGTCGACGTCGACG	*Kpn*I
AHA1844CTDrev	GCGCGGTACCTTAGGCCTCGCCAATGCCCGCACT	*Kpn*I
AHA1844for2	GCGCTCTAGAAAGTGCGGGCATTGGCGAGGC	*Xba*I
AHA1846fullfor	GCGCTCTAGAGATGTCACAGAATCAACAAGG	*Xba*I
AHA1846-Crev2	GCGCGGTACCCATAGCAGCATGTTCAGCTA	*Kpn*I
AHA1846-Nrev	GCGCGGTACCCAGCTACCGGAAGCGCCCTGTT	*Kpn*I
AHA1846for	GCGCTCTAGAGGCCCGCTACCAGAACCAGTTG	*Xba*I
AHA1846NTDrev	GCGCGGTACCTTACTGCTGCTCAATCTTGAGCAC	*Kpn*I
AHA1846NTDrev2	GCGCGGTACCTGCTGCTCAATCTTGAGCAC	*Kpn*I
AHA1846fullCTDfor	GCGCTCTAGAGTGGCAGCAACATCAGCAGGA	*Xba*I
AHA1846-Crev2	GCGCGGTACCCATAGCAGCATGTTCAGCTA	*Kpn*I

Primers used in protein overexpression experiments

Primer name	Sequence (5´ to 3´)	Restriction site
TssDforpACYC.NtermVSVGTag.Mod	GCGCCCATGGCATACACTGATATCGAAATGAACCGCCTGGGTAAGATGTTACATATGCACTTGCAGTTTGGTAGT	*Nco*I
TssDforpACYC.NtermVSVGTag.Mod	GCGCCCATGGCATACACTGATATCGAAATGAACCGCCTGGGTAAGATGTTAAATATGTACTTGCAGTTTGGTAGT	*Nco*I
TssDRevpACYC.NtermVSVGTag	GCGCAGATCTTTAGACCGCGTAGGTCTTGT	*Bgl*II
TssEforpET.CtermVSVGTag	GCGCCCATGGCAATGAAACGATTCGAACCC	*Nco*I
TssE.CtermVSVGTagRev	GCGCAGATCTTTACTTACCCAGGCGGTTCATTTCGATATCAGTGTACTGCATGCGTGCCGCGCGTG	*Bgl*II
TssIforpACYC.NtermVSVGTag	GCGCCATATGGCATACACTGATATCGAAATGAACCGCCTGGGTAAGATGGATGCGCACAGCATGAT	*Nde*I
TssIRevpACYC.NtermVSVGTag	GCGCAGATCTAGTCTATGCTTCCGACGAAG	*Bgl*II
TssHforpET.NtermVSVGTag	GCGCCCATGGCATACACTGATATCGAAATGAACCGCCTGGGTAAGATGAGCACGCCTCTGAAGAC	*Nco*I
TssHRevpET.NtermVSVGTag	GCGCAGATCTTCATTCGACGGTGTACGCGA	*Bgl*II
pACYCtssEfor	GCGCTCAT**G**AAACGATTCGAACCCAG	*Bsp*HI
pACYCtssErevHis	GCGCGGATCCTTAATGATGGTGATGGTGATGCTGCATGCGTGCCGCGCGTG	*Bam*HI
pACYCtssEforHis	GCGCGGATCCGATGAAACGATTCGAACCCAG	*Bam*HI
pACYCtssErev	GCGCAAGCTTCGGCTTGATTTACTGCATGC	*Hind*III
pACYC-ACTDfor	GCGCGGATCCACCTCCGGTGACGCAGACGAT	*Bam*HI

348

pACYC-ACTDrev	GCGCGGTACCCGACTGCTCGTCGGGCCGCA	*Kpn*I
pACYC-AHA1844.for	GCGCAGATCTAATGAGCTATCAACACCCCTG	*Bgl*II
pACYC-AHA1844.rev	GCGCAAGCTTTCATTTCGACAACGGCGCCG	*Hind*III
TssIforpET	GCGCCCATGGATGCGCACAGCATGAT	*Nco*I
gp27CtermHisTagrev	GCGCAGATCTTTAATGGTGATGGTGATGATGGTGATGGTG ATGTTCGATGGCAACTCGATA	*Bgl*II
gp5CtermHisTagrev	GCGCAGATCTTTAATGGTGATGGTGATGATGGTGATGGTG ATGAAACTCCTGGGACCGGTAGC	*Bgl*II
pET14b-iotAfor	GCGCCATATGCCGATCAATCTCCCCGA	*Nde*I
pET14b-iotArev	GCGCGGATCCTGCGTTTACGACTGCTCGTC	*Bam*HI
pET14b-ACTDfor	GCGCCATATGCCTCCGGTGACGCAGACGAT	*Nde*I
pET14b-iotAfor2	GCGCAAGCTTGCGCATATGCCGATCAATCTCCCCGA	*Nde*I
tssACtermHisTag	GCGCGGATCCTTAATGGTGATGGTGATGATGGTGATGGTGATGC GACTGCTCGTCGGGCCGCA	*Bam*HI
DecaHisFor	CATGGGCAGCAGCCATCACCATCACCATCATCACCATCACCATA GCAGCGGCCTGGTGCCGCGCGGCAGCCA	*Nde*I
DecaHisRev	TATGGCTGCCGCGCGGCACCAGGCCGCTGCTATGGTGATGGTGA TGATGGTGATGGTGATGGCTGCTGCC	*Xba*I
tssAnonHisFor	GCGCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATAC CATGCCGATCAATCTCCCCGA	*Xba*I
pET14b-TssA-NTDrev	GCGCGGATCCTTAGCCGGTATAGCCCTGTTCGC	*Bam*HI
TssAlinkerHis1for	GCGCATCACCATCACCATCACGCGCAGGCCCAGCCCGA	*Nde*I
TssAlinkerHis1rev	GCGGTGATGGTGATGGTGATGCGCCTGCGTCGGCGCGC	*Bam*HI
TssAlinkerHis2for	CAGCATCACCATCACCATCACCAGCAGACCGCTTCGCGTCCTCC	*Nde*I
TssAlinkerHis2rev	CTGCTGATGGTGATGGTGATGCTGCACGTGAGTCTCCTCGG	*Bam*HI
pACYC-T7-1for	GGATCTCGACGCTCTCCCT	
pACYC-T7-1rev	TTGTACACGGCCGCATAATC	
pACYC-T7-2for	TTGTACACGGCCGCATAATC	
pACYC-T7-2rev	ACCGCTGAGCAATAACTAGC	
T7for	ACGACTCACTATAGGGAGAC	
T7rev	GCTAGTTATTGCTCAGCGGT	

8.2. Appendix 2. Nucleotide sequences of *tss* genes from J2315 and H111 strains of *B. cenocepacia*.

A. *tssH* **gene (J2315)**

```
ATGAGCACGCCTCTGAAGACCCTGATCACGAAACTGAACCCGCTGTGCCGCCACGCGACC
GAGCGGGCGGCGAGCGCGTGCCTGGCGCGCGGCCACTACGAGGTCGATCTGGAACACCTG
TTCCTCGCGTTGCTCGACGAGGCGACGGGCGACCTGCCGCTCGCGTTGCGCGCGAGCCGC
GTCGATCCGCATGCGCTGCACGCCGATCTCGAACGCGAGCTCACGCGCCTGAAGACCGGC
AACACGCGCACGCCGGTGTTCTCCGTGCACCTGATCGCGCTGTTCGAGCAAGCATGGCTG
ATCGCGTCGCTCGATTCGCAGCTCGGCCGCATCCGGTCGGGGCACCTGCTGCTCGCGCTG
CTGACCGCGCCGGATCTCGCGCAGTTCGCGCAACGGATGTCGTCGCGCTTCGCGGAAATG
AACGTGACGGACCTGAAGCACAAGTTCGACGAAATCATGGCCGGTTCGAGCGAAGCCGAG
CCGCGCCAGACGGACGACGAGGGCAGCGACGTCGCGCCCGCCGCCGACGGCATGGCACCC
GCGGCCGGCCCGTCGAAGACGCCCGCGCTCGACACGTACACGACCAACCTCACGCAGCGC
GCCCGCGACGGCAAGATCGATCCGGTGATCGGCCGCGAAGCGGAGATCCGCCAGGCGATC
GACATCCTGATGCGCCGTCGCCAGAACAACCCGATCATGACCGGCGAGGCCGGCGTCGGC
AAAACGGCGGTCGTCGAAGGGCTCGCGCTGCGGATCGCGGCCGACGACGTGCCGCCGCCG
TTGCGCGGCGTCGCGCTGCACGTGCTCGACATGGGGCTGCTGCAGGCCGGCGCGAGCGTG
AAGGGCGAGTTCGAGAACCGGCTGAAGAGCGTGATCGACGAGGTGAAGAAGAGCGCGCAT
CCGATCATCCTGTTCATCGACGAGGCGCACACGATCATCGGCGCGGGCGGCCAGGCCGGC
CAGAACGACGCGGCGAACCTGCTGAAGCCGGCGCTCGCGCGCGGCGAGCTGCGCACGATC
GCCGCGACGACGTGGAGCGAATACAAGAAGTATTTCGAGAAGGATGCGGCGCTCGCGCGG
CGCTTCCAGGTCGTGAAGGTCGAGGAGCCGAGCGAGCCGCTCGCGGCCGCGATGCTGCGC
GGGATGTCGGGCCTGATGGAGAAGCACTTCAACGTGCGGATCCTCGACGATGCGATCACC
GAGGCCGTGCGCCTGTCGCATCGCTACATCAGCGGCCGTCAGCTGCCGGACAAGGCGATC
AGCGTGCTCGACACCGCCTGCGCGAAGGTCGCGCTCGCGCACAGCGCGACACCGGCGGCG
ATCGACGACACCAAGAAGCGCATCGAGCGCATCGATGCGGAAATCGCGTCGCTCGAGCGC
GAGGTGGCGGGCGGCGCGGCGCACGACGAGCGGCTCGGCGAGCTGCGCGGCGCGCGCGAC
ACGGCGCTCGAACAACTGGCAAAGGACGAGGAACGCTACGAAGCCGAGCGCGTGATCGTC
GCCGAGATCACCGAGCTGCGCGATGCGCTCGACCGCGCACGCGGCCCGTCGGAAGACGGC
CAGCCGGTCGACGTGCAGGCCACCCGCGACAAGCTCGCGGAGCGCGTCGCGGCGCTGCAC
GCGCTGCAAGGGAGCGAGCCGATGGTGCCGCTGCAGGTCGACGGCCACGTGGTCGCCGAG
ATCGTCGCCGCGTGGACCGGCATTCCGCTCGGCCGGATGGTGAAGGACGAGATCGACACC
GTGCTGAACCTGCAGCCGCTGCTGAGCGCGCGCGTGATCGGCCAGGACCATGCGCTGGAG
GCGATCGCGCAGCGCGTGCGCACCGCGACCGCGAACCTCGAGGATCCGAACAAGCCGCGC
GGCGTGTTCATGTTCGTCGGGCCGTCGGGCGTCGGCAAGACCGAGACGGCGCTGGCGCTG
GCCGACATCCTGTACGGCGGCGAGCGCAAGATGGTCACGATCAACATGAGCGAGTACCAG
GAAGCGCACAGCGTGTCGGGCCTGAAGGGCTCGCCGCCGGGCTACGTCGGGTACGGCGAA
GGCGGCGTGCTGACCGAGGCCGTGCGCCGCAACCCGTATTCCGTCGTGCTGCTCGACGAA
GTCGAGAAGGCGCACCCCGACGTGCTCGAAATGTTCTTCCAGGTGTTCGACAAGGGCACG
ATGGACGATGCGGAAGGGCGCGAGATCGACTTTCGCAACACGCTGATCATCCTGACCTCG
AACGTCGGCTCGGCCGCGGTGATGCAGGCGTGCCTGAACAAGCCGGCCGAGGAACTGCCC
GACCCGGACACGCTCGCCGAGACGCTGCGTCCGCAGTTGTACAAGACCTTCAAGCCGGCG
TTCCTCGGCCGGATGAAGGTCGTGCCGTACTACCCGATTTCCGACGACGTGCTGGCCGAG
ATCATCGAGCTGAAGCTCGAACGCATCCGCCGCCGGATCGAAGCGAACCACAAGGCCGCG
TTCGAGTGGGACGAGTCGCTCGTCGACGCGGTGCTCGCGCGCTGCACCGAGGTCGATTCG
GGCGCCCGCAACGTCGACCACATCCTGAACGGCACGCTGCTGCCGGAGATCGCGGGCCAC
GTGCTGGGCCGGATCGCCGACGGCGCGGCCATCGCGCGCATCGCGGTGCGCGCGGACGAG
GCCGGCGAATTCGCGTACACCGTCGAATGA
```

B. *tssH* **gene (H111)**

```
ATGAGCACGCCTCTGAAGACCCTGATCACGAAACTGAACCCGCTGTGCCGCCACGCGACC
GAGCGGGCGGCGAGCGCGTGCCTGGCGCGCGGCCACTACGAGGTCGATCTGGAACACCTG
TTCCTCGCGTTGCTCGACGAGGCGACGGGCGACCTGCCGCTCGCGTTGCGCGCGAGCCGC
GTCGATCCGCATGCGCTGCATGCCGATCTCGAACGCGAGCTCACGCGCCTGAAGACCGGC
AACACGCGCACGCCGGTGTTCTCCGTCCATCTGATCGCGCTGTTCGAGCAGGCATGGCTG
ATCGCGTCGCTCGATTCGCAGCTCGGCCGCATCCGGTCGGGGCACCTGCTGCTCGCGCTG
CTGACCGCGCCGGATCTCGCGCAGTTCGCGCAACGGATGTCGTCGCGCTTCGCGGAAATG
AACGTGACGGACCTGAAGCACAAGTTCGACGAAATCATGGCCGGTTCGAGCGAAGCCGAG
CCGCGCCAGACGGACGACGAGGGCAGCGACGTCGCGCCCGCCGCCGACGGCATGGCACCC
GCGGCCGGCCCGTCGAAGACGCCCGCGCTCGACACGTACACGACCAACCTCACGCAGCGC
GCCCGCGACGGCAAGATCGATCCGGTGATCGGCCGCGAAGCGGAGATCCGCCAGGCGATC
GACATCCTGATGCGCCGTCGCCAGAACAACCCGATCATGACCGGCGAGGCCGGCGTCGGC
AAAACGGCGGTCGTCGAAGGGCTCGCGCTGCGGATCGCGGCCGACGACGTGCCGCCGCCG
TTGCGCGGCGTCGCGCTGCACGTGCTCGACATGGGGCTGCTGCAGGCCGGCGCGAGCGTG
AAGGGCGAGTTCGAGAACCGGCTGAAGAGCGTGATCGACGAGGTGAAGAAGAGCGCGCAT
CCGATCATCCTGTTCATCGACGAGGCGCACACGATCATCGGCGCGGGCGGCCAGGCCGGC
CAGAACGACGCGGCGAACCTGCTGAAGCCGGCGCTCGCGCGCGGCGAGCTGCGCACGATC
GCCGCGACGACGTGGAGCGAATACAAGAAGTATTTCGAGAAGGATGCGGCGCTCGCGCGG
CGCTTCCAGGTCGTGAAGGTCGAGGAGCCGAGCGAGCCGCTCGCGGCCGCGATGCTGCGC
GGGATGTCGGGCCTGATGGAGAAGCACTTCAACGTGCGGATCCTCGACGATGCGATCACC
GAGGCCGTGCGCCTGTCGCATCGCTACATCAGCGGCCGTCAGCTGCCGGACAAGGCGATC
AGCGTGCTCGACACCGCATGCGCGAAGGTCGCGCTCGCGCACAGCGCGACACCGGCGGCG
ATCGACGACACCAAGAAGCGCATCGAGCGCATCGATGCGGAAATCGCGTCGCTCGAGCGC
GAGGCGGCGGGCGGCGCGGCGCACGACGAGCGGCTTGGCGAGCTGCGCGGCGCGCGCGAC
ACGGCGCTCGAACAACTGGCAAAGGACGAGGAACGCTACGAAGCCGAGCGCGTGATCGTC
GCCGAGATCACCGAGCTGCGCGATGCGCTCGACCGCGCACGCGGCCCGTCGGAAGACGGC
CAGCCGGTCGACGTGCAGGCCACCCGCGACAAGCTCGCGGAGCGCGTCGCGGCGCTGCAC
GCGCTGCAAGGGGGCGAGCCGATGGTGCCGCTGCAGGTCGACGGCCACGTGGTCGCCGAG
ATCGTCGCCGCGTGGACCGGCATTCCGCTCGGCCGGATGGTGAAGGACGAGATCGACACC
GTGCTGAACCTGCAGCCGCTGCTGACCGCGCGCGTGATCGGCCAGGACCATGCGCTGGAG
GCGATCGCGCAGCGCGTGCGCACCGCGACCGCGAACCTCGAGGATCCGAACAAGCCGCGC
GGCGTGTTCATGTTCGTCGGGCCGTCGGGCGTCGGCAAGACCGAGACGGCGCTGGCGTTG
GCCGACATCCTGTACGGCGGCGAGCGCAAGATGGTCACGATCAACATGAGCGAGTACCAG
GAAGCGCACAGCGTGTCGGGCCTGAAGGGCTCGCCGCCGGGCTACGTCGGTTACGGCGAA
GGCGGCGTGCTGACCGAGGCCGTGCGCCGCAACCCGTATTCCGTCGTGCTGCTCGACGAA
GTCGAGAAGGCGCACCCCGACGTGCTCGAAATGTTCTTCCAGGTGTTCGACAAGGGCACG
ATGGACGATGCGGAAGGGCGCGAGATCGACTTTCGCAACACGCTGATCATCCTGACCTCG
AACGTCGGCTCGGCCGCGGTGATGCAGGCGTGCCTGAACAAGCCGGCCGAGGAACTGCCC
GACCCGGACACGCTCGCCGAGACGCTGCGTCCGCAGTTGTACAAGACCTTCAAGCCGGCG
TTCCTCGGCCGGATGAAGGTCGTGCCGTACTACCCGATTTCCGACGACGTGCTGGCCGAG
ATCATCGAGCTGAAGCTCGAACGCATCCGCCGCCGGATCGAAGCGAACCACAAGGCCGCG
TTCGAGTGGGACGAGTCGCTCGTCGACGCGGTGCTCGCGCGCTGCACCGAGGTCGACTCG
GGCGCCCGCAACGTCGACCACATCCTGAACGGCACGCTGCTGCCGGAGATCGCGGGCCAC
GTGCTGGGCCGGATCGCCGACGGCGCGGCCATCGCGCGCATCGCGGTGCGCGCGGACGAG
GCCGGCGAATTCGCGTACACCGTCGAATGA
```

Nucleotide sequences of *tssH* from J2315 and H111. A. Complete sequence of J2315 *tssH* gene, including stop codon. Position of primers used to amplify t*ssH*, the forward primer highlighted in green colour, C-tssH reverse primer is highlighted in yellow colour and the region highlighted in grey colour is shared between the two reverse primers, N-tssH reverse is highlighted in light blue colour. B. nucleotide sequence of region of H111 *tssH* cloned into BACTH plasmids. Base pair changes compared to J2315 *tssH* are highlighted in pink colour.

A. *tssI* (J2315)

ATGGATGCGCACAGCATGATCGGGGCACTCACGGGAGGGCTGATTCAGCAGGAACGGCTT
CTCAAGCTCGATACACCGCTTGGTAGTAACGTCTTGATCCCGCATCGCGTGTTCGGGCAG
TCTCGTATCGGTCGGGATTACCTGTTCATGGTCGACTGTGTGTCGACATCGAATGATGTC
CAGCTCAAGGCACTGATTTCGCAGCCCATCACGCTGTGGATCCAGCAAACCGACAAGTCC
TATCTTCCTCACCACGGCTACATTCATACCGCGAGAAAACTCGGTGTGGATGGCGGACTC
GCGTGTTATCAGCTCAGTTTCTCGTCATGGCTGCATTTCCTGAAATTCCGACGGGACCAG
CGCCACTGGCAAGACAAGACCGTCGACGCGATCATTGCCGACGTATTCGATGCGCATCCG
CAGGCGCGCGGAATGTATCGCTTTGAATTGTCGCAACCGCTGCCCTCGCGTTCCTATTGC
CGTCAGGACGAAACCGACTGGAATTTCGTGCATCGGCTGCTCGAATCCGAAGGTCTGTAC
GGCGTCTGGAGACAAGCGCAGGACGGCAAATCGCACACGCTCGTGATGACCGACCGTCTG
CAGACGCTCGAGCCGCTTGCGCCGGCAACGGTCGAGTTTTCGCGGGCGGGCGTTCATGGC
GAAGTGGACGCCTTGACGCAGTGGGCCGGTACCCGGACGCTGCAGTCCGTCCTGCTGTCG
ACGCGAACGTTCGACTACAAGAATCCGTCGACCCCTGCCAACCCGAAGGCGACGTCGGTG
CCCACGATGGCGAACCAGGGCGATCTGCCGTCGCAGGCCGAGGTGTATGAATACACGGGC
GGCTATACGTACCCGGAACAGGAGCGCGGCGATCATCTCGGCAAGATCCGCATGGAGGAA
TGGGAGTCGCGCGAGCCAAGCGCTACGATGGTGTCGGCGGGGTGCGCCGCATCGATGCGGGG
CGGCGATTCACGCTGACGGGGCATCCCGAGCACGATCAGGACGCGGCCGATCACCGGGAA
TTCGCGGTGATCGAGGTCGAGTGGATCATCGTGAACAACGTGCCGCTTTCCGGGCACGAA
GCCAACTATCCACACAGCCTCTATTCGGCAGTGAAACAGGCCGATGCGGACGACCCCGAC
AAGCTGTTCACCGTATCGCATGACGACGGCTCGACTGGCTTCTATCGAGTTGCCATCGAA
GCGCAGCGTACGTCCGTGCCATATCGCAGCCCGCTCGAACACCGGAAACCGGAAGCCAAG
CTGGAATCGGCCATCGTGGCCGGCCCCAAAGGGCAGCAGGCGTATACCGACAGCCTGAAC
CGGATCAAGGTGCTGTTCATCTGGGATCGGCACAACGAAGGTGACGAGCGGGCGTCGTGC
TGGGTACGTGTCGCGCAATCCGACACCGGCGACGGGTACGGTGGCGTTCACATGCCGCGC
GCAGGCGAGGAACTGCTGATCGGCCATATCGGGAACGACATCGACCGTCCGATTGCATTG
CACCGCGTCTACAACGGGGCCGCGAAGCCGCGATGGCATTCGAATGCGCTGCTGTCGGGC
TACCGGTCCCAGGAGTTTTCGGGAAGCGGTTACAACCAGATGGTCATGGACGACTCCACG
GGCCAGAACCGCGTGCATCTGTATAGCAGCAGCACGAATGCCGCGTTGCACCTGGGCTAC
CTGATCGATCAGAACGACAACGCACGGGGCGGTTATCTGGGCAAAGGCTTCGACTTGAGC
TCCGAAGCCTATGGCGCGCTGCGGGCCGGGCGAGGACTGTTCGTGACGACGCACCCGGTG
GCCAGCCAGCCACTCGATGCCAGCCTGGCGACGCGTCAATTGAACGGCTCCGCGAATATC
CTCGATGGCTTGTCGCAGGCAAGCGAAACGTCGAAGGCGGAAAGCCTGAAGGATGGCCAC
GACACCTTCAAGGCCTTCGCGGATGCGACGCAGCACAGCGAATCGGGTGCGACGGGCGGT
GGCGGTGTGACGGCCGGCGGAGGGACGGGCGACGCCAGTGCGTTCAAGCAGCCGGTGATC
CTCATGGCTGCGCCTGCCGGCATCGGCCTGTCGACGCAGAAGTCGACGCACGTTGCGGCG
GATCAGCACGTCAATATCGTCAGCGGGCAAAGCACGCACGTGGCGACCGGCAAATCGCTG
GTCGCCAGCATTACGGAGAAACTGAGCCTGTTCGTGCAGAATGCCGGGATGAAGCTGTTT
GCCGCGAAGGGGAAGATCGAGGTTCAGACTCATGCGGACAATGTCGAAGTGACCGCGCAG
AAGTCGTTGTTGCTGGCCTCCGTGACCGAGAAGGTCCAGGCATCGGCGCAGCAGGAAATC
CTGTTGACGTCGGGCGGCGCGTACATTCGGCTCAAAGGGGGCGACATCGAGATTCATGCG
CCAGGAAAGATCGACATCAAGGGCGCGCAGCATGCGTTCAGCGGCCCTGCGCGGATGGAT
GTGACCCATCCCGCGTTCAAGGATTTGCCGACGCGTCGGCTGATGCTGAATACCATGGCC
TCGCCGTCCGCGACCCGTGTCGTTCCGGTCGGCATGCCGTACAAGCTTTACGCGGACGGG
GCGCTGGTCAAGCAAGGGGTGTTCGACAAGACCGGGCAGTTGCCGATCGATCATCAGGTC
ACCACGCAGAAGTACACGCTGGAGATGGCGAACGGGGTGAAGCACGAGATTCCGGTGCCG
GGTGAGTATCGGGATGCCGAGAACGGGAAACTTGCGAATCAGGGGTTTCAGTATCACGAA
GGTGCACCGGACGATGGTGCTGCACCAGATCGGGCCACGCATCGGCAAATCTATAATGAA
CTGCTCAGTCCTTCGTCGGAAGCATAGACT

B. *tssI*(715j)

```
ATGGATGCGCACAGCATGATCGGGGCACTCACGGGAGGGCTGATTCAGCAGGAACGGCTT
CTCAAGCTCGATACACCGCTTGGTAGTAACGTCTTGATCCCGCATCGCGTGTTCGGGCAG
TCTCGTATCGGTCGGGATTACCTGTTCATGGTCGACTGTGTGTCGACATCGAATGATGTC
CAGCTCAAGGCACTGATTTCGCAGCCCATCACGCTGTGGATCCAGCAAACCGACAAGTCC
TATCTTCCTCACCACGGCTACATTCATACCGCGAGAAACTCGGTGTGGATGGCGGACTC
GCGTGTTATCAGCTCAGTTTCTCGTCATGGCTGCATTTCCTGAAATTCCGACGGGACCAG
CGCCACTGGCAAGACAAGACCGTCGACGCGATCATTGCCGACGTATTCGATGCGCATCCG
CAGGCGCGCGGAATGTATCGCTTTGAATTGTCGCAACCGCTGCCCTCGCGTTCCTATTGC
CGTCAGGACGAAACCGACTGGAATTTCGTGCATCGGCTGCTCGAATCCGAAGGTCTGTAC
GGCGTCTGGA{A}ACAAGCGCAGGACGGCAAATCGCACACGCTCGTGATGACCGACCGTCTG
CAGACGCTCGAGCCGCTTGCGCCGGCAACGGTCGAGTTTTCGCGGGCGGGCGTTCATGGC
GAAGTGGACGCCTTGACGCAGTGGGCCGGTACCCGGACGCTGCAGTCCGTCCTGCTGTCG
ACGCGAACGTTCGACTACAAGAATCCGTCGACCCCTGCCAACCCGAAGGCGACGTCGGTG
CCCACGATGGCGAACCAGGGCGATCTGCCGTCGCAGGCCGAGGTGTATGAATACACGGGC
GGCTATCGTACCCGGAACAGGAGCGCGGCGATCATCTCGGCAAGATCCGCATGGAGGAA
TGGGAGTCGCGAGCCAAGCGCTACGATGGTGTCGGCGGGGTGCGCCGCATCGATGCGGGG
CGGCGATTCACGCTGACGGGGCATCCCGAGCACGATCAGGACGCGGCCGATCACCGGGAA
TTCGCGGTGATCGAGGTCGAGTGGATCATCGTGAACAACGTGCCGCTTTCCGGGCACGAA
GCCAACTATCCA{T}ACAGCCTCTATTC{T}GCAGTGAAACAGGCCGATGCGGACGACCCCGAC
AAGCTGTTCACCGTATCGCATGACGACGGCTCGACTGGCTTCTATCGAGTTGCCATCGAA
GCGCAGCGTACGTCCGTGCCATATCGCAGCCCGCTCGAACACCGGAAACCGGAAGCCAAG
CTGGAATCGGCCATCGTGGCCGGCCCCAAAGGGCAGCAGGCGTATACCGAC{G}CCTGAAC
CGGATCAAGGTGCTGTTCATCTGGGATCGGCACAACGAAGGTGACGAGCGGGCGTCGTGC
TGGGTACGTGTCGCGCAATCCGA{T}ACCGGCGACGGGTACGGTGGCGTTCACATGCCGCGC
GCAGGCGAGGAACTGCTGATCGGCCATATCGGGAACGACATCGACCGTCCGATTGC{G}TTG
CACCGCGTCTACAACGGGGCCGCGAAGCCGCGATGGCATTCGAATGCGCTGCTGTCGGGC
TACCGGTCCCAGGAGTTTTCGGGAAGCGGTTACAACCAGATGGTCATGGACGACTCCACG
GGCCAGAACCGCGTGCATCTGTATAGCAGCAGCACGAATGCCGCGTTGCACCTGGGCTAC
CTGATCGATCAGAACGACAACGCACGGGGCGGTTATCTGGGCAAAGCTTCGACTTGAGC
TCCGA{G}GCCTATGGCGCGCTGCGGGCCGGGCGAGGACTGTTCGTGACGACGCACCCGGTG
GCCAGCCAGCCACTCGATGCCAGCCTGGCGACGCGTCAATTGAACGGCTCCGCGAATATC
CTCGATGGCTTGTCGCAGGCAAGCGAAACGTCGAAGGCGGAAAGCCTGAAGGATGGCCAC
GACACCTTCAAGGCCTTCGCGGATGCGACGCAGCACAGCGAATCGGGTGCGACGGCGGT
GGCGGTGTGACGGCCGGCGGAGGGACGGGCGACGCCAGTGCGTTCAAGCAGCCGGTGATC
CTCATGGCTGCGCCTGCCGGCATCGGCCTGTCGACGCAGAAGTCGACGCACGTTGCGGCG
GATCAGCACGTCAATATCGTCAGCGGGCAAAGCACGCACGTGGCGACCGGCAAATCGCTG
GTCGCCAGCATTACGGAGAAACTGAGCCTGTTCGTGCAGAATGCCGGGATGAAGCTGTTT
GCCGCGAAGGGGAAGATCGAGGTTCAGACTCATGCGGACAATGTCGAAGTGACCGCGCAG
AAGTCGTTGTTGCTGGCCTCCGTGACCGAGAAGGTCCAGGC{G}TCGGCGCAGCAGGAAATC
CTGTTGACGTCGGGCGG{T}GCGTACATTCGGCTCAAAGGGGGCGACATCGAGATTCATGCG
CC{G}GGAAAGATCGACATCAAGGGCGCGCAGCATGCGTTCAGCGGCCCTGCGCGGATGGAT
GTGACCCATCCCGCGTTCAAGGATTTGCCGACGCG{G}CGGCTGATGCTGAATACCATGGCC
TCGCCGTC{G}GCGACCCGTGTCGTTCCGGTCGGCATGCCGTACAAGCTTTACGCGGACGGG
GCGCTGGTCAAGCAAGGGGTGTTCGACAAGACCGGGCAGTTGCCGATCGATCATCAGGTC
ACCACGCAGAAGTACACGCTGGAGATGGCGAACGGGGTGAAGCACGAGATTCCGGTGCCG
GGTGAGTATCGGGATGCCGAGAACGGGAAACTTGCGAATCAGGGGTTTCAGTATCACGAA
GGTGCACCGGACGATGGTGCTGCACCAGATCGGGCCACGCATCGGCAAATCTATAATGAA
CTGCTCAGTCCTTCGTCGGAAGCA
```

Nucleotide sequences of *tssI* from J2315 and 715j. A. Complete sequence of J2315 *tssI* gene, including stop codon. Position of primers used to amplify t*ssI*, the forward primer highlighted in green colour, C-tssI reverse primer is highlighted in yellow colour and the region highlighted in grey colour is shared between the two reverse primers, N-tssI reverse is highlighted in light blue colour. B. nucleotide sequence of region of 715j *tssI* cloned into BACTH plasmids. Base pair changes compared to J2315 *tssI* are highlighted in pink colour.

A. *tssJ* gene (J2315)

ATGATTCGCTACGCGCTGCCGCTGCTTGCGTGCGCCCTTCTGGCCGGATGCGCGGCCGCT
CCGCTGCTCGGCTCGGCCGCGAGCGCCGTGATGTCCGCGGCCGGCATCGGCAAGCCCGAG
GTGCCCGACGCGCAGAAGCCGCCGCGCAACATAGGCCTCACGCTGGCCGCCGCACCGAAC
CTGAATGCCGCGACCGACAACAAGCCGCTTGCGCTCGTCGTGCGCCTCTATGCGCTGAAA
GACCCGACCTCGTTCCAGCAAGCGCCGTTCGACGCATTTACCGACCCGGCCAAGGAAAAA
ACCGCACTCGGCGCCGATCTGCTGAACGTGCGTGAAATCACGCTGATTCCGGGCCAGCGC
TATACGGCAACCGAGAAGGTATCGCGCGAAGCGCAGGCATTCGGCATTGTCGCGCTGTTC
CGCGATCCCGCACTGCAACGCTGGAAACTGACGTTCGATCCGGTGAAGTCGGAGAAATCC
GGCATAATCATCGGCCTGCATAATTGCGCGATGACGGTCACCGACGGCACGGTCATCGCA
CCGCAACAGGGTGCGCCTTCGCAACCGCTGAATATGCTGTCTTCGGTCTCCTGCGGTTAA

B. *tssJ* gene (H111)

CTTCTGGCCGGATGCGCGGCCGCTCCGCTGCTCGGCTCGGCCGCGAGCGCCGTGATGTCC
GCGGCCGGCATCGGCAAGCCCGAGGTACCCGACGCGCAGAAGCCGCCGCGCAACATCGGC
CTCACGCTGGCCGCCGCACCGAACCTGAATGCCGCGACCGACAACAAGCCGCTCGCGCTC
GTCGTGCGCCTCTATGCGCTGAAGGACCCGACCTCGTTCCAGCAAGCGCCGTTCGACGCA
TTTACCGACCCGGCCAAGGAAAAAACCGCACTCGGCGCCGATCTGCTGAACGTGCGTGAA
ATCACGCTGATTCCGGGCCAGCGCTATACGGCAACCGAGAAGGTATCGCGCGAAGCGCAG
GCATTCGGCATTGTCGCGCTGTTCCGCGATCCCGCACTGCAACGCTGGAAACTGACGTTC
GATCCGGTGAAGTCGGAGAAATCCGGCATAATCATCGGCCTGCATAATTGCGCGATGACG
GTCACCGACGGCACGGTCATCGCACCGCAACAGGGTGCGCCTTCGCAACCGCTGAATATG
CTGTCTTCGGTCTCCTGCGGTTAA

Nucleotide sequences of *tssJ* from J2315 and H111. A. Complete sequence of J2315 *tssJ* gene, including stop codon. Region encoding predicted N-terminal signal sequence is shown in red font. Position of primers used to amplify t*ssJ*, the forward primer highlighted in green colour, C-tssJ reverse primer is highlighted in yellow colour and the region highlighted in grey colour is shared between the two reverse primers, N-tssJ reverse is highlighted in light blue colour. B. nucleotide sequence of region of H111 *tssJ* cloned into BACTH plasmids. Base pair changes compared to J2315 *tssJ* are highlighted in pink colour.

A. *tssK* gene (J2315)

ATGAGTTATTCGGCCAAGGTGCTGTGGGGGGAAGGGCTGTTCCTCCGGCCTCAACACTTT
CAGCGGCAGGACGCGTACCACGAGGCGCGCCTGTTCGAGTCCATCCAGGCGATCCAGCCA
TACAACTGGGGCGTGCGCTCGGTGCGCATCGACCGCGACGCGCTCGGCAGCAACGTGCTG
CGGGTGGCCGAACTCGCGCTCGTGTTCCCGGACGGCGCGCTGTATGCCGCGCCGCAGGCC
GACGACCTGCCGCCGCCGATCGCACTGGATACGTTGCCCGACGGCATCAACGAATTCGTG
TTCTATCTCGCGCTGCACCCGCTGCGCGAGAACGGCACCAACTACAGCGACGATCCGGCC
GCCGGCTTCATGACGCGGTTCGTCAGCGAGCAGACGAGCGTCGCCGACAACTTCACCGAC
GCCGCCGAAGCCGACATCACGTTCCTGAAGACGCAGGTCAAGCTGATCGCGCACAGCGAG
CCGCGCGACCAGCTGCTGTCGGTGCCGCTCGTACGCGTGCGCCGCACCGCGACGTCCGGC
TTCGAGATCGACGACAGCTTCGTGCCGCCGTGCCTCGCGATCGAAGCGTCGCCGATCCTG
CACCAGCGGCTGCGCCAGCTGGTCGACGCGCTGCAGGCGAAGGTGAACGCGCTGTACGGC
TTTCACCGCGAGCCGTCGAAGAACATCATCGAGTTCCGCTCGGGCGACATCGCGTCGTTC
TGGCTGCTGCACACGGCGAACGCCGCATTCGCGACGCTCGCGCACCTGCACCAGCATGCG
GCGCTGCATCCGGAGCGGCTGTTCCAGGAGCTGCTGCGCCTCGCCGGCCAGTTGATGACG
TTCTCGAAGGGCTATACGCTCGCCGACCTGCCCGTGTACCGCCACGACGATCCGGGCCCG
AGCTTCGCGCGTCTCGACCTGATGCTGCGCGAACTGCTCGACACCGTGATCTCGACGCGC
TATTTCGCGATCACGCTCGACGAGGTGCGCCCGTCGTTCCACCTCGGCCGCCTCGATTCC
GGCAAGATCGACGACAAGACCGAGTTCTACCTCGCAGTGTCCGCGGACATGCCGAGCGTC
GAGCTCGTCGATGCGGTGCCGGCCCGCTTCAAGGTCGGCGCGCCGGACGACGTCGACAAG
CTCGTGCTGTCGGCGATGCCCGGCGTGCGGCTGTCGTACACGCCGCAGGTGCCGCCCGCG
ATTCCGGTGCGCCCCGGCGCGTGCTACTTCGCGCTCGATTCGCGCAGCCCGCTGTACGAG
CGCATGCTGCAGGCCCAGTCCGCGATGATCTACGCGCCGACTGGCATCAATGACCTGAAA
TTCGAACTGATCGCGGTCACATCATGA

B. *tssK* gene (H111)

ATGAGTTATTCGGCCAAGGTGCTGTGGGGGGAAGGGCTGTTCCTCCGGCCTCAACACTTT
CAGCGGCAGGACGCGTACCACGAGGCGCGCCTGTTCGAGTCCATCCAGGCGATCCAGCCA
TACAACTGGGGCGTGCGCTCGGTGCGCATCGACCGCGACGCGCTCGGCAGCAATGTGCTG
CGGGTGGCCGAACTCGCGCTCGTGTTCCCGGACGGCGCGCTGTATGCCGCGCCGCAGGCC
GACGACCTGCCGCCGCCGATCGCACTGGATACGTTGCCCGACGGCATCAACGAATTCGTG
TTCTATCTCGCGCTGCACCCGCTGCGCGAGAACGGCACCAACTACAGCGACGATCCGGCC
GCCGGCTTCATGACGCGGTTCGTCAGCGAGCAGACGAGCGTCGCCGACAACTTCACCGAC
GCCGCCGAAGCCGACATCACGTTCCTGAAGACGCAGGTCAAGCTGATCGCACACAGCGAG
CCGCGCGACCAGCTGCTGTCGGTGCCGCTCGTACGCGTGCGCCGCACCGCGACGTCCGGC
TTCGAGATCGACGACAGCTTCGTGCCGCCGTGCCTCGCGATCGAAGCGTCGCCGATCCTG
CACCAGCGGCTGCGCCAGCTGGTCGACGCGCTGCAGGCGAAGGTGAACGCGCTGTACGGC
TTTCACCGCGAGCCGTCGAAGAACATCATCGAGTTCCGCTCGGGCGACATCGCGTCGTTC
TGGCTGCTGCACACGGCGAACGCCGCATTCGCGACGCTCGCGCACCTGCACCAGCATGCG
GCGCTGCATCCGGAGCGACTGTTCCAGGAGCTGCTGCGCCTCGCCGGCCAGTTGATGACG
TTCTCGAAGGGCTATACGCTCGCCGACCTGCCCGTGTACCGCCACGACGATCCGGGCCCG
AGCTTCGCGCGTCTCGACCTGATGCTGCGCGAACTGCTCGACACCGTGATCTCGACGCGC
TATTTCGCGATCACGCTCGACGAGGTGCGCCCGTCGTTCCACCTCGGCCGCCTCGATTCC
GGCAAGATCGACGACAAGACCGAGTTCTACCTCGCAGTGTCCGCGGACATGCCGAGCGTC
GAGCTCGTCGATGCGGTGCCGGCCCGCTTCAAGGTCGGCGCGCCGGACGACGTCGACAAG
CTCGTGCTGTCGGCAATGCCCGGCGTGCGGCTGTCGTACACGCCGCAGGTGCCGCCCGCG
ATTCCGGTGCGCCCCGGCGCGTGCTACTTCGCGCTCGATTCGCGCAGCCCGCTGTACGAG
CGCATGCTGCAGGCCCAGTCCGCGATGATCTACGCGCCGACTGGCATCAATGACCTGAAA
TTCGAACTGATCGCGGTCACATCATGA

Nucleotide sequences of *tssK* from J2315 and H111. A. Complete sequence of J2315 *tssK* gene, including stop codon. Position of primers used to amplify t*ssK*, the forward primer highlighted in green colour, C-tssK reverse primer is highlighted in yellow colour and the region highlighted in grey colour is shared between the two reverse primers, N-tssK reverse is highlighted in light blue colour. B. nucleotide sequence of region of H111 *tssK* cloned into BACTH plasmids. Base pair changes compared to J2315 *tssK* are highlighted in pink colour.

A. *tssL* gene (J2315)

ATGAGCTACGCGCCTTCCTTGTTCGGCGACAACGCGCCGGCACCGCTGCATACCCCGGCG
TCGACGGACTCCGCGTTCCAGGCCCGTTCGCTGCTCGATTTGCTGTACGACGGGTTCTTC
ATGCTGTTCCTGCTCAAGAACGGACGGGAGCCGGACAGCGCCAGCGAGTTCAGCACGCGG
ATCCAGGAATTCCTGTCGGATTTCGAGCGCGGCGCGAAGAAGCTGAATATCGCCGCCGAG
GACGTGTACGGCGCGAAGTTCGCGTTCTGCGCGGCCATCGACGAGATGGTGCTGTCGTCG
CAGTTCAAGATCCGCGCGGACTGGGAACGCCGGCCGCTGCAGCTCGTGTTGTTCGGCGAG
CAGCTCGCCGGCGAGAAGTTCTTCCAGTATCTCGAGGAAGGTCGTGCGCAGGGTGCGGCA
CGGCTGCAGTCGCTCGAGGTGTTCCATATGTGCCTGCTGCTCGGGTTCCAGGGCAAGTAC
CTGCTCGAAGGGCCGGAAAAGCTCGCGTACCTGACCGCGCGGCTGGGCGACGAGATCGCG
CACATGAAGGGCAAGCGCGCGCCGTTCGCGCCGCACTGGCCGCTGCCGGACCAGATCGCG
CACCGCTTGAAGCGCGAGGTGCCGGTGTGGGCGATCGGGACCGTGTTCGCGCTGGTGGCG
CTGCTCGGGTACCTCGGGCTCAACACGTACCTGAAGGACAAGACGCTGCAAGCGCTCGCG
CCGTATTCGCAGGTGATCAAGGTCGGGCCGGAGTCGGCCAACCTGACGATCTCGTTGCCG
TAA

B. *tssL* gene (H111)

ATGAGCTACGCGCCTTCCTTGTTCGGCGACAACGCGCCGGCACCGCTGCACACCCCGGCG
TCGACGGACTCCGCGTTCCAGGCCCGTTCGCTGCTCGATTTGCTGTACGACGGGTTCTTC
ATGCTGTTCCTGCTCAAGAACGGACGGGAGCCGGACAGCGCCAGCGAGTTCAGCACGCGG
ATCCAGGAATTCCTGTCGGATTTCGAGCGCGGCGCGAAGAAGCTGAATATCGCCGCCGAG
GACGTGTACGGCGCGAAGTTCGCGTTCTGCGCGGCCATCGACGAGATGGTGCTGTCGTCG
CAGTTCAAGATCCGCGCGGACTGGGAACGCCGGCCGCTGCAGCTCGTGTTGTTCGGCGAG
CAGCTCGCCGGCGAGAAGTTCTTCCAGTATCTCGAGGAAGGTCGTGCGCAGGGTGCGGCA
CGGCTGCAGTCGCTCGAGGTGTTCCATATGTGCCTGCTGCTCGGGTTCCAGGGCAAGTAC
CTGCTCGAAGGGCCGGAAAAGCTCGCTTACCTGACCGCGCGGCTGGGCGACGAGATCGCG
CACATGAAGGGCAAGCGCGCGCCGTTCGCGCCGCACTGGCCGCTGCCGGACCAGATCGCG
CACCGCTTGAAG

Nucleotide sequences of *tssL* from J2315 and H111. A. Complete sequence of J2315 *tssL* gene, including stop codon. Region encoding predicted trans-membrane domain is shown in red font. Position of primers used to amplify t*ssL*, the forward primer highlighted in green colour, N-tssL reverse is highlighted in light blue colour. B. nucleotide sequence of region of H111 *tssL* cloned into BACTH plasmids. Base pair changes compared to J2315 *tssL* are highlighted in pink colour.

A. *tssM*_{NTD} gene (J2315)

A. *tssM*_NTD gene (J2315)

```
ATGCAACGCATCCTCAACGTGCTGACTCATCCGCGTACGCTCTCGATCGTCGGGATCGTC
GCGCTAGCGGCCATTCTGTTCATCGTCGCCGACATGTTGCAGCTGCCGCTCCTGTGGGCG
GCGACCGCGTTCGCGGCGATCCTCGCGCTGTGGCTCGTGGTCGCGCTGTGGCGCCGCTGG
CGCGTGAAGCGTGCGAACCAGCAGCTCGGCCAGGTGCTGGAAGAGCAGGCGGAAACCGGC
AAAATCGCCGCGCCGGCCGCTGCCGCGCTCGCACCGGACGCGAAGACGGCCGACCTCGAC
GTGCTGCGCACGCGCCTGTCGGATGCGGTGAAGACGATCAAGACGTCGAAGATCGGCCAG
GTGTCGGGCGGCTCCGCGCTGTACGAACTGCCGTGGTACATCGTGATCGGCAACCCGGCC
GCGGGCAAGAGCAGCGCCGTGCTCAATTCCGGCCTGCAGTTCCCGTTCGCGGACAAGAAC
AGCGCCGTGATCCACGGCATCGGCGGCACGCGTAACTGCGACTGGTTCTTCACGACCGAA
GGGATCCTGCTCGACACGGCCGGCCGCTATTCGGTGCACGAGGAAGACCGCAGCGAATGG
CTCGGCTTCCTCGGGCTGCTCAAGCGCTACCGTCCGAAGGCGCCGATCAACGGCATCATC
GTCACGGCCAGCATCGCCGAACTCACCGGCAATCGCCCCGAATTCGCGATCAACCTCGCG
AAGAACCTCCGTCAGCGCGTTCAGGAGCTGACGGAAAAGCTGGAAGTGTTCGCGCCGGTC
TACGTGATGTTCACGAAGGCCGACCTGATCACCGGCTTCACCGAATTCTTCAGCAGCAGC
GACAAGCACGAGTACGACCGCGTGTGGGGCGCCACCCTGCCCTACGAGCCCGACGACAAG
CGCGACGTCGTCGCGCAGTTCGACACGCACTTCGAGGAACTCTACGAAGGGCTGAAGGAA
ATCAGCGTCGCGCAGCTGTCGCTGTCGCGCGGCAACCAGCTGTCGCCGGGCCAGTTGAGC
TTCCCGCTCGAATTCTCGACGATCAAGCCGTCGCTGCGCGCGTTCCTCGCGACGCTGTTC
GAGAACAACCCGTTCCAGTACAAGCCGATCTTCCGCGGCTTCTACTTCACCAGCGCGCTG
CAGGAAGGCGAAACCAACAGCGCGGCCGCGCAGCGCATCGCGCACCGCTTCGGCCTCGAC
GGGAATGCGCTGCCGAAACCGCACAGCGCGTTCTCGAAGAACGGCTTCTTCCTGCGCGAC
CTGTTCTCGAAGGTGATCTTCGCGGACCGCCAGACGGTGCGCCAGTTCGCGAGCCCGACC
AAGACGCGGCTGCGCTACGCGACCTTCTT
```

B. *tssM*_NTD gene (H111)

```
GTGAAGCGTGCGAACCAGCAGCTCGGCCAGGTGCTGGAAGAGCAGGCGGAAACCGGCAAG
ATCGCCGCGCCGGCCGCTGCCGCGCTCGCACCGGACGCGAAGACCGCCGACCTCGACGTG
CTGCGCACGCGCCTGTCGGATGCGGTGAAGACGATCAAGACGTCGAAGATCGGCCAGGTG
TCGGGCGGCTCCGCGCTGTACGAACTGCCGTGGTACATCGTGATCGGCAACCCGGCCGCG
GGCAAGAGCAGCGCCGTGCTCAATTCCGGTCTGCAGTTCCCGTTCGCGGACAAGAACAGC
GCCGTGATCCACGGCATCGGCGGCACGCGCAACTGCGACTGGTTCTTCACGACCGAAGGG
ATCCTGCTCGACACGGCCGGCCGCTATTCGGTGCACGAGGAAGACCGCAGCGAGTGGCTC
GGCTTCCTCGGGCTGCTCAAGCGCTACCGTCCGAAGGCGCCGATCAACGGCATCATCGTC
ACGGCCAGCATCGCCGAACTCACCGGCAACCGCCCCGAATTCGCGATCAACCTCGCGAAG
AACCTCCGTCAGCGCGTTCAGGAGCTGACGGAAAAGCTGGAAGTGTTCGCGCCGGTCTAC
GTGATGTTCACGAAGGCCGACCTGATCACCGGCTTCACCGAATTCTTCAGCAGCAGCGAC
AAGCACGAGTACGACCGCGTGTGGGGCGCCACCCTGCCCTACGAGCCCGACGACAAGCGC
GACGTCGTCGCGCAGTTCGACACGCACTTCGAGGAACTCTACGAAGGGCTGAAGGAAATC
AGCGTCGCGCAGCTGTCGCTGTCGCGCGGCAACCAGCTGTCGCCGGGCCAGTTGAGCTTC
CCGCTCGAATTCTCGACGATCAAGCCGTCGCTGCGCGCGTTCCTCGCGACGCTGTTCGAG
AACAACCCGTTCCAGTATAAGCCGATCTTCCGCGGCTTCTACTTCACCAGCGCGCTGCAG
GAAGGCGAAACCAACAGCGCGGCCGCGCAGCGCATCGCGCACCGCTTCGGCCTCGACGGG
AATGCGCTGCCGAAACCGCACAGCGCGTTCTCGAAGAACGGCTTCTTCCTGCGCGACCTG
TTCTCGAAGGTGATCTTCGCGGACCGCCAGACGGTGCGCCAGTTCGCGAGCCCGACCAAG
ACGCGGCTGCGCTACGCGACCTTCTT
```

Nucleotide sequences of *tssM*_NTD from J2315 and H111. A. N-terminal region sequence of J2315 *tssM*_NTD gene. Region encoding predicted trans-membrane domain is shown in red font. Position of primers used to amplify *tssM*_NTD, the forward primer highlighted in green colour, N-*tssM*_NTD reverse is highlighted in light blue colour. B. nucleotide sequence of region of H111 *tssM*_NTD cloned into BACTH plasmids. Base pair changes compared to J2315 *tssM*_NTD are highlighted in pink colour.

A. *tssM*_{CTD} **gene (J2315)**

A. *tssM*$_{CTD}$ **gene (J2315)**

```
CGGCTTCGTCGCGGCGCTCGCGCTCGCGCTGGGCGGCTGGACCTGGTCGACGATCGGCAA
CCAGCAGCTGGTCGCGAACGTGCAGGCCGACCTCGACAACGTCACGCGCCTGCAGCAGGG
CCGCAACGACCTGCAGTCGCGCCTGCAGGCGATGGACATCCTCGAGGACCGGATCGAGCA
GCTCGAACAGTTCCGCCGCGACAAGCCGATGTCGGTGTCGCTCGGCCTGTACCAGGGCGA
CCGGCTCGAGCAGCATCTGCTGACCGAGTACTACAACGGCGTGCGCCAGATCCTGCTCGC
GCCGGTCTCGCAAAACCTCGCGTCGTTCATGAAAGACGTGAACGCGCACCCCGAGCAGCT
CGTGCCGATGACGCGCCCGCCCGAATCGGGCGCCGTGCCGGGCGGCACGATACCGGTCTC
GACGAACGCGGCAGGCGCCGCCCCGCAGGCGGGCGCCGCGCTGGCCGCGTCCGGCACCGC
GCCGGCCACCGCGCCGCAACAGGCGGCGGCCCCGCAAGGCGGCCTCTACAGCGACGCGTC
GCCAACCAACGTGCAGGATGCGTACAACGCGCTCAAGACGTACCTGATGCTGTCCGACAA
GCGTCACGTCGAACAGGCGCACCTGACCGATCAACTCGCCCGCTTCTGGCGCGGCTGGCT
CGAGACGAATCGCGGCAACATGCCGCGCGACGAGATGATCCGGAGCGCCGAGCGGATGAT
CTCGTTCTACCTGTCCCGCGTGAACGACAACGACTGGCCGATGATCGACGCGAACCTCGC
GCTCGTCGACCAGACCCGCGAGAACCTGCGCCGCGTGGTGCGCGGCATGCCGGCCCGCCA
GCGCGTCTACGAGGAAATCAAGGCCCGCGCGTCGACCCGCTTCGCGCCGATGACCATCGC
GCGCATCGTCGGCGACGGCAACCAGGGGCTCGTGGCAGGCAGCTACGCGATTCCGGGCAC
GTTCACGCGTGAAGCGTGGTTCGACTACGTGCAGCCGGCGATCCGCGATGCGGCGACCAA
GGAGCTGCAGGCGAAGGACTGGGTGCTCAACACGTCGACGCAGGACGACCTGACGCTCGA
GGGCAGCCCCGAGCAGATCCAGAAGACGCTCGTCGGCATGTACAAGACCGAGTACGCGCA
GCACTGGCAGAAGTTCATGCAGGGCATCGCGGTACAGGGCTTCACCAGCTTCGGCCAGGC
CGTCGACGGGATGAACCGCCTCGGCGACCCGCAGGATTCGCCGATCCGCAAGATCCTCGA
GACCGCGTACGACCAGACGTCGTGGGACAACCCGTCGCTCGCGAACGTGACAATCAAGAA
GGCGCAGACGGGCGTCGTGAACTGGGTCAAGCAGCTGTTCTCGCGCTCGCAGGCCGGCCA
GGTGGCGGCAGCCAACATCGACATCAACGGCAATCCGGCCGAGGTGCCGATGGGTCCGAT
CGGCCAGGAGTTCGTCGGGCTCGCGCGGATCGTCGCGACGCACGACGGCACGTCGATGCT
GAAGGGCTACATGGAGTCGCTGTCGAAGGTCCGCACGCGCTTCAACGTGATCAAGAACCA
GGGCGACCCGGGGCCGGGCGCGCGCCAGCTGATGCAGCAGACGCTCGACGGCAGCGGTTC
GGAACTTGCCGATTCGCTGAAGCTCGTCGACGAGCAGATGCTGACGGGCCTGACCGATTC
GCAACGCAAGTCGCTGCGTCCGCTGCTGGTGCGGCCGCTGATGCAGGCGTTCGCGGTCGT
GATCCAGCCGGCCAGCGCCGAAGTCAACAAGGTGTGGAACGCGCAGGTCTACCAGCCGTT
CCAGAGCTCGCTCGCGACGAAGTACCCGTTCGCCGCGAACGCGAAGGTCGAGGCCGGCGC
CGGCGAGATCGCGCAGGTGTTCGGTCCGGACGGCTCGATCGCGAAGTTCGTCGGGACGAC
GCTCGGGCCGCTCGCGGTGCGCCGCGGCGACACGCTCGCCGCCCGCACGTGGGGCGACAT
GGGCATCGGACTCACGCCGGACTTCACCAACGGCTTCGCCCGCTGGGTCGCGCCGCTCGC
GGGCGGCGCGGCCGGCAGCGCAGCCGCGTCGGCCGAGCCGCAGACCGTGTTCCAGATCCT
GCCGCAACCGAGCACGGGCACGACGGAATACACGATCGCGATCGACGGCCAGCAGCTGCG
CTACCGCAACACGCCGCCGCAGTGGACCAACTTCGTGTGGCCGAACCCGCAGGGTTCGCC
GGGCGCGACGCTGTCCGCGACGACCTTCGACGGCCGCACGGTGCAGCTCGTCAACGAGCC
GGGCCGCTACGGTCTCGAGAAGCTGATCAACTCGGCGCAGCGCAAGCGCCGTCCGGATGG
CACCTTCGACCTGACGTGGGCACAGGGCAGCGTGAACGTGTCGGTGACGATGCGCATCAT
CAGCACGTCGCAACCGTCGGCCGGTGGCGGCGACCAGCCGCAGCAGCAAAGCCTGCGCGG
CCTGCAGCTGCCGTCGTCGGTCGCCGACGCGAGCGCGGGCGCCGCGCAGAACGCGACGCA
AGCCGGCACCGGCACGCCGGGCGCCGCGCCGGCGGTCGCGGCCGCCAACGCAACGAATGC
ACAGAGGGCGCAATGAC
```

B. *tssM*_{CTD} **gene (H111)**

```
ACGATCGGCAACCAGCAGCTGGTCGCGAACGTGCAGGCCGACCTCGACAACGTCACGCGC
CTGCAGCAGGGCCGCAACGACCTGCAGTCGCGCCTGCAGGCGATGGACATCCTCGAGGAC
CGGATCGAGCAGCTCGAACAGTTCCGCCGCGACAAGCCGATGTCGGTGTCGCTCGGCCTG
TACCAGGGCGACCGGCTCGAGCAGCATCTGCTGACCGAGTACTACAACGGCGTGCGCCAG
ATCCTGCTCGCGCCGGTCTCGCAAAACCTCGCGTCGTTCATGAAGGATGTGAACGCGCAC
CCCGAGCAGCTCGTGCCGATGACGCGCCCGCCCGAATCGGGCGCCGTGCAGGGCGGCACG
ATACCGGTCTCGACGAACGCGGCAGGCACCGCGCCGGCCGCGTCCGGCACCGCGCCGGCC
ACCGCACCGCAACAGGCGGCGGCCCCGCAAGGCGGCCTCTACAGCGACGCGTCGCCAACC
AACGTGCAGGATGCGTACAACGCGCTCAAGACGTACCTGATGCTGTCCGACAAGCGTCAC
GTCGAACAGGCGCACCTGACCGACCAGCTCGCCCGCTTCTGGCGCGGCTGGCTCGAGACG
AATCGCGGCAACATGCCGCGCGACGAGATGATCCGGAGCGCCGAGCGGATGATCTCGTTC
TACCTGTCCCGCGTGAACGACAACGACTGGCCGATGATCGACGCGAACCTCGCGCTCGTC
GACCAGACCCGCGAGAACCTGCGCCGCGTGGTGCGCGGCATGCCGGCCCGCCAGCGCGTC
TACGAGGAAATCAAGGCCCGCGCGTCGACCCGCTTCGCGCCGATGACCATCGCGCGCATC
GTCGGCGACGGCAACCAGGGGCTCGTGGCAGGCAGCTACGCGATTCCGGGCACGTTCACG
CGTGAAGCGTGGTTCGACTACGTGCAGCCGGCGATCCGCGATGCGGCGACCAAGGAGCTG
CAGGCGAAGGACTGGGTGCTCAACACGTCGACGCAGGACGACCTGACGCTCGAGGGCAGC
CCCGAGCAGATCCAGAAGACGCTCGTCGGCATGTACAAGACCGAGTACGCGCAGCACTGG
CAGAAGTTCATGCAGGGCATCGCGGTACAGGGCTTCACCAGCTTCGGCCAGGCCGTCGAC
GGGATGAACCGCCTCGGCGACCCGCAGGATTCGCCGATCCGCAAGATCCTCGAGACCGCG
TACGACCAGACGTCGTGGGACAACCCGTCGCTCGCGAACGTGACGATCAAGAAGGCGCAG
ACGGGCGTCGTGAACTGGGTCAAGCAGCTGTTCTCGCGCTCGCAGGCCGGCCAGGTGGCG
GCAGCCAACATCGACATCAACGGCAATCCGGCCGAGGTGCCGATGGGTCCGATCGGCCAG
GAGTTCGTCGGGCTCGCGCGGATCGTCGCGACGCACGACGGCACGTCGATGCTGAAGGGC
TACATGGAGTCGCTGTCGAAGGTCCGCACGCGCTTCAACGTGATCAAGAACCAGGGCGAC
CCGGGGCCGGGCGCGCGCCAGCTGATGCAGCAGACGCTCGACGGCAGCGGTTCGGAACTC
GCCGATTCGCTGAAGCTCGTCGACGAGCAGATGCTGACGGGCCTGACCGATTCGCAACGC
AAGTCGCTGCGTCCGCTGCTGGTGCGGCCGCTGATGCAGGCGTTCGCGGTCGTGATCCAG
CCGGCCAGCGCCGAAGTCAACAAGGTGTGGAACGCGCAGGTCTACCAGCCGTTCCAGAGC
TCGCTCGCGACCAAGTACCCGTTCGCCGCGAACGCGAAGGTCGAGGCCGGCGCCGGCGAG
ATCGCGCAGGTGTTCGGTCCGGACGGCTCGATCGCGAAGTTCGTCGGCACGACGCTCGGG
CCGCTCGCGGTGCGCCGCGGCGACACGCTCGCCGCCCGCACGTGGGGCGACATGGGCATC
GGACTCACGCCGGACTTCACCAACGGCTTCGCCCGCTGGGTCGCGCCGCTCGCGGGCGGC
GCGGCCGGCAGCGCAGCCGCGTCGGCCGAGCCGCAGACCGTGTTCCAGATCCTGCCGCAA
CCGAGCACGGGCACGACGGAATACACGATCGCGATCGACGGCCAGCAGCTGCGCTACCGC
AACACGCCGCCGCAGTGGACCAACTTCGTGTGGCCGAACCCGCAGGGTTCGCCGGGCGCG
ACGCTGTCCGCGACGACCTTCGACGGCCGCACGGTGCAGCTCGTCAACGAGCCGGGCCGC
TACGGTCTCGAGAAGCTGATCAACTCGGCGCAGCGCAAGCGCCGTCCGGATGGCACCTTC
GACCTGACGTGGGCACAGGGCAGCGTGAACGTGTCGGTGACGATGCGCATCATCAGCACG
TCGCAGCCGTCGGCCGGTGGCGGCGACCAGCCGCAGCAGCAAAGCCTGCGCGGCCTGCAG
CTGCCGTCGTCGGTCGCCGACGCGAGCGCGGGCGCCGCGCAGAACGCGACGCAAGCCGGC
ACCGGCACGCCGGGCGCCGCGCCGGCGGTCGCGGCCGCCAACGCAACGAAtGCACAGAgg
GCGCAA
```

Nucleotide sequences of *tssM*_{CTD} from J2315 and H111. A. C-terminal region sequence of J2315 *tssM*_{CTD} gene, including stop codon. Region encoding predicted trans-membrane domain is shown in red font. Position of primers used to amplify *tssM*_{CTD}, the forward primer highlighted in green colour, C-*tssM*_{CTD} reverse primer is highlighted in yellow colour and the region highlighted in grey colour is shared between the two reverse primers, N-*tssM*_{CTD} reverse is highlighted in light blue colour. B. nucleotide sequence of region of H111 *tssM*_{CTD} cloned into BACTH plasmids. Base pair changes compared to J2315 *tssM*_{CTD} are highlighted in pink colour.

8.3. Appendix 3. T6SS Proteins sequences of J2315 and H111 strains of *B. cenocepacia*.

A. Putative Type VI secretion system protein TssH (J2315)

```
MSTPLKTLITKLNPLCRHATERAASACLARGHYEVDLEHLFLALLDEATG
DLPLALRASRVDPHALHADLERELTRLKTGNTRTPVFSVHLIALFEQAWL
IASLDSQLGRIRSGHLLLALLTAPDLAQFAQRMSSRFAEMNVTDLKHKFD
EIMAGSSEAEPRQTDDEGSDVAPAADGMAPAAGPSKTPALDTYTTNLTQR
ARDGKIDPVIGREAEIRQAIDILMRRQNNPIMTGEAGVGKTAVVEGLAL
RIAADDVPPPLRGVALHVLDMGLLQAGASVKGEFENRLKSVIDEVKKSAH
PIILFIDEAHTIIGAGGQAGQNDAANLLKPALARGELRTIAATTWSEYKK
YFEKDAALARRFQVVKVEEPSEPLAAAMLRGMSGLMEKHFNVRILDDAIT
EAVRLSHRYISGRQLPDKAISVLDTACAKVALAHSATPAAIDDTKKRIER
IDAEIASLERE V AGGAAHDERLGELRGARDTALEQLAKDEERYEAERVIV
AEITELRDALDRARGPSEDGQPVDVQATRDKLAERVAALHALQG S EPMVP
LQVDGHVVAEIVAAWTGIPLGRMVKDEIDTVLNLQPLL S ARVIGQDHALE
AIAQRVRTATANLEDPNKPRGVFMFVGPSGVGKTETALALADILYGGERK
MVTINMSEYQEAHSVSGLKGSPPGYVGYGEGGVLTEAVRRNPYSVVLLDE
VEKAHPDVLEMFFQVFDKGTMDDAEGREIDFRNTLIILTSNVGSAAVMQA
CLNKPAEELPDPDTLAETLRPQLYKTFKPAFLGRMKVVPYYPISDDVLAE
IIELKLERIRRRIEANHKAAFEWDESLVDAVLARCTEVDSGARNVDHILN
GTLLPEIAGHVLGRIADGAAIARIAVRADEAGEFAYTVE
```

B. Putative Type VI secretion system protein TssH (H111)

```
MSTPLKTLITKLNPLCRHATERAASACLARGHYEVDLEHLFLALLDEATG
DLPLALRASRVDPHALHADLERELTRLKTGNTRTPVFSVHLIALFEQAWL
IASLDSQLGRIRSGHLLLALLTAPDLAQFAQRMSSRFAEMNVTDLKHKFD
EIMAGSSEAEPRQTDDEGSDVAPAADGMAPAAGPSKTPALDTYTTNLTQR
ARDGKIDPVIGREAEIRQAIDILMRRQNNPIMTGEAGVGKTAVVEGLAL
RIAADDVPPPLRGVALHVLDMGLLQAGASVKGEFENRLKSVIDEVKKSAH
PIILFIDEAHTIIGAGGQAGQNDAANLLKPALARGELRTIAATTWSEYKK
YFEKDAALARRFQVVKVEEPSEPLAAAMLRGMSGLMEKHFNVRILDDAIT
EAVRLSHRYISGRQLPDKAISVLDTACAKVALAHSATPAAIDDTKKRIER
IDAEIASLERE A AGGAAHDERLGELRGARDTALEQLAKDEERYEAERVIV
AEITELRDALDRARGPSEDGQPVDVQATRDKLAERVAALHALQG G EPMVP
LQVDGHVVAEIVAAWTGIPLGRMVKDEIDTVLNLQPLL T ARVIGQDHALE
AIAQRVRTATANLEDPNKPRGVFMFVGPSGVGKTETALALADILYGGERK
MVTINMSEYQEAHSVSGLKGSPPGYVGYGEGGVLTEAVRRNPYSVVLLDE
VEKAHPDVLEMFFQVFDKGTMDDAEGREIDFRNTLIILTSNVGSAAVMQA
CLNKPAEELPDPDTLAETLRPQLYKTFKPAFLGRMKVVPYYPISDDVLAE
IIELKLERIRRRIEANHKAAFEWDESLVDAVLARCTEVDSGARNVDHILN
GTLLPEIAGHVLGRIADGAAIARIAVRADEAGEFAYTVE
```

Amino acid sequence of TssH from J2315 and H111. A. Amino acid sequence of TssH from J2315. B. Amino acid sequence encoded by region of H111 *tssH* cloned into BACTH plasmids. Amino acids that are different in J2315 and H111 are highlighted in pink.

A. Putative Type VI secretion system protein TssI (J2315)

```
MDAHSMIGALTGGLIQQERLLKLDTPLGSNVLIPHRVFGQSRIGRDYLFM
VDCVSTSNDVQLKALISQPITLWIQQTDKSYLPHHGYIHTARKLGVDGGL
ACYQLSFSSWLHFLKFRRDQRHWQDKTVDAIIADVFDAHPQARGMYRFEL
SQPLPSRSYCRQDETDWNFVHRLLESEGLYGVWRQAQDGKSHTLVMTDRL
QTLEPLAPATVEFSRAGVHGEVDALTQWAGTRTLQSVLLSTRTFDYKNPS
TPANPKATSVPTMANQGDLPSQAEVYEYTGGYTYPEQERGDHLGKIRMEE
WESRAKRYDGVGGVRRIDAGRRFTLTGHPEHDQDAADHREFAVIEVEWII
VNNVPLSGHEANYPHSLYSAVKQADADDPDKLFTVSHDDGSTGFYRVAIE
AQRTSVPYRSPLEHRKPEAKLESAIVAGPKGQQAYTDSLNRIKVLFIWDR
HNEGDERASCWVRVAQSDTGDYGGVHMPRAGEELLIGHIGNDIDRPIAL
HRVYNGAAKPRWHSNALLSGYRSQEFSGSGYNQMVDDSTGQNRVHLYSS
STNAALHLGYLIDQNDNARGGYLGKGFDLSSEAYGALRAGRGLFVTTHPV
ASQPLDASLATRQLNGSANILDGLSQASETSKAESLKDGHDTFKAFADAT
QHSESGATGGGGVTAGGGTGDASAFKQPVILMAAPAGIGLSTQKSTHVAA
DQHVNIVSGQSTHVATGKSLVASITEKLSLFVQNAGMKLFAAKGKIEVQT
HADNVEVTAQKSLLLASVTEKVQASAQQEILLTSGGAYIRLKGGDIEIHA
PGKIDIKGAQHAFSGPARMDVTHPAFKDLPTRRLMLNTMASPSATRVVPV
GMPYKLYADGALVKQGVFDKTGQLPIDHQVTTQKYTLEMANGVKHEIPVP
GEYRDAENGKLANQGFQYHEGAPDDGAAPDRATHRQIYNELLSPSSEA
```

B. Putative Type VI secretion system protein TssI (715j)

```
MDAHSMIGALTGGLIQQERLLKLDTPLGSNVLIPHRVFGQSRIGRDYLFM
VDCVSTSNDVQLKALISQPITLWIQQTDKSYLPHHGYIHTARKLGVDGGL
ACYQLSFSSWLHFLKFRRDQRHWQDKTVDAIIADVFDAHPQARGMYRFEL
SQPLPSRSYCRQDETDWNFVHRLLESEGLYGVWRKAQDGKSHTLVMTDRL
QTLEPLAPATVEFSRAGVHGEVDALTQWAGTRTLQSVLLSTRTFDYKNPS
TPANPKATSVPTMANQGDLPSQAEVYEYTGGYTYPEQERGDHLGKIRMEE
WESRAKRYDGVGGVRRIDAGRRFTLTGHPEHDQDAADHREFAVIEVEWII
VNNVPLSGHEANYPYSLYSAVKQADADDPDKLFTVSHDDGSTGFYRVAIE
AQRTSVPYRSPLEHRKPEAKLESAIVAGPKGQQAYTDGLNRIKVLFIWDR
HNEGDERASCWVRVAQSDTGDYGGVHMPRAGEELLIGHIGNDIDRPIAL
HRVYNGAAKPRWHSNALLSGYRSQEFSGSGYNQMVDDSTGQNRVHLYSS
STNAALHLGYLIDQNDNARGGYLGKGFDLSSEAYGALRAGRGLFVTTHPV
ASQPLDASLATRQLNGSANILDGLSQASETSKAESLKDGHDTFKAFADAT
QHSESGATGGGGVTAGGGTGDASAFKQPVILMAAPAGIGLSTQKSTHVAA
DQHVNIVSGQSTHVATGKSLVASITEKLSLFVQNAGMKLFAAKGKIEVQT
HADNVEVTAQKSLLLASVTEKVQASAQQEILLTSGGAYIRLKGGDIEIHA
PGKIDIKGAQHAFSGPARMDVTHPAFKDLPTRRLMLNTMASPSATRVVPV
GMPYKLYADGALVKQGVFDKTGQLPIDHQVTTQKYTLEMANGVKHEIPVP
GEYRDAENGKLANQGFQYHEGAPDDGAAPDRATHRQIYNELLSPSSEA
```

Amino acid sequence of TssI from J2315 and 715j. A. Amino acid sequence of TssI from J2315. B. Amino acid sequence encoded by region of 715j *tssI* cloned into BACTH plasmids. Amino acids that are different in J2315 and H111 are highlighted in pink.

A. Putative Type VI secretion system protein TssJ (J2315)

MIRYALPLLACALLAGC<mark>AAAPLLGSAASAVMSAAGIGKPEVPDAQKPPRN
IGLTLAAAPNLNAATDNKPLALVVRLYALKDPTSFQQAPFDAFTDPAKEK
TALGADLLNVREITLIPGQRYTATEKVSREAQAFGIVALFRDPALQRWKL
TFDPVKSEKSGIIIGLHNCAMTVTDGTVIAPQQGAPSQPLNMLSSVSCG</mark>

B. Putative Type VI secretion system protein TssJ (H111)

LLAGC<mark>AAAPLLGSAASAVMSAAGIGKPEVPDAQKPPRNIGLTLAAAPNLN
AATDNKPLALVVRLYALKDPTSFQQAPFDAFTDPAKEKTALGADLLNVRE
ITLIPGQRYTATEKVSREAQAFGIVALFRDPALQRWKLTFDPVKSEKSGI
IIGLHNCAMTVTDGTVIAPQQGAPSQPLNMLSSVSCG</mark>

Amino acid sequence of TssJ from J2315 and H111. A. Amino acid sequence of TssJ from J2315 with the predicted periplasmic domain highlighted in green (TssJ is predicted to be outer membrane-anchored lipoprotein). Predicted signal sequence is shown in red font. B. Amino acid sequence encoded by region of H111 *tssJ* cloned into BACTH plasmids.

A. Putative Type VI secretion system protein TssK (J2315/H111)

```
MSYSAKVLWGEGLFLRPQHFQRQDAYHEARLFESIQAIQPYNWGVRSVRI
DRDALGSNVLRVAELALVFPDGALYAAPQADDLPPPIALDTLPDGINEFV
FYLALHPLRENGTNYSDDPAAGFMTRFVSEQTSVADNFTDAAEADITFLK
TQVKLIAHSEPRDQLLSVPLVRVRRTATSGFEIDDSFVPPCLAIEASPIL
HQRLRQLVDALQAKVNALYGFHREPSKNIIEFRSGDIASFWLLHTANAAF
ATLAHLHQHAALHPERLFQELLRLAGQLMTFSKGYTLADLPVYRHDDPGP
SFARLDLMLRELLDTVISTRYFAITLDEVRPSFHLGRLDSGKIDDKTEFY
LAVSADMPSVELVDAVPARFKVGAPDDVDKLVLSAMPGVRLSYTPQVPPA
IPVRPGACYFALDSRSPLYERMLQAQSAMIYAPTGINDLKFELIAVTS
```

Amino acid sequence of TssK from J2315 and H111, all H111 *tssK* cloned into BACTH plasmids.

A. Putative Type VI secretion system protein TssL (J2315)

```
MSYAPSLFGDNAPAPLHTPASTDSAFQARSLLDLLYDGFFMLFLLKNGRE
PDSASEFSTRIQEFLSDFERGAKKLNIAAEDVYGAKFAFCAAIDEMVLSS
QFKIRADWERRPLQLVLFGEQLAGEKFFQYLEEGRAQGAARLQSLEVFHM
CLLLGFQGKYLLEGPEKLAYLTARLGDEIAHMKGKRAPFAPHWPLPDQIA
HRLKREVPVWAIGTVFALVALLGYLGLNTYLKDKTLQALAPYSQVIKVGP
ESANLTISLP
```

B. Putative Type VI secretion system protein TssL (H111)

```
MSYAPSLFGDNAPAPLHTPASTDSAFQARSLLDLLYDGFFMLFLLKNGRE
PDSASEFSTRIQEFLSDFERGAKKLNIAAEDVYGAKFAFCAAIDEMVLSS
QFKIRADWERRPLQLVLFGEQLAGEKFFQYLEEGRAQGAARLQSLEVFHM
CLLLGFQGKYLLEGPEKLAYLTARLGDEIAHMKGKRAPFAPHWPLPDQIA
HRLK
```

Amino acid sequence of TssL from J2315 and H111. A. Amino acid sequence of TssL from J2315 with the predicted cytoplasmic domain highlighted in green. Predicted TMD is shown in red font. B. Amino acid sequence encoded by region of H111 *tssL* cloned into BACTH plasmids.

A. Putative Type VI secretion system protein TssM (J2315)

MQRILNVLTHPRTLSIVGIVALAAILFIVADMLQLPLLWAATAFAAILAL
WLVVALWRRWRVKRANQQLGQVLEEQAETGKIAAPAAAALAPDAKTADLD
VLRTRLSDAVKTIKTSKIGQVSGGSALYELPWYIVIGNPAAGKSSAVLNS
GLQFPFADKNSAVIHGIGGTRNCDWFFTTEGILLDTAGRYSVHEEDRSEW
LGFLGLLKRYRPKAPINGIIVTASIAELTGNRPEFAINLAKNLRQRVQEL
TEKLEVFAPVYVMFTKADLITGFTEFFSSSDKHEYDRVWGATLPYEPDDK
RDVVAQFDTHFEELYEGLKEISVAQLSLSRGNQLSPGQLSFPLEFSTIKP
SLRAFLATLFENNPFQYKPIFRGFYFTSALQEGETNSAAAQRIAHRFGLD
GNALPKPHSAFSKNGFFLRDLFSKVIFADRQTVRQFASPTKTRLRYATFF
GFVAALALALGGWTWSTIGNQQLVANVQADLDNVTRLQQGRNDLQSRLQA
MDILEDRIEQLEQFRRDKPMSVSLGLYQGDRLEQHLLTEYYNGVRQILLA
PVSQNLASFMKDVNAHPEQLVPMTRPPESGAVPGGTIPVSTNAAGAAPQA
GAALAASGTAPATAPQQAAAPQGGLYSDASPTNVQDAYNALKTYLMLSDK
RHVEQAHLTDQLARFWRGWLETNRGNMPRDEMIRSAERMISFYLSRVNDN
DWPMIDANLALVDQTRENLRRVVRGMPARQRVYEEIKARASTRFAPMTIA
RIVGDGNQGLVAGSYAIPGTFTREAWFDYVQPAIRDAATKELQAKDWVLN
TSTQDDLTLEGSPEQIQKTLVGMYKTEYAQHWQKFMQGIAVQGFTSFGQA
VDGMNRLGDPQDSPIRKILETAYDQTSWDNPSLANVTIKKAQTGVVNWVK
QLFSRSQAGQVAAANIDINGNPAEVPMGPIGQEFVGLARIVATHDGTSML
KGYMESLSKVRTRFNVIKNQGDPGPGARQLMQQTLDGSGSELADSLKLVD
EQMLTGLTDSQRKSLRPLLVRPLMQAFAVVIQPASAEVNKVWNAQVYQPF
QSSLATKYPFAANAKVEAGAGEIAQVFGPDGSIAKFVGTTLGPLAVRRGD
TLAARTWGDMGIGLTPDFTNGFARWVAPLAGGAAGSAAASAEPQTVFQIL
PQPSTGTTEYTIAIDGQQLRYRNTPPQWTNFVWPNPQGSPGATLSATTFD
GRTVQLVNEPGRYGLEKLINSAQRKRRPDGTFDLTWAQGSVNVSVTMRII
STSQPSAGGGDQPQQQSLRGLQLPSSVADASAGAAQNATQAGTGTPGAAP
AVAAANATNAQRAQ

**B. Putative Type VI secretion system protein TssM_{NTD} (H111)**

```
VKRANQQLGQVLEEQAETGKIAAPAAAALAPDAKTADLDVLRTRLSDAVK
TIKTSKIGQVSGGSALYELPWYIVIGNPAAGKSSAVLNSGLQFPFADKNS
AVIHGIGGTRNCDWFFTTEGILLDTAGRYSVHEEDRSEWLGFLGLLKRYR
PKAPINGIIVTASIAELTGNRPEFAINLAKNLRQRVQELTEKLEVFAPVY
VMFTKADLITGFTEFFSSSDKHEYDRVWGATLPYEPDDKRDVVAQFDTHF
EELYEGLKEISVAQLSLSRGNQLSPGQLSFPLEFSTIKPSLRAFLATLFE
NNPFQYKPIFRGFYFTSALQEGETNSAAAQRIAHRFGLDGNALPKPHSAF
SKNGFFLRDLFSKVIFADRQTVRQFASPTKTRLRYATF
```

C. Putative Type VI secretion system protein TssM_{CTD} (H111)

```
TIGNQQLVANVQADLDNVTRLQQGRNDLQSRLQAMDILEDRIEQLEQFRRDKPMSVSLGL
YQGDRLEQHLLTEYYNGVRQILLAPVSQNLASFMKDVNAHPEQLVPMTRPPESGAVQGGT
IPVSTNAAGTAPAASGTAPATAPQQAAAPQGGLYSDASPTNVQDAYNALKTYLMLSDKRH
VEQAHLTDQLARFWRGWLETNRGNMPRDEMIRSAERMISFYLSRVNDNDWPMIDANLALV
DQTRENLRRVVRGMPARQRVYEEIKARASTRFAPMTIARIVGDGNQGLVAGSYAIPGTFT
REAWFDYVQPAIRDAATKELQAKDWVLNTSTQDDLTLEGSPEQIQKTLVGMYKTEYAQHW
QKFMQGIAVQGFTSFGQAVDGMNRLGDPQDSPIRKILETAYDQTSWDNPSLANVTIKKAQ
TGVVNWVKQLFSRSQAGQVAAANIDINGNPAEVPMGPIGQEFVGLARIVATHDGTSMLKG
YMESLSKVRTRFNVIKNQGDPGPGARQLMQQTLDGSGSELADSLKLVDEQMLTGLTDSQR
KSLRPLLVRPLMQAFAVVIQPASAEVNKVWNAQVYQPFQSSLATKYPFAANAKVEAGAGE
IAQVFGPDGSIAKFVGTTLGPLAVRRGDTLAARTWGDMGIGLTPDFTNGFARWVAPLAGG
AAGSAAASAEPQTVFQILPQPSTGTTEYTIAIDGQQLRYRNTPPQWTNFVWPNPQGSPGA
TLSATTFDGRTVQLVNEPGRYGLEKLINSAQRKRRPDGTFDLTWAQGSVNVSVTMRIIST
SQPSAGGGDQPQQQSLRGLQLPSSVADASAGAAQNATQAGTGTPGAAPAVAAANATNAQR
AQ
```

Amino acid sequence of TssM from J2315 and H111. A. Amino acid sequence of full length of TssM from J2315. Amino acid sequence of TssM_{NTD} (the predicted cytoplasmic domain) is highlighted in green. Amino acid sequence of TssM_{CTD} (the predicted periplasmic domain) is highlighted in yellow. Predicted TMD are shown in red font. B. Amino acid sequence encoded by region of H111 *tssM*_{NTD} cloned into BACTH plasmids. C. Amino acid sequence encoded by region of H111 *tssM*_{CTD} cloned into BACTH plasmids. Amino acids that are different in J2315 and H111 are highlighted in pink.

8.4. Appendix 4. DNA ladders used in this study.
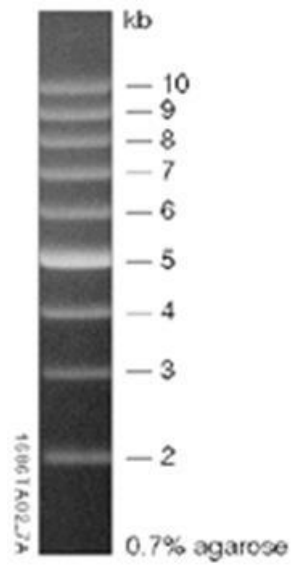
Promega Supercoiled DNA ladder   **Yorkshire Bio Q-step 4 DNA ladder**
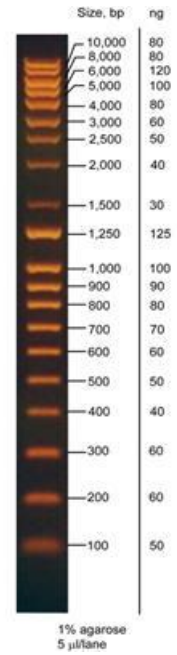
8.5. Appendix 5. Protein markers used in this study.
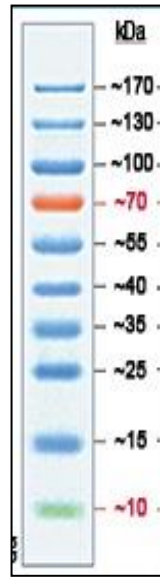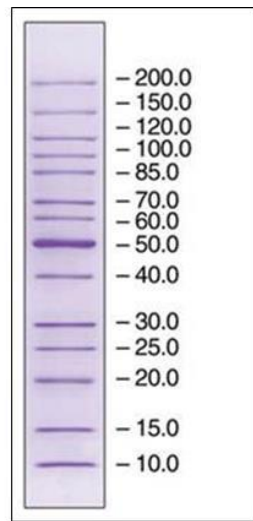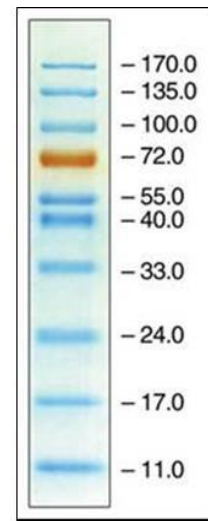

Fermentas Prestained Pageruler protein ladder

Fisher BioReagents* EZ-Run*
Rec Unstained Protein Ladder

Fisher BioReagents* EZ-Run*
Prestained protein ladder

9. Bibliography

AKSYUK, A. A., LEIMAN, P. G., KUROCHKINA, L. P., SHNEIDER, M. M., KOSTYUCHENKO, V. A., MESYANZHINOV, V. V. & ROSSMANN, M. G. (2009) The tail sheath structure of bacteriophage T4: a molecular machine for infecting bacteria. *EMBO J,* 28**,** 821-9.

ALOULOU, A., ALI, Y. B., BEZZINE, S., GARGOURI, Y. & GELB, M. H. (2012) Phospholipases: an overview. *Methods Mol Biol,* 861**,** 63-85.

ARTS, J., DE GROOT, A., BALL, G., DURAND, E., EL KHATTABI, M., FILLOUX, A., TOMMASSEN, J. & KOSTER, M. (2007) Interaction domains in the Pseudomonas aeruginosa type II secretory apparatus component XcpS (GspF). *Microbiology,* 153**,** 1582-92.

ASCHTGEN, M. S., BERNARD, C. S., DE BENTZMANN, S., LLOUBES, R. & CASCALES, E. (2008) SciN is an outer membrane lipoprotein required for type VI secretion in enteroaggregative Escherichia coli. *J Bacteriol,* 190**,** 7523-31.

ASCHTGEN, M. S., GAVIOLI, M., DESSEN, A., LLOUBES, R. & CASCALES, E. (2010a) The SciZ protein anchors the enteroaggregative Escherichia coli Type VI secretion system to the cell wall. *Mol Microbiol,* 75**,** 886-99.

ASCHTGEN, M. S., THOMAS, M. S. & CASCALES, E. (2010b) Anchoring the type VI secretion system to the peptidoglycan: TssL, TagL, TagP... what else? *Virulence,* 1**,** 535-40.

ASCHTGEN, M. S., ZOUED, A., LLOUBES, R., JOURNET, L. & CASCALES, E. (2012) The C-tail anchored TssL subunit, an essential protein of the enteroaggregative Escherichia coli Sci-1 Type VI secretion system, is inserted by YidC. *MicrobiologyOpen,* 1**,** 71-82.

AUBERT, D., MACDONALD, D. K. & VALVANO, M. A. (2010) BcsKC is an essential protein for the type VI secretion system activity in Burkholderia cenocepacia that forms an outer membrane complex with BcsLB. *J Biol Chem,* 285**,** 35988-98.

AUBERT, D. F., FLANNAGAN, R. S. & VALVANO, M. A. (2008) A novel sensor kinase-response regulator hybrid controls biofilm formation and type VI secretion system activity in Burkholderia cenocepacia. *Infect Immun,* 76**,** 1979-91.

BALLISTER, E. R., LAI, A. H., ZUCKERMANN, R. N., CHENG, Y. & MOUGOUS, J. D. (2008) In vitro self-assembly of tailorable nanotubes from a simple protein building block. *Proc Natl Acad Sci U S A,* 105**,** 3733-8.

BASLER, M., PILHOFER, M., HENDERSON, G. P., JENSEN, G. J. & MEKALANOS, J. J. (2012) Type VI secretion requires a dynamic contractile phage tail-like structure. *Nature,* 483**,** 182-6.

BATTESTI, A. & BOUVERET, E. (2012) The bacterial two-hybrid system based on adenylate cyclase reconstitution in Escherichia coli. *Methods,* 58**,** 325-34.

BERGGARD, T., LINSE, S. & JAMES, P. (2007) Methods for the detection and analysis of protein-protein interactions. *Proteomics,* 7**,** 2833-42.

BERKS, B. C. (1996) A common export pathway for proteins binding complex redox cofactors? *Mol Microbiol,* 22**,** 393-404.

BERKS, B. C., SARGENT, F. & PALMER, T. (2000) The Tat protein export pathway. *Mol Microbiol,* 35**,** 260-74.

BINGLE, L. E., BAILEY, C. M. & PALLEN, M. J. (2008) Type VI secretion: a beginner's guide. *Curr Opin Microbiol,* 11**,** 3-8.

BLADERGROEN, M. R., BADELT, K. & SPAINK, H. P. (2003) Infection-blocking genes of a symbiotic Rhizobium leguminosarum strain that are involved in temperature-dependent protein secretion. *Mol Plant Microbe Interact,* 16**,** 53-64.

BLEVES, S., LAZDUNSKI, A. & FILLOUX, A. (1996) Membrane topology of three Xcp proteins involved in exoprotein transport by Pseudomonas aeruginosa. *J Bacteriol,* 178**,** 4297-300.

BLOCK, H., MAERTENS, B., SPRIESTERSBACH, A., BRINKER, N., KUBICEK, J., FABIS, R., LABAHN, J. & SCHAFER, F. (2009) Immobilized-metal affinity chromatography (IMAC): a review. *Methods Enzymol,* 463**,** 439-73.

BONEMANN, G., PIETROSIUK, A., DIEMAND, A., ZENTGRAF, H. & MOGK, A. (2009) Remodelling of VipA/VipB tubules by ClpV-mediated threading is crucial for type VI protein secretion. *EMBO J,* 28**,** 315-25.

BONEMANN, G., PIETROSIUK, A. & MOGK, A. (2010) Tubules and donuts: a type VI secretion story. *Mol Microbiol,* 76**,** 815-21.

BOYER, F., FICHANT, G., BERTHOD, J., VANDENBROUCK, Y. & ATTREE, I. (2009) Dissecting the bacterial type VI secretion system by a genome wide in silico analysis: what can be learned from available microbial genomic resources? *BMC Genomics,* 10**,** 104.

BRADFORD, M. M. (1976) A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Analytical Biochemistry,* 72**,** 248-254.

BRAUN, T. F., KHUBBAR, M. K., SAFFARINI, D. A. & MCBRIDE, M. J. (2005) Flavobacterium johnsoniae gliding motility genes identified by mariner mutagenesis. *Journal of bacteriology,* 187**,** 6943-52.

BROK, R., VAN GELDER, P., WINTERHALTER, M., ZIESE, U., KOSTER, A. J., DE COCK, H., KOSTER, M., TOMMASSEN, J. & BITTER, W. (1999) The C-terminal domain of the Pseudomonas secretin XcpQ forms oligomeric rings with pore activity. *J Mol Biol,* 294**,** 1169-79.

BROMS, J. E., ISHIKAWA, T., WAI, S. N. & SJOSTEDT, A. (2013) A functional VipA-VipB interaction is required for the type VI secretion system activity of Vibrio cholerae O1 strain A1552. *BMC Microbiol,* 13**,** 96.

BROMS, J. E., LAVANDER, M. & SJOSTEDT, A. (2009) A Conserved {alpha}-Helix Essential for a Type VI Secretion-Like System of Francisella tularensis.

BROOKS, T. M., UNTERWEGER, D., BACHMANN, V., KOSTIUK, B. & PUKATZKI, S. (2013) Lytic activity of the Vibrio cholerae type VI secretion toxin VgrG-3 is inhibited by the antitoxin TsaB. *J Biol Chem,* 288**,** 7618-25.

BRUSER, T., DEUTZMANN, R. & DAHL, C. (1998) Evidence against the double-arginine motif as the only determinant for protein translocation by a novel Sec-independent pathway in Escherichia coli. *FEMS Microbiol Lett,* 164**,** 329-36.

BUCHAN, D. W., WARD, S. M., LOBLEY, A. E., NUGENT, T. C., BRYSON, K. & JONES, D. T. (2010) Protein annotation and modelling servers at University College London. *Nucleic Acids Res,* 38**,** W563-8.

BURKHOLDER, W. (1950).

BURTNICK, M. N., BRETT, P. J., HARDING, S. V., NGUGI, S. A., RIBOT, W. J., CHANTRATITA, N., SCORPIO, A., MILNE, T. S., DEAN, R. E., FRITZ, D. L., PEACOCK, S. J., PRIOR, J. L., ATKINS, T. P. & DESHAZER, D. (2011) The cluster 1 type VI secretion system is a major virulence determinant in Burkholderia pseudomallei. *Infect Immun,* 79**,** 1512-25.

BUSCH, A. & WAKSMAN, G. (2012) Chaperone-usher pathways: diversity and pilus assembly mechanism. *Philos Trans R Soc Lond B Biol Sci,* 367**,** 1112-22.

CARRUTHERS, M. D., NICHOLSON, P. A., TRACY, E. N. & MUNSON, R. S., JR. (2013) Acinetobacter baumannii utilizes a type VI secretion system for bacterial competition. *PLoS One,* 8**,** e59388.

CASABONA, M. G., VANDENBROUCK, Y., ATTREE, I. & COUTE, Y. (2013) Proteomic characterization of Pseudomonas aeruginosa PAO1 inner membrane. *Proteomics,* 13**,** 2419-23.

CASADABAN, M. J. & COHEN, S. N. (1980) Analysis of gene control signals by DNA fusion and cloning in Escherichia coli. *J Mol Biol,* 138**,** 179-207.

CASCALES, E. (2008) The type VI secretion toolkit. *EMBO Rep,* 9**,** 735-41.

CASCALES, E. & CAMBILLAU, C. (2012) Structural biology of type VI secretion systems. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences,* 367**,** 1102-11.

CHADDOCK, A. M., MANT, A., KARNAUCHOV, I., BRINK, S., HERRMANN, R. G., KLOSGEN, R. B. & ROBINSON, C. (1995) A new type of signal peptide: central role of a twin-arginine motif

in transfer signals for the delta pH-dependent thylakoidal protein translocase. *EMBO J,* 14, 2715-22.

CHOUDHURY, D., THOMPSON, A., STOJANOFF, V., LANGERMANN, S., PINKNER, J., HULTGREN, S. J. & KNIGHT, S. D. (1999) X-ray structure of the FimC-FimH chaperone-adhesin complex from uropathogenic Escherichia coli. *Science,* 285, 1061-6.

CHOW, J. & MAZMANIAN, S. K. (2010) A pathobiont of the microbiota balances host colonization and intestinal inflammation. *Cell Host Microbe,* 7, 265-76.

CHRISTIE, P. J. & CASCALES, E. (2005) Structural and dynamic properties of bacterial type IV secretion systems (review). *Mol Membr Biol,* 22, 51-61.

CHRISTIE, P. J. & VOGEL, J. P. (2000) Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. *Trends Microbiol,* 8, 354-60.

COENYE, T., VANDAMME, P., GOVAN, J. R. & LIPUMA, J. J. (2001) Taxonomy and identification of the Burkholderia cepacia complex. *J Clin Microbiol,* 39, 3427-36.

COMPANT, S., NOWAK, J., COENYE, T., CLEMENT, C. & AIT BARKA, E. (2008) Diversity and occurrence of Burkholderia spp. in the natural environment. *FEMS Microbiol Rev,* 32, 607-26.

CRISTOBAL, S., DE GIER, J. W., NIELSEN, H. & VON HEIJNE, G. (1999) Competition between Sec- and TAT-dependent protein translocation in Escherichia coli. *EMBO J,* 18, 2982-90.

DALBOGE, H., BAYNE, S. & PEDERSEN, J. (1990) In vivo processing of N-terminal methionine in E. coli. *FEBS Lett,* 266, 1-3.

DAS, S. & CHAUDHURI, K. (2003) Identification of a unique IAHP (IcmF associated homologous proteins) cluster in Vibrio cholerae and other proteobacteria through in silico analysis. *In Silico Biol,* 3, 287-300.

DE BRUIN, O. M., DUPLANTIS, B. N., LUDU, J. S., HARE, R. F., NIX, E. B., SCHMERK, C. L., ROBB, C. S., BORASTON, A. B., HUEFFER, K. & NANO, F. E. (2011) The biochemical properties of the Francisella pathogenicity island (FPI)-encoded proteins IglA, IglB, IglC, PdpB and DotU suggest roles in type VI secretion. *Microbiology,* 157, 3483-91.

DEAN, P. (2011) Functional domains and motifs of bacterial type III effector proteins and their roles in infection. *Fems Microbiology Reviews,* 35, 1100-1125.

DELEPELAIRE, P. (2004) Type I secretion in gram-negative bacteria. *Biochim Biophys Acta,* 1694, 149-61.

DESVAUX, M., HEBRAUD, M., TALON, R. & HENDERSON, I. R. (2009) Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue. *Trends Microbiol,* 17, 139-45.

DESVAUX, M., PARHAM, N. J. & HENDERSON, I. R. (2004) The autotransporter secretion system. *Res Microbiol,* 155, 53-60.

DI GUAN, C., LI, P., RIGGS, P. D. & INOUYE, H. (1988) Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein. *Gene,* 67, 21-30.

DONG, T. G., HO, B. T., YODER-HIMES, D. R. & MEKALANOS, J. J. (2013) Identification of T6SS-dependent effector and immunity proteins by Tn-seq in Vibrio cholerae. *Proc Natl Acad Sci U S A,* 110, 2623-8.

DOUZI, B., FILLOUX, A. & VOULHOUX, R. (2012) On the path to uncover the bacterial type II secretion system. *Philos Trans R Soc Lond B Biol Sci,* 367, 1059-72.

DRIESSEN, A. J., FEKKES, P. & VAN DER WOLK, J. P. (1998) The Sec system. *Curr Opin Microbiol,* 1, 216-22.

DUDLEY, E. G., THOMSON, N. R., PARKHILL, J., MORIN, N. P. & NATARO, J. P. (2006) Proteomic and microarray characterization of the AggR regulon identifies a pheU pathogenicity island in enteroaggregative Escherichia coli. *Mol Microbiol,* 61, 1267-82.

DURAND, E., ZOUED, A., SPINELLI, S., WATSON, P. J., ASCHTGEN, M. S., JOURNET, L., CAMBILLAU, C. & CASCALES, E. (2012) Structural characterization and oligomerization

of the TssL protein, a component shared by bacterial type VI and type IVb secretion systems. *The Journal of biological chemistry,* 287**,** 14157-68.

ECONOMOU, A. (1999) Following the leader: bacterial protein export through the Sec pathway. *Trends Microbiol,* 7**,** 315-20.

ENGLISH, G., TRUNK, K., RAO, V. A., SRIKANNATHASAN, V., HUNTER, W. N. & COULTHURST, S. J. (2012) New secreted toxins and immunity proteins encoded within the Type VI secretion system gene cluster of Serratia marcescens. *Mol Microbiol,* 86**,** 921-36.

ESTRELA, A. B., HECK, M. G. & ABRAHAM, W. R. (2009) Novel approaches to control biofilm infections. *Curr Med Chem,* 16**,** 1512-30.

EUROMEDEX (2000) Bacterial Adenylate Cyclase Two-Hybrid System Kit Manual.

FELISBERTO-RODRIGUES, C., DURAND, E., ASCHTGEN, M. S., BLANGY, S., ORTIZ-LOMBARDIA, M., DOUZI, B., CAMBILLAU, C. & CASCALES, E. (2011) Towards a structural comprehension of bacterial type VI secretion systems: characterization of the TssJ-TssM complex of an Escherichia coli pathovar. *PLoS Pathog,* 7**,** e1002386.

FILLOUX, A. (2013) Microbiology: a weapon for bacterial warfare. *Nature,* 500**,** 284-5.

FISHER, A. C., KIM, J. Y., PEREZ-RODRIGUEZ, R., TULLMAN-ERCEK, D., FISH, W. R., HENDERSON, L. A. & DELISA, M. P. (2008) Exploration of twin-arginine translocation for expression and purification of correctly folded proteins in Escherichia coli. *Microb Biotechnol,* 1**,** 403-15.

FLANNAGAN, R. S., JAUMOUILLE, V., HUYNH, K. K., PLUMB, J. D., DOWNEY, G. P., VALVANO, M. A. & GRINSTEIN, S. (2012) Burkholderia cenocepacia disrupts host cell actin cytoskeleton by inactivating Rac and Cdc42. *Cell Microbiol,* 14**,** 239-54.

FOLKESSON, A., LOFDAHL, S. & NORMARK, S. (2002) The Salmonella enterica subspecies I specific centisome 7 genomic island encodes novel protein families present in bacteria living in close contact with eukaryotic cells. *Res Microbiol,* 153**,** 537-45.

FRONZES, R., CHRISTIE, P. J. & WAKSMAN, G. (2009) The structural biology of type IV secretion systems. *Nat Rev Microbiol,* 7**,** 703-14.

GALAN, J. E. & COLLMER, A. (1999) Type III secretion machines: bacterial devices for protein delivery into host cells. *Science,* 284**,** 1322-8.

GALAN, J. E. & WOLF-WATZ, H. (2006) Protein delivery into eukaryotic cells by type III secretion machines. *Nature,* 444**,** 567-73.

GHOSH, P. (2004) Process of protein transport by the type III secretion system. *Microbiol Mol Biol Rev,* 68**,** 771-95.

GILBERT, W. & MULLER-HILL, B. (1966) Isolation of the lac repressor. *Proc Natl Acad Sci U S A,* 56**,** 1891-8.

GOLOMB, M. & CHAMBERLIN, M. (1974) Characterization of T7-specific ribonucleic acid polymerase. IV. Resolution of the major in vitro transcripts by gel electrophoresis. *The Journal of biological chemistry,* 249**,** 2858-63.

GREENFIELD, N. & FASMAN, G. D. (1969) Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry,* 8**,** 4108-16.

GREENFIELD, N. J. (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc,* 1**,** 2876-90.

HACHANI, A., LOSSI, N. S., HAMILTON, A., JONES, C., BLEVES, S., ALBESA-JOVE, D. & FILLOUX, A. (2011) Type VI secretion system in Pseudomonas aeruginosa: secretion and multimerization of VgrG proteins. *J Biol Chem,* 286**,** 12317-27.

HARDIE, K. R., SEYDEL, A., GUILVOUT, I. & PUGSLEY, A. P. (1996) The secretin-specific, chaperone-like protein of the general secretory pathway: separation of proteolytic protection and piloting functions. *Mol Microbiol,* 22**,** 967-76.

HAYES, C. S., AOKI, S. K. & LOW, D. A. (2010) Bacterial contact-dependent delivery systems. *Annu Rev Genet,* 44**,** 71-90.

HAZES, B. & FROST, L. (2008) Towards a systems biology approach to study type II/IV secretion systems. *Biochim Biophys Acta,* 1778**,** 1839-50.

HENDERSON, I. R., NAVARRO-GARCIA, F., DESVAUX, M., FERNANDEZ, R. C. & ALA'ALDEEN, D. (2004) Type V protein secretion pathway: the autotransporter story. *Microbiol Mol Biol Rev,* 68**,** 692-744.

HIREL, P. H., SCHMITTER, M. J., DESSEN, P., FAYAT, G. & BLANQUET, S. (1989) Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid. *Proc Natl Acad Sci U S A,* 86**,** 8247-51.

HOLDEN, M. T., SETH-SMITH, H. M., CROSSMAN, L. C., SEBAIHIA, M., BENTLEY, S. D., CERDENO-TARRAGA, A. M., THOMSON, N. R., BASON, N., QUAIL, M. A., SHARP, S., CHEREVACH, I., CHURCHER, C., GOODHEAD, I., HAUSER, H., HOLROYD, N., MUNGALL, K., SCOTT, P., WALKER, D., WHITE, B., ROSE, H., IVERSEN, P., MIL-HOMENS, D., ROCHA, E. P., FIALHO, A. M., BALDWIN, A., DOWSON, C., BARRELL, B. G., GOVAN, J. R., VANDAMME, P., HART, C. A., MAHENTHIRALINGAM, E. & PARKHILL, J. (2009) The genome of Burkholderia cenocepacia J2315, an epidemic pathogen of cystic fibrosis patients. *J Bacteriol,* 191**,** 261-77.

HOOD, R. D., SINGH, P., HSU, F., GUVENER, T., CARL, M. A., TRINIDAD, R. R., SILVERMAN, J. M., OHLSON, B. B., HICKS, K. G., PLEMEL, R. L., LI, M., SCHWARZ, S., WANG, W. Y., MERZ, A. J., GOODLETT, D. R. & MOUGOUS, J. D. (2010) A type VI secretion system of Pseudomonas aeruginosa targets a toxin to bacteria. *Cell Host Microbe,* 7**,** 25-37.

HSU, F., SCHWARZ, S. & MOUGOUS, J. D. (2009) TagR promotes PpkA-catalysed type VI secretion activation in Pseudomonas aeruginosa. *Mol Microbiol,* 72**,** 1111-25.

ISLES, A., MACLUSKY, I., COREY, M., GOLD, R., PROBER, C., FLEMING, P. & LEVISON, H. (1984) Pseudomonas cepacia infection in cystic fibrosis: an emerging problem. *J Pediatr,* 104**,** 206-10.

IZARD, J. W. & KENDALL, D. A. (1994) Signal peptides: exquisitely designed transport promoters. *Mol Microbiol,* 13**,** 765-73.

JACKSON, A. P., THOMAS, G. H., PARKHILL, J. & THOMSON, N. R. (2009) Evolutionary diversification of an ancient gene family (rhs) through C-terminal displacement. *BMC Genomics,* 10**,** 584.

JACOB-DUBUISSON, F., FERNANDEZ, R. & COUTTE, L. (2004) Protein secretion through autotransporter and two-partner pathways. *Biochim Biophys Acta,* 1694**,** 235-57.

JACOB-DUBUISSON, F., LOCHT, C. & ANTOINE, R. (2001) Two-partner secretion in Gram-negative bacteria: a thrifty, specific pathway for large virulence proteins. *Mol Microbiol,* 40**,** 306-13.

JOHNSTON, R. B., JR. (2001) Clinical aspects of chronic granulomatous disease. *Curr Opin Hematol,* 8**,** 17-22.

JONES.R.A (2013) Characterisation of Putative Type VI Secretion System Effector Proteins from Burkholderia cenocepacia. PhD thesis

infection and immunity. sheffield.

KANAMARU, S., LEIMAN, P. G., KOSTYUCHENKO, V. A., CHIPMAN, P. R., MESYANZHINOV, V. V., ARISAKA, F. & ROSSMANN, M. G. (2002) Structure of the cell-puncturing device of bacteriophage T4. *Nature,* 415**,** 553-7.

KAPITEIN, N., BONEMANN, G., PIETROSIUK, A., SEYFFER, F., HAUSSER, I., LOCKER, J. K. & MOGK, A. (2013) ClpV recycles VipA/VipB tubules and prevents non-productive tubule formation to ensure efficient type VI protein secretion. *Mol Microbiol,* 87**,** 1013-28.

KAPUST, R. B. & WAUGH, D. S. (1999) Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci,* 8**,** 1668-74.

KARIMOVA, G., PIDOUX, J., ULLMANN, A. & LADANT, D. (1998) A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc Natl Acad Sci U S A,* 95**,** 5752-6.

KARIMOVA, G., ULLMANN, A. & LADANT, D. (2000) A bacterial two-hybrid system that exploits a cAMP signaling cascade in Escherichia coli. *Methods Enzymol,* 328**,** 59-73.

KELLERMANN, O. K. & FERENCI, T. (1982) Maltose-binding protein from Escherichia coli. *Methods Enzymol,* 90 Pt E**,** 459-63.

KNOBLAUCH, N. T., RUDIGER, S., SCHONFELD, H. J., DRIESSEN, A. J., SCHNEIDER-MERGENER, J. & BUKAU, B. (1999) Substrate specificity of the SecB chaperone. *J Biol Chem,* 274**,** 34219-25.

KOOI, C. & SOKOL, P. A. (2009) Burkholderia cenocepacia zinc metalloproteases influence resistance to antimicrobial peptides. *Microbiology,* 155**,** 2818-25.

KOROTKOV, K. V. & HOL, W. G. (2008) Structure of the GspK-GspI-GspJ complex from the enterotoxigenic Escherichia coli type 2 secretion system. *Nat Struct Mol Biol,* 15**,** 462-8.

KOROTKOV, K. V., JOHNSON, T. L., JOBLING, M. G., PRUNEDA, J., PARDON, E., HEROUX, A., TURLEY, S., STEYAERT, J., HOLMES, R. K., SANDKVIST, M. & HOL, W. G. (2011) Structural and functional studies on the interaction of GspC and GspD in the type II secretion system. *PLoS Pathog,* 7**,** e1002228.

KOSKINIEMI, S., LAMOUREUX, J. G., NIKOLAKAKIS, K. C., T'KINT DE ROODENBEKE, C., KAPLAN, M. D., LOW, D. A. & HAYES, C. S. (2013) Rhs proteins from diverse bacteria mediate intercellular competition. *Proc Natl Acad Sci U S A,* 110**,** 7032-7.

KOSTAKIOTI, M., NEWMAN, C. L., THANASSI, D. G. & STATHOPOULOS, C. (2005) Mechanisms of protein export across the bacterial outer membrane. *Journal of bacteriology,* 187**,** 4306-14.

KOSTYUCHENKO, V. A., CHIPMAN, P. R., LEIMAN, P. G., ARISAKA, F., MESYANZHINOV, V. V. & ROSSMANN, M. G. (2005) The tail structure of bacteriophage T4 and its mechanism of contraction. *Nat Struct Mol Biol,* 12**,** 810-3.

KOSTYUCHENKO, V. A., LEIMAN, P. G., CHIPMAN, P. R., KANAMARU, S., VAN RAAIJ, M. J., ARISAKA, F., MESYANZHINOV, V. V. & ROSSMANN, M. G. (2003) Three-dimensional structure of bacteriophage T4 baseplate. *Nat Struct Biol,* 10**,** 688-93.

KUNG, V. L., KHARE, S., STEHLIK, C., BACON, E. M., HUGHES, A. J. & HAUSER, A. R. (2012) An rhs gene of Pseudomonas aeruginosa encodes a virulence protein that activates the inflammasome. *Proc Natl Acad Sci U S A,* 109**,** 1275-80.

LADANT, D. (1988) Interaction of Bordetella pertussis adenylate cyclase with calmodulin. Identification of two separated calmodulin-binding domains. *J Biol Chem,* 263**,** 2612-8.

LADANT, D. & ULLMANN, A. (1999) Bordatella pertussis adenylate cyclase: a toxin with multiple talents. *Trends Microbiol,* 7**,** 172-6.

LAUE, T., SHAH, B., RIDGEWAY, T. AND PELLETIER, S. (1992) Analytical Ultracentrifugation in Biochemistry and Polymer Science. Cambridge: Royal Society of Chemistry.

LEBOWITZ, J., LEWIS, M. S. & SCHUCK, P. (2002) Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Sci,* 11**,** 2067-79.

LEE, J. & KIM, S. H. (2009) High-throughput T7 LIC vector for introducing C-terminal poly-histidine tags with variable lengths without extra sequences. *Protein Expr Purif,* 63**,** 58-61.

LEE, P. A., TULLMAN-ERCEK, D. & GEORGIOU, G. (2006) The bacterial twin-arginine translocation pathway. *Annu Rev Microbiol,* 60**,** 373-95.

LEE, S., SOWA, M. E., WATANABE, Y. H., SIGLER, P. B., CHIU, W., YOSHIDA, M. & TSAI, F. T. (2003) The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state. *Cell,* 115**,** 229-40.

LEIMAN, P. G., ARISAKA, F., VAN RAAIJ, M. J., KOSTYUCHENKO, V. A., AKSYUK, A. A., KANAMARU, S. & ROSSMANN, M. G. (2010) Morphogenesis of the T4 tail and tail fibers. *Virol J,* 7**,** 355.

LEIMAN, P. G., BASLER, M., RAMAGOPAL, U. A., BONANNO, J. B., SAUDER, J. M., PUKATZKI, S., BURLEY, S. K., ALMO, S. C. & MEKALANOS, J. J. (2009) Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc Natl Acad Sci U S A,* 106**,** 4154-9.

LI, M., LE TRONG, I., CARL, M. A., LARSON, E. T., CHOU, S., DE LEON, J. A., DOVE, S. L., STENKAMP, R. E. & MOUGOUS, J. D. (2012) Structural basis for type VI secretion effector recognition by a cognate immunity protein. *PLoS pathogens,* 8**,** e1002613.

LIN, R. J., CAPAGE, M. & HILL, C. W. (1984) A repetitive DNA sequence, rhs, responsible for duplications within the Escherichia coli K-12 chromosome. *J Mol Biol,* 177**,** 1-18.

LOFERER, H., HAMMAR, M. & NORMARK, S. (1997) Availability of the fibre subunit CsgA and the nucleator protein CsgB during assembly of fibronectin-binding curli is limited by the intracellular concentration of the novel lipoprotein CsgG. *Mol Microbiol,* 26**,** 11-23.

LOSSI, N. S., DAJANI, R., FREEMONT, P. & FILLOUX, A. (2011) Structure-function analysis of HsiF, a gp25-like component of the type VI secretion system, in Pseudomonas aeruginosa. *Microbiology,* 157**,** 3292-305.

LOSSI, N. S., MANOLI, E., FORSTER, A., DAJANI, R., PAPE, T., FREEMONT, P. & FILLOUX, A. (2013) The HsiB1C1 (TssB-TssC) complex of the Pseudomonas aeruginosa type VI secretion system forms a bacteriophage tail sheathlike structure. *J Biol Chem,* 288**,** 7536-48.

LOSSI, N. S., MANOLI, E., SIMPSON, P., JONES, C., HUI, K., DAJANI, R., COULTHURST, S. J., FREEMONT, P. & FILLOUX, A. (2012) The archetype Pseudomonas aeruginosa proteins TssB and TagJ form a novel subcomplex in the bacterial type VI secretion system. *Mol Microbiol,* 86**,** 437-56.

LOWE, C. A., ASGHAR, A. H., SHALOM, G., SHAW, J. G. & THOMAS, M. S. (2001) The Burkholderia cepacia fur gene: co-localization with omlA and absence of regulation by iron. *Microbiology,* 147**,** 1303-14.

MA, A. T., MCAULEY, S., PUKATZKI, S. & MEKALANOS, J. J. (2009a) Translocation of a Vibrio cholerae type VI secretion effector requires bacterial endocytosis by host cells. *Cell Host Microbe,* 5**,** 234-43.

MA, L. S., LIN, J. S. & LAI, E. M. (2009b) An IcmF family protein, ImpLM, is an integral inner membrane protein interacting with ImpKL, and its walker a motif is required for type VI secretion system-mediated Hcp secretion in Agrobacterium tumefaciens. *Journal of bacteriology,* 191**,** 4316-29.

MACINTYRE, D. L., MIYATA, S. T., KITAOKA, M. & PUKATZKI, S. (2010) The Vibrio cholerae type VI secretion system displays antimicrobial properties. *Proc Natl Acad Sci U S A,* 107**,** 19520-4.

MAHENTHIRALINGAM, E., BALDWIN, A. & DOWSON, C. G. (2008) Burkholderia cepacia complex bacteria: opportunistic pathogens with important natural biology. *J Appl Microbiol,* 104**,** 1539-51.

MAHENTHIRALINGAM, E., BALDWIN, A. & VANDAMME, P. (2002) Burkholderia cepacia complex infection in patients with cystic fibrosis. *J Med Microbiol,* 51**,** 533-8.

MAHENTHIRALINGAM, E., URBAN, T. A. & GOLDBERG, J. B. (2005) The multifarious, multireplicon Burkholderia cepacia complex. *Nat Rev Microbiol,* 3**,** 144-56.

MAINA, C. V., RIGGS, P. D., GRANDEA, A. G., 3RD, SLATKO, B. E., MORAN, L. S., TAGLIAMONTE, J. A., MCREYNOLDS, L. A. & GUAN, C. D. (1988) An Escherichia coli vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein. *Gene,* 74**,** 365-73.

MCBRIDE, M. J. & ZHU, Y. (2013) Gliding motility and Por secretion system genes are widespread among members of the phylum bacteroidetes. *Journal of bacteriology,* 195**,** 270-8.

MELCHER, K. (2004) New chemical crosslinking methods for the identification of transient protein-protein interactions with multiprotein complexes. *Curr Protein Pept Sci,* 5**,** 287-96.

MIYATA, S. T., KITAOKA, M., BROOKS, T. M., MCAULEY, S. B. & PUKATZKI, S. (2011) Vibrio cholerae requires the type VI secretion system virulence factor VasX to kill Dictyostelium discoideum. *Infect Immun,* 79**,** 2941-9.

MOGK, A., SCHLIEKER, C., STRUB, C., RIST, W., WEIBEZAHN, J. & BUKAU, B. (2003) Roles of individual domains and conserved motifs of the AAA+ chaperone ClpB in oligomerization, ATP hydrolysis, and chaperone activity. *J Biol Chem,* 278**,** 17615-24.

MOHANTY, A. K. & WIENER, M. C. (2004) Membrane protein expression and production: effects of polyhistidine tag length and position. *Protein Expr Purif,* 33**,** 311-25.

MOTA, L. J., SORG, I. & CORNELIS, G. R. (2005) Type III secretion: the bacteria-eukaryotic cell express. *FEMS Microbiol Lett,* 252**,** 1-10.

MOUGOUS, J. D., CUFF, M. E., RAUNSER, S., SHEN, A., ZHOU, M., GIFFORD, C. A., GOODMAN, A. L., JOACHIMIAK, G., ORDONEZ, C. L., LORY, S., WALZ, T., JOACHIMIAK, A. & MEKALANOS, J. J. (2006) A virulence locus of Pseudomonas aeruginosa encodes a protein secretion apparatus. *Science,* 312**,** 1526-30.

MOUGOUS, J. D., GIFFORD, C. A., RAMSDELL, T. L. & MEKALANOS, J. J. (2007) Threonine phosphorylation post-translationally regulates protein secretion in Pseudomonas aeruginosa. *Nat Cell Biol,* 9**,** 797-803.

MULDER, D. T., COOPER, C. A. & COOMBES, B. K. (2012) Type VI secretion system-associated gene clusters contribute to pathogenesis of Salmonella enterica serovar Typhimurium. *Infect Immun,* 80**,** 1996-2007.

MURDOCH, S. L., TRUNK, K., ENGLISH, G., FRITSCH, M. J., POURKARIMI, E. & COULTHURST, S. J. (2011) The opportunistic pathogen Serratia marcescens utilizes type VI secretion to target bacterial competitors. *J Bacteriol,* 193**,** 6057-69.

NAGAI, H. & KUBORI, T. (2011) Type IVB Secretion Systems of Legionella and Other Gram-Negative Bacteria. *Front Microbiol,* 2**,** 136.

NAKANE, D., SATO, K., WADA, H., MCBRIDE, M. J. & NAKAYAMA, K. (2013) Helical flow of surface protein required for bacterial gliding motility. *Proc Natl Acad Sci U S A,* 110**,** 11145-50.

NOUWEN, N., RANSON, N., SAIBIL, H., WOLPENSINGER, B., ENGEL, A., GHAZI, A. & PUGSLEY, A. P. (1999) Secretin PulD: association with pilot PulS, structure, and ion-conducting channel formation. *Proc Natl Acad Sci U S A,* 96**,** 8173-7.

NUCCIO, S. P. & BAUMLER, A. J. (2007) Evolution of the chaperone/usher assembly pathway: fimbrial classification goes Greek. *Microbiol Mol Biol Rev,* 71**,** 551-75.

OGURA, T. & WILKINSON, A. J. (2001) AAA+ superfamily ATPases: common structure--diverse function. *Genes Cells,* 6**,** 575-97.

PALLEN, M., CHAUDHURI, R. & KHAN, A. (2002) Bacterial FHA domains: neglected players in the phospho-threonine signalling game? *Trends Microbiol,* 10**,** 556-63.

PALLERONI, N. J., KUNISAWA, R., CONTOPOULOU, R. & DOUDOROFF, M. (1973) Nucleic Acid Homologies in the Genus Pseudomonas. *INTERNATIONAL JOURNAL of SYSTEMATIC BACTERIOLOGY,* 23**,** 333-339.

PARKE, J. L. & GURIAN-SHERMAN, D. (2001) Diversity of the Burkholderia cepacia complex and implications for risk assessment of biological control strains. *Annu Rev Phytopathol,* 39**,** 225-58.

PARSONS, D. A. & HEFFRON, F. (2005) sciS, an icmF homolog in Salmonella enterica serovar Typhimurium, limits intracellular replication and decreases virulence. *Infect Immun,* 73**,** 4338-45.

PEABODY, C. R., CHUNG, Y. J., YEN, M. R., VIDAL-INGIGLIARDI, D., PUGSLEY, A. P. & SAIER, M. H., JR. (2003) Type II protein secretion and its relationship to bacterial type IV pili and archaeal flagella. *Microbiology,* 149**,** 3051-72.

PELL, L. G., KANELIS, V., DONALDSON, L. W., HOWELL, P. L. & DAVIDSON, A. R. (2009) The phage lambda major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. *Proc Natl Acad Sci U S A,* 106**,** 4160-5.

PEZOA, D., YANG, H. J., BLONDEL, C. J., SANTIVIAGO, C. A., ANDREWS-POLYMENIS, H. L. & CONTRERAS, I. (2013) The type VI secretion system encoded in SPI-6 plays a role in

gastrointestinal colonization and systemic spread of Salmonella enterica serovar Typhimurium in the chicken. *PLoS One,* 8**,** e63917.

PIETROSIUK, A., LENHERR, E. D., FALK, S., BONEMANN, G., KOPP, J., ZENTGRAF, H., SINNING, I. & MOGK, A. (2011) Molecular basis for the unique role of the AAA+ chaperone ClpV in type VI protein secretion. *J Biol Chem,* 286**,** 30010-21.

POHLNER, J., HALTER, R., BEYREUTHER, K. & MEYER, T. F. (1987) Gene structure and extracellular secretion of Neisseria gonorrhoeae IgA protease. *Nature,* 325**,** 458-62.

POHLSCHRODER, M., PRINZ, W. A., HARTMANN, E. & BECKWITH, J. (1997) Protein translocation in the three domains of life: variations on a theme. *Cell,* 91**,** 563-6.

POTTS, J. T., JR., YOUNG, D. M. & ANFINSEN, C. B. (1963) Reconstitution of fully active RNase S by carboxypeptidase-degraded RNase S-peptide. *J Biol Chem,* 238**,** 2593-4.

POTVIN, E., LEHOUX, D. E., KUKAVICA-IBRULJ, I., RICHARD, K. L., SANSCHAGRIN, F., LAU, G. W. & LEVESQUE, R. C. (2003) In vivo functional genomics of Pseudomonas aeruginosa for high-throughput screening of new virulence factors and antibacterial targets. *Environ Microbiol,* 5**,** 1294-308.

PUGSLEY, A. P. (1993) The complete general secretory pathway in gram-negative bacteria. *Microbiol Rev,* 57**,** 50-108.

PUKATZKI, S., MA, A. T., REVEL, A. T., STURTEVANT, D. & MEKALANOS, J. J. (2007) Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *Proc Natl Acad Sci U S A,* 104**,** 15508-13.

PUKATZKI, S., MA, A. T., STURTEVANT, D., KRASTINS, B., SARRACINO, D., NELSON, W. C., HEIDELBERG, J. F. & MEKALANOS, J. J. (2006) Identification of a conserved bacterial protein secretion system in Vibrio cholerae using the Dictyostelium host model system. *Proc Natl Acad Sci U S A,* 103**,** 1528-33.

RAO, V. A., SHEPHERD, S. M., ENGLISH, G., COULTHURST, S. J. & HUNTER, W. N. (2011) The structure of Serratia marcescens Lip, a membrane-bound component of the type VI secretion system. *Acta Crystallogr D Biol Crystallogr,* 67**,** 1065-72.

REIK, R., SPILKER, T. & LIPUMA, J. J. (2005) Distribution of Burkholderia cepacia complex species among isolates recovered from persons with or without cystic fibrosis. *J Clin Microbiol,* 43**,** 2926-8.

RICHARDS, F. M. (1955) An assay for ribonuclease activity using a cyclic phosphate substrate. *C R Trav Lab Carlsberg Chim,* 29**,** 315-21.

RICHARDS, F. M. & VITHAYATHIL, P. J. (1959) The preparation of subtilisn-modified ribonuclease and the separation of the peptide and protein components. *J Biol Chem,* 234**,** 1459-65.

RICHARME, G. & CALDAS, T. D. (1997) Chaperone properties of the bacterial periplasmic substrate-binding proteins. *J Biol Chem,* 272**,** 15607-12.

ROEST, H. P., MULDERS, I. H. M., SPAINK, H. P., WIJFFELMAN, C. A. & LUGTENBERG, B. J. J. (1997) A Rhizobium leguminosarum Biovar trifolii Locus Not Localized on the Sym Plasmid Hinders Effective Nodulation on Plants of the Pea Cross-Inoculation Group.

ROSALES-REYES, R., AUBERT, D. F., TOLMAN, J. S., AMER, A. O. & VALVANO, M. A. (2012a) Burkholderia cenocepacia type VI secretion system mediates escape of type II secreted proteins into the cytoplasm of infected macrophages. *PLoS One,* 7**,** e41726.

ROSALES-REYES, R., SKELDON, A. M., AUBERT, D. F. & VALVANO, M. A. (2012b) The Type VI secretion system of Burkholderia cenocepacia affects multiple Rho family GTPases disrupting the actin cytoskeleton and the assembly of NADPH oxidase complex in macrophages. *Cell Microbiol,* 14**,** 255-73.

ROSSMANN, M. G., MESYANZHINOV, V. V., ARISAKA, F. & LEIMAN, P. G. (2004) The bacteriophage T4 DNA injection machine. *Curr Opin Struct Biol,* 14**,** 171-80.

RUSSELL, A. B., HOOD, R. D., BUI, N. K., LEROUX, M., VOLLMER, W. & MOUGOUS, J. D. (2011) Type VI secretion delivers bacteriolytic effectors to target cells. *Nature,* 475**,** 343-7.

RUSSELL, A. B., LEROUX, M., HATHAZI, K., AGNELLO, D. M., ISHIKAWA, T., WIGGINS, P. A., WAI, S. N. & MOUGOUS, J. D. (2013) Diverse type VI secretion phospholipases are functionally plastic antibacterial effectors. *Nature,* 496**,** 508-12.

RUSSELL, A. B., SINGH, P., BRITTNACHER, M., BUI, N. K., HOOD, R. D., CARL, M. A., AGNELLO, D. M., SCHWARZ, S., GOODLETT, D. R., VOLLMER, W. & MOUGOUS, J. D. (2012) A widespread bacterial type VI secretion effector superfamily identified using a heuristic approach. *Cell Host Microbe,* 11**,** 538-49.

S.SHASTRI (2011) Chacterisation of the type VI secretion system of Burkholderia cenocepacia.

SAJJAN, S. U., CARMODY, L. A., GONZALEZ, C. F. & LIPUMA, J. J. (2008) A type IV secretion system contributes to intracellular survival and replication of Burkholderia cenocepacia. *Infect Immun,* 76**,** 5447-55.

SANDKVIST, M. (2001) Type II secretion and pathogenesis. *Infect Immun,* 69**,** 3523-35.

SARGENT, F., BERKS, B. C. & PALMER, T. (2006) Pathfinders and trailblazers: a prokaryotic targeting system for transport of folded proteins. *FEMS Microbiol Lett,* 254**,** 198-207.

SARGENT, F., BOGSCH, E. G., STANLEY, N. R., WEXLER, M., ROBINSON, C., BERKS, B. C. & PALMER, T. (1998) Overlapping functions of components of a bacterial Sec-independent protein export pathway. *EMBO J,* 17**,** 3640-50.

SATCHELL, K. J. (2009) Bacterial martyrdom: phagocytes disabled by type VI secretion after engulfing bacteria. *Cell Host Microbe,* 5**,** 213-4.

SATO, K., YUKITAKE, H., NARITA, Y., SHOJI, M., NAITO, M. & NAKAYAMA, K. (2013) Identification of Porphyromonas gingivalis proteins secreted by the Por secretion system. *FEMS Microbiol Lett,* 338**,** 68-76.

SAUER, F. G., FUTTERER, K., PINKNER, J. S., DODSON, K. W., HULTGREN, S. J. & WAKSMAN, G. (1999) Structural basis of chaperone function and pilus biogenesis. *Science,* 285**,** 1058-61.

SAUER, R. T., BOLON, D. N., BURTON, B. M., BURTON, R. E., FLYNN, J. M., GRANT, R. A., HERSCH, G. L., JOSHI, S. A., KENNISTON, J. A., LEVCHENKO, I., NEHER, S. B., OAKES, E. S., SIDDIQUI, S. M., WAH, D. A. & BAKER, T. A. (2004) Sculpting the proteome with AAA(+) proteases and disassembly machines. *Cell,* 119**,** 9-18.

SAUVONNET, N., VIGNON, G., PUGSLEY, A. P. & GOUNON, P. (2000) Pilus formation and protein secretion by the same machinery in Escherichia coli. *EMBO J,* 19**,** 2221-8.

SCHELL, M. A., ULRICH, R. L., RIBOT, W. J., BRUEGGEMANN, E. E., HINES, H. B., CHEN, D., LIPSCOMB, L., KIM, H. S., MRAZEK, J., NIERMAN, W. C. & DESHAZER, D. (2007) Type VI secretion is a major virulence determinant in Burkholderia mallei. *Mol Microbiol,* 64**,** 1466-85.

SCHLIEKER, C., ZENTGRAF, H., DERSCH, P. & MOGK, A. (2005) ClpV, a unique Hsp100/Clp member of pathogenic proteobacteria. *Biol Chem,* 386**,** 1115-27.

SCHUCK, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J,* 78**,** 1606-19.

SCHUCK, P. (2003) On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal Biochem,* 320**,** 104-24.

SEGAL, G., FELDMAN, M. & ZUSMAN, T. (2005) The Icm/Dot type-IV secretion systems of Legionella pneumophila and Coxiella burnetii. *FEMS Microbiol Rev,* 29**,** 65-81.

SETTLES, A. M., YONETANI, A., BARON, A., BUSH, D. R., CLINE, K. & MARTIENSSEN, R. (1997) Sec-independent protein translocation by the maize Hcf106 protein. *Science,* 278**,** 1467-70.

SEXTON, J. A., MILLER, J. L., YONEDA, A., KEHL-FIE, T. E. & VOGEL, J. P. (2004) Legionella pneumophila DotU and IcmF are required for stability of the Dot/Icm complex. *Infect Immun,* 72**,** 5983-92.

SHALOM, G., SHAW, J. G. & THOMAS, M. S. (2007) In vivo expression technology identifies a type VI secretion system locus in Burkholderia pseudomallei that is induced upon invasion of macrophages. *Microbiology,* 153**,** 2689-99.

SHEAHAN, K. L., CORDERO, C. L. & SATCHELL, K. J. (2004) Identification of a domain within the multifunctional Vibrio cholerae RTX toxin that covalently cross-links actin. *Proc Natl Acad Sci U S A,* 101**,** 9798-803.

SHEVCHIK, V. E., ROBERT-BAUDOUY, J. & CONDEMINE, G. (1997) Specific interaction between OutD, an Erwinia chrysanthemi outer membrane protein of the general secretory pathway, and secreted proteins. *EMBO J,* 16**,** 3007-16.

SHNEIDER, M. M., BUTH, S. A., HO, B. T., BASLER, M., MEKALANOS, J. J. & LEIMAN, P. G. (2013) PAAR-repeat proteins sharpen and diversify the type VI secretion system spike. *Nature,* 500**,** 350-3.

SILVERMAN, J. M., AGNELLO, D. M., ZHENG, H., ANDREWS, B. T., LI, M., CATALANO, C. E., GONEN, T. & MOUGOUS, J. D. (2013) Haemolysin Coregulated Protein Is an Exported Receptor and Chaperone of Type VI Secretion Substrates. *Mol Cell*.

SILVERMAN, J. M., BRUNET, Y. R., CASCALES, E. & MOUGOUS, J. D. (2012) Structure and regulation of the type VI secretion system. *Annual review of microbiology,* 66**,** 453-72.

SOLBIATI, J., CHAPMAN-SMITH, A., MILLER, J. L., MILLER, C. G. & CRONAN, J. E., JR. (1999) Processing of the N termini of nascent polypeptide chains requires deformylation prior to methionine removal. *J Mol Biol,* 290**,** 607-14.

SOUSA, S. A., RAMOS, C. G. & LEITAO, J. H. (2010) Burkholderia cepacia Complex: Emerging Multihost Pathogens Equipped with a Wide Range of Virulence Factors and Determinants. *Int J Microbiol,* 2011.

SPEERT, D. P. (2001) Understanding Burkholderia cepacia: Epidemiology, Genomovars, and Virulence *Infect Med,* 18.

STATHOPOULOS, C., HENDRIXSON, D. R., THANASSI, D. G., HULTGREN, S. J., ST GEME, J. W., 3RD & CURTISS, R., 3RD (2000) Secretion of virulence determinants by the general secretory pathway in gram-negative pathogens: an evolving story. *Microbes Infect,* 2**,** 1061-72.

STUDIER, F. W. & MOFFATT, B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of molecular biology,* 189**,** 113-30.

SUAREZ, G., SIERRA, J. C., EROVA, T. E., SHA, J., HORNEMAN, A. J. & CHOPRA, A. K. (2010) A type VI secretion system effector protein, VgrG1, from Aeromonas hydrophila that induces host cell toxicity by ADP ribosylation of actin. *Journal of bacteriology,* 192**,** 155-68.

SUAREZ, G., SIERRA, J. C., SHA, J., WANG, S., EROVA, T. E., FADL, A. A., FOLTZ, S. M., HORNEMAN, A. J. & CHOPRA, A. K. (2008) Molecular characterization of a functional type VI secretion system from a clinical isolate of Aeromonas hydrophila. *Microb Pathog,* 44**,** 344-61.

TERPE, K. (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol,* 60**,** 523-33.

THANASSI, D. G. (2002) Ushers and secretins: channels for the secretion of folded proteins across the bacterial outer membrane. *J Mol Microbiol Biotechnol,* 4**,** 11-20.

THANASSI, D. G. & HULTGREN, S. J. (2000) Multiple pathways allow protein secretion across the bacterial outer membrane. *Curr Opin Cell Biol,* 12**,** 420-30.

THANASSI, D. G., STATHOPOULOS, C., KARKAL, A. & LI, H. (2005) Protein secretion in the absence of ATP: the autotransporter, two-partner secretion and chaperone/usher pathways of gram-negative bacteria (review). *Mol Membr Biol,* 22**,** 63-72.

THOMAS, M. S. (2007) Iron acquisition mechanisms of the Burkholderia cepacia complex. *Biometals,* 20**,** 431-52.

TOMICH, M., GRIFFITH, A., HERFST, C. A., BURNS, J. L. & MOHR, C. D. (2003) Attenuated virulence of a Burkholderia cepacia type III secretion mutant in a murine model of infection. *Infect Immun,* 71**,** 1405-15.

TOMLIN, K. L., COLL, O. P. & CERI, H. (2001) Interspecies biofilms of Pseudomonas aeruginosa and Burkholderia cepacia. *Can J Microbiol,* 47**,** 949-54.

VAN DEN BERG, B., CLEMONS, W. M., JR., COLLINSON, I., MODIS, Y., HARTMANN, E., HARRISON, S. C. & RAPOPORT, T. A. (2004) X-ray structure of a protein-conducting channel. *Nature,* 427**,** 36-44.

VAN DER MEEREN, R., WEN, Y., VAN GELDER, P., TOMMASSEN, J., DEVREESE, B. & SAVVIDES, S. N. (2012) New insights into the assembly of bacterial secretins: structural studies of the periplasmic domain of XcpQ from Pseudomonas aeruginosa. *J Biol Chem,* 288**,** 1214-25.

VANDAMME, P., HENRY, D., COENYE, T., NZULA, S., VANCANNEYT, M., LIPUMA, J. J., SPEERT, D. P., GOVAN, J. R. & MAHENTHIRALINGAM, E. (2002) Burkholderia anthina sp. nov. and Burkholderia pyrrocinia, two additional Burkholderia cepacia complex bacteria, may confound results of new molecular diagnostic tools. *FEMS Immunol Med Microbiol,* 33**,** 143-9.

VANDAMME, P., HOLMES, B., COENYE, T., GORIS, J., MAHENTHIRALINGAM, E., LIPUMA, J. J. & GOVAN, J. R. (2003) Burkholderia cenocepacia sp. nov.--a new twist to an old story. *Res Microbiol,* 154**,** 91-6.

VANDAMME, P., HOLMES, B., VANCANNEYT, M., COENYE, T., HOSTE, B., COOPMAN, R., REVETS, H., LAUWERS, S., GILLIS, M., KERSTERS, K. & GOVAN, J. R. (1997) Occurrence of multiple genomovars of Burkholderia cepacia in cystic fibrosis patients and proposal of Burkholderia multivorans sp. nov. *Int J Syst Bacteriol,* 47**,** 1188-200.

VANLAERE, E., BALDWIN, A., GEVERS, D., HENRY, D., DE BRANDT, E., LIPUMA, J. J., MAHENTHIRALINGAM, E., SPEERT, D. P., DOWSON, C. & VANDAMME, P. (2009) Taxon K, a complex within the Burkholderia cepacia complex, comprises at least two novel species, Burkholderia contaminans sp. nov. and Burkholderia lata sp. nov. *Int J Syst Evol Microbiol,* 59**,** 102-11.

VANLAERE, E., LIPUMA, J. J., BALDWIN, A., HENRY, D., DE BRANDT, E., MAHENTHIRALINGAM, E., SPEERT, D., DOWSON, C. & VANDAMME, P. (2008) Burkholderia latens sp. nov., Burkholderia diffusa sp. nov., Burkholderia arboris sp. nov., Burkholderia seminalis sp. nov. and Burkholderia metallica sp. nov., novel species within the Burkholderia cepacia complex. *Int J Syst Evol Microbiol,* 58**,** 1580-90.

VANRHEENEN, S. M., DUMENIL, G. & ISBERG, R. R. (2004) IcmF and DotU are required for optimal effector translocation and trafficking of the Legionella pneumophila vacuole. *Infect Immun,* 72**,** 5972-82.

VERMIS, K., COENYE, T., LIPUMA, J. J., MAHENTHIRALINGAM, E., NELIS, H. J. & VANDAMME, P. (2004) Proposal to accommodate Burkholderia cepacia genomovar VI as Burkholderia dolosa sp. nov. *Int J Syst Evol Microbiol,* 54**,** 689-91.

VON HEIJNE, G. (1985) Signal sequences. The limits of variation. *J Mol Biol,* 184**,** 99-105.

VOULHOUX, R., BALL, G., IZE, B., VASIL, M. L., LAZDUNSKI, A., WU, L. F. & FILLOUX, A. (2001) Involvement of the twin-arginine translocation system in protein secretion via the type II pathway. *EMBO J,* 20**,** 6735-41.

WANG, J., LI, C., YANG, H., MUSHEGIAN, A. & JIN, S. (1998) A novel serine/threonine protein kinase homologue of Pseudomonas aeruginosa is specifically inducible within the host infection site and is required for full virulence in neutropenic mice. *J Bacteriol,* 180**,** 6764-8.

WANG, M., LUO, Z., DU, H., XU, S., NI, B., ZHANG, H., SHENG, X., XU, H. & HUANG, X. (2011) Molecular characterization of a functional type VI secretion system in Salmonella enterica serovar Typhi. *Curr Microbiol,* 63**,** 22-31.

WHITNEY, J. C., CHOU, S., RUSSELL, A. B., BIBOY, J., GARDINER, T. E., FERRIN, M. A., BRITTNACHER, M., VOLLMER, W. & MOUGOUS, J. D. (2013) Identification, structure and function of a novel type VI secretion peptidoglycan glycoside hydrolase effector-immunity pair. *J Biol Chem*.

WILDERMAN, P. J., VASIL, A. I., JOHNSON, Z. & VASIL, M. L. (2001) Genetic and biochemical analyses of a eukaryotic-like phospholipase D of Pseudomonas aeruginosa suggest horizontal acquisition and a role for persistence in a chronic pulmonary infection model. *Mol Microbiol,* 39**,** 291-303.

WILLIAMS, A. J., NORCROSS, A. J., CHANDLER, K. A. & BINGLEY, P. J. (2006) Non-specific binding to protein A Sepharose and protein G Sepharose in insulin autoantibody assays may be reduced by pre-treatment with glycine or ethanolamine. *J Immunol Methods,* 314**,** 170-3.

WILLIAMS, S. G., VARCOE, L. T., ATTRIDGE, S. R. & MANNING, P. A. (1996) Vibrio cholerae Hcp, a secreted protein coregulated with HlyA. *Infect Immun,* 64**,** 283-9.

WORRALL, L. J., LAMEIGNERE, E. & STRYNADKA, N. C. (2011) Structural overview of the bacterial injectisome. *Curr Opin Microbiol,* 14**,** 3-8.

WYCKOFF, H. W., HARDMAN, K. D., ALLEWELL, N. M., INAGAMI, T., JOHNSON, L. N. & RICHARDS, F. M. (1967a) The structure of ribonuclease-S at 3.5 A resolution. *J Biol Chem,* 242**,** 3984-8.

WYCKOFF, H. W., HARDMAN, K. D., ALLEWELL, N. M., INAGAMI, T., TSERNOGLOU, D., JOHNSON, L. N. & RICHARDS, F. M. (1967b) The structure of ribonuclease-S at 6 A resolution. *J Biol Chem,* 242**,** 3749-53.

YABUUCHI, E., KOSAKO, Y., OYAIZU, H., YANO, I., HOTTA, H., HASHIMOTO, Y., EZAKI, T. & ARAKAWA, M. (1992) Proposal of Burkholderia gen. nov. and transfer of seven species of the genus Pseudomonas homology group II to the new genus, with the type species Burkholderia cepacia (Palleroni and Holmes 1981) comb. nov. *Microbiol Immunol,* 36**,** 1251-75.

YANISCH-PERRON, C., VIEIRA, J. & MESSING, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene,* 33**,** 103-19.

YAP, M. L., MIO, K., ALI, S., MINTON, A., KANAMARU, S. & ARISAKA, F. (2010) Sequential assembly of the wedge of the baseplate of phage T4 in the presence and absence of gp11 as monitored by analytical ultracentrifugation. *Macromol Biosci,* 10**,** 808-13.

YAP, M. L., MIO, K., LEIMAN, P. G., KANAMARU, S. & ARISAKA, F. (2009) The baseplate wedges of bacteriophage T4 spontaneously assemble into hubless baseplate-like structure in vitro. *J Mol Biol,* 395**,** 349-60.

ZHANG, H., GAO, Z. Q., WANG, W. J., LIU, G. F., XU, J. H., SU, X. D. & DONG, Y. H. (2013) Structure of the type VI effector-immunity complex (Tae4-Tai4) provides novel insights into the inhibition mechanism of the effector by its immunity protein. *J Biol Chem,* 288**,** 5928-39.

ZHENG, J., HO, B. & MEKALANOS, J. J. (2011) Genetic analysis of anti-amoebae and anti-bacterial activities of the type VI secretion system in Vibrio cholerae. *PLoS One,* 6**,** e23876.

ZHENG, J. & LEUNG, K. Y. (2007) Dissection of a type VI secretion system in Edwardsiella tarda. *Mol Microbiol,* 66**,** 1192-206.

ZIEMER, M. A., MASON, A. & CARLSON, D. M. (1982) Cell-free translations of proline-rich protein mRNAs. *J Biol Chem,* 257**,** 11176-80.

ZOUED, A., DURAND, E., BEBEACUA, C., BRUNET, Y. R., DOUZI, B., CAMBILLAU, C., CASCALES, E. & JOURNET, L. (2013) TssK is a trimeric cytoplasmic protein interacting with components of both phage-like and membrane anchoring complexes of the Type VI secretion system. *J Biol Chem*.