

1 Literature Review

1.1 Overview of Bacterial Toxin-antitoxin (TA) Loci

TA loci were first discovered on low copy number plasmids (Ogura and Hiraga, 1983) where they function to stabilize their host plasmids via the post-segregational killing (PSK) of host cells that have missegregated the plasmid during cell division (Jaffé *et al.*, 1985; Van Melderen *et al.*, 1994; Eberl *et al.*, 1994; Sobecky *et al.*, 1996). Homologues of plasmid-encoded TA loci were subsequently discovered on the prokaryotic chromosome (Masuda *et al.*, 1993; Aizenman *et al.*, 1996). A typical TA locus consists of two genes: the toxin-encoding gene and a gene encoding the cognate antitoxin. The toxin gene encodes a stable toxin protein whereas the antitoxin gene encodes either an untranslated antisense ribonucleic acid (RNA) which binds to the toxin messenger ribonucleic acid (mRNA) and inhibits toxin translation (known as type I TA systems) (Gerdes and Wagner 2007) or a labile antitoxin protein which binds to the toxin protein to neutralize its toxic effect (type II TA systems) (Hayes, 2003). This literature review will only be focusing on type II proteic TA systems. A TA locus is typically bi-cistronic and constitutes an operon with the antitoxin gene usually preceding the toxin gene (Engelberg-Kulka and Glaser, 1999). However, there are some exceptions such as the omega-epsilon-zeta (ω - ϵ - ζ) TA locus which consists of three components: the ω protein that regulates the transcription of the ϵ antitoxin and the corresponding ζ toxin (de la Hoz *et al.*, 2000). In the case of the *higBA* locus, the *higB* toxin gene is located upstream of the *higA* antitoxin gene (Tian *et al.*, 1996a). In most cases, the toxin gene encodes a well-folded and stable toxin protein whereas the antitoxin gene encodes a half-structured and labile antitoxin protein (Van Melderen *et al.*, 1996; Oberer *et al.*, 2002). The N-terminus of the antitoxin usually consists of a deoxyribonucleic acid (DNA)-binding domain while the C-terminus of the antitoxin is less ordered in structure as well as highly flexible when not bound to the toxin and thus,

is likely responsible for its instability and susceptibility to degradation by host proteases (Madl *et al.*, 2006; Oberer *et al.*, 2002; Phillips, 1994; Raumann *et al.* 1994). However, the unstructured C-terminus has propensities for defined conformational structure upon binding to its cognate toxin protein (Oberer *et al.*, 2007).

The antitoxin protein is usually smaller than the toxin protein except for a few like the antitoxins encoded by the *higBA* and *hipBA* TA loci (Engelberg-Kulka and Glaser, 1999; Tian *et al.*, 2006b; Black *et al.*, 1991). The distance between both toxin and antitoxin-encoding genes varies among the various TA loci. However, in most cases, both genes overlap by one nucleotide (in which case the stop codon of the antitoxin gene overlaps with the start codon of the toxin gene) or four nucleotides (Pandey and Gerdes, 2005). Overlapping genes are common in prokaryotes and often suggests translational coupling (Ruiz-Echevarría *et al.*, 1991; 1995a). Studies have shown that both TA genes are usually co-transcribed and the proteins are co-expressed to form a tight complex (Engelberg-Kulka and Glaser, 1999). The binding of the antitoxin protein to the toxin protein neutralizes the toxic effect of the toxin protein and in return the antitoxin protein is protected from proteolytic attack (Madl *et al.*, 2006; Roberts *et al.*, 1994).

A typical TA operon is negatively autoregulated at the transcriptional level by its own antitoxin protein and co-regulated by its toxin protein (Engelberg-Kulka and Glaser, 1999). In some cases, the DNA binding site of the antitoxin is found to overlap with the -10 element or -35 element of its promoter or even the transcriptional start site and therefore implies that the repression is due to the hindrance of binding of RNA polymerase to the promoter region (Eberl *et al.*, 1992; Magnuson *et al.*, 1996; Tian *et al.*, 2001). The toxin protein itself does not bind to the operator site but in complex with the antitoxin, improves the binding of the antitoxin to the operator and thus enhances further transcriptional repression (Engelberg-Kulka and Glaser, 1999). For some TA

families, their promoters comprise a few copies of direct or indirect repeats and palindrome sequences. Consequently, the toxin protein promotes cooperative binding of more antitoxin proteins to most if not all the copies of the repeats or palindrome sequences by bridging the antitoxin proteins (Afif *et al.*, 2001; Monti *et al.*, 2007; Madl *et al.*, 2006; Roberts *et al.*, 1993; Overgaard *et al.*, 2009; Magnuson and Yarmolinsky, 1998). However, there are some exceptions where the antitoxin protein alone is sufficient for the autorepression, such as *parDE* (Davis *et al.*, 1992; Eberl *et al.*, 1992; Roberts *et al.*, 1993) and *higBA* of *Vibrio cholerae* (Budde *et al.*, 2007). In the case of the ω - ϵ - ζ TA locus, the regulatory role is not played by the ϵ antitoxin but the ω regulator (de la Hoz *et al.*, 2000). The ratio of toxin and antitoxin is found to be crucial in the regulation instead of the amount of individual proteins. When the antitoxin is in excess, a TA complex will be formed and bound to the promoter to repress its expression. Conversely, when toxin is in excess, the formation of the TA complex will be disrupted thereby giving RNA polymerase access to the promoter, resulting in transcriptional activation of the TA operon (Madl *et al.*, 2006; Overgaard *et al.*, 2008).

The function of plasmid-encoded TA loci is clear: plasmid stability and maintenance by PSK. Once a cell acquires a plasmid which harbours a TA locus, the daughter cells that do not inherit the plasmid copy will be killed via PSK as any remaining antitoxin is not replenished and the antitoxin is rapidly degraded by the host protease thus freeing the toxin protein to exert its killing effect. Therefore, plasmid-encoded TA loci were also known as addiction modules (Yarmolinsky, 1995), as once the cells acquire a TA-encoded plasmid, the cells cannot survive if the plasmid is lost, thus causing the host cells to be “addicted” to the presence of the plasmid. In addition, plasmid-encoded TA loci also function in plasmid-plasmid competition. A conjugative plasmid that harbours a TA locus could outcompete another conjugative plasmid belonging to the same incompatibility group (i.e., identical replicon) since the absence

of a TA locus meant no PSK mechanism in place to prevent the loss of the plasmid following cell division (Cooper and Heinemann, 2000).

The function of chromosomally-encoded TA loci is less clear-cut and still open to debate. Chromosomally-encoded TA loci were first proposed to function in bacterial programmed cell death (PCD) when the *Escherichia coli*-encoded *mazEF* locus was found to mediate cell death under a wide variety of stresses (Engelberg-Kulka *et al.*, 2006). The reasoning behind this hypothesis is that bacteria in the natural environment exist as multicellular colonies displaying coordinated multicellular processes (Shapiro, 1998); therefore, PCD in bacteria allows surviving cells to benefit from scavenging nutrients from dead siblings. However, MazF-induced cell lysis and death remains to be established due to conflicting reports from other laboratories (Van Melderen and Saavedra De Bast, 2009). Research into another *E. coli*-encoded TA locus, *relBE*, showed that activation of the RelE toxin under conditions of nutritional stress led to cell growth arrest (i.e., cell stasis) instead of cell death (Christensen and Gerdes, 2003). TA loci were thus proposed to be stress-response modules that conferred quality control to gene expression and also as regulators of the global rate of translation. As a consequence of toxin activation, cells will remain in a dormant state until favorable growth conditions return (Christensen and Gerdes, 2003). However, growth inhibition was subsequently shown to be independent of the presence of the *relBE*, *mazEF*, *chpBA*, *yefM-yoeB* and *dinJ-yefQ* TA loci in *E. coli*. Moreover, in a competition experiment, the *E. coli* wild-type strains did not have any advantage in their recovery from stressful conditions when compared to a derivative strain that was devoid of the five chromosomal TA loci (Tsilibaris *et al.*, 2007). This casts another doubt that TA loci could confer bacterial fitness and competitiveness under unfavourable conditions.

Interestingly, a solitary *mazF* homologue (*mazF-mx*) was found in *Myxococcus xanthus* which is controlled by a key developmental regulator MrpC via positive

regulation of *mazF-mx* expression and which also served as the antitoxin to MazF-mx (Nariya and Inuoye, 2008). During nutrient starvation, 80% of *M. xanthus* forms fruiting-bodies by lysis and the remaining 20% develops into myxospores. It was found that *mazF-mx* plays an important role in this regulation cascade as $\Delta mazF-mx$ mutants had a drastic reduction in myxospore formation (Nariya and Inuoye, 2008). Besides integrating into the host developmental regulatory network like *mazF-mx*, TA loci also play a part in dynamic genome evolution. Certain TA loci are highly mobile and are preferentially associated with superintegrons (SIs) and genomic islands (Pandey and Gerdes, 2005). It was postulated that TA loci enable these 'selfish' entities to be stabilized, making them refractory to gene efflux (Van Melderen and Saavedra De Bast, 2009). Thus, in specific locations like plasmids or genomic islands, TA loci can function to maintain these structures within the host by PSK. However if the TA locus is integrated into the core genome, they may accumulate mutations and ultimately lose their addictive characteristics through genetic drift (Van Melderen and Saavedra De Bast, 2009). On the other hand, chromosomally-encoded TA loci were also proposed to serve as anti-addiction modules by ejecting any foreign plasmids with a TA locus that is sufficiently identical with the host-encoded homologue. The chromosomally-encoded TA locus can prevent PSK when the foreign plasmid is not inherited following cell division as the chromosomally-encoded antitoxin can counteract the plasmid-encoded toxin (Saavedra De Bast, 2008). In summary, perhaps the functions of TA loci cannot be generalized as they may have multiple biological roles that are dependent on the nature of the toxin, their location on the genome and other yet-to-be-discovered factors (Van Melderen and Saavedra De Bast, 2009).

An exhaustive search of 126 complete prokaryotic genome sequences (110 bacteria and 16 archaea) using standard BLASTP and TBLASTN with cut-off *E*-value of 10^{-4} had identified 671 intact TA loci and 37 solitary toxin genes belonging to the

seven known TA families (*ccd*, *parD/pem*, *vapBC*, *phd/doc*, *parDE*, *higBA* and *relBE*) (Pandey and Gerdes, 2005). *vapBC* was found in remarkably high numbers in both bacterial and archaeal genomes. Similarly, the *relBE* locus was abundant in bacteria and in most of the archaea. The rest of the TA loci were present in both Gram-positive and Gram-negative bacteria except for the *ccd* locus, which was just confined to Gram-negative bacteria. Strikingly, almost all of the TA loci were found in free-living slowly growing organisms and the 31 organisms that were devoid of TA loci were obligate host-associated organisms which thrive in a constant environment and encounter minimal metabolic stress (Pandey and Gerdes, 2005). The few obligate intracellular cells that harbour TA loci were thought to be still undergoing reductive evolution. Interestingly, the survey also revealed that 30 organisms have eight or more TA loci, e.g. *Nitrosomonas europaea* has 43 intact TA loci and two solitary toxin genes. One possible explanation is that the number of TA loci is correlated with the cell growth rate, as slow growing free-living organisms characterized by slow translation rates benefit from having abundant TA loci (Pandey and Gerdes, 2005). The benefit of an organism having multiple TA loci in their genome is still unclear as the wild-type *E. coli* strain has not gained any fitness or competitiveness on the recovery against its derivative that is devoid of five TA loci under various stress conditions (Tsilibaris *et al.*, 2007). The presence of multiple TA loci in an organism also raises the possibility of whether these TA loci can cross-talk (Santos-Sierra *et al.*, 1997). Overexpression of the *Enterococcus faecium* Txe toxin in *E. coli* was partially alleviated by expression of the heterologous *E. coli* YefM protein and likewise, synthesis of the of *E. faecium* Axe (Txe cognate antitoxin) alleviated the toxic effect of *E. coli* YoeB (YefM cognate toxin), indicating the ability of both TA loci to interact with each other *in vivo* (Grady and Hayes, 2003). Similarly, the *Mycobacterium tuberculosis*-encoded YefM antitoxin was able to counteract the *E. coli* YoeB toxin (Kumar *et al.*, 2008). However, there are also

cases where the TA homologues cannot cross-complement each other such as both HigBA homologues in *V. cholera* (Christensen-Dalsgaard and Gerdes, 2006). There is other evidence that proves certain TA loci do interact regardless in a direct or indirect manner: Doc toxin protein by itself does not cleave translated mRNA like the RelE toxin protein, but mediates inhibition of translation by activating endogenous RelE via Lon-dependent decay of the RelB antitoxin protein (Garcio-Pino *et al.*, 2008a).

Some TA loci are located nonrandomly and are clustered in a particular chromosome region: *V. cholera* harbours 13 TA loci all of which resided in the mega-integron of Chromosome II. Analysis of the TA loci and the *attC* sites suggested that TA loci from *V. cholera* are mobile cassettes that undergo frequent movement and gene duplication. Observation of other cases also raises the possibility that TA loci are mobile elements that undergo rapid evolution and horizontal gene transfer (Pandey and Gerdes, 2005).

1.2 Proteic TA Loci and Their Classification

Type II proteic TA loci have been classified into nine families based on toxin amino sequence homology (Table 1) (Pandey and Gerdes, 2005; Van Melderen and Saavedra De Bast, 2009). Some toxins from different families share similar modes of action despite the differences in their amino acid sequences. The details of each toxin family are described in the following sections.

Table 1: The nine toxins with their cognate antitoxins. The target of the toxin and the cellular process that affected by their activities are shown (Table adapted and modified from Van Melderen and Saavedra De Bast, 2009).

Toxin	Antitoxin	Target	Activity	Cellular Process
CcdB	CcdA	DNA gyrase	Generates double strand breaks by trapping DNA-gyrase complex.	Replication
RelE	RelB	Translating ribosome	Induces mRNA cleavage at ribosomal A-site	Translation
MazF	MazE	RNAs	Endoribonuclease	Translation
ParE	ParD	DNA gyrase	Generates double strand breaks by poison DNA gyrase	Replication
Doc	Phd	Translating ribosome	Induces mRNAs cleavage by activating RelBE TA locus	Translation
VapC	VapB	RNAs	Endoribonuclease	Not determined
ζ	ε	Not determined	Phosphotransferase	Not determined
HipA	HipB	EF-Tu	Protein kinase	Translation
HigB	HigA	Translating ribosome	mRNA cleavage	Translation

1.2.1 The *ccdAB* TA locus

ccd was the first TA locus to be discovered and is located on the F plasmid, which is a low copy number plasmid in *E. coli* (Ogura and Hiraga, 1983; Jaffé *et al.*, 1985). The *ccd* family is mostly confined to γ -proteobacteria and only a small number of chromosomally-encoded homologues have been identified and studied so far (Saavedra De Bast *et al.*, 2008; Wilbaux *et al.*, 2007). Originally, *ccd* stood for couples cell division but today, *ccd* is an acronym for control cell death (Jaffé *et al.*, 1985). The *ccd* locus consists of two genes, *ccdB* (also known as *G*, *lynB* or *letD*) which encodes a 101-amino acid residue, 11.7 kDa stable toxin, CcdB, and its counterpart *ccdA* (also known as *H*, *lynA* or *letA*) which encodes a 72-amino acid residue, 8.3 kDa unstable antitoxin, CcdA (Bex *et al.*, 1983; Karoui *et al.*, 1983; Miki *et al.*, 1984; Sommer *et al.*, 1985).

The CcdB monomer consists of a major five-stranded N-terminal β -sheet and a C-terminal α -helix. In between two strands of the main β -sheet, a minor three-stranded β -sheet is inserted that sticks out of the molecule as a wing (Loris *et al.*, 1999). The wing sheet of CcdB contains loop residues 41-50, which are protected from cleavage by the LysC protease in the presence of CcdA (Van Melderen *et al.*, 1996). Therefore, this subdomain is postulated to be the recognition site for the CcdA antitoxin (Loris *et al.*, 1999). On the other hand, CcdA consists of two structural modules that have two distinct functions: (i) the N-terminal domain (CcdA_N) forms a well-structured dimeric ribbon-helix-helix arrangement that is responsible for DNA binding; and (ii) the C-terminal region (CcdA_C) is highly flexible in the unbound state and becomes structured upon toxin binding. The lack of structure in CcdA_C is most likely responsible for its instability (Madl *et al.*, 2006). Under normal growth conditions, the CcdA antitoxin inhibits the toxic activity of CcdB toxin by forming a tight complex (Dao-Thi *et al.*, 2000). The relative instability of the CcdA antitoxin protein is due to its susceptibility to

degradation by the *E. coli* Lon ATP-dependent protease and is the basis for the killing of host cells that have missegregated the F plasmid (Van Melderren *et al.*, 1994; 1996). When a bacterium loses the F plasmid, the labile CcdA is not replenished by *de novo* synthesis leaving the stable CcdB toxin unopposed in the cytoplasm (Van Melderren *et al.*, 1994). CcdB is known to exert its lethal effect like quinolone, by trapping DNA-gyrase complexes (Maxwell, 1999; Bernard and Couturier, 1992; Bernard *et al.*, 1993). This causes an obstruction to cellular polymerases (Bernard *et al.*, 1993), leading to double-strand DNA breaks, followed by induction of the SOS response (Karoui *et al.*, 1983) and ultimately cell death.

The *ccd* TA locus is negatively autoregulated at the level of transcription by binding of a CcdA₂-CcdB₂ complex to the operator-promoter region of the operon (Tam and Kline, 1989a; 1989b; De Feyter *et al.*, 1989; Salmon *et al.*, 1994). The antitoxin-DNA interaction is described by the solution structure of CcdA bound to a 12 base pair (bp) operator-promoter-DNA stretch comprising a 6 bp palindromic sequence. CcdA binds to its target DNA by inserting its N-terminal β -sheet into the major groove of DNA. The residues Arg4, Thr6 and Thr8 are involved in the base-specific interactions whereby Arg4 formed most of the hydrogen bonds with a high affinity site (6 bp palindrome 5'-GTATAC-3') (Madl *et al.*, 2006). Moreover, two additional palindromes of 4 bp: 5'-TATA-3', which have lower affinity for CcdA, have also been identified upstream of the high affinity site. The cooperative binding of three CcdA dimers to DNA increased the overall binding strength and specificity for the CcdA operator region. CcdA binds to all these three sites and the bridging between the CcdA dimers is through Val22. The interaction between CcdB and CcdA dimers involves the CcdA_C region, with one dimer of CcdB bridging two dimers of CcdA, which are ideally spaced along the DNA stretch (Madl *et al.*, 2006). When CcdA levels are equal or in excess to those of CcdB, a complex composed of a dimer of CcdA and a dimer of CcdB (CcdA₂-

CcdB₂) is formed (Van Melderren *et al.*, 1996). The affinity of CcdA for its cognate DNA approaches a maximum by formation of a DNA-(CcdA₂)₃-(CcdB₂)₃ complex, which represses transcription *in vivo* (Madl *et al.*, 2006). On the other hand, when CcdB levels exceed those of CcdA, a CcdA₂-(CcdB₂)₂ complex is formed and sterically loosens the protein complex with the DNA (Van Melderren *et al.*, 1996; Madl *et al.*, 2006). As a result, cooperative DNA-binding is not achieved; the binding affinity of CcdA to the operator site decreases and therefore more antitoxin is transcribed. The biological relevance is that when the toxin is in excess, the repression is alleviated and synthesis of the antitoxin is achieved. By ensuring a ratio of antitoxin:toxin of greater than one, this mechanism could prevent the harmful effect of free CcdB toxins (Madl *et al.*, 2006).

1.2.2 The *parDE* TA locus

The 3.2 kb *par* locus was discovered on the 60 kb broad host range, low copy number plasmid pRK2 (also known as RP4) in Gram-negative bacteria. It is known to function to maintain the stability of the RK2 plasmid in its host cells (Gerlitzet *al.*, 1990; Roberts *et al.*, 1990; Saurugger *et al.*, 1986). The RK2/RP4 *par* region comprises five genes organized in two divergently transcribed operons: *parCBA* and *parDE* (Roberts and Helinski, 1992). The *parCBA* operon encodes a system for the resolution of plasmid multimers through site-specific recombination and also a partition system which ensures each daughter cell receives a plasmid copy during cell division whereas the *parDE* operon serves as a backup mechanism that stabilizes the plasmid via PSK of plasmid-free daughter cells (Eberl *et al.*, 1994; Sobecky *et al.*, 1996).

ParE (103 amino acid residues and 12 kDa) is the positively charged toxin, and ParD (83 amino acid residues and 9 kDa) constitutes the negatively charged antidote that neutralizes ParE by forming a tight complex that is also effective in autorepression of the *parDE* operon at the level of transcription (Roberts and Helinski, 1992). In contrast to other TA loci that require the TA complex for full repression of the operon, ParD alone is sufficient for autorepression (Davis *et al.*, 1992; Eberl *et al.*, 1992). ParE is not required for the binding of ParD to the promoter (Roberts *et al.*, 1993), but ParE binds to the *parDE* promoter only in the presence of the ParD protein (Johnson *et al.*, 1996). It is still not known if ParE participates in autoregulation of the operon (Gerdes, 2000).

ParD exists as a homodimer (ParD₂ParE₂) in solution which is stable at high temperature and is able to refold after heat-induced denaturation (Oberer *et al.*, 1999; 2002). Similar to CcdA, the ParD antitoxin consists of two structurally distinct moieties, which are a well-ordered N-terminal half that displays a ribbon-helix-helix fold, which is a member of MetJ/Arc structural superfamily, and an unstructured and highly flexible

C-terminal half (Oberer *et al.*, 2002; Phillips, 1994; Raumann *et al.*, 1994). The unstructured C-terminal half of ParD antitoxin is vulnerable to degradation as unfolded proteins are far more susceptible to proteolytic attack than densely packed regions of well-ordered domains. Unlike the antitoxin CcdA, which is degraded by the ATP-dependent Lon protease in the absence of the corresponding CcdB toxin (Tsuchimoto *et al.*, 1992; Van Melderen *et al.*, 1994), mutated Lon protein in *E. coli* failed to affect the ability of the *parDE* PSK system to stabilize a mini-RK2 plasmid (Roberts *et al.*, 1994). This indicated that the presumed differential decay rate of the ParD antitoxin is dependent on another protease or another mechanism.

Each of the ParD monomers is comprised of 8% β -strands and 37% α -helices (A, B, and C), that are interacted by turns and linker regions (Oberer *et al.* 1999; 2002). The ParD dimer consists of two intertwined monomers, and the N-terminal β -strands of both monomers form an antiparallel β -ribbon across the local twofold axis and two subsequent helices. The β -strands and the helices B and B' contain hydrophobic residues which contribute to the hydrophobic interactions at the interface of the dimer and the hydrophobic residues from the helices A and A' form the intramolecular contacts. The positively charged area at the N-terminus of ParD is proposed to bind to the target DNA whereas the negatively charged region of the C-terminal tail is responsible for the interaction with the basic ParE toxin. Most of the unfolded proteins or domains do not exist as statistically random coils but have a tendency for defined conformational structures upon binding to the molecular target (Oberer *et al.*, 2007). ParD seems to exhibit such features with the less structured C-terminal domain with its propensities for the formation of α -helical regions after binding with its cognate ParE toxin (Oberer *et al.*, 2007).

The transcriptional initiation site of *parDE* was identified using primer extension and a σ^{70} -like promoter structure was located 7 bp upstream of the identified

transcription initiation site (Eberl *et al.*, 1992). Inspection of the DNA sequence revealed a few copies of direct and inverted repeats overlapping the -10 element that could be involved in DNA-repressor interaction (Eberl *et al.*, 1992). By comparing the structure of ParD antitoxin with the protein-DNA structures of other members of the same fold family, it is deduced that at least two ParD dimers bind to the DNA target by inserting their β -sheets to the major groove of the double-stranded DNA (Oberer *et al.*, 2007). ParD dimers repress transcription of *parDE* by binding to a discrete 48 bp sequence within the *parDE* promoter region (Davis *et al.*, 1992; Eberl *et al.*, 1992; Roberts *et al.*, 1993). According to the concentration-dependent model of DNA binding, Oberer and colleagues (2007) proposed that the sites at the inverted and direct repeats are occupied by two or three ParD dimers in a cooperative manner, and the binding of the additional dimers to the flanking regions only occurs when ParD is excess (Roberts *et al.*, 1993). This extensive covering of the whole promoter area and transcriptional start site is important to fine-tune the regulation. Similar to CcdA, the hydrophobic region at the terminal of helix A is the likely potential attachment site for dimer-dimer interactions of the DNA-bound ParD complex (Madl *et al.*, 2006).

ParE toxin causes extensive cell filamentation (Roberts *et al.*, 1994; Johnson *et al.*, 1996). *In vitro*, ParE hinders *E. coli* gyrase in the presence of ATP and converts the supercoiled plasmid DNA to a singly cleaved linear form (Jiang *et al.*, 2002). Nonetheless, it remains to be elucidated whether the inhibition of gyrase-catalysed formation of negative supercoils or the formation of a cleavable complex with gyrase that serves as a barrier to vital DNA processes and/or gives rise to DNA breaks, which leads to the cell killing activity of the ParE toxin. However, addition of the ParD antitoxin can prevent and reverse the inactivation of gyrase by ParE (Jiang *et al.*, 2002).

1.2.3 The *mazEF* and the *parD/pem* TA loci

1.2.3.1 The *mazEF* TA locus

The *E. coli mazEF* (also known as *chpA*) locus was the first characterized regulatable prokaryotic chromosomal addiction module (Aizenman *et al.*, 1996). *mazE* (also known as *chpAI*) and *mazF* (also known as *chpAK*) are located in the *E. coli rel* operon, downstream of the *relA* gene. The *mazEF* locus has all the properties required for an addiction module: MazE is a labile protein degraded by the ATP-dependent ClpPA serine protease and it protects bacterial cells from the toxic effect of the more stable MazF protein. Induction of *mazF* expression causes a reduction in the cell count of about three orders of magnitude after about 20 minutes (min). In M9 medium, induction of *mazF* causes cell lysis. The expression of *mazEF* was shown to be regulated by the cellular levels of guanosine 5'-diphosphate 3'-diphosphate (ppGpp). During amino acid starvation, increased levels of ppGpp lead to inhibition of transcription of *mazEF* triggering PCD (Aizenman *et al.*, 1996).

However, a subsequent study by Pederson *et al.* (2002) had precluded the MazF-mediated cell killing. Instead, overproduction of MazF was found to induce a bacteriostatic condition in which the cells are still viable but unable to proliferate. MazF inhibited both translation and replication and this static condition can be fully reversed by expression of the cognate MazE antitoxin at a later time (Pederson *et al.*, 2002). Therefore, another function of this chromosomal TA locus was proposed, which is to modulate the global rates of macromolecule synthesis during nutritional stress conditions (Pederson *et al.*, 2002). On the other hand, using a similar ectopic overexpression, Amitai *et al.* (2004) showed that overexpression of MazE could reverse MazF lethality but only over a short window of time. The size of that window was also dependent on the nature of the medium in which MazF is overexpressed (Amitai *et al.*, 2004). “A point of no return” occurred sooner in minimal M9 medium than it did in rich

Luria-Bertani medium. In spite of the ability to reverse the inhibitory effect of MazF on translation completely, MazE overproduction could not reverse the bacteriocidal effect of MazF (Amitai *et al.*, 2004).

E. coli mazEF-mediated cell death could be triggered by various stressful conditions. Some antibiotics like rifampicin, chloramphenicol, and spectinomycin, which are general inhibitors of transcription and/or translation, could hinder the expression of the labile MazE antitoxin, thus leading to cell death mediated by the relatively more stable MazF toxin (Sat *et al.*, 2001). Besides, other stressful conditions such as high temperatures, DNA damage (due to thymine starvation, ultraviolet (UV) irradiation, nalidixic acid and mitomycin C exposure) and oxidative stress also induced *mazEF*-mediated cell death (Sat *et al.*, 2003; Hazan *et al.*, 2004). However, *mazEF*-mediated cell death only occurs during logarithmic growth and not during stationary growth, and in a window of mildly stressful conditions. Within this window of conditions, $\Delta mazEF$ derivatives of *E. coli* are relatively resistant to the stresses, but extreme stressful conditions induce cell death in all the $\Delta mazEF$ cells. It is still not clear whether the $\Delta mazEF$ cells die because of the induction of some other TA loci or through the inactivation of some essential component. In agreement with Aizenman *et al.* (1996), *mazEF*-mediated cell death was found to require an intact *relA* gene as *relA1* mutated cells survived under all these treatment conditions (Hazan *et al.*, 2004). Intriguingly, overproduction of MazF kills both *relA*⁺ and *relA1* cells (Hazan *et al.*, 2004). Thus, it seems that ppGpp triggers the *mazEF*-mediated death rather than affecting the toxic action of MazF. Moreover, nonidentical kinetics were obtained for resuscitation of MazE from MazF in various stress conditions. The *mazEF*-mediated cell death could be reversed by the MazE antitoxin within a wider window of time when *mazEF* was triggered by the inhibition of translation (120 to 150 min) than by the inhibition of transcription (30 to 60 min) or by DNA damage (60 to 90 min); but later

on, the ability of MazE antitoxin to reverse the *mazF*-mediated lethality was decreased gradually or drastically in all the stressful conditions studied. Most importantly, in all cases, the MazF lethality was not able to be reversed 210 min after MazF induction, and this irreversible loss of viability is presented as the basic characteristic of cell death (Kolodkin-Gal and Engelberg-Kulka, 2006).

Purified MazF-His₆ is an endoribonuclease, which preferentially cleaves single-stranded RNA between A and C residues at the ACA recognition sequence in a manner independent of ribosomes and thus, inhibits protein synthesis (Zhang *et al.*, 2003b). MazF cleaves RNA at the 5' end of ACA sequences, yielding a 2',3'-cyclic phosphate at one side and a free 5'-OH group at the other. Using DNA-RNA chimeric substrates containing XACA, the 2'-OH group of residue X was found to be absolutely essential for MazF cleavage while all the other residues may be deoxyriboses. Therefore, MazF exhibits exquisite site specificity and has utility as an RNA-restriction enzyme for RNA structural studies or as an mRNA interferase to regulate cell growth in prokaryotic and eukaryotic cells (Zhang *et al.*, 2005). Purified MazF inhibits protein synthesis in both prokaryotic and eukaryotic cell-free systems, and this inhibition is released by MazE (Zhang *et al.*, 2003b). On the other hand, Munoz-Gomez and colleagues (2004) had reported that native MazF protein cleaves both single- and double-stranded RNA with a marked preference for 5'-(U/A)AC-3' sequences in single-stranded RNA. The reason for this discrepancy between the results reported by Munoz-Gomez *et al.* (2004) and Zhang *et al.* (2003b) are unclear. However, Munoz-Gomez and colleagues (2004) used native MazF protein instead of a C-terminal His-tagged MazF (Zhang *et al.*, 2003b), which perhaps could potentially affect the binding of MazF to its natural target. Moreover, *in vivo* experiments carried out by overexpressing the native *mazF* gene indicated that the RNA cleavage by MazF occurred not only at 5'-NAC-3' (Christensen *et al.*, 2003; Zhang *et al.*, 2003b), thus supporting the view that the intrinsic activity of

MazF is subjected to additional controls. The cleavage of double-stranded RNA could be due to the local melting of RNA upon binding of MazF to distal sequences that permits subsequent cleavage, or that the binding of MazF to the double-stranded RNA specifically cleaves the target strand. The relaxed substrate specificity of MazF and the ribosome-independent cleavage manner suggest that the ribonuclease activity of MazF may not be restricted to targeting mRNA, but rather involves the regulation of a large panel of RNA-dependent cellular process (Munoz-Gomez *et al.*, 2004).

There are two promoters, designated P₂ and P₃, which are 13 nucleotides apart upstream of the *mazE* reading frame (Figure 1). The P₂ promoter is about 10-fold stronger than the P₃ promoter. β -galactosidase assays showed that induction of expression of MazE in *trans* led to moderate (40%) inhibition of the *mazEF* promoter activity whereas induction of the expression of the MazEF complex in *trans* led to a much higher level (up to 90%) of inhibition (Marianovsky *et al.*, 2001). The MazEF complex could bind to an “alternating palindrome” found within the promoter region. This alternating palindrome, which is the operator of *mazEF*, could exist in one of two alternative states: its middle part, designated “a”, complemented with either of the outer parts designated “b” or “c” (Figure 1) (Marianovsky *et al.*, 2001). The numerous mutations that were introduced into the “alternating palindrome” did not at all affect the binding efficiency of the MazEF complex, suggesting that the secondary structure of the regulating region is more important than its DNA sequence *per se*. The role of the additional d-e palindrome is unclear. The duplication of the structural elements (promoters or binding sites for auto-regulation) assures that *mazEF* regulation will be adequate even in the case that one of these elements may be destroyed. Besides being autoregulated by its own protein, the expression from the *mazEF* promoters was activated 1.6-fold by factor for inversion stimulation (FIS) protein with the FIS binding site located immediately upstream of the palindrome sequences (Figure 1). In rich

medium, FIS concentration is very high in the early exponential phase, but sharply decreases towards stationary phase (Marianovsky *et al*, 2001). FIS is known to act as a homodimer (Koch and Kahmann, 1986). It has been shown that by binding to the DNA region upstream from a promoter, this homodimer causes the DNA to bend, thus increasing the binding efficiency of the RNA polymerase (Pan *et al.*, 1996). Thus, positive regulation of the *mazEF* promoter by FIS must be maximal under conditions of rapid growth on rich media. The combined presence of two promoters, a complicated palindrome structure, and the FIS binding site permits regulation of expression that is simultaneously safe and dynamic, enabling quick responses to changes in physiological conditions (Marianovsky *et al*, 2001).

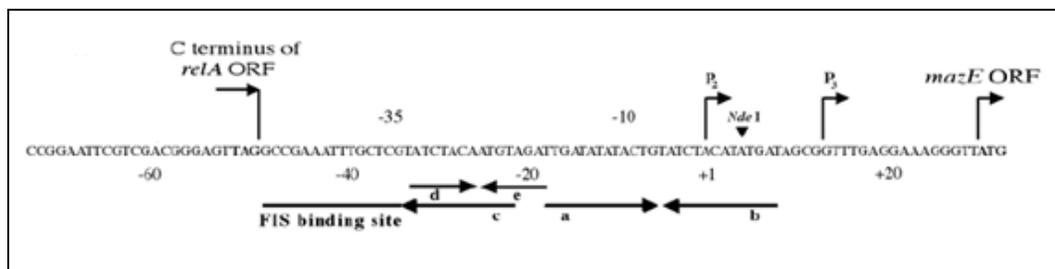


Figure 1: Sequence of the *mazEF* promoter region. The arrows indicate the transcriptional starting sites P_2 and P_3 , the start codon of *mazE* and the stop codon of *relA*. The transcriptional starting point of the P_2 promoter has been designated “+1.” The FIS binding site is underlined and indicated. The arrows labeled d-e, a-b or a-c denote the alternating palindrome sequences that could be formed. (Figure is adapted from Marianovsky *et al*, 2001).

The elucidated crystal structure of the MazEF complex indicated that it forms a stable two:four heterohexamer in solution consisting of alternating MazF and MazE homodimers ($MazF_2$ - $MazE_2$ - $MazF_2$) (Kamada *et al.*, 2003). Electrophoretic mobility shift assays (EMSA) showed that MazF enhanced the binding of purified $(His)_6$ -MazE to the *mazEF* operator site. The binding of MazE to the *mazEF* operator site was through its N-terminal domain, as mutations at the K7A, R8A, S12A, and R16A, which were the conserved amino acid residues at the MazE N-terminal region, disrupted the DNA binding ability of both $(His)_6$ -MazE and the $MazE$ - $MazF$ - $(His)_6$ complex. In

addition, site-directed mutations at the C-terminal domain of MazE (L55A and L58A) disrupted MazE-MazF complex formation (Zhang *et al.*, 2003a), but both N- and C-terminal regions of MazF were postulated to be involved in the interaction with MazE (Zhang *et al.*, 2003a).

Kolodkin-Gal and colleagues (2007) recently showed that the *mazEF*-mediated cell death is related to quorum sensing. Bacteria communicate among themselves via quorum-sensing signal molecules that are also called “autoinducers” (Fuqua *et al.*, 1994). Quorum sensing allows bacteria to sense the presence of each other and to modulate expression of certain genes in response to population density. Kolodkin-Gal *et al.* (2007) showed that *mazEF*-mediated cell death is a population phenomenon that depended on the density of the bacterial culture and death occurred in a dense culture but not in a diluted one. *mazEF*-mediated cell death requires a quorum-sensing peptide that was termed the extracellular death factor (EDF) (Kolodkin-Gal *et al.*, 2007). Structural analysis reveals that EDF is a linear pentapeptide, Asn-Asn-Trp-Asn-Asn (NNWNN), and each of the five amino acids of EDF is important for its activity with the N- and C-terminal residues being the most crucial (Kolodkin-Gal *et al.*, 2007). Surprisingly, from a database search of the *E. coli* genome, only five open reading frames (ORF) with predicted peptide similar to NNWNN were found. The deletion of only two genes prevented the production of an active EDF: *zwf* encoding NNWDN (D = Asp) and *ygeO* encoding NNWN (Kolodkin-Gal *et al.*, 2007). Based on preliminary results, the *zwf* gene product (glucose-6-phosphate-dehydrogenase), with the amino acid sequence NNWDN may be the precursor of EDF, while YgeO has only a secondary role (Kolodkin-Gal *et al.*, 2007). A subsequent amidation step may generate the full NNWNN sequence. On the other hand, although ClpXP is not involved in *mazEF* activation, ClpXP is shown to be crucial for EDF production (Kolodkin-Gal and Engelberg-Kulka, 2008). Therefore, ClpXP is suggested to serve as the Zwf-cleaving

protease which is involved in the generation of EDF (Kolodkin-Gal and Engelberg-Kulka, 2008).

1.2.3.2 The *parD/pem* TA locus

The *parD* (*kis-kid*) TA locus was discovered by serendipity in 1987 during an attempt to isolate conditional replication mutants of a low-copy number R1 plasmid devoid of the *parA* and *parB* maintenance system of enterobacteria (Bravo *et al.*, 1987; Diago-Navarro *et al.*, 2010). A mutation in the *kis* antitoxin gene caused the death of the cells by the action of the *kid* gene (Bravo *et al.*, 1988). The *parD* operon encodes an 85-amino acid residue, 10 kDa Kis (Killer suppressor) antitoxin as well as a 110-amino acid residue, 12 kDa Kid (Killer determinant) toxin (Bravo *et al.*, 1987). Mutation in the gene of the plasmid replication protein, *repA*, that reduced the copy number of the plasmid by two-fold, lead to derepression of *parD* system, which thus increased high plasmid stability. The activation of this plasmid stability operon could be prevented by a suppressor mutation in *copB*, a copy number control gene of plasmid R1, that increased the efficiency of replication of the *repA* mutant (Ruiz-Echevarría *et al.*, 1995b). An identical TA locus, called *pem*, was discovered by another group of researchers in 1988 on plasmid R100, where the antitoxin (PemI) is found to be less stable than the toxin (PemK) due to the susceptibility of PemI to the cleavage by Lon protease (Tsuchimoto *et al.*, 1988; 1992). The PemK toxin inhibited cell growth of the host cells while PemI antitoxin suppressed the inhibitory effect of PemK (Tsuchimoto *et al.*, 1992). The same group also identified two homologues on the chromosome of *E. coli*, named *chpA* (also known as *mazEF*) and *chpB* (Masuda *et al.*, 1993), which are similar to *pem* structurally and functionally (Masuda *et al.*, 1993). Therefore, it was not surprising that the ChpAI and ChpBI antitoxins were able to neutralize the Kid toxin (Santos-Sierra *et al.*, 1997;

1998). Moreover, MazE was also demonstrated to neutralize Kid toxicity to a certain extent (Kamphuis *et al.*, 2007b).

The Kid toxin is a dimer in both solution and in crystal structure. It forms a two-fold symmetry resembling the DNA gyrase inhibitory protein CcdB despite the lack of sequence similarity (11% amino acid sequence similarity). The structure of each monomer is dominated by anti-parallel β -strands flanked by two main α -helices. The loops at the N-terminal region of both monomers, also known as ‘dorsal loops’, are linked by a salt bridge between Glu18 and Arg85. This salt bridge in addition to other charged residues located in the core of the dimer are important for Kid toxicity (Hargreaves *et al.*, 2002; Santos-Sierra *et al.*, 2003). Although Kid and CcdB toxins have strikingly similar tertiary structures, their toxin target is totally different. Comparative genetics studies showed that both toxins used different regions to reach their target (Loris *et al.*, 1999; Santos-Sierra *et al.*, 2003).

Kid preferentially cleaves RNA at the 5' side of the A residue in the nucleotide sequence 5'-UA(A/C)-3' of single-stranded regions (Kamphuis *et al.*, 2007a), although cleavage of the single stranded region and at the 3' side of the A base have also been observed (Munoz-Gomez *et al.*, 2005; Zhang *et al.*, 2004a). The cleavage of a five-nucleotide RNA fragment AUACA by the Kid toxin yielded two fragments with a 2':3'-cyclic phosphate group and a free 5'-OH group. This cleavage mechanism is similar to RNases A and T1, which involves the uracil 2'-OH group (Kamphuis *et al.*, 20067a). Thus, Kid likely inhibits RNA-regulated cellular process in general (Munoz-Gomez *et al.*, 2005).

Sequences upstream of the *parD* operon consisted of two imperfect repeats (regions I and II) of 33 bp each. Region I contains a perfect internal repeat of 9 bp and its left half overlaps with the extended -10 element as well as the transcriptional start site. Region II is located upstream of the -35 element and is separated by 33 bp from

region I. Region II also contains an imperfect inverted repeat of 9 bp (Monti *et al.*, 2007). Kis interacts with the *parD* promoter with low affinity. Kid alone does not bind to the promoter region, but it enhances binding of Kis to the promoter. Kis and the Kis-Kid complex interact with both Regions I and II but with higher affinity to Region I. The binding of Kis and the Kis-Kid complex to Region I which overlaps with the -10 element suggests that these proteins negatively autoregulate their own promoter. The lower affinity of Kis and the Kis-Kid complex to Region II is perhaps due to the four non-conserved residues within the element (Monti *et al.*, 2007). The affinity of the binding was found to be dependent on the molar ratio of Kis and Kid (Kamphuis *et al.*, 2006). Kis and Kid can form multiple complexes with different stoichiometries depending on the molar ratio of Kis and Kid. When Kid is in excess of Kis, a Kid₂-Kis₂-Kid₂ hexamer is most abundant and the binding of this complex to the operator-promoter is weak. When an equal molar ratio of Kis and Kid were added to the promoter site, a tight cooperative binding between the DNA and the Kid₂-Kis₂-Kid₂-Kis₂ octamer was observed. Kis-Kid hexamer interacts with the two half-sites of Regions I and II using one Kis dimer whereas the Kis-Kid octamer can interact with the two half-sites using two Kis dimers. Thus, this explains the more efficient binding of the Kis-Kid octamer to the operator-promoter. The protection pattern was similar when Kis is in excess of Kid. Addition of extra Kid will weaken the interaction of the operator-promoter-Kis-Kid octamer complex (Monti *et al.*, 2007).

The toxicity of Kid was also tested on eukaryotic cells. Kid was demonstrated to inhibit cell proliferation in *Saccharomyces cerevisiae*, *Xenopus laevis* and human cells, while expression of the Kis antitoxin protected against Kid toxicity (de la Cueva-Méndez *et al.*, 2003). In addition, regulated expression of Kid can trigger apoptosis in human cells, and may be a useful approach to kill tumour cells (de la Cueva-Méndez *et al.*, 2003). Besides somatic cells, Kid also inhibited the growth of embryonic cells. This

approach could have a value in studies of differentiation, organogenesis or degenerative disorders (de la Cueva-Méndez *et al.*, 2003).

1.2.4 The *vapBC* TA locus

The *vapBC* (virulence associated protein) locus was first identified on a *Salmonella dublin* virulence plasmid where mutations and inactivation of *vapB* were found to result in impaired plasmid stability and loss of plasmid-mediated virulence (Pullinger and Lax, 1992). Subsequently, chromosomal *vapBC* loci from *Dichelobacter nodosus* and *Leptospira interrogans*, the etiologic agent of leptospirosis, have also been reported to be functional TA loci (Bloomfield *et al.*, 1997; Zhang *et al.*, 2004b). Another enteric virulence plasmid, pMYS6000 of *Shigella flexneri*, was also stabilized by a *vapBC* homologue designated *mvpAT* (maintenance of virulence plasmid) (Sayeed *et al.*, 2000). A survey of sequenced microbial genomes indicated more than 250 *vapBC* homologues in 126 prokaryotic organisms and they are surprisingly abundant in the archaea where more than 20 *vapBC* loci have been identified in *Archaeoglobus fulgidus* and *Sulfolobus tokodaii* (Pandey and Gerdes, 2005). The hyperthermophilic crenarchaeon *Sulfolobus solfataricus* encodes at least 26 *vapBC* loci in its genome. VapC is a PilT N-terminus (PIN) domain protein with putative ribonuclease activity, while the VapB antitoxin is a proteolytically labile protein, which purportedly functions to silence the VapC toxin when associated as a cognate pair. Global transcriptional analysis of *S. solfataricus* heat-shock response dynamics (temperature shift from 80°C to 90°C) revealed that several *vapBC* genes were triggered by the thermal shift, suggesting a role in the heat-shock-response. Indeed, knocking out a specific *vapBC* locus in *S. solfataricus* substantially changed the transcriptome and, in one case, rendered the crenarchaeon heat-shock-labile (Cooper *et al.*, 2009).

Most VapB antitoxins contain a SpoVT/AbrB DNA binding domain and, as such, belong to the superfamily of transcriptional regulators of the same name. AbrB, which has been studied extensively in *Bacillus subtilis* and *Bacillus anthracis*, is a transition-state regulator (Strauch *et al.*, 2005; Vaughn *et al.*, 2000; Vaughn *et al.*,

2001). SpoVT, an AbrB homologue, was shown to regulate expression of at least 15 genes, probably via DNA-binding interactions with target promoters (Bagyan *et al.*, 1996). The VapC toxin is characterized by a PIN domain, a domain homologous with the N-terminal domain of the pilin biogenesis protein Pil-T (Anantharaman and Aravind, 2003). The PIN domain protein family has similarities to the nuclease domains of *Taq* polymerase, T4 RNase H, and the 5'-3' flap endonucleases (Arcus *et al.*, 2005). In eukaryotes, PIN domain proteins are ribonucleases involved in nonsense-mediated mRNA decay and RNA interference (Clissold and Ponting, 2000). PIN domains could provide clues to the cellular targets of VapC toxins, but this connection has yet to be made experimentally. Generally, VapC toxins are putative ribonucleases and they have been found to exhibit endonuclease activity in mycobacteria and exonuclease activity in the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum* (Arcus *et al.*, 2004; 2005), although their precise specificity is not clear. Furthermore, a VapC from *Haemophilus influenzae* was determined to be ribonucleolytic, degrading free RNA *in vitro* (Daines *et al.*, 2007).

Nontypeable *H. influenzae* carries two chromosomal *vapBC* operons designated *vapBC-1* and *vapBC-2* (Daines *et al.*, 2007), but only *vapBC-1* was studied in detail. Expression of the VapC-1 toxin in *E. coli* led to growth inhibition which was counteracted by co-expression of the VapB-1 antitoxin. VapC-1 displayed MazF-like activity, that is active on free RNA *in vitro*, and purified VapB-1 was able to inhibit this activity (Daines *et al.*, 2007). Interestingly, no *mazEF* homologue is found in *H. influenzae*, whereas no *vapBC* homologue is found in *E. coli*. This mutual exclusivity was further supported by a survey of 23 separate genomes whereby 20 of these genomes were discovered to contain either a *mazEF* locus or a *vapBC* locus but not both (Zhang *et al.*, 2004b). The VapC-1 toxin contains a PIN domain that is about 100 amino acids and with two nearly invariant Asp residues, which have been shown to play an

important role for metal ion coordination in other PIN domain-containing ribonucleases proteins (Fatica *et al.*, 2004).

A *vapBC* homologue in the soil bacterium *Sinorhizobium meliloti*, designated *ntrPR*, was the first characterized functional TA locus in Rhizobiaceae: expression of the NtrR toxin in *E. coli* led to growth inhibition which was alleviated by the NtrP antitoxin; both genes were co-transcribed and the *ntrPR* operon was negatively autoregulated by the NtrP-NtrR protein complex (Bodogai, *et al.*, 2006). *S. meliloti* develops symbiotic interactions with leguminous host plants under conditions of nitrogen deprivation. This results in the formation of root nodules in which *S. meliloti* reduces atmospheric nitrogen to ammonia for the utilization of the host plant, which in turn, provides carbon sources and energy for the bacteria (Bodogai *et al.*, 2006). *S. meliloti* infecting the host plant root encounters stressful conditions which elicit physiological changes necessary for the transition from the free-living to the symbiotic state. Therefore the *ntrPR* module was suggested to contribute to the adjustment of metabolic processes under such conditions (Bodoga, *et al.*, 2006). The upstream region of *ntrP* contains a palindrome sequence, 5'-CAGATCATATC-TGATCTG-3', and a direct repeat, 5'-GGCATATACATTTA-GGCATATACA-3' (Figure 2). The DNA binding site of the NtrP-NtrR protein is delimited to the direct repeat sequences but the perfect palindrome sequence was not protected by the NtrP-NtrR complex in the DNase I footprint experiment (Figure 2). It was postulated that the repression was due to interference of RNA polymerase binding as the second repeat overlaps the transcription start site (Bodogai *et al.*, 2006).

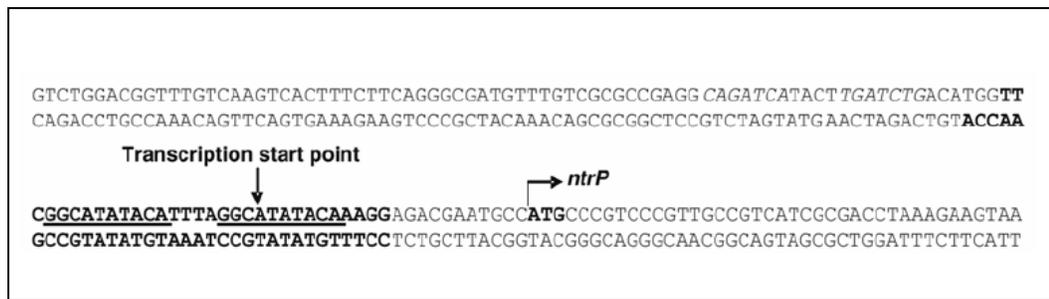


Figure 2: The nucleotide sequence of the *ntrPR* promoter. The transcription start site (+1) and the start codon of *ntrP* are indicated by arrows. Bold letters marks the nucleotides protected by protein binding. The two direct repeats are underlined. The palindrome sequence upstream of the protected region is shown in italic letters (Figure is adapted from Bodogai *et al.*, 2006).

A genetic screen for *Neisseria gonorrhoeae* led to the identification of a VapBC-like TA locus named *fit* (fast intracellular trafficking) (Merz *et al.*, 1996; Merz and So, 2000), which encodes two proteins, FitA and FitB, that play a role in transcellular trafficking and intracellular growth across polarized epithelial monolayers (Hopper *et al.*, 2000). *N. gonorrhoeae* is able to survive and grow within epithelial cells (Hopper *et al.*, 2000), and can also traverse the epithelial monolayer to infect the stomal tissue of the subepithelium (Merz and So, 2000). The immune response to the bacteria in the subepithelium produces the inflammation and purulent discharge characteristic of gonorrhea (Edwards and Apicella, 2004; Holmes *et al.*, 1971). The key to spread of gonococcal disease is where *N. gonorrhoeae* establish a carrier state in which an asymptomatic individual harbors culturable and transmissible bacteria. The organism resides within the epithelial cells instead of crossing into the subepithelium, thus evading the host immune response. The gene products that affect *N. gonorrhoeae* intracellular growth and transcytosis are therefore important for the maintenance of gonococci in the human population. An *N. gonorrhoeae* mutant that lacks *fitAB* has a normal growth rate extracellularly, but displayed accelerated growth during intracellular replication with a concomitant increase in the rate of traversing a monolayer of polarized epithelial cells. Thus, either FitA or FitB, or their complex, was hypothesized

to slow intracellular replication and intracellular trafficking of *N. gonorrhoeae*, inferring a role in *N. gonorrhoeae* virulence (Wilbur *et al.*, 2005).

The *fitA* and *fitB* reading frames are tandemly arranged with a +1 base overlap, i.e. the last base of the *fitA* stop codon serves as the first base of the *fitB* start codon (Wilbur *et al.*, 2005). FitA, the antitoxin protein, binds DNA through its ribbon-helix-helix motifs at its N-terminal and autoregulates its own promoter (Mattison *et al.*, 2006). FitA forms a homodimer in aqueous solution and binds poorly to the DNA sequence TGCTATCA, termed the Fit interaction sequence (FitIS) upstream of the *fitA* ORF. The FitB toxin protein interacts with FitA and increases its DNA-binding affinity and in turn, FitA greatly enhances the solubility of FitB. The FitAB complex binds with a 2.5-fold higher affinity to DNA fragments containing two copies of the FitIS sequence (TGCTATCA-N₁₂-TGATAGCA) than to fragments with a single copy upstream of the *fitA* ORF. The sequence TGCTATCA (FitIS) occurs 14 times in the *N. gonorrhoeae* genome, and the palindromic sequence TGATATCA, termed the Fit perfect palindrome (FitPP) occurs four times. Besides the two that are located within the *fitAB* promoter, 12 are located within other open reading frames and six are found in predicted intergenic regions. These observations raise the question of whether FitAB interacts with these sequences and, if so, whether such an interaction reflects a more global transcriptional regulatory function (Wilbur *et al.*, 2005).

Crystal structure of the FitAB complex was resolved as a tetramer of FitAB heterodimers. The FitB toxin formed a compact domain with an $\alpha/\beta/\alpha$ -fold and showed high degree of structural homology to PIN domain proteins with conservation of the four acidic residues Asp5, Glu42, Asp104 and Asp122 that formed the active site for exonuclease activity. An Arg residue of the FitA antitoxin, Arg68, is located within the FitB acidic pocket, interacting with the carboxyl groups of residues Asp5, Glu42 and

Asp104, and blocking access of potential substrates to the FitB active site (Mattison *et al.*, 2006).

The VapC toxin homologue from the hyperthermophilic crenarchaeon *P. aerophilum*, designated PAE0151, also forms a single compact domain with a stacked $\alpha/\beta/\alpha$ -fold and the conserved four acidic residues Asp6, Glu42, Asp100 and Asp118 (Bunker *et al.*, 2008). The PAE0151 protein was purified without its antitoxin counterpart and its crystal structure revealed a dimer. A partially occupied Mn^{2+} ion with an incomplete coordination sphere was found coordinated by Asp118 and three water molecules, two of which are hydrogen-bonded to Asp100. This is not surprising since PAE0151, as a PIN-domain protein, is predicted to be a divalent metal-dependent ribonuclease (Bunker *et al.*, 2008). Like FitAB, PAE0151 shares minimal sequence similarity but relatively high structural similarity with other PIN-domain proteins (consensus RMSD = 0.92 Å over 52 residues of the alignment across nine structures), particularly at the active site (Bunker *et al.*, 2008). However, ribonuclease activity was not demonstrated for both FitB and PAE0151 (Mattison *et al.*, 2006; Bunker *et al.*, 2008). PAE0152, the product of the reading frame adjacent to PAE0151, has a predicted ribbon-helix-helix motif and is postulated to be the cognate antitoxin for PAE0151 although no sequence similarity with VapB is detected. When mixed *in vitro*, purified PAE0151 and PAE0152 were able to form a protein complex and size exclusion chromatography indicated the formation of a PAE0151₂PAE0152₂ heterotetramer (Bunker *et al.*, 2008).

1.2.5 The *relBE* and *yefM-yoeB* TA loci

1.2.5.1 The *relBE* TA locus

The *relBE* TA locus was originally found on the genome of *E. coli* K12 (Bech *et al.*, 1985; Gotfredsen and Gerdes, 1998). By database searching, *relBE* homologues were found ubiquitously in the genomes of Gram-negative as well as Gram-positive bacteria and more strikingly also in the genome of archaea, and in the enterotoxin-encoding *E. coli* plasmid P307 (Gotfredsen and Gerdes, 1998). The ubiquity of these genes is most likely due to vertical transmission as perhaps an ancestor of *relBE* had existed early in evolution before the branching of bacteria and archaea (Gronlund and Gerdes, 1999). However, the genes may also be horizontally transferred as the *relBE* in *V. cholera* is located within an integron (Mazel *et al.*, 1998) and the *relBE* in *Bacillus thuringiensis* is located on the Tn3-like transposon Tn5401 (Baum, 1994). Most of the RelE toxin homologues are basic and are between 85 and 102 amino acid residues in length. In contrast, the RelB antitoxin homologues are acidic and are between 68 and 98 amino acid residues in length (Gronlund and Gerdes, 1999).

The *E. coli relB* operon encodes three genes: *relB*, *relE* and *relF* (Bech *et al.*, 1985). *relF* encodes a *hok* homologue (renamed as *hokD*) which leads to rapid cessation of cell growth, respiration arrest and collapse of the cell membrane potential (Gerdes *et al.*, 1986). The *relBE* locus on the chromosome of *E. coli* and its homologue, *relBE*_{P307} on *E. coli* plasmid P307, have similar characteristics: *relE/relE*_{P307} encodes a cytotoxic protein which is lethal or inhibitory to the host cells and *relB/relB*_{P307} encodes an antitoxin protein that prevents the lethal effect by direct protein-protein interaction (Gotfredsen and Gerdes, 1998; Galvani *et al.*, 2001); the RelB/ RelB_{P307} antitoxin autoregulates the *relBEF* operon at the transcriptional level and the RelE/ RelE_{P307} toxin serves as a co-repressor; the *relBE* operon stabilizes a mini-R1 (for *relBE*) or mini-P307 replicon (for *relBE*_{P307}) test plasmid in a *E. coli* strain that had been deleted of the

chromosomal *relBEF* operon (Gotfredsen and Gerdes, 1998; Gronlund and Gerdes, 1999). In addition, the RelB_{P307} antitoxin is degraded by Lon protease (Gronlund and Gerdes, 1999) which is induced under adverse conditions such as temperature upshift, carbon starvation and the stringent response (Chung and Goldberg, 1981; Goff *et al.*, 1984; Cashel *et al.*, 1996).

In prokaryotes, the stringent response is defined as the pleiotropic physiological changes triggered by amino acid starvation (Cashel *et al.*, 1996). Amino acid starvation leads to the accumulation of uncharged transfer ribonucleic acids (tRNA) that bind at the A site of the ribosome and stall translation, thereby triggering RelA (ppGpp synthetase I), which in turn results in increased concentration of ppGpp (Cashel *et al.*, 1996). The primary effect of the increase in ppGpp concentration will be the instantaneous downregulation of the transcription of components of the translational apparatus such as stable RNA molecules (ribosomal ribonucleic acid (rRNA) and tRNA) and translation factors, by direct interaction between ppGpp and RNA polymerase. In contrast, the stringent response also provokes stimulation of gene expression. The transcription of amino acid biosynthetic operons will be increased and the building blocks needed for this synthesis are derived from the protein turnover (Cashel *et al.*, 1996). This response is regulated by PolyP, which is generated from Pi by polyphosphate kinase (*ppk*) due to the accumulation of ppGpp. The increase in PolyP activates Lon protease towards a subset of idling ribosome proteins, ultimately leading to the generation of free amino acids, which can be used for *de novo* protein synthesis (Kuroda *et al.*, 2001). Concomitantly, transcription of *relBE* is strongly activated during amino acid starvation (Christensen *et al.*, 2001). This activation is independent of ppGpp but dependent on Lon protease (Christensen and Gerdes, 2004). Consistently, RelB is degraded by Lon protease. A mutation on *relB* confers a 'delayed-relaxed response' characterized by continued RNA synthesis after a lag period of ~10 min

following the onset of amino acid starvation. The *relB101* mutation decreased the stability of RelB, causing its rapid degradation by Lon protease which, in turn, activates RelE (Christensen and Gerdes, 2004). Activated RelE reduces consumption of charged tRNA due to its inhibition of translation activity. This thereby shuts down the activated RelA, and the ppGpp concentration level returns to its prestarvation stage, hence leading to resumption of stable RNA synthesis characteristic of the delayed-relaxed response (Christensen and Gerdes, 2004).

RelE has been shown as a global inhibitor of translation which is activated during nutritional stress (Christensen *et al.*, 2001). The purified RelE protein exhibited ribosome-binding activity *in vitro* supported the observation that RelB is a translation inhibitor (Galvani *et al.*, 2001). RelE does not cleave free RNA but mediated cleavage of mRNA positioned at the ribosomal A site with high codon specificity (Pedersen *et al.*, 2003; Neubauer *et al.*, 2009), except for its homologue in *E. coli*, known as the *ygiUT* loci, where the YgiU toxin cleaved free RNA site-specifically at GC[A/U] (Christensen-Dalsgaard *et al.*, 2010). RelE inhibits translation by mediating cleavage of translated RNAs at both sense (UCG and CAG) and stop codons (UAG>UAA>UGA). Resuscitation of RelE-inhibited cells requires transfer-messenger ribonucleic acid (tmRNA) (also called SsrA or 10 SaRNA) (Muto *et al.*, 1998; Karzai *et al.*, 2000). *E. coli* tmRNA has two known functions: it releases stalled ribosomes from damaged mRNAs and it tags the nascent polypeptides from such ribosomes for proteolysis. tmRNA has a domain that folds very similarly to a part of tRNA^{Ala} and is aminoacylated at its CCA 3' end. It is postulated that tmRNA enters the empty A-site of stalled ribosomes and adds its charged Ala to the nascent polypeptide. tmRNA also encodes a short tag sequence ANDENYALAA (Tu *et al.*, 1995; Keiler *et al.*, 1996). A process called *trans*-translation occurs where the translation shifts from the original mRNA to the tmRNA that contains the ANDENYALAA reading frame. Consequently,

the stalled ribosome is rescued from the damaged mRNA, and the nascent polypeptide along with the tmRNA-encoded peptide tag added to its C-terminus is susceptible to various proteases, resulting in rapid degradation of these proteins (Keiler *et al.*, 1996; Gottesman *et al.*, 1998; Herman *et al.*, 1998). Therefore, tmRNA rescues ribosomes stalled on mRNAs cleaved by RelE. However, tmRNA is also cleaved by RelE, therefore the neutralization of RelE by RelB is important, so that the replenishment of tmRNA can lead to a rapid restart of translation and rescue of cell viability (Christensen and Gerdes, 2003). RelE proteins from Gram-positive bacteria and archaea cleave tmRNA with a pattern similar to that of *E. coli* RelE, suggesting that the function and target of RelE may be conserved across the prokaryotic domains (Christensen and Gerdes, 2003).

The structure of RelBE from the hyperthermophilic archeon *Pyrococcus horikoshii* OT3 (*aRelBE*) showed that the aRelB/RelE complex forms a 2:2 heterotetramer, where aRelE folds into an α/β structure whereas aRelB is present as a polypeptide chain that lacks any distinct hydrophobic core and wraps around aRelE by more than one turn (Takagi *et al.*, 2005). The lack of tertiary structure suggests that aRelB is susceptible to the degradation by ATP-dependent Lon protease. Mutation of Arg85 in the C-terminal region of aRelE abolishes the protein synthesis inhibitory activity whereas mutation of Arg40, Leu48, Arg58 or Arg65 moderately decreases the inhibitory activity (Takagi *et al.*, 2005). All these residues are exposed and located on one side of the nuclease, which interacts with the mRNA, rRNA or both, and taken together with the fact that aRelB does not directly interact with these exposed residues of aRelE, this suggests that aRelB might prevent aRelE from binding to the ribosome by wrapping up the toxin and therefore making it too large to enter the A site (Takagi *et al.*, 2005; Wilson and Nierhaus, 2005). By using circular dichroism and infrared spectroscopy, the *E. coli* RelB and RelE components were determined and found to

share similar secondary structure content with their orthologous archeal protein (Cherny *et al.*, 2007). In addition, RelB exhibits high thermostability with a T_m of 58.5°C, and has a considerable heat resistance, whereas RelE displays lower thermostability with a T_m of 52.5°C, and exhibits exceptional sensitivity to heat. However, the RelBE complex displays higher thermostability (T_m of 64.3°C) and 80% of the RelBE complex remained soluble and was able to be refolded even after exposure to 90°C suggesting that RelB thermodynamically stabilized RelE (Cherny *et al.*, 2007). *In vivo* interaction and protein footprinting experiments showed that the C-terminal region of RelB is responsible for the interaction with RelE. This region is also found to be the protease sensitive part in its free state and upon binding with RelE, is protected from proteolysis (Cherny *et al.*, 2007). All these results indicated that *E. coli* RelB is similar to the ParD and CcdA antitoxins in both folding and thermodynamic properties (Cherny *et al.*, 2007). In the free RelE^{R81A/R83A} structure, helix $\alpha 4$ at the C-terminus adopts a closed conformation contacting with the β -sheet core and adjacent loops. However in the RelE^{R81A/R83A} - RelB_C (Lys47-Leu79) complex, helix $\alpha 3^*$ of RelB_C displaces $\alpha 4$ of RelE^{R81A/R83A} from the binding site on the β -sheet core (Li *et al.*, 2009). The released helix $\alpha 4$ becomes unfolded, adopting an open conformation with increased mobility. This helix replacement results in neutralization of a conserved positively charged cluster of RelE by acidic residues from $\alpha 3^*$ of RelB and disrupts the geometry of critical residues, including Arg81 and Tyr87, in a putative active site of RelE toxin (Li *et al.*, 2009). This indicates that RelB neutralizes the toxic activity of RelE by direct interaction with RelE.

The binding of RelB and RelE to their own promoter has been studied in detail. The *relBE* operator (*relO*) consists of two 6 bp inverted repeats, located within each end of larger 26 bp inverted repeats (Overgaard *et al.*, 2009). A RelB dimer recognizes and binds weakly to a hexad repeat in the palindromic operator though an N-terminal

ribbon-helix-helix motif and RelE enhances the affinity of adjacent bound RelB dimers for the operator elements (Li *et al.*, 2008; Overgaard *et al.*, 2009). By mutational analysis and footprinting assays, RelB was shown to occupy four hexad repeats within *relO* with the core sequences [A/T]TGT[A/C]A. The spacing between each half-site was found to be essential for the cooperative interactions of the RelB dimers stabilized by the co-repressor RelE. The flexible C-terminus of RelB is required for RelB dimers to dimerize (Overgaard *et al.*, 2009). Similar to the *ccdAB* TA locus, transcription of *relBE* is controlled by the RelB/RelE ratio rather than the absolute amounts of the proteins. In an excess of RelB to RelE, two heterotrimeric RelB₂RelE complexes bind cooperatively to the two operator sites of *relO* and repress its transcription (Overgaard *et al.*, 2008). However, when RelE is in excess, cooperative binding of the RelB₂RelE complex to the operator DNA will be counteracted and *relBE* transcription will be stimulated (Overgaard *et al.*, 2008). Therefore RelE regulates *relBE* transcription by conditional cooperativity.

The genome of *M. tuberculosis* harbours three homologues of *relBE* termed *relBE*, *relFG* and *relJK* (Yang *et al.*, 2010). Using a bacterial two-hybrid technique, the RelB, RelF and RelJ antitoxins were found to interact physically with their respective cognate toxins, RelE, RelG and RelK. Surprisingly, surface plasmon resonance assay showed interactions between RelB and RelG, RelB and RelK, RelF and RelE as well as RelF and RelK. In addition, RelB could neutralize its cognate toxin, RelE, and RelG as well. However, RelF could not counteract the toxic effect of RelE although they interacted physically with each other, but on the other hand RelF could enhance the toxicity of RelE. Similar observations were reported for the RelB antitoxin and the RelK toxin where RelB could not neutralize yet could increase the toxicity of RelK. RelF had no effect on RelK although interaction between them had been observed (Yang *et al.*, 2010). EMSA was carried out to determine the regulation of these *M.*

tuberculosis relBE homologues. Only the RelJ antitoxin was found to bind strongly to the promoter region whereas RelB and RelF required their cognate toxins, RelE and RelG, respectively, for binding to the promoter regions with the optimal ratio of two antitoxin:one toxin. Similar patterns were observed for RelB and RelG. Conversely, RelK improved the binding of RelB to its promoter when increasing amounts of RelK were added. On the other hand, the binding of RelF to its promoter region was stimulated by RelE but not RelK (Yang *et al.*, 2010).

The toxicity of RelE in eukaryotic cells has also been explored. Expression of RelB was highly toxic to the yeast *S. cerevisiae*, however, RelB expression could counteract the toxic effect of RelE to some extent (Kristoffersen *et al.*, 2000). Therefore, this TA locus has a potential application as containment control in eukaryotic cells, especially in bioprocess industries in which cell proliferation needs to be controlled in case of the escape of genetically modified cells (Kristoffersen *et al.*, 2000). The *relBE* TA locus has also been tested on human osteosarcoma cells. Induction of RelE caused growth retardation and led to cell death (Yamamoto *et al.*, 2002). Under the light microscope, the morphology of the tetracycline-induced RelE-transformed cells includes membrane budding, chromatin condensation and fragmentation and reduction in cell volume, which all are the characteristics of apoptotic cell death (Yamamoto *et al.*, 2002). RelE lacks any eukaryotic homologue but it has been demonstrated that it efficiently and specifically cleaves mRNA in the A site of eukaryote ribosomes (Andreev *et al.*, 2008). This cleavage mechanism is similar to that in bacteria and therefore suggests the feasibility of A-site cleavage of mRNA for regulatory purposes in eukaryotes (Andreev *et al.*, 2008).

1.2.5.2 The *yefM-yoeB* TA loci

The YefM-YoeB TA loci in the *E. coli* chromosome was originally identified from the homology to the Axe (antitoxin)-Txe (toxin) pair of the *E. faecium* pRUM multidrug-resistant plasmid (Grady and Hayes, 2003). YefM belongs to the Phd family of antitoxins, whereas YoeB is a homologue of the RelE toxin rather than Doc, the cognate toxin of Phd (Kumar *et al.*, 2008). YoeB is a well-folded protein whereas YefM is a natively unfolded antitoxin, lacking in secondary structure even at low temperatures or in the presence of a stabilizing agent (Cherny and Gazit, 2004). Due to the unfolded state of the protein, a linear determinant rather than a conformational one was presumably being recognized by its toxin counterpart. Arg19 was found to be essential for the binding of YefM to the YoeB toxin (Cherny and Gazit, 2004). In contrast, in the crystallographic and biophysical studies of the YefM homologue found in the genome of *M. tuberculosis*, YefM was not intrinsically unfolded and instead formed a well-defined structure with significant secondary and tertiary structure conformations (Kumar *et al.*, 2008). However the C-terminal polypeptides region of the *M. tuberculosis*-encoded YefM is highly pliable and less conserved. As the C-terminus of YefM is involved in the recognition of its counterpart toxin, the divergence suggests the antitoxin evolved to recognize its cognate toxin specifically. Although there is a low level of sequence conservation between the C-terminus of YefM in *E. coli* and *M. tuberculosis*, *M. tuberculosis* YefM could however neutralize the *E. coli* YoeB toxin in an *E. coli* strain that is devoid of the *yefM-yoeB* locus (Kumar *et al.*, 2008). However, *E. coli* YefM cannot counteract its counterpart YoeB homologue in *Streptococcus pneumoniae* and vice versa (Christensen-Dalsgaard and Gerdes, 2006).

Fluorescence anisotropy studies showed that purified YefM and YoeB formed a 1:2 stoichiometry complex (Cherny *et al.*, 2005). No confirmation change was observed after YefM-YoeB complex formation, but 50% increment in proteolytic stability was

displayed when YefM complexed with YoeB (Cherny *et al.*, 2005). However, Kamada and Hanaoka (2005) reported a different structure of the YefM-YoeB protein complex. Elucidated crystal structures of the YefM-YoeB complex showed a 2:1 YefM:YoeB ratio instead (Kamada and Hanaoka, 2005). The YoeB monomer forms a compact globular structure, consisting of a five-stranded β sheet and two α helices. The N-terminus of two YefM monomers form a symmetrical dimer and one of the C-terminus interacts with an atypical microbial ribonuclease (RNase) fold of the YoeB monomer whereas the other C-terminus is structurally disordered. A conformational rearrangement of the RNase catalytic site of YoeB, induced by interaction with YefM, was shown when compared with YoeB alone (Kamada and Hanaoka, 2005).

YoeB was shown to inhibit translation initiation. *In vivo* primer extension experiments showed that YoeB causes cleavage of mRNAs at three bases downstream of the initiation codons (Zhang and Inouye, 2009). YoeB alone does not have endoribonuclease activity and was found to be associated with 50S subunits but not tRNA^{fMet}. Using tetracycline and hygromycin B, YoeB was found to bind to the 50S ribosomal subunit in 70S ribosomes and interacts with the ribosomal A site leading to cleavage of the mRNA (Zhang and Inouye, 2009). Thus the 3' end portion of the mRNA was released from ribosomes and translation initiation was inhibited (Zhang and Inouye, 2009). In addition, the YoeB toxin has also been shown to have *in vitro* RNase activity that preferentially cleaves at the 3' end of purine ribonucleotides (Kamada and Hanaoka, 2005).

Investigations into the regulation of the *E. coli yefM-yoeB* locus using β -galactosidase assays indicated the presence of a strong promoter upstream of *yefM*. YefM is a transcriptional autorepressor in *trans* and YoeB serves as a co-repressor in *trans* (Kedzierska *et al.*, 2007). The operator site 5' of the *E. coli yefM-yoeB* locus comprises adjacent long and short palindromes with core 5'-TGTACA-3' motifs.

DNaseI footprinting assays showed that *E. coli* YefM binds to the long palindrome, followed sequentially by short palindrome recognition when the amount of YefM added was increased. The preferential binding of YefM to the long palindrome might be facilitated by the additional palindromic nucleotides that flank the hexameric core sequence which is absent in the short repeat. The *E. coli* YefM-YoeB complex recognizes both motifs more avidly with less amounts of protein needed, implying that YoeB improves DNA binding by YefM either by enhancing the stability of YefM or altering the YefM conformation to one which is more favorable for DNA binding (Kedzierska *et al.*, 2007). In addition, scanning mutagenesis demonstrated that the short repeat is crucial for correct interaction of the *E. coli* YefM-YoeB with its operator site *in vivo* and *in vitro*. Circular dichroism spectra indicated that YefM and the YefM-YoeB complex undergo structural transitions when bound to DNA. In contrast, the operator DNA does not undergo major structural shifts when bound by its cognate proteins. However, the presence of a DNaseI hypersensitive cleavage site in the YefM- and YefM-YoeB-operator complexes suggests that the operator within the nucleoprotein complex may undergo deformation as DNA conformational changes such as DNA bending, major groove opening and kinking are not uncommon in repressor-operator interactions (Kedzierska *et al.*, 2007).

1.2.6 The *phd-doc* TA locus

Bacteriophage P1 lysogenizes *E. coli* as a low-copy-number plasmid (Ikeda and Tomozawa, 1968). There are a few mechanisms that contribute to the maintenance of the P1 plasmid prophage: replication (Chattoraj *et al.*, 1984), dimer resolution (Austin *et al.*, 1981), plasmid partitioning (Abeles *et al.*, 1985; Austin *et al.*, 1982) and PSK of plasmid-free segregants by TA loci (Lehnherr *et al.*, 1993). There are two P1-encoded TA proteins: a stable 126-amino acid toxin, Doc (causing death on curing) and an unstable 73-amino acid antitoxin, Phd (prevents host death while the plasmid is retained). The corresponding genes form an operon in which *phd*, the antitoxin gene, precedes *doc*, the toxin gene, with the last codon of the antitoxin gene overlapping with the start codon of the toxin gene (Lehnherr *et al.*, 1993). The *phd-doc* locus stabilizes P1 approximately sevenfold (Lehnherr *et al.*, 1993). The Phd antitoxin is degraded by the ATP-dependent ClpXP protease (Lehnherr and Yarmolinsky, 1995).

The *phd-doc* TA locus is negatively autoregulated by both the Phd antitoxin and the Phd-Doc TA complex. Transcriptional fusion assays showed that the *lacZ* reporter was repressed about 10-fold when *phd* was expressed from an exogenous promoter and when *doc* is co-expressed with *phd*, repression of the *lacZ* fusion was enhanced more than 100-fold (Magnuson *et al.*, 1996). The DNA site bound by Phd is 23 bp in length and includes two 10-bp subsites that are roughly palindromic and separated by 3 bp (Magnuson *et al.*, 1996). The palindromic sites are located between the -10 region of the putative promoter and the start codon of *phd*. DNase I footprinting showed that Phd was bound to a perfect 10-bp palindromic DNA sequence and, at higher concentrations, an adjacent, imperfect palindrome (Magnuson *et al.*, 1996). Two Phd monomers bound cooperatively to each left and right operator subsites; Doc did not bind to DNA but improves Phd-DNA binding affinity and mediates cooperative interaction between these two adjacent Phd-binding sites (Magnuson and Yarmolinsky, 1998).

A nontoxic mutant version of Doc interacts physically with Phd in a Phd-Doc trimeric complex and this interaction is probably the molecular basis for the antitoxic effect of Phd (Gazit and Sauer, 1999b; Magnuson and Yarmolinsky, 1998). Using fluorescence resonance energy transfer, the *in vitro* half-life of the trimeric complex was shown to be less than 1 second (s), perhaps indicating the presence of small amounts of free Doc protein *in vivo*. The Phd-Doc interaction was thus viewed as buffering the free concentration of Doc instead of eliminating free Doc from the cell. Therefore, in cells containing the Phd antitoxin, low concentrations of free Doc toxin are likely present both transiently and in steady state, making it improbable for mechanisms of single-hit Doc toxicity (Gazit and Sauer, 1999b).

Expression of the Doc toxin results in rapid cell growth arrest and marks inhibition of translation without perturbation of transcription or replication. However, the Doc toxin does not act like other toxins such as RelE or YoeB, which cleaves mRNA and results in translation inhibition. Instead, Doc toxin mimics the aminoglycoside antibiotic hygromycin B by interacting with 30S ribosomal subunits, stabilizing polysomes, increasing mRNA half-life and resulting in inhibition of translation elongation. The ability of hygromycin B to compete with Doc to bind to the 30S ribosomal subunit, together with the loss of Doc toxicity in a hygromycin B-resistant bacterial strain indicates that the Doc/hygromycin B binding site includes the highly conserved 16S rRNA helix 44 at the 30S–50S interface of the ribosome containing the P and A sites essential for protein translation (Liu *et al.*, 2008).

In contrast, Garcio-Pino *et al.* (2008a) observed the destabilization of two different model mRNAs (*lpp* and *dksA*) after induction of *doc* in *E. coli* strain MG1655. Primer extension analysis after *doc* induction of *lpp* and *dksA* mRNAs revealed cleavage patterns very similar to that induced by RelE. Non-translated versions of the *lpp* and *dksA* mRNAs (in which the start codons were changed to AAG) were not

affected by the induction of *doc*. Thus, like RelE-induced cleavage, Doc-induced mRNA cleavage appeared to be depended on translation (Garcio-Pino *et al.*, 2008a). The combined results obtained from Liu's group and Garcio-Pino's group gave an insight that the cleavage of the translated mRNAs was perhaps due to the activation of an endogenous TA locus, rather than a direct effect of Doc itself. A previous report (Hazan *et al.*, 2001) indicated that Doc induced MazF activity. However, Garcio-Pino and colleagues (2008a) did not observe mRNA cleavage at ACA sites, the signature sequence of MazF-mediated mRNA cleavage (Zhang *et al.*, 2003b) or any influence of deleting *mazEF* on the cleavage pattern (Garcio-Pino *et al.*, 2008a). Instead, expression of Doc in a strain that lacked *relBE* (MG1655 Δ *relBE*) failed to induce Doc-mediated mRNA cleavage, indicating that ectopic production of Doc activates endogenous RelE (Garcio-Pino *et al.*, 2008a). RelE had previously been described to be activated during nutritional stress due to Lon-mediated degradation of RelB antitoxin (Christensen *et al.*, 2001). The Doc-induced cleavage sites are also dependent on Lon (Garcio-Pino *et al.*, 2008a). These results are consistent with the proposal that the Doc-mediated inhibition of translation led to RelE activation via Lon-dependent decay of RelB. Thus, the mRNA decay seen after induction of *doc* appeared to be an indirect consequence of Lon-dependent activation of RelE (Garcio-Pino *et al.*, 2008a).

Mutant DocH66Y was crystallized in complex with a peptide encompassing the C-terminal 22 amino acid residues of Phd with SeMet substituted for Leu52 and Leu70 (Phd^{52-73Se}). The Doc protein showed an all- α -helical fold consisting of six α -helices (Garcia-Pino *et al.*, 2008b). In solution, Phd exists predominantly in an unfolded, random-coil conformation, and DNA binding stabilizes the native Phd fold (Gazit and Sauer, 1999a). Far UV circular dichroism experiments showed that Phd^{52-73Se} is intrinsically unstructured in its isolated state, but gains an appreciable amount of α -helix upon binding to Doc. The C-terminal segment of Phd^{52-73Se} is highly hydrophilic and

provided only a single hydrophobic residue (Leu70 in Phd, SeMet70 in Phd^{52-73Se}) to the binding interface. This residue makes extensive contacts with a small hydrophobic cavity on the Doc surface. The Phd:Doc contact surface in this region showed a high degree of charge complementarities with several negatively charged side chains of Phd^{52-73Se} (Glu55, Asp61 and Asp64) interacting favorably with positive residues in the Phd binding groove of DocH66Y (Arg19 and Arg85). Most striking in this region of Phd^{52-73Se} are the interactions involving Asn67. This residue is completely buried in the interface, its side chain protruding inside a small hydrophilic pocket where it makes complementary hydrogen bonds with the side chain of Asn16 and Asn78 of Doc. It should be noted here that although the C-terminus of Phd^{52-73Se} is adjacent to the surface cluster of conserved residues, the conserved sequence motif of Doc is not part of the Phd binding site. This indicates that Phd^{52-73Se} counteracts the toxic activity of Doc either by inducing a conformational change in Doc or by sterically preventing Doc to interact with the ribosome (Garcia-Pino *et al.*, 2008a).

1.2.7 The *higBA* TA locus

The *higBA* (host inhibition of growth) TA locus is found on a low copy kanamycin-resistant plasmid, Rts1, which was originally isolated from *Proteus vulgaris* (Tian *et al.*, 1996b). Rts1 is 217 kb in size and expresses a pleiotropic temperature-sensitive phenotype (Terawaki *et al.*, 1968; 1981). The HigB toxin inhibits segregation of plasmid free cells, and the HigA antitoxin suppresses the HigB function both in *cis* and in *trans* (Tian *et al.*, 1996b). The *hig* phenotype was observed in an *E. coli* strain, SM8, where growth arrest was observed at 42°C when *hig* was cloned into the temperature-sensitive plasmid mini-Rts1 and introduced into SM8, and no growth inhibition was observed in its *lon*- derivative, SM32 (Tian *et al.*, 1996b). This suggests that the Lon protease is likely involved in the degradation of the HigA antitoxin protein (Van Melderren *et al.*, 1994; 1996).

The *higBA* locus is unique when compared with other proteic killer genes described so far because *higBA* exhibits a reversed gene order where the toxin gene (*higB*) is located upstream of the antitoxin gene (*higA*) (Tian *et al.*, 1996a). Besides, the size of the HigB toxin (92 amino acid residues) is also smaller than the antitoxin HigA (104 amino acid residues) (Tian *et al.*, 1996b). The *hig* locus harbours two promoters: a stronger one, termed *Phig*, is located upstream of *higB* as well as a weaker one, *PhigA*, which is identified upstream of *higA* and lies within the *higB* coding region. HigA serves as a repressor to repress the *Phig* promoter whereas HigB serves as a co-repressor to further repress the promoter activity. However, HigA or HigAB complex has no effect on *PhigA* (Tian *et al.*, 1996a). EMSA using the purified HigA* protein (HigA derivative generated by thrombin cleavage that contained two additional amino acid, Gly and Ser, in place of the first Met residue) demonstrated that HigA binds specifically to the *Phig* promoter region (Tian *et al.*, 2001).

The HigA-binding sequence was investigated using DNase I footprinting assay to be a 56-bp sequence that completely covered the -35 and -10 boxes of *Phig* (Tian *et al.*, 2001). Two inverted repeats with a consensus sequence of GTATTACACA(T/C)CGTGTAATAC were found in this region. The inverted repeats are 9-bp sequence in distance, and the center of which was most sensitive to DNase I cleavage. The presence of two inverted repeats in the binding sequence and the identification of a dimeric form of HigA by cross-linking experiments suggest that the protein binds to the *Phig* region as a dimer (Tian *et al.*, 2001).

The *V. cholerae* genome encodes 13 TA loci (seven *relBE*, three *parDE*, two *higBA* and one *phd/doc*), all of which are located within the SI on chromosome II (Pandey and Gerdes, 2005). Each of the 13 TA loci has closely linked *attC* sites, suggesting that the TA loci can be acquired as bona fide integron cassettes. The two *higBA* TA loci of *V. cholerae* encode functional toxins, HigB-1 and HigB-2, and their ectopic expression inhibited the growth of *E. coli*, whereas the functional antitoxins, HigA-1 and HigA-2, counteracted the toxicity of the cognate toxins (Christensen-Dalsgaard and Gerdes, 2006). Ectopic expression of the HigB toxin for three hours led to cell stasis but without any loss of cell viability. However, these two *higBA* loci do not cross-talk, which could be due to the low degree of similarity between the two loci (HigB-1 and HigB-2 are 26% similar while HigA-1 and HigA-2 are 17% similar). Both *higBA* loci stabilize a test-plasmid very efficiently in *E. coli*, raising the possibility that the loci contribute to maintain genetic stability of the *V. cholerae* SI (Christensen-Dalsgaard and Gerdes, 2006).

Promoter analysis in *V. cholerae* and *E. coli* showed that the two *higBA* loci were both transcribed into bi-cistronic mRNAs. Inspection of the DNA sequences upstream of the +1 site of both *higBA* loci revealed the presence of putative -35 and -10 boxes (Figure 3). Interestingly, the transcriptional start site of *higBA-2* coincided

with the A base of the *higB* AUG start codon, suggesting that the *higBA-2* mRNA is leaderless (Figure 3). This suggestion is in agreement with the lack of a functional Shine–Dalgarno sequence upstream of *higB-2* and no other possible in-frame start codons in the *higB-2* reading frame. Leaderless mRNAs are uncommon in Gram-negative bacteria (Moll *et al.*, 2002), suggesting that the leaderless *higBA-1* mRNA may have a specialized function. Otherwise, the leaderless mRNA could have evolved as a consequence of the compact genetic organization of the SI (Christensen-Dalsgaard and Gerdes, 2006). On the other hand, *lacZ* fusions showed very low activity or no significant transcription initiation in the region upstream of (or within) the *higA* genes (Christensen-Dalsgaard and Gerdes, 2006). These data indicated that if a *higA*-specific promoter is present, it has a very minimal promoter activity or under different growth conditions from those that activate the *higB* promoter (Budde *et al.*, 2007). Although toxins have been reported to act as corepressors for other TA loci, including the Rts1 *higBA*, the *V. cholerae* HigB showed no indication of corepression (Budde *et al.*, 2007), which is similar to the *parDE* TA locus.

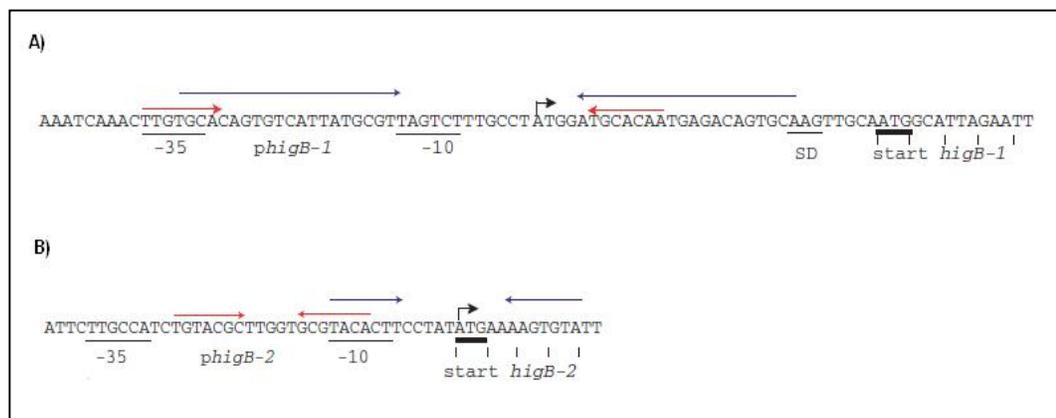


Figure 3: The promoter region of the *higBA-1* (A) and *higBA-2* (B) loci. Arrows pointing rightwards indicate transcription start points. The -10 and -35 sequences are underlined. Start codons, ATG, for both cases are also indicated. The start codon of *higB-2* (B) coincides with the transcriptional start site, indicating that the resulting mRNA is leaderless. Opposing arrows in red and blue indicate inverted repeats that may function as operator elements for HigA and/or HigBA complex binding. (Figure adapted and modified from Christensen-Dalsgaard and Gerdes, 2006)

Transcription of the two *higBA* loci was strongly induced by amino acid starvation in *V. cholerae* and *E. coli* (Christensen-Dalsgaard and Gerdes, 2006) which presumably reflected inactivation of the repressor/antitoxin HigA and a corresponding activation of HigB. This transcription pattern of the *hig* loci of *V. cholerae* was very similar to those of *relBE* and *mazEF* of *E. coli* which were all induced by amino acid starvation and the toxin genes encode mRNA cleaving enzymes that can reduce the global rate-of-translation. In turn, the drain on charged tRNA is reduced due to the reduced translation-rate and thus the translational error-rate is reduced as well (Wagner *et al.*, 1982; Sørensen, 2001; Elf *et al.*, 2003). Thus, these TA loci can be seen to function as translational quality control elements.

The HigB toxins inhibit translation by cleavage of mRNA preferentially within the translated part of a model mRNA and only when the mRNA is translatable. HigB cleavage patterns are similar but not identical to that of RelE. The HigB-mediated cleavages are mostly located between the second and third base of a codon (HigB of Rts1: AA↓A, AA↓A; HigB-1: AA↓A, AC↓U, AA↓A, GC↓G, GC↓A; HigB-2: AA↓A, AC↓U, AA↓A, CU↓G, CU↓G, GC↓G, CU↓G, CU↓G, CU↓G); two cleavage sites mediated by HigB-1 were detected between two codons (GUA↓AUC, AGC↓AAC); whereas only one cleavage site was observed between the first and second base of a codon (HigB-1: U↓GC) (Christensen-Dalsgaard and Gerdes, 2006). The HigB toxin is associated with the 50S ribosomal subunit and it is this HigB-ribosome complex which cleaves within mRNA coding regions at all AAA triplet sequences, both in and out-of-frame. HigB appeared to be responsible for the mRNA cleavage activity of the HigB-ribosome complex since a HigB H92Q mutant lacked mRNA cleavage activity but remained associated with the ribosome (Hurley and Woychick, 2009). Interestingly, the cleavage specificity of HigB on plasmid Rts1 coincided with the sequence (AAA, Lys)

of either the most abundant, or the second most abundant codon in its *Proteus* host (Hurley and Woychick, 2009).

1.2.8 The *hipBA* TA locus

The *hip* (high persistence) locus, located at 33.8 min in the *E. coli* chromosome, consists of an operon containing two genes, with the *hipB* gene preceding the *hipA* gene (Moyed and Bertrand, 1986; Black *et al.*, 1991). The last codon of *hipB* overlaps with the start codon of *hipA* by 1 bp suggesting translation coupling of both genes (Black *et al.*, 1991). *hipA* encodes a ~50 kDa (440 amino acid residues) toxin protein whereas *hipB* encodes a ~10 kDa (88 amino acid residues) antitoxin protein. The HipB antitoxin acts as a transcription repressor and autoregulates its own operon (Black *et al.*, 1991). In addition, HipB also forms a tight complex with HipA toxin in one-to-one molar ratio to neutralize the toxic effect of HipA (Black *et al.*, 1994).

The region upstream of *hipB* contains the -35 and -10 promoter elements and several inverted repeats (Figure 4). There are three similar inverted repeats with conserved sequence TATCCN₈GGATA which are designated O1, O2 and O3 (Figure 4). There is another degenerate version of an inverted repeat sequence further upstream of O3, with the non-conserved half-site, designated as O4 (Figure 4). A longer inverted repeat and a putative integration host factor (IHF) binding site were also identified further upstream of O4 (Figure 4) (Black *et al.*, 1994). Gel shift and DNaseI footprinting assays showed that HipB binds to these four operator sites (O1, O2, O3 and O4) simultaneously (only one retarded band observed in gel shift assay), in a cooperative manner, but not the longer inverted repeats (Black *et al.*, 1994). The binding sites contained the -35 and -10 elements as well as the *hip* transcriptional start site, thus inferring that HipB likely represses transcription by steric occlusion of RNA polymerase. Methylation protection experiments showed that interaction of HipB within the individual operator sites appeared to occur in adjacent major grooves. The binding sites of HipB also occurred on the same face of the DNA helix (Black *et al.*, 1994). Sequence analysis revealed that HipB contains a helix-turn-helix, Cro-like DNA

binding motif at its N-terminal region and exists as a dimer in solution from gel filtration and cross-linking studies. All these findings inferred that HipB binds to each of the four operator sites as an oligomer of at least two subunits in two adjacent helical turns of the major grooves along one face of the DNA helix, like most prokaryotic DNA-binding proteins that harbour helix-turn-helix motif (Black *et al.*, 1994).

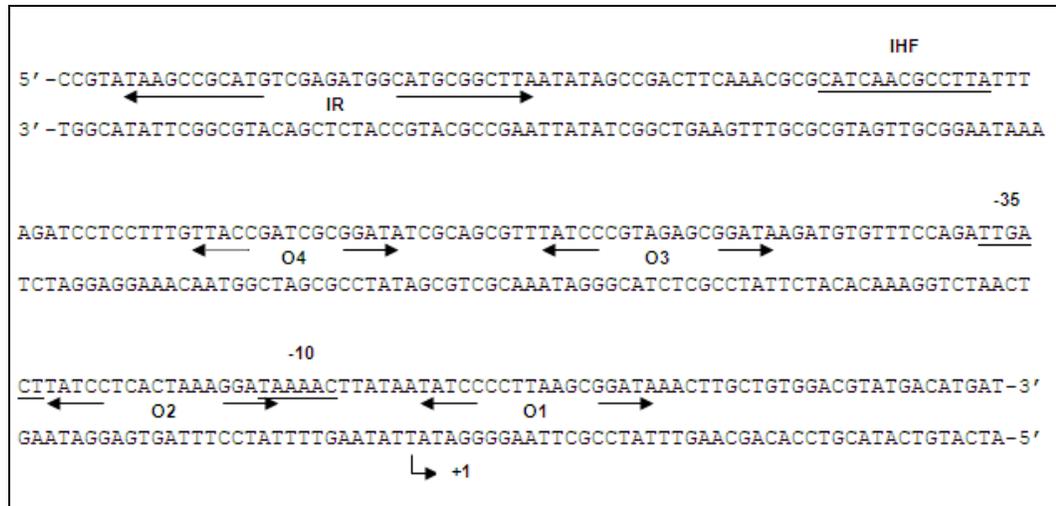


Figure 4: Nucleotide sequence of the *hip* regulatory region. The transcriptional start site is labeled as + 1. The -10 and -35 regions of the promoter are underlined. Regions of inverted repeats are indicated with pairs of opposing arrows and labeled as O1, O2, O3, O4 and IR. The site that is similar to the IHF consensus binding site is underlined and indicated. (Figure adapted from Black *et al.*, 1994)

Promoter-*lacZ* fusion experiments indicated that HipB represses the *hip* promoter in *trans*. Like most other TA toxins, HipA does not appear to bind to the *hip* promoter region directly but indirectly via HipB as a component of a HipA-HipB complex. Although insertional inactivation of *hipA* slightly increases the transcription of the *hip* promoter, expression of *hipA* and *hipB* in *trans* from a *lac* promoter has a similar extent of repression on the *hip* promoter when compared to the expression of *hipB* in *trans* from a *lac* promoter. The binding affinity of the HipA-HipB complex to the *hip* regulatory region is similar to HipB alone. The DNaseI protection sites of the HipA-HipB complex are similar to HipB but slightly larger and eight sites were DNaseI

hypersensitive which were not found when HipB was used. Multiple retarded bands in gel shift assays indicated that the HipA-HipB complex interacts with the regulatory region in a manner different from that of HipB (Black *et al.*, 1994). Gel shift and DNaseI footprinting assays also confirmed that IHF binds to the putative IHF site found upstream of the operator sites (Black *et al.*, 1994). This inferred that the regulation of the *hip* operon perhaps has additional complexity that is yet to be discovered.

Overexpression of *hipA* inhibits cell growth for a transient period and a small fraction of cells outgrow 6 hours (h) after *hipA* induction. These cells could have acquired mutations or became insensitive to HipA or the inducer. Pulse-chase experiments showed that the HipA toxin inhibits macromolecular synthesis with the effect on DNA synthesis secondary to the effects on protein and RNA synthesis *in vivo*. Moreover, HipA inhibits protein synthesis *in vivo* but not *in vitro*. Prolonged exposure of cells to HipA was bacteriostatic rather than bacteriocidal. Expression of *hipA* conferred a transient dormant state (persistence), whereas a majority of the cells remained in a prolonged dormant state that can be reversed by the expression of HipB under appropriate conditions (Korch and Hill, 2006). Residues Gly22 and Asp291 of HipA are highly conserved among the HipA homologues. A G22S mutation render the HipA protein non-toxic whereas a D291A mutation confers high persistence. The combination of both G22S and D291A mutations (*hipA7*), led to the production of non-toxic proteins which were cold sensitive and produced persister cells at increasing frequency at high cell densities, possibly due to entry to the stationary phase or through action of a quorum-sensing system (Korch *et al.*, 2003). HipB plays no direct role in developing the persistency state. The production of high frequency persister cells was found to rely on the products of the *relA* and *spoT* genes, suggesting that *hipA7* facilitates the establishment of the persistent state through the production of (p)ppGpp (Korch *et al.*, 2003). HipA expressed in excess of HipB generated high frequency of

persisters and the ability of the *hipA7* mutants to generate high frequency of persisters was equal to that conferred by the wild-type cells, thus suggesting that persistency is distinct from the macromolecular inhibition (Korch and Hill, 2006).

Prolonged cultivation of an *E. coli* derivative that was disrupted in six TA loci (*chapBIK*, *dinJ-yafQ*, *hipBA*, *mazEF*, *relBE* and *yefM-yoeB*) in M9 medium showed that the Δ *hipAB* cells exhibited a significantly longer life span than the rest of the disruptants and wild-type cells. In wild-type cells, *hipA* expression was detected at 15 h (log phase) and reached maximum expression at 40 h (stationary phase). It is believed that HipA is activated when the transcription level of *hipA* reaches maximum through degradation or inactivation of HipB (Kawano *et al.*, 2009). Ectopic expression of *hipA* inhibits macromolecular synthesis including RNA and proteins (Korch and Hill, 2006). Both RNA and protein contents at 40 h were higher in the Δ *hipAB* cells as compared to the wild-type cells suggesting that HipA in the wild-type cells activated and inhibited macromolecular synthesis from about 40 h. The HipA toxin is thought to be bacteriostatic instead of bacteriocidal. The dead cells discovered during prolonged cultivation were not killed by HipA, but rather by other stresses, and HipA only expedited cell death via inhibition of macromolecular synthesis during prolonged stationary phase (Kawano *et al.*, 2009). Oxidative stress exists during prolonged cultivation (Dukan and Nyström, 1999). Under anaerobic conditions, both wild-type and Δ *hipAB* cells maintained the same number of cells whereas under aerobic conditions, both cell numbers dropped with the drop in wild-type cells being more drastic. Similar results were obtained when the cells were treated with H₂O₂. These results indicated that Δ *hipAB* cells were more susceptible to oxidative stress; perhaps active macromolecular synthesis in the Δ *hipAB* cells functioned as an antioxidant mechanism for higher resistance to H₂O₂ thus enabling cells to survive better in long-term stationary phase (Kawano *et al.*, 2009).

Sequence analysis showed that the HipA toxin is a member of the phosphatidylinositol 3/4-kinase superfamily. HipA is autophosphorylated *in vitro* in the presence of ATP and purified HipA carries a single phosphate on Ser150, thus HipA was deduced to be a serine kinase and autophosphorylates *in vivo* as well (Correia *et al.*, 2006). Overexpression of HipA led to cell growth cessation and increased persister formation. Mutation of D309Q in the conserved kinase active site, D332Q in the Mg²⁺ binding site or S150A, respectively, caused the mutant protein to lose its function in inhibiting cell growth upon overexpression. Cells with wild-type HipA are highly tolerant to cefotaxim (a cell wall synthesis inhibitor), ofloxacin (a DNA gyrase inhibitor), topoisomerase IV and mitomycin C (an antibiotic that kills non-growing cells by forming DNA adducts). The mutant protein does not protect the cells from these antibiotics (Correia *et al.*, 2006). Taken together, these results indicated that the protein kinase activity of HipA is essential for persister formation.

1.2.9 The ω - ε - ζ TA locus

The broad host-range and low copy number *inc18* family pSM19035 plasmid was originally isolated from *Streptococcus pyogenes*. The pSM19035 plasmid consisted of SegA and SegB regions, which encode products that prevent the formation of plasmid-free cells. SegA encodes a site-specific recombinase (gene β), which resolves plasmid oligomers into monomers and thus maximizes random plasmid segregation between the daughter cells (Rojo and Alonso, 1994) whereas SegB, which includes the δ , ω , ε and ζ genes, renders plasmids ~1000 times more stable than expected for random segregation (Ceglowski *et al.*, 1993a; b). The ω , ε and ζ genes form an operon with two discrete promoters, $P\omega$ and $P\varepsilon$; the $P\omega$ promoter is involved in transcribing ω , ω and ε , as well as a full transcript including ω , ε and ζ ; whereas $P\varepsilon$ drives the transcription of ε alone as well as the ε and ζ genes (Ceglowski *et al.*, 1993a; de la Hoz *et al.*, 2000). The organization of the ω - ε - ζ operon is unique among all the known TA loci. Neither the ε antitoxin protein nor the ζ toxin protein is involved in their own transcription, and the transcription of the ω - ε - ζ operon is regulated by the third element, the ω gene product (de la Hoz *et al.*, 2000). The 71-amino acid residue ω protein coordinates regulation of genes involved in replication, transfer and stable maintenance of plasmids, whereas the ε and ζ gene products are strictly involved in the killing of plasmid-free segregants (de la Hoz *et al.*, 2000).

Investigations using *lacZ* transcription fusion assays showed that the promoter activity for $P\omega$ is moderate and low for $P\varepsilon$. Transcription from the promoters of the ω - ε - ζ operon is repressed by the ω protein, but transcription of the ε - ζ operon from $P\varepsilon$ is not directly controlled by the ω protein (de la Hoz *et al.*, 2000). The ω gene is located in between the δ and ε genes. The ω protein belongs to homodimeric ribbon-helix-helix repressors typified by a central, antiparallel β -sheet for DNA major groove binding (Weihofen *et al.*, 2006). The promoter region of the ω gene is embedded in a set of two

types of 7-bp conserved repeats, 5'-ATCACAA-3' and 5'-ATCACTT-3'. There are seven copies of both types of heptamers in both head-to-tail and head-to-head configurations in the region upstream of the ω gene. EMSA showed that the amount of protein ω -DNA (region upstream of ω) complex formed increased exponentially with increasing amounts of ω protein added (de la Hoz *et al.*, 2000). The exponential increase in complex formation suggests that the ω protein binds to its target sequences in a cooperative manner. Transcription of the $P\omega$ promoter is repressed in the presence of ω proteins but the binding of ω proteins to its DNA target does not affect the binding of σ^A RNA polymerase to the $P\omega$ promoter region (de la Hoz *et al.*, 2000). Therefore the repression is not caused by the inhibition of the binding of σ^A RNA polymerase to the $P\omega$ promoter region by the ω proteins.

The ζ gene encodes a 287-amino acid residue toxin protein, which is much larger than other proteic plasmid addiction systems, which are usually composed of 100 amino acid residues. A PROSITE (Bairoch *et al.*, 1997) search showed that ζ featured a Walker A motif (Walker *et al.*, 1982), also known as a P-loop (Murzin *et al.*, 1995) that is typical for ATP/GTP-binding proteins, and was thus postulated to act as a phosphotransferase. The ε gene encodes a 90-amino acid residue ε antitoxin that prevents the toxic action of ζ (Zielenkiewicz and Ceglowski, 2005). According to analytical centrifugation and gel filtration, the ε and ζ proteins co-purified as a stable $\varepsilon_2\zeta_2$ heterotetramer in solution (Camacho *et al.*, 2002). The crystallized $\varepsilon_2\zeta_2$ heterotetramer formed a dumbbell-shape and the NCS axis cuts the handgrip that is formed by the ε_2 dimer and stabilized mainly by hydrophobic interactions; the ζ_2 proteins forming the outer parts of the dumbbell interact with the ε_2 dimer but not with each other (Meinhart *et al.*, 2003). The N-terminal part of the ζ toxin and the Walker A motif, which is also found in the N-terminal of ζ were found to be responsible for the toxicity (Zielenkiewicz *et al.*, 2009). The sequence changes within the NTP binding

Walker motif A located between amino acid 39 and 47 led to the complete abolishment of ζ activity (Nowakowska *et al.*, 2005). In $\varepsilon_2\zeta_2$, the toxin ζ is inactivated by the binding of the N-terminal helix of the antitoxin ε to the N-terminal part as well as its ATP/GTP binding motif (Meinhart *et al.*, 2003; Zielenkiewicz *et al.*, 2009).

The ζ protein has lower thermodynamic stability than the ε protein, both in the free state or in the complex. Proteolytic studies showed that ζ is more stable in the $\varepsilon_2\zeta_2$ complex than in the free state. *In vivo*, the ε protein in *B. subtilis* was shown to have a short half life of ~18 min whereas the half life for ζ is longer (>60 min) (Camacho *et al.*, 2002). Preliminary experiments with the use of *B. subtilis* protease-deficient mutants indicated an involvement of ClpP protease with the chaperone ClpX in the degradation of the ε antitoxin (Zielenkiewicz and Ceglowski, 2005). However, Lioy *et al.* (2006) had shown that depletion of the ε_2 antitoxin was compromised in the absence of the LonA protease and, to a minor extent, in the absence of the ClpX chaperone.

The ε - ζ TA cassette was found to stabilize plasmids in Gram-negative *E. coli* less efficiently than in Gram-positive *B. subtilis* (Zielenkiewicz and Ceglowski, 2005). Expression of the toxin ζ was bactericidal for *B. subtilis* cells. Microscopic examination of cells subjected to ζ activity suggested some changes of morphology as affected cells were smaller, shorter and finer, as well as showing massive cell lysis (Zielenkiewicz and Ceglowski, 2005). In agreement with this study, Lioy and colleagues (2006) showed that ζ induces a viable but non-culturable state in majority of the cells and killed a fraction of the population (~20% for normal level of ζ and ~50% for very high level of ζ induction). Sequence analysis of cells that survived at the time of the lowest optical density (OD)₆₀₀ due to the killing by overproduced ζ indicates that the ζ gene sequence of these survivors had been altered (Zielenkiewicz and Ceglowski, 2005). However in *E. coli* cells, overproduction of ζ was bacteriostatic rather than bacteriocidal. Induction of ζ caused growth inhibition for the first few hours and growth restoration was

observed after ~3 h. Microscopic examination of the ζ -induced cultures revealed massive filamentation of cells during the 2.5-h period after ζ induction, and the number of filaments decreased gradually after growth restoration (Zielenkiewicz and Ceglowski, 2005). A decrease in colony-forming unit for cells subjected to the action of ζ was mainly correlated with their tendency to form filaments. Cell filamentation is one of the phenomena related to the SOS response. However, by using *recA* and *recA*⁺ strains, cell filamentation upon expression of ζ was found to be not connected with the SOS response. Thus, the presence of filaments suggested that in *E. coli*, ζ may act at the stage of cell division (Zielenkiewicz and Ceglowski, 2005). On the other hand, Lioy *et al.* (2006) showed that bacteriostasis induced by ζ overexpression was fully reversible by subsequent expression of the ϵ antitoxin during a time window of 240 min but after prolonged exposure to ζ , expression of ϵ could not reverse the growth-arrested state. ‘High overexpression’ of ζ even reduced the time window of partial reversibility by the ϵ antitoxin to 60 min (Lioy *et al.*, 2006). The target of ζ is still unknown but it is believed that ζ interacts with its specific target and reversibly inhibits cell proliferation, and further accumulation of ζ might lead to cell death due to pleiotropic effects (Lioy *et al.*, 2006). The toxicity of ζ has also been demonstrated in eukaryotic cells. Expression of the ζ protein was toxic to *S. cerevisiae* in a dose dependent manner and the toxic effects were alleviated by the co-expression of ϵ in the same cell (Zielenkiewicz *et al.*, 2009).

1.3 TA Loci in the Genome of *S. pneumoniae*

According to The United Nations Children's Fund (UNICEF) and the World Health Organization (WHO), pneumonia kills more children than any other illness – more than Acquired immune deficiency syndrome (AIDS), malaria and measles combined (Wardlaw *et al.*, 2006). Over two million children die from pneumonia every year, accounting for almost one in five under-five deaths worldwide. It is known that the leading cause of pneumonia among children especially in the developing world is *S. pneumoniae* (Wardlaw *et al.*, 2006).

S. pneumoniae are Gram-positive, lancet-shaped cocci which are arranged singly or in short chains. Individual cells are between 0.5 and 1.25 μm in diameter (Alonso de Velasco *et al.*, 1995). *S. pneumoniae* causes respiratory diseases and is a major cause of morbidity and mortality especially in the elderly, very young and immunocompromised patients (Yokota *et al.*, 2002; Camps Serra *et al.*, 2008; Wardlaw *et al.*, 2006). Besides pneumonia, *S. pneumoniae* is also the causal agent of many other human diseases, such as bacteremia, meningitis, otitis media, osteomyelitis, septic arthritis, endocarditis, peritonitis, pericarditis and sinusitis (Di Guilmi and Dessen, 2002; Kaplan and Mason, 1998).

Significantly expanding treatment coverage is crucial to reduce the pneumonia death rate among children under five in the developing world. Nonetheless, high levels of antibiotic resistance to first-line treatments, notably cotrimoxazole, have been reported in many places of the world. The use of antibiotics to treat children infected with pneumonia could make managing antibiotic resistance become more difficult in the future (Wardlaw *et al.*, 2006). Therefore, there is a need to look for alternative ways to combat pneumonia and TA loci have the potential as targets for novel antimicrobial agents. Comprehensive genome searches for TA loci in prokaryotes have identified multiple copies within the genome of a single organism. Some of these toxin or

antitoxin homologues appeared to be stand-alone genes without their counterparts (Makarova *et al.*, 2009). However, these toxin or antitoxin homologues are yet to be proven as bona fide TA loci *in vivo*. About 40-50 toxin and/or antitoxin genes have been identified from the genome sequence of *S. pneumoniae*. Nevertheless, only three to nine copies appeared as complete TA pairs (Table 2 and Table 3) (Makarova *et al.*, 2009). Three of these pneumococcal TA loci, namely *relBE2Spn*, *pezAT* and *yefM-yoeB* were recently shown to be functional TA loci (Nieto *et al.*, 2006, 2007; Khoo *et al.*, 2007).

Table 2: The representation of solo toxins/antitoxins and TA pairs in *S. pneumoniae*. The total number of protein coding genes is indicated (Table adapted and modified from Makarov *et al.*, 2009).

Species	Number of proteins	T/A	Number of TA
<i>S. pneumoniae</i> CGSP14	2206	50	5
<i>S. pneumoniae</i> D39	1914	40	3
<i>S. pneumoniae</i> G54	2115	47	5
<i>S. pneumoniae</i> Hungary19A-6	2155	51	9
<i>S. pneumoniae</i> R6	2043	45	5
<i>S. pneumoniae</i> TIGR4	2105	48	7

Two *E. coli* K-12 *relBE* homologues, termed *relBE1Spn* and *relBE2Spn*, were initially identified from the genome sequence of *S. pneumoniae* (Gerdes, 2000). Cloning of *relE1Spn* and *relE2Spn* genes in expression vectors had demonstrated that overexpression of RelE1Spn was innocuous in *E. coli* whereas RelE2Spn exhibited growth arrest in both *E. coli* and *S. pneumoniae* (Nieto *et al.*, 2006). However for *S. pneumoniae*, the full toxicity of RelE2Spn was only achieved when the chromosomal copy of the RelB2Spn antitoxin was inactivated (Nieto *et al.*, 2006). This indicated that the presence of a single copy of the RelB2Spn antitoxin in the genome was enough to counteract the cell growth arrest provoked by overexpression of the cognate RelE2Spn toxin. Cell growth arrest caused by expression of the RelE2Spn toxin could be reverted

by expression of the cognate RelB2Spn antitoxin, although prolonged exposure to the toxin led to cell death (Nieto *et al.*, 2006).

Table 3: List of GenBank identifiers (GI numbers) for all predicted and known toxins and antitoxins of *S. pneumoniae* (Table adapted and modified from: Makarova *et al.*, 2009).

Species	GI numbers	T/A	GI numbers	T/A
<i>S. pneumoniae</i> CGSP14	182683256	RHH	182683257	RelE
<i>S. pneumoniae</i> CGSP14	182683808	AbrB	182683809	Fic
<i>S. pneumoniae</i> CGSP14	182684048	RHH	182684049	RelE
<i>S. pneumoniae</i> CGSP14	182684684	RelE	182684685	PHD
<i>S. pneumoniae</i> CGSP14	182684880	COG2856	182684881	Xre
<i>S. pneumoniae</i> D39	116515838	RelE	116516949	PHD
<i>S. pneumoniae</i> D39	116516209	RelE	116517043	RHH
<i>S. pneumoniae</i> D39	116516886	RHH	116516116	RelE
<i>S. pneumoniae</i> G54	194396825	COG2856	194397882	Xre
<i>S. pneumoniae</i> G54	194397028	RelE	194397339	RHH
<i>S. pneumoniae</i> G54	194398013	RHH	194398587	RelE
<i>S. pneumoniae</i> G54	194398053	HicB	194397600	HicA
<i>S. pneumoniae</i> G54	194398155	AbrB	194398735	Fic
<i>S. pneumoniae</i> Hungary19A-6	169832735	COG2856	169834505	Xre
<i>S. pneumoniae</i> Hungary19A-6	169832832	Xre	169833276	Bro
<i>S. pneumoniae</i> Hungary19A-6	169832892	RelE	169833374	PHD
<i>S. pneumoniae</i> Hungary19A-6	169832920	RelE	169833661	RHH
<i>S. pneumoniae</i> Hungary19A-6	169832994	AbrB	169833045	Fic
<i>S. pneumoniae</i> Hungary19A-6	169833524	HicB	169834067	HicA
<i>S. pneumoniae</i> Hungary19A-6	169833989	COG2856	169832661	Xre
<i>S. pneumoniae</i> Hungary19A-6	169834001	Bro	169834410	Xre
<i>S. pneumoniae</i> Hungary19A-6	169834473	RHH	169832900	RelE
<i>S. pneumoniae</i> R6	15902296	RHH	15902297	RelE
<i>S. pneumoniae</i> R6	15903146	RelE	15903147	RHH
<i>S. pneumoniae</i> R6	15903627	RelE	15903628	PHD
<i>S. pneumoniae</i> R6	15903654	HicB	15903655	HicA
<i>S. pneumoniae</i> R6	15903794	COG2856	15903795	Xre
<i>S. pneumoniae</i> TIGR4	15900209	RHH	15900210	RelE
<i>S. pneumoniae</i> TIGR4	15900771	AbrB	15900772	Fic
<i>S. pneumoniae</i> TIGR4	15901009	RelE	15901010	Xre
<i>S. pneumoniae</i> TIGR4	15901085	RelE	15901086	RHH
<i>S. pneumoniae</i> TIGR4	15901572	RelE	15901573	PHD
<i>S. pneumoniae</i> TIGR4	15901615	HicB	15901616	HicA
<i>S. pneumoniae</i> TIGR4	15901759	COG2856	15901760	Xre

The *pezAT* (for pneumococcal epsilon-zeta) TA locus was first identified on the genome of *S. pneumoniae* TIGR4 based on the sequence homology of the PezT toxin

with the ζ toxin of the *S. pyogenes* plasmid pSM19035-encoded ϵ - ζ TA locus (Khoo *et al.*, 2007). Overexpression of the PezT toxin from plasmid pET11a in *E. coli* resulted in growth inhibition with reduction in cell viability for the first 3 h. However, restoration of cell growth was observed ~3-h-post-induction with gradual increase in both the A_{600} value as well as viable cell count. Similar results were obtained by growing the cells in minimum medium instead of Luria-Bertani broth. Sequencing of the plasmid DNA isolated from cells that recovered from stasis did not reveal any mutations in the *pezT* gene nor in the T7 promoter of the pET11a expression vector (Khoo *et al.*, 2007). These results indicated that overexpression of the PezT toxin was bacteriostatic instead of bacteriocidal in *E. coli* cells. Nevertheless, growth inhibition was not observed when the PezA antitoxin was expressed either *in cis* or *in trans* (Khoo *et al.*, 2007). The growth profile of *E. coli* cells overexpressing PezT was remarkably similar to the *E. coli* cells that overexpressed the ζ toxin (Zielenkiewics and Ceglowski, 2005). Khoo *et al.* (2007) also showed that PezT interacted with PezA both *in vitro* as well as *in vivo*. Therefore, *pezAT* was inferred to be a bona fide functional TA locus.

The *E. coli yefM-yoeB* TA locus was identified on the basis of its similarity with the *axe-txe* TA locus of plasmid pRUM in *E. facium* (Grady and Hayes, 2003). A homologue of the *E. coli yefM-yoeB* or the *E. facium axe-txe* TA locus is also present in the genome of *S. pneumoniae* and was designated *yefM-yoeB_{Spn}* (Nieto *et al.*, 2007). Overproduction of *yefM-yoeB_{Spn}* in *E. coli* led to severe inhibition of cell growth and reduction in cell viability. In addition, this toxicity is more pronounced in an *E. coli* B strain than two other tested *E. coli* K-12 strains (Nieto *et al.*, 2007). The cytotoxic effect of the YoeB_{Spn} toxin was alleviated by coinduction of its cognate YefM_{Spn} antitoxin. Recovery of cell viability was observed by overexpressing the YefM_{Spn} antitoxin for the cultures treated with the YoeB_{Spn} toxin for 6 h (Nieto *et al.*, 2007). Hence, *yefM-yoeB_{Spn}* was also shown to be a functional TA locus.

1.4 Objectives of this Project

As both *pezAT* and *yefM-yoeB_{Spn}* had been demonstrated to be functional TA loci from the genome of *S. pneumoniae*, the overall objective of this project was to determine the regulatory mechanisms for both TA loci. Specific objectives for each TA locus were as follows.

1.4.1 The *pezAT* TA locus

1. To determine the promoter activity and the mode of regulation of the pneumococcal *pezAT* TA locus in *E. coli* using transcriptional fusions with promoter-less *lacZ* gene;
2. To determine if the *pezA* antitoxin and *pezT* toxin genes are co-transcribed using reverse transcriptase polymerase chain reaction (RT-PCR);
3. To determine the transcriptional start sites of the *pezAT* TA locus; and
4. To determine the DNA binding sites of the PezA antitoxin and the PezAT TA complex using EMSA.

1.4.2 The *yefM-yoeB_{Spn}* TA locus

1. To determine the promoter activity and the mode of regulation of the pneumococcal *yefM-yoeB_{Spn}* locus in *E. coli* using transcriptional fusions with promoter-less *lacZ* gene;
2. To determine if the *yefM_{Spn}* antitoxin and *yoeB_{Spn}* toxin genes are co-transcribed;
3. To determine the transcriptional start sites of the *yefM-yoeB_{Spn}* TA locus; and
4. To determine DNA binding sites of the YefM_{Spn} antitoxin and the YefM-YoeB_{Spn} TA complex using EMSA, DNaseI footprinting and hydroxyl radical footprinting assays.

2 Materials and Methods

2.1 Bioinformatics analyses

All pairwise alignments were done using Bioedit Version 7.0.4.1 (www.mbio.ncsu.edu/BioEdit/bioedit.html), whereas multiple sequence alignments were conducted using ClustalW2, which was accessed at <http://www.ebi.ac.uk/Tools/clustalw2/index.html>. The helix-turn-helix motifs within the N-terminal portion of PezA were identified by InterProScan (<http://www.ebi.ac.uk/Tools/InterProScan/>). The molecular weight and pI of the proteins were estimated using Protparam (<http://expasy.org/tools/protparam.html>). The secondary structure of boxA-C element, the putative transcription terminators of the gene upstream of *yefM_{Spn}* and sequence immediately after the *yoeB_{Spn}* stop codon were predicted using mfold (<http://mfold.rna.albany.edu/>). The amino acid sequences of the possible reading frames of the C-terminus of *yoeB_{Spn}* as well as between the *yoeB_{Spn}* stop codon and *spr1584* were determined using the Translate tool at <http://expasy.org/tools/dna.html>. Putative promoters were identified by comparing the sequence with the σ^{70} -type promoter consensus sequence (the -35 sequence 5'-TTGACA-3' and the -10 sequence 5'-TATAAT-3' with a 17 bp spacer between them) (Horwitz and Loeb, 1988; Morrison and Jaurin, 1990).

2.2 Bacterial Growth Media and Conditions

The bacterial strains and plasmids used in this study are listed in Table 4. *E. coli* strains were grown in Luria-Bertani medium at 37°C (for cells that harboured pGEM-T Easy, pQF52 and pET28a recombinant plasmids) or at 30°C (for cells that harboured pLNBAD recombinant plasmids). Liquid cultures were grown in an orbital shaking incubator at 250 rpm. For solid media, Bacto-agar was added at 15% (w/v). When required, growth medium was supplemented with antibiotics at the following concentrations: ampicillin, 100 µg/ml (for cells that harboured pGEM-T Easy and pQF52 recombinant plasmids); chloramphenicol, 20 µg/ml (for cells that harboured pLNBAD recombinant plasmids) and kanamycin, 50 µg/ml (for cells that harboured pET28a recombinant plasmids). For storage, a single colony of bacteria was grown overnight in the Luria-Bertani broth supplemented with appropriate antibiotics. The cells were harvested by centrifugation at $2,500 \times g$ for 5 min, resuspended in sterile Luria-Bertani broth containing 15% glycerol and stored at -80°C. *S. pneumoniae* R6 were grown in AGCH media (Lacks, 1968) at 37°C, supplemented with 1% sucrose or maltose and 0.25% yeast extract. Antibiotics were added at the following concentrations: tetracycline, 1 µg/ml and chloramphenicol, 2 µg/ml.

Table 4: Bacterial strains and plasmid vectors used in this study

Strains/Plasmids	Relevant Genotypes	Sources/References
Strains		
<i>S. pneumoniae</i>	Local clinical isolates	Isolates from Faculty of Medicine, Universiti Putra Malaysia
<i>S. pneumoniae</i> R6	Wild type	Lacks <i>et al.</i> , 1986
<i>E. coli</i> DH5 α	<i>supE44</i> Δ <i>lacU169</i> (ϕ 80 <i>lacZ</i> Δ M15) <i>hsdR17</i> <i>recA1</i> <i>endA1</i> <i>gyrA96</i> <i>thi-1</i> <i>relA1</i>	Sambrook <i>et al.</i> , 1989
<i>E. coli</i> TOP10	F ⁻ <i>mcrA</i> Δ (<i>mrr-hsdRMS-mcrBC</i>) ϕ 80 <i>lacZ</i> Δ M15 Δ <i>lacX74</i> <i>recA1</i> <i>araD139</i> Δ (<i>araA-leu</i>)7697 <i>galU</i> <i>galK</i> <i>rpsL</i> (Str ^R) <i>endA1</i> <i>nupG</i>	Invitrogen TM , USA
<i>E. coli</i> XL10-Gold	Tet ^R Δ (<i>mcrA</i>)183 Δ (<i>mcrCB-hsdSMR-mrr</i>)173 <i>endA1</i> <i>supE44</i> <i>thi-1</i> <i>recA1</i> <i>gyrA96</i> <i>relA1</i> <i>lac</i> Hte [F ⁻ <i>proAB</i> <i>lacI</i> ^f Z Δ M15 Tn10 (Tet ^R) Amy Cam ^R]	Stratagene, USA
<i>E. coli</i> BL21-CodonPlus(DE3)-RIL	<i>E. coli</i> B F ⁻ <i>ompT</i> <i>hsdS</i> (rB ⁻ mB ⁻) <i>dcm</i> ⁺ Tet ^f <i>gal</i> λ (DE3) <i>endA</i> Hte [<i>argU</i> <i>ileY</i> <i>leuW</i> Cam ^f]	Stratagene, USA
Plasmids		
pGEM-T Easy	Ampicillin resistant; α - <i>lac</i> ; multiple cloning site; 3.0 kb T-tailed cloning vector	Promega, USA
pET28a	Kanamycin resistant; multiple cloning site; contains T7 promoter	Novagen, USA
pQF52	Ampicillin resistant; multiple cloning site; contains promoter-less <i>lacZ</i> ; medium copy number	McLean <i>et al.</i> , 1997
pLNBAD	Chloramphenicol resistant; multiple cloning site; contains P _{BAD} promoter	A gift from Ramon Díaz-Orejas (Department of Molecular Microbiology and Infection Biology, Centro de Investigaciones Biológicas, Spain)

2.3 Construction of Recombinant Plasmids

2.3.1 Construction of pGEM-T Easy, pQF52, pLNBAD and pET28a recombinant plasmids

Genomic DNA extracted from isolates of *S. pneumoniae* from the collection of the Faculty of Medicine, Universiti Putra Malaysia, were used as templates to amplify *pezAT* and *yefM-yoeB_{Spn}* genes along with the upstream sequence using specific primers (Table 6). For the *pezAT* locus, the DNA fragment encompassing the *pezA* and *pezT* reading frames along with 191 bp of sequence immediately upstream of *pezA* was amplified using primer pairs that were designed based on the sequences of the genome in *S. pneumoniae* TIGR4. For the *yefM-yoeB_{Spn}* locus, the 762 bp DNA fragment encompassing both *yefM_{Spn}* and *yoeB_{Spn}* reading frames along with 237 bp of the region upstream of *yefM_{Spn}* was amplified using primer pairs that were designed based on the sequences of the genome in *S. pneumoniae* R6. The sizes of the PCR products were verified using gel electrophoresis and the desired DNA bands were excised from the gel and purified. The purified PCR products were ligated directly into pGEM-T Easy vectors (Appendix Figure A1) resulting in recombinant plasmids pGEMT_PpezApezT and pGEMT_P1P2yefMyoeB, respectively (Table 5). The pGEM-T Easy recombinant plasmids were then transformed into *E. coli* DH5 α . Both constructs were then used as the respective template for subsequent cloning of *pezAT* or *yefM-yoeB_{Spn}* along with their related upstream sequences into pQF52, pLNBAD and pET28a vectors. Both pGEM-T Easy constructs were also used to verify if the TA genes were co-transcribed and to locate the transcriptional start site of the TA genes.

For cloning into either pQF52, pLNBAD or pET28a vectors, purified PCR products containing the necessary DNA fragments were digested with specific restriction enzymes and ligated into similarly digested pQF52 (Appendix Figure A2), pLNBAD (Appendix Figure A3) or pET28a (Appendix Figure A4), respectively (Table

2.2.1). For the pQF52 recombinant plasmids, the PCR-amplified products were cloned into the multiple cloning sites upstream of the promoter-less *lacZ* gene. The pQF52 plasmid (used as control) and the resulting constructed recombinant plasmids were transformed into *E. coli* DH5 α (for the *pezAT* TA locus) or *E. coli* TOP10 (for the *yefM-yoeB_{Spn}* TA locus) and subjected to β -galactosidase assays to study the promoter activities. To investigate the promoter activity when YefM_{Spn} and YefM-YoeB_{Spn} were expressed in *trans*, the arabinose-inducible pLNBAD expression vector was used to harbour various fragments of *yefM_{Spn}* and *yoeB_{Spn}*. The PCR-amplified products were cloned into the region downstream of the P_{BAD} promoter of pLNBAD and then co-transformed together with the pQF52-derived recombinants harbouring various fragments of the *yefM_{Spn}* upstream regulatory regions into *E. coli* TOP10. The pLNBAD plasmid is a derivative of pLN135 plasmid (Lemonnier *et al.*, 2003) and it harbours a pSC101 *ori* and a chloramphenicol resistance gene whereas pQF52 plasmid, a medium copy number plasmid, harbours a ColE1 *ori* and ampicillin resistance (McLean *et al.*, 1997). Therefore the two plasmids are compatible and could be co-transformed into the same cell. However, as pLNDAB is a heat sensitive plasmid (Lemonnier *et al.*, 2003), the *E. coli* cells harbouring both pQF52-derived and pLNBAD-derived recombinants were grown at 30°C instead of 37°C, which is the optimal growth temperature for *E. coli*. The pET28a-derivative recombinants which harboured *pezA*, *pezAT*, *yefM_{Spn}* or *yefM-yoeB_{Spn}* reading frames were transformed into *E. coli* BL21-CodonPlus(DE3)-RIL and subsequently used for overexpression and protein purification.

Table 5: Recombinant plasmids constructed in this study

Recombinant constructs	<i>E. coli</i> host strains	Primer pairs	Descriptions
pGEMT_PpezApezT	DH5 α	PpezApezT-F & PpezApezT-R	pGEM-T Easy with <i>pezAT</i> and 191 bp upstream of <i>pezA</i> .
pGEMT_P1P2yefMyoeB	DH5 α	P1P2yefMyoeB-F & P1P2 yefMyoeB-R	pGEM-T Easy with <i>yefM_{Spn}-yoeB_{Spn}</i> and 237 bp upstream of <i>yefM_{Spn}</i> .
pQF_P	DH5 α	P-F & P-R	191 bp upstream of <i>pezA</i> cloned into <i>NcoI-BamHI</i> site of pQF52, preceding the promoter-less <i>lacZ</i> gene.
pQF_PpezA	DH5 α	P-F & <i>pezA</i> -R	<i>pezA</i> along with 191 bp of its upstream sequence cloned into <i>NcoI-BamHI</i> site of pQF52, preceding the promoter-less <i>lacZ</i> gene.
pQF_PpezApezT	DH5 α	P-F & <i>pezT</i> -R	<i>pezAT</i> along with 191 bp of its upstream sequence cloned into <i>NcoI-BamHI</i> site of pQF52, preceding the promoter-less <i>lacZ</i> gene.
pQF_CpezA	DH5 α	CpezA-F & <i>pezA</i> -R	194 bp upstream of <i>pezT</i> cloned into <i>NcoI-BamHI</i> site of pQF52, preceding the promoter-less <i>lacZ</i> gene.
pQF_P1P2	TOP10	P1-F & P2-R	237 bp upstream of <i>yefM_{Spn}</i> cloned into <i>NcoI-BamHI</i> site of pQF52, preceding the promoter-less <i>lacZ</i> gene.
pQF_P1P2yM	TOP10	P1-F & <i>yefM</i> -R	<i>yefM_{Spn}</i> along with 237 bp of its upstream sequence cloned into <i>NcoI-BamHI</i> site of pQF52, preceding the promoter-less <i>lacZ</i> gene.
pQF_P1P2yMyB	TOP10	P1-F & <i>yoeB</i> -R	<i>yefM-yoeB_{Spn}</i> along with 237 bp of its upstream sequence cloned into <i>NcoI-BamHI</i> site of pQF52, preceding the promoter-less <i>lacZ</i> gene.
pQF_P1	TOP10	mP1-F & mP1-R; mP2-F & mP2-R	237 bp upstream of <i>yefM_{Spn}</i> cloned into <i>NcoI-BamHI</i> site of pQF52, preceding the promoter-less <i>lacZ</i> gene. The -10 region of P _{<i>yefM2</i>} and the nucleotides flanking this region were substituted from 5'-tgTATAATa-3' to 5'-ctgcAg-3'.

Table 5: (continued)

Recombinant constructs	<i>E. coli</i> host strains	Primer pairs	Descriptions
pQF_P1yM	TOP10	mP1-F & mP1-R; mP2-F & yefM-R	<i>yefM_{Spn}</i> along with 237 bp of its upstream sequence cloned into <i>NcoI-BamHI</i> site of pQF52, preceding the promoter-less <i>lacZ</i> gene. The -10 region of P _{<i>yefM2</i>} and the nucleotides flanking this region were substituted from 5'-tgTATAATa-3' to 5'-ctgcAg-3'.
pQF_P1yMyB	TOP10	mP1-F & mP1-R; mP2-F & yoeB-R	<i>yefM-yoeB_{Spn}</i> along with 237 bp of upstream sequence cloned into <i>NcoI-BamHI</i> site of pQF52, preceding the promoter-less <i>lacZ</i> gene. The -10 region of P _{<i>yefM2</i>} and the nucleotides flanking this region were substituted from 5'-tgTATAATa-3' to 5'-ctgcAg-3'.
pQF_P2	TOP10	P2-F & mP2-R	87 bp upstream of <i>yefM_{Spn}</i> cloned into <i>NcoI-BamHI</i> site of pQF52, preceding the promoter-less <i>lacZ</i> gene.
pQF_P2yM	TOP10	P2-F & yefM-R	<i>yefM_{Spn}</i> along with 87 bp of its upstream sequence cloned into <i>NcoI-BamHI</i> site of pQF52, preceding the promoter-less <i>lacZ</i> gene.
pQF_P2yMyB	TOP10	P2-F & yoeB-R	<i>yefM-yoeB_{Spn}</i> along with 87 bp of its upstream sequence cloned into <i>NcoI-BamHI</i> site of pQF52, preceding the promoter-less <i>lacZ</i> gene.
pQF_nP	TOP10	P2-F & mP1-R; mP2-F & mP2-R	87 bp upstream of <i>yefM_{Spn}</i> cloned into <i>NcoI-BamHI</i> site of pQF52, preceding the promoter-less <i>lacZ</i> gene. The -10 region of P _{<i>yefM2</i>} and the nucleotides flanking this region were substituted from 5'-tgTATAATa-3' to 5'-ctgcAg-3'.
pQF_nPyM	TOP10	P2-F & mP1-R; mP2-F & yefM-R	<i>yefM_{Spn}</i> along with 87 bp of its upstream sequence cloned into <i>NcoI-BamHI</i> site of pQF52, preceding the promoter-less <i>lacZ</i> gene. The -10 region of P _{<i>yefM2</i>} and the nucleotides flanking this region were substituted from 5'-tgTATAATa-3' to 5'-ctgcAg-3'.

Table 5: (continued)

Recombinant constructs	<i>E. coli</i> host strains	Primer pairs	Descriptions
pQF_nPyMyB	TOP10	P2-F & mP1-R; mP2-F & yoeB-R	<i>yefM-yoeB_{Spn}</i> along with 87 bp of upstream sequence cloned into <i>NcoI-BamHI</i> site of pQF52, preceding the promoter-less <i>lacZ</i> gene. The -10 region of P _{<i>yefM2</i>} and the nucleotides flanking this region were substituted from 5'-tgTATAATa-3' to 5'-ctgcAg-3'.
pQF_yM	TOP10	yefM-F & yefM-R	<i>yefM_{Spn}</i> cloned into <i>NcoI-BamHI</i> site of pQF52, preceding the promoter-less <i>lacZ</i> gene.
pQF_CyM	TOP10	CyefM-F & yefM-R	195 bp upstream of <i>yoeB_{Spn}</i> cloned into <i>NcoI-BamHI</i> site of pQF52, preceding the promoter-less <i>lacZ</i> gene.
pQF_yB	TOP10	yoeB-F & yoeB-R	<i>yoeB_{Spn}</i> cloned into <i>NcoI-BamHI</i> site of pQF52, preceding the promoter-less <i>lacZ</i> gene.
pQF_CyB	TOP10	CyoeB-F & yoeB-R	159 bp upstream of <i>yoeB_{Spn}</i> stop codon cloned into <i>NcoI-BamHI</i> site of pQF52, preceding the promoter-less <i>lacZ</i> gene.
pQF_M1yM	TOP10	yefM7-F & yefM7-R	<i>yefM_{Spn}</i> along with 237 bp of its upstream sequence cloned into <i>NcoI-BamHI</i> site of pQF52, preceding the promoter-less <i>lacZ</i> gene. Using site-directed mutagenesis, two base pairs were substituted, which led to the 7 th codon of YefM _{Spn} , 'S', replaced by an amber stop codon.
pQF_M1yMyB	TOP10	yefM7-F & yefM7-R	<i>yefM-yoeB_{Spn}</i> along with 237 bp of its upstream sequence cloned into <i>NcoI-BamHI</i> site of pQF52, preceding the promoter-less <i>lacZ</i> gene. Using site-directed mutagenesis, two base pairs were substituted, which led to the 7 th codon of YefM _{Spn} , 'S', replaced by an amber stop codon.

Table 5: (continued)

Recombinant constructs	<i>E. coli</i> host strains	Primer pairs	Descriptions
pQF_M2yM	TOP10	yefM273233-F & yefM273233-R	<i>yefM_{Spn}</i> along with 237 bp of its upstream sequence cloned into <i>NcoI-BamHI</i> site of pQF52, preceding the promoter-less <i>lacZ</i> gene. Using site-directed mutagenesis, four base pairs were substituted, which led to the 27 th , 32 nd and 33 rd codons of YefM _{Spn} , 'L', 'K' and 'N', replaced by amber, ochre and amber stop codons respectively.
pQF_M2yMyB	TOP10	yefM273233-F & yefM273233-R	<i>yefM_{Spn}-yoeB_{Spn}</i> along with 237 bp of its upstream sequence cloned into <i>NcoI-BamHI</i> site of pQF52, preceding the promoter-less <i>lacZ</i> gene. Using site-directed mutagenesis, four base pairs were substituted, which led to the 27 th , 32 nd and 33 rd codons of YefM _{Spn} , 'L', 'K' and 'N', replaced by amber, ochre and amber stop codons respectively.
pLN_yM	TOP10	pLNyefM-F & pLNyefM-R	<i>yefM_{Spn}</i> cloned into <i>NdeI-HindIII</i> site of pLNBAD, downstream of the P _{BAD} promoter.
pLN_yMyB	TOP10	pLNyefM-F & pLNyoeB-R	<i>yefM-yoeB_{Spn}</i> cloned into <i>NdeI-HindIII</i> site of pLNBAD, downstream of the P _{BAD} promoter.
pLN_CyMyB	TOP10	pLNCyefM-F & pLNyoeB-R	159 bp upstream of <i>yoeB_{Spn}</i> along with <i>yoeB_{Spn}</i> cloned into <i>NdeI-HindIII</i> site of pLNBAD, downstream of the P _{BAD} promoter.
pET28a_HisPezA	TOP10	pETpezA-F & pETpezA-R	<i>pezA</i> cloned into <i>NdeI-BamHI</i> site of pET28a, in frame with (His) ₆ at N-terminal region of <i>pezA</i> .
pET28a_HisPezAPezT	BL21-CodonPlus(DE3)-RIL	pETpezA-F & pETpezT-R	<i>pezAT</i> cloned into <i>NdeI-BamHI</i> site of pET28a, in frame with (His) ₆ at N-terminal region of <i>pezA</i> .

Table 5: (continued)

Recombinant constructs	<i>E. coli</i> host strains	Primer pairs	Descriptions
pET28a_HisYefMHis	BL21-CodonPlus(DE3)-RIL	pETyefM-F & pETyefM-R	<i>yefM_{Spn}</i> , start from GTG as annotated in <i>S. pneumoniae</i> R6, which is 36 bp upstream of ATG start site annotated in <i>S. pneumoniae</i> TIGR4, cloned into <i>NdeI</i> - <i>Bam</i> HI site of pET28a, in frame with (His) ₆ at N-terminal region of <i>yefM_{Spn}</i> .
pET28a_HisYefMYoeB	BL21-CodonPlus(DE3)-RIL	pETyefM-F & pETyoeB-R	<i>yefM-yoeB_{Spn}</i> , start from GTG as annotated in <i>S. pneumoniae</i> R6, which is 36 bp upstream of ATG start site annotated in <i>S. pneumoniae</i> TIGR4, cloned into <i>NdeI</i> - <i>Bam</i> HI site of pET28a, in frame with (His) ₆ at N-terminal region of <i>yefM_{Spn}</i> .
Recombinant constructs	<i>S. pneumoniae</i> host strains	Primer pairs	Descriptions
pYFS10	R6	yefM _N & yefM _C	<i>yefM_{Spn}</i> along with 72 bp of its upstream sequence cloned into <i>Bam</i> HI- <i>Eco</i> RI site of pNM220 (Nieto <i>et al.</i> , 2007).
pEMB13	R6	yefM _N His & yoeB _C His	<i>yefM-yoeB_{Spn}</i> cloned into <i>NheI</i> and <i>XhoI</i> site of pET24b digested with (Nieto <i>et al.</i> , 2007).

2.3.2 Primer design

The primers that were used in this study are listed in Table 6. The primers were designed with reference to the sequences of the *S. pneumoniae* TIGR4 genome (NCBI accession number: NC_003028) for the *pezAT* locus and *S. pneumoniae* R6 genome (NCBI accession number: NC_003098) for the *yefM-yoeB_{Spm}* locus. Designed primers were then checked for melting temperature estimation, presence of secondary structures and possible homodimer formation using the OligoTech version 1.0 program that is freely available in the web at <http://www.oligotetc.com/oligotech/oligotech.exe>. The optimal primer design was chosen for each set and these primers were commercially synthesized.

Primers were designed for the following purposes: PCR amplification, which was used for cloning, site-directed mutagenesis, EMSA, DNaseI footprinting assay, hydroxyl radical footprinting assay, quantitative real-time RT-PCR to determine the levels of specific transcripts in the cell, and 5'-Rapid Amplification of cDNA Ends (5'-RACE) to determine the transcriptional start site. Primer pairs for PCR were designed 20-30 nucleotides in length and with similar melting temperature ($\pm 3^{\circ}\text{C}$). The GC content was ensured to be between 40-60% with C and G nucleotides distributed uniformly throughout the primer. High stability, i.e., G:C clamps, in the 5'-end central regions of the primer confer hybridization stability with the target sequence. On the other hand, more than three G or C nucleotides at 3'-end of the primer were avoided as it can result in nonspecific priming, since the potential to misprime at nontarget sites was increased (Rychlik, 1995). The primer was carefully designed not to be self-complementary or complementary to any other primer in the reaction mixture to avoid thermostable secondary structures or hairpins (with estimated melting temperature $< 35^{\circ}\text{C}$), primer-dimers and homodimers under the PCR annealing conditions (Rychlik and Rhodes, 1989; Rychlik, 1995; Skerra, 1992). For cloning, some primers were

designed with certain restriction sites incorporated into the 5' ends so that the resulting amplified products could be cleaved by the particular restriction enzyme and subsequently cloned into the cloning vector. For site-directed mutagenesis, primers were designed between 25 and 45 nucleotides in length with melting temperature $\geq 78^{\circ}\text{C}$. The desired mutations were designed optimally in the middle of the primer with ~10-15 nucleotides of correct sequence on both sides. Both mutagenic primers contained the desired mutation and annealed to the same sequence on opposite strands of the plasmid. The mutagenic primers were designed to have a minimum GC content of 40% and to terminate in one or more C or G bases.

For quantitative real-time RT-PCR, primers were designed such that the amplicon size did not exceed 750 bp in length with optimal results obtained for amplicons up to 500 bp or less. For 5'-RACE, two antisense gene-specific primers were designed. For gene-specific primer 1, the melting temperature would have to be appropriate for the relatively low temperature (42°C) of the complementary deoxyribonucleic acid (cDNA) synthesis reaction and a short primer (16 to 20 nucleotides) facilitates the efficient separation of gene-specific primer 1 from the cDNA product. Efficient recovery of cDNA from the S.N.A.P. column (provided in the 5'-RACE system (InvitrogenTM, USA)) also required a product of at least 200 nucleotides in length. Thus, gene-specific primer 1 was designed to anneal at least 300 nucleotides from the mRNA 5'-end so that the cDNA can be easily purified using a S.N.A.P. column. Gene-specific primer 2 was designed to anneal to sequences located 3' (with respect to cDNA) of gene-specific primer 1 and to anneal either immediately adjacent to gene-specific primer 1 or at sequences located further upstream of gene-specific primer 1 within the cDNA product. In practice, primers with melting temperatures between $60-75^{\circ}\text{C}$ usually can function effectively in 5'-RACE (Frohman, 1990).

The anchor primers provided in the 5'-RACE system (InvitrogenTM, USA) contained 3' sequences complementary to the homopolymeric tail and additional 5' sequences that encoded an adapter region, comprising of restriction endonuclease sites and other functional sequences which would facilitate the cloning and characterization of 5'-RACE products. Normally, homopolymer primers create melting temperatures that are either higher [poly (dG)•poly (dC)] or lower [poly (dA)•poly (dT)] than a typical gene-specific primer. They also can have poor specificity that can lead to mispriming at internal sequences (Invitrogen instruction manual). To minimize these problems, the abridged anchor primers (Figure 5) were designed with the selective placement of deoxyinosine residues in the poly (dG) portion. This design eliminates the need to use the mixtures of anchor and adapter primers described in the original method (Frohman, 1990). Deoxyinosine has the capacity to base-pair with all four bases; however, it does so with varying affinities. The order of stabilities for the different combinations, from greatest to least stable, are as follows: I:C, I:A, I:T, and I:G. I:C pairs were found to be slightly less stable than A:T pairs (Martin *et al.*, 1985). The selective placement of deoxyinosine residues in the 3' region of the anchor primer maintains low stability on the primer's 3'-end ($\Delta G = -8.2$ kCal/mol) and creates a melting temperature for the 16-base anchor region (66.6°C) which is comparable to that of a typical 20-mer primer with 50% GC content (Rychlik and Rhodes, 1989; Rychlik, 1995). This maximizes specific priming from the oligo-dC tail, minimizes priming at internal C-rich regions of the cDNA, and establishes a relationship of a “balanced” melting temperature for the anchor region to that of gene-specific primer 2, which is required for efficient PCR (Frohman, 1990).

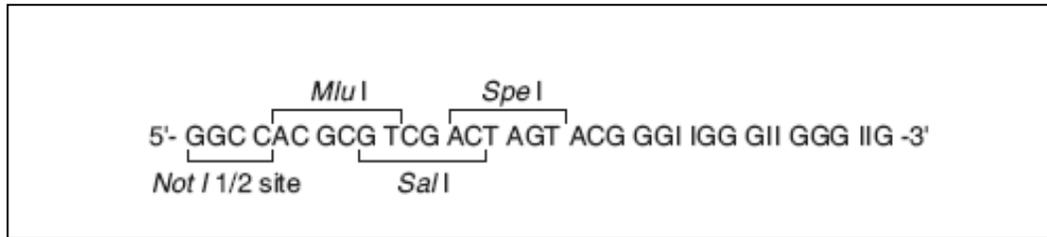


Figure 5: Sequence of the 5'-RACE abridged anchor primer (Figure adapted and modified from: 5'-RACE system, Version 2.0 instruction manual, Invitrogen™, USA).

Table 6: List of primers used in this study

Primers name	Primers sequence (5' to 3')	Restriction sites incorporated at 5' end
A) Primers for cloning into pGEM-T Easy vector:		
i) Genes of interest: <i>pezAT</i>		
PpezApezT-F	<u>GGATCC</u> ATCGTTATAGATGTATTTG	<i>Bam</i> HI
PpezApezT-R	<u>GGATCC</u> TATTTTTCAAGTAATTCATTAAG	<i>Bam</i> HI
ii) Genes of interest: <i>yefM-yoeB_{Spn}</i>		
P1P2yefMyoeB-F	<u>GGATCC</u> GCTACTGAAAGGTAGGCT	<i>Bam</i> HI
P1P2yefMyoeB-R	<u>GGATCC</u> TAGTAATGATCTTTAAAG	<i>Bam</i> HI
B) Primers for cloning into pQF52 vector:		
i) Genes of interest: <i>pezAT</i>		
P-F	GAAAC <u>CATGG</u> ATCGTTATAGATGTATTTGTAG	<i>Nco</i> I
P-R	GGGGATCCAGCAA <u>ACTCCTTT</u> ATTTTTATA	<i>Bam</i> HI
pezA-R	CTTGGATCC <u>TATCAGTATAATCTTGG</u> ATTTC	<i>Bam</i> HI
pezT-R	CCCGGATCC <u>TATTTTTCAAGTAATTC</u>	<i>Bam</i> HI
CpezA-F	TTATCCATGG <u>CGTTATCAAGATAGTCAGG</u>	<i>Nco</i> I
ii) Genes of interest: <i>yefM-yoeB_{Spn}</i>		
P1-F	TAT <u>TCCATGGG</u> TCTACTGAAAGGTAGGC	<i>Nco</i> I
mP1-F	CCGTCCATGGG <u>TCTACTGAAA</u>	<i>Nco</i> I
P2-R	CCGGATCCAACCATA <u>CCTCGTTTTAGC</u>	<i>Bam</i> HI
mP2-R	CCGGATCCAACCATA <u>CCTCG</u>	<i>Bam</i> HI
yefM-R	GGGGATCCTTCTTTTTTAGTGTTTTTTTATC	<i>Bam</i> HI
yoeB-R	GGGGATCCTTAGTAATGATCTTTAAAG	<i>Bam</i> HI
CyefM-F	GTTACCATGGGATGAATTTGAGCCTTTG	<i>Nco</i> I
mP1-R	GACTGCAGCGAAAAGAAAGAAATTG	<i>Pst</i> I
mP2-F	GACTGCAGGTGGAAAAGAGCT	<i>Pst</i> I
P2-F	GCCCATGGGATAGAAAAGAAAATCT	<i>Nco</i> I
yefM-F	CTTTCCATGGAAGCAGTCCTTTAC	<i>Nco</i> I
yoeB-F	CGTACCATGGATGCTACTCAAGTTTACAGA	<i>Nco</i> I
CyoeB-F	GGAACCATGGAAGGATATTCAACGCGAT	<i>Nco</i> I
yoeB-R	CCGGATCCTTAGTAATGATCTTTAAAGG	<i>Bam</i> HI

C) Primers for cloning into pLNBAD vector:

Genes of interest: *yefM-yoeB_{Spn}*

pLNyefM-F	<u>GACATATGATGGAAGCAGTCCTTTACTCAA</u>	<i>NdeI</i>
pLNyefM-R	<u>CCCAAGCTTTCACCTCAATCACATG</u>	<i>HindIII</i>
pLNyoeB-R	<u>CGCAAGCTTTTAGTAATGATCTTTAAAGGACAA</u>	<i>HindIII</i>
pLNCyefM-F	<u>GACATATGCCAGATGAGGACATTGTA</u>	<i>NdeI</i>

D) Primers for cloning into pET28a vector:

i) Genes of interest: *pezAT*

pETpezA-F	<u>CATATGATTGGAAAGAACATAAAATCC</u>	<i>NdeI</i>
pETpezA-R	<u>CTCGAGAATGGCCACCA</u>	<i>XhoI</i>
pETpezT-R	<u>CTCGAGTTATTTTTCAAGTAATTCATTAAG</u>	<i>XhoI</i>

ii) Genes of interest: *yefM-yoeB_{Spn}*

pETyefM-F	<u>CATATGGTGATAATAGTGGAAAAGAGCTAAAAAC</u>	<i>NdeI</i>
pETyefM-R	<u>GGATCCCTACTCCTCAATCACATGG</u>	<i>BamHI</i>
pETyoeB-R	<u>GGATCCCTTAGTAATGATCTTTAAAG</u>	<i>BamHI</i>

E) Primers for site-directed mutagenesis of pQF52 recombinant plasmids:

Genes of interest: *yefM-yoeB_{Spn}*

yefM7-F	<u>GAAGCAGTCCTTTACTAGACATTCGAAATCATTAAA</u>
yefM7-R	<u>TTTAAATGATTCGGAATGTCTAGTAAAGGACTGCTTC</u>
yefM273233-F	<u>TGAGCCTTAGACGGTGGTCAATTAATAGCCAGATGAGGAC</u>
yefM273233-R	<u>GTCTCATCTGGCTATTAATTGACCACCGTCTAAGGCTCA</u>

F) Primers for RT-PCR:

i) Genes of interest: *pezAT*

RTpezA-F	<u>GTCAGAAGTTTAAATGTATCTTATGTCG</u>
RTpezA-R	<u>CTCTAACATACGTTCAATTCATCC</u>
RTpezT-F	<u>CCAAGATTACTGATAGTGAATTCAAAC</u>
RTpezT-R	<u>CTTGCTGCAGTTCTAAATAGTGTG</u>

ii) Genes of interest: *yefM-yoeB_{Spn}*

RTyefM-F	<u>GCAGTCCTTTACTCAACATTCCG</u>
RTyefM-R	<u>CATTCCTCGTAAAACCTTGTCTGATAAC</u>
RTyoeB-F	<u>CTACTCAAGTTTACAGAAGATGCCTG</u>
RTyoeB-R	<u>CGCTATCTCCATCCATCATATAAATC</u>

G) Primers for quantitative real-time RT-PCR:

Genes of interest: *pezAT*

16SrRNA-F	<u>CGACCAGGGCTACACA</u>
16SrRNA-R	<u>CTGATCCACGATTACT</u>
lacZ-F	<u>CCTTGCAGCACATCCCCCTTTCGCC</u>
lacZ-R	<u>GTAACCGTGCATCTGCCAGTTTGAGGG</u>

H) Primers for the determination of transcriptional start site using 5'-RACE:

i) Genes of interest: *pezAT*

pezAGSP1*	<u>GCCACTATATCTCTCTATTTTC</u>
pezAGSP2*	<u>CAAGGATTAGACTCATCA</u>

ii) Genes of interest: *yefM-yoeB_{Spn}*

yefMGSP1* CATTCCCTCGTAAAACCTTGTCTGATAAC
yefMGSP2* CTGTCCCATTTCGCTCTTTGAAAG

I) Primers for primer extension analysis:

yefM_N CGCGGATCCGCTTGTACAAGTTCCTGACAATTC *Bam*HI
yefM_c CTGGAATTCGTTTTGCCAGTAGCAATAATCTGC *Eco*RI
yefM_NHis GCTCTAGAATGGTTATGGAAGCAGTCCTT *Xba*I
yoeB_cHis GAACTCGAGGTAATGATCTTTAAAGGACAAG *Ava*I
yefM-near GACCACCGTCAAAGGCTCA
yefM-far GCTTCCATAACCATAACCTCG

J) Primers for EMSA and footprinting assays:

i) Genes of interest: *pezAT*

PS-F GTATTTGTAGATATCGTTTGC
PS-R GGATTTTATGTTCTTTCCAATC

ii) Genes of interest: *yefM-yoeB_{Spn}*

PS1-F CAAGGAATGGTGCATGCC
PS2-F CGTCAACGTCGCCTTGC
PS2-R GTCCTTTAAATGATTTTCGGAATG

*GSP denotes gene-specific primer

2.3.3 PCR

PCR was carried out in a final reaction volume of 50 µl containing the following final concentration: 50 pg to 1 ng of DNA template, 0.2 µM of each primer (forward and reverse), 1 × reaction buffer provided, 2-4 mM MgSO₂/MgCl₂, 0.2 mM of each dNTP (dATP, dCTP, dGTP and dTTP) and 1.25 U of *Pfu* DNA polymerase (Fermentas, Lithuania) or Expand High Fidelity Enzyme mix (mixture of *Taq* DNA polymerase and *Tgo* DNA polymerase) (Roche, Switzerland). The reaction mix was brought to the final volume of 50 µl with sterile distilled water. PCR was performed in a thermal cycler according to the manufacturer instructions: initial denaturation of the DNA template at 95°C for 2 min, denaturation of the DNA template at 95°C for 1 min, annealing of the primers to the DNA template at 5°C lower than the melting temperature of the primer-template DNA duplex for 1 min, followed by an extension time of 2 min/kb (for *Pfu* DNA polymerase) or initial denaturation of the DNA template at 94°C for 2 min,

denaturation of the DNA template at 94°C for 15 s, annealing of the primers to the DNA template at 5°C lower than the melting temperature of the primer-template DNA duplex for 30 s, followed by an extension time of 2 min/kb (for Expand High Fidelity Enzyme mix) to be amplified at 72°C. This set of conditions (except for initial denaturation of the DNA template) was repeated for 29 cycles (for *Pfu* DNA polymerase) or 14 cycles (for Expand High Fidelity Enzyme mix), followed by a final extension at 72°C for 5 min.

2.3.4 Agarose gel electrophoresis

DNA fragments (PCR products or digested DNAs) were separated by agarose gel electrophoresis. First, 5 µl of DNA samples (PCR products or digested DNAs) were mixed with 1 µl of 6 × bromophenol blue loading dye onto an agarose gel along with an appropriate DNA size marker (100 bp or 1 kb DNA ladder, Promega, USA). To separate DNA fragments smaller than 250 bp, 1.5% (w/v) of agarose gel was used; otherwise 1% (w/v) of agarose gel was used. The loaded gel was subjected to electrophoresis in 1 × TBE buffer with a voltage of 100 V applied until the dye front had reached the edge of the gel. The gel was then stained in water containing 0.1 µg/ml ethidium bromide for approximately 20 min. DNA fragments on the gel were visualized by exposing the gel to long-wave Ultraviolet light and the image photographed using the ChemiImager™ System (Alpha Innotech, USA).

2.3.5 Purification of DNA fragments from agarose gels

DNA fragments (PCR products) were purified from agarose gels using the GFX PCR DNA and Gel Band Purification kit (General Electric Healthcare, UK). Under long-wave Ultraviolet light, the PCR-amplified DNA bands were excised from the agarose gel with a clean scalpel and transferred to a 1.5 ml microfuge tube. For each 10

mg of gel slice, 10 μ l of capture buffer was added (maximum column capacity was 300 μ l of capture buffer added to a 300 mg gel slice). The tube was then mixed by vortexing vigorously and incubated at 60°C for 5-15 min until the agarose was completely dissolved. The mixture was then transferred to the GFX column that was placed onto a collection tube and left at room temperature for 1 min, after which the column was subjected to centrifugation at full speed for 30 s in a bench top microcentrifuge. The flow-through was discarded by emptying the collection tube and the GFX column was placed back inside the collection tube. The GFX column was then washed by addition of 500 μ l wash buffer and centrifuged at full speed for 30 s. The collection tube was discarded and the GFX column was transferred to a fresh 1.5 ml microfuge tube. Elution was carried out using 10-50 μ l of elution buffer (either 1 \times TE buffer, pH 8.0, or sterile distilled water) which was directly applied to the top of the glass fiber matrix of the GFX column and incubated at room temperature for 1 min. DNA was then recovered by centrifugation at full speed for 1 min.

2.3.6 Restriction enzyme digestion

Restriction enzymes from New England Biolabs, USA, were used to digest plasmids or PCR products for recombinant plasmid construction. Digestions were carried out in sterile microfuge tubes containing 500 ng to 1 μ g of DNA, 1 \times reaction buffer, 1 \times bovine serum albumin (if needed), 1 μ l of restriction enzyme(s) as recommended by the manufacturer, and sterile distilled water to bring the final reaction volume to 20 μ l. Reactions were incubated at 37°C for 1 h. Double digestion was carried out in the same reaction mixture by using a common buffer as recommended by the manufacturer. The digested plasmids or PCR products were purified using GFX PCR DNA and Gel Band Purification kit (General Electric Healthcare, UK) prior to ligation. 500 μ l of capture buffer was added to the GFX column followed by mixing of

the digested DNA solution with the capture buffer in the column. The rest of the protocols were as described in Section 2.3.5.

2.3.7 Ligation

Cloning of *Taq* polymerase-amplified PCR products (amplified using Expand High Fidelity Enzyme mix, from Roche, Switzerland) into the T/A cloning vector pGEM-T Easy (Promega, USA) to produce pGEMT_PpezApezT and pGEMT_P1P2yefMyoeB recombinant plasmids (Table 5) utilized the single A base overhang added to the 3' termini of *Taq*-amplified DNA to anneal with the single T base overhang at the 5' ends of the pGEM-T Easy vector. In this instance, ligation was carried out in a reaction volume of 10 μ l containing 1 \times ligation buffer, 3 Weiss units of T4 DNA ligase, 50 ng of pGEM-T Easy vector and the purified PCR products in a insert:vector molar ratio of 3:1, and sterile distilled water to make up to the total reaction volume. For cloning of purified PCR products into the pQF52, pLNBAD or pET28a vectors, the reaction mixture contained 1 \times ligation buffer, 1 μ l of T4 DNA ligase (New England Biolabs, USA), digested purified PCR products and the relevant plasmid DNA in a molar ratio of 3:1. The overall concentration of vector and insert was between 1-10 μ g/ml for efficient ligation. Ligation was performed at 4°C overnight. After ligation, the reaction mixture was transformed into chemically competent *E. coli* DH5 α cells (for *pezAT* TA locus), *E. coli* TOP10 cells (for *yefM-yoeB_{Spn}* TA locus) or *E. coli* BL21-CodonPlus(DE3)-RIL cells (for pET28a recombinant plasmids).

2.3.8 Preparation of chemically-induced *E. coli* competent cells

One colony of *E. coli* DH5 α , *E. coli* TOP10 or *E. coli* BL21-CodonPlus(DE3)-RIL was inoculated into 10 ml sterile Luria-Bertani broth supplemented with appropriate antibiotics and incubated overnight at 37°C in a shaking incubator at 250 rpm. A 400 μ l aliquot of the overnight culture was inoculated into 40 ml sterile Luria-

Bertani broth in a 200 ml flask and the culture allowed to grow at 37°C with shaking at 250 rpm until an OD₆₀₀ of 0.5 was reached. The cells were then placed on ice before harvested in a refrigerated centrifuge at 6,500 × *g* and at 4°C for 10 min. The supernatant was discarded and the cell pellet was then gently resuspended in 20 ml of ice-cold 0.1 M CaCl₂. The resuspended cells were kept on ice for 30 min before they were harvested by centrifugation at 6,500 × *g* and at 4°C for 10 min. The resulting cell pellet was then resuspended in 4 ml of CaCl₂ solution (85 mM CaCl₂ containing 15% glycerol) using chilled tips and 200 µl of the resuspended competent cells were then aliquoted into pre-chilled microfuge tubes to be stored at -80°C until use.

2.3.9 Transformation of chemically-induced *E. coli* competent cells

Frozen competent *E. coli* cells in microfuge tubes were taken out from -80°C storage and thawed on ice for 15 min. The ligation reaction mixture (Section 2.3.7) was added to the competent cells and incubated on ice for 30 min. The cells were then subjected to heat shock by placing the microfuge tubes onto a 42°C heating block for 30 s and then, immediately transferring the tubes back on ice to incubate for a further 2 min. A 700 µl aliquot of sterile Luria-Bertani broth was added into each microfuge tube which was then placed in a 30°C (for pLNBAD recombinant plasmids) or 37°C (for pGEM-T Easy, pQF52 or pET28a recombinant plasmids) incubator and shaken at 250 rpm for 1 h. The transformed cells were then plated onto Luria-Bertani agar plates supplemented with appropriate antibiotics as well as 80 µg/ml of 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-gal) when necessary (for blue-white selection) and incubated overnight at 37°C.

2.3.10 Screening of transformed cells

To screen the transformants for those that harbored the required recombinant plasmid, selected cells (antibiotic selection and/or blue-white selection using X-gal) were re-streaked on Luria-Bertani agar plates supplemented with appropriate antibiotics and incubated overnight at 30°C (for pLNBAD recombinant plasmids) or 37°C (for pGEM-T Easy, pQF52 or pET28a recombinant plasmids). Plasmid DNA was extracted (Section 2.3.11) from these cells and digested using the appropriate restriction enzyme(s) (Section 2.3.6). The resulting DNA fragments were separated using gel electrophoresis and their sizes were estimated. The correct recombinant plasmids were sequenced and the sequence results were analyzed using Bioedit Version 7.0.4.1, which can be accessed at www.mbio.ncsu.edu/BioEdit/bioedit.html.

2.3.11 Plasmid DNA extraction

Plasmid DNA was extracted using the Wizard® *Plus* SV Minipreps DNA Purification System (Promega, USA). One full loop of bacterial culture was thoroughly resuspended in 250 µl of cell resuspension solution in a microfuge tube before adding 250 µl of cell lysis solution and mixed by inverting the tube four times. This was followed by the addition of 10 µl of alkaline protease solution and incubating the tube at room temperature for 5 min before mixing in 350 µl of neutralization solution. The tube was then centrifuged at 14,000 × *g* for 10 min at room temperature.

The cleared lysate obtained after centrifugation was transferred into a spin column that was placed on top of a collection tube which was then subjected to centrifugation at 14,000 × *g* for 1 min at room temperature. The flow-through was discarded and the column was reinserted into the collection tube. The column was then washed with 750 µl of wash solution followed by a final wash using 250 µl of wash solution. For the first wash, the column along with the collection tube was centrifuged

at $14,000 \times g$ for 1 min whereas for the final wash, centrifugation was carried out at $14,000 \times g$ for 2 min.

For elution, the spin column was transferred to a sterile 1.5 ml microfuge tube. Then, 50-100 μ l of nuclease-free water was added to the spin column, incubated for 1 min at room temperature and centrifuged at $14,000 \times g$ for 1 min. The column was discarded and the eluted DNA was stored at -20°C until use.

2.4 Site-Directed Mutagenesis

Substitution of a few nucleotides was done using pQF_P1P2yM and pQF_P1P2yMyB as templates to introduce an amber stop codon at the 7th codon of YefM_{S_{pn}} to yield pQF_M1yM and pQF_M1yMyB (Table 5). Another two mutant recombinants were constructed from pQF_P1P2yM and pQF_P1P2yMyB by replacing the 27th, 32nd and 33rd codon of YefM_{S_{pn}} with amber, ochre and amber stop codons respectively yielding plasmids pQF_M2yM and pQF_M2yMyB (Table 5). The substitution of nucleotides was done using the QuikChange II XL Site-Directed Mutagenesis Kit (Stratagene, USA). The protocol needs a supercoiled double-stranded DNA vector with an insert of interest as template and two synthetic oligonucleotide primer pairs, which contains the desired mutation (Table 6). The oligonucleotide primers, each complementary to opposite strands of the vector, were extended during temperature cycling by *PfuUltra* HF DNA polymerase, which was provided in the kit, without primer displacement. Extension of the oligonucleotide primers generated a mutated plasmid containing staggered nicks. Following temperature cycling, the parental DNA was then digested with *DpnI*. The *DpnI* endonuclease, which targeted 5'-Gm6ATC-3' sequences, is specific for methylated and hemimethylated DNA and was therefore used to select for mutation containing synthesized DNA (Nelson and McClelland, 1992). DNA isolated from almost all *E. coli* strains is dam methylated (dam⁺) and therefore susceptible to *DpnI* digestion. The nicked vector DNA incorporating the desired mutations was then transformed into *E. coli* XL10-Gold ultracompetent cells for nick repair.

The reaction mixture contained 10 ng of double-stranded DNA template, 1 × reaction buffer, 125 ng of each forward and reverse oligonucleotide primers, 1 µl of dNTP mix which was provided in the kit, 3 µl of QuickSolution, distilled water added to a final volume of 50 µl and finally 1 µl (2.5 U) of *PfuUltra* HF DNA polymerase.

The reaction mixture was mixed well and then subjected to the following temperature cycling: initial denaturation of the template at 95°C for 1 min, denaturation at 95°C for 50 s, annealing of the oligonucleotide primers to the DNA template at 60°C for 50 s, followed by an extension time of 1 min/kb at 72°C. This set of conditions (except for initial denaturation of the DNA template) was repeated for 18 cycles, followed by a final extension at 68°C for 7 min. After that, 1 µl (10 U) of *DpnI* was then added to the amplification reaction and the reaction was then incubated at 37°C for 1 h. Prior to transformation, 2 µl of a β-mercapthoethanol mix, which was provided in the kit, was added into 45 µl of the *E. coli* XL10-Gold ultracompetent cells to increase the transformation efficiency. The ultracompetent cells were then gently mixed with 2µl of the *DpnI* treated DNA, followed by heat-pulse in a 42°C heat block for 30 s. The reaction tube was then placed on ice for 2 min prior to the addition of 0.5 ml of preheated (42°C) NZY⁺ broth, and then incubated at 37°C for 1 h with shaking at 250 rpm. The cells were spread on Luria-Bertani agar plates supplemented with Ampicillin and 80 µg/ml of X-gal. The transformation plates were placed overnight at 37°C and the blue colonies were selected for screening (Section 2.3.10).

2.5 Protein Purification

2.5.1 Purification of PezA and PezAT protein complex

E. coli BL21-CodonPlus(DE3)-RIL harboring the pET28a_HisPezA and pET28a_HisPezAPezT recombinants (Table 5) were inoculated in 10 ml of sterile Luria-Bertani broth supplemented with 50 µg/ml Kanamycin and 34 µg/ml Chloramphenicol and grown overnight at 37°C and with shaking at 250 rpm. The overnight cell cultures were diluted to 10 × with fresh sterile Luria-Bertani broth supplemented with 50 µg/ml kanamycin. The cells were grown at 37°C and with shaking at 250 rpm until OD₆₀₀ ~ 0.6, followed by induction with 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) for 4 h at 28°C. The cells were then harvested by centrifugation at 6,500 × g at 4°C for 10 min and the resulting cell pellet was resuspended in 5 ml of phosphate buffered saline (140 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, pH 7.4) containing a protease inhibitor cocktail (General Electric Healthcare, UK) to a final concentration of 10 µl/ml and lysozyme to a final concentration of 1 mg/ml. The cell suspension was incubated on ice for 30 min prior to sonication.

All purification steps were performed at 4°C, unless otherwise stated. The cells were lysed by sonication on ice using a 36 × 5 s burst at 3 W with a 10 s cooling period between each burst. The cell debris and unbroken cells were cleared by centrifugation at 14,000 × g for 20 min. The supernatant was collected and the pellet was resuspended in 1 ml of phosphate buffered saline (pH 7.4) and resonicated. The whole process was repeated twice. The collected supernatants were pooled together and subjected to ultracentrifugation at 105,000 × g for 20 min to remove cell membrane and insoluble fractions. The pellet was discarded and the supernatant was labeled as crude lysate.

The crude lysate was subsequently precipitated by ammonium sulfate fractionation. Ammonium sulfate was slowly added with constant stirring to a final

concentration of 40% (fraction I) and equilibrated for 1.5 h. The turbid solution was centrifuged at $17,000 \times g$ for 20 min. The proteins in the supernatant were precipitated again by the addition of ammonium sulfate to final concentrations of 70% (fraction II) and 100% (fraction III) respectively. The pellet resulting from each fraction was saved and resuspended in 1 ml of phosphate buffered saline (pH 7.4), and dialyzed against binding buffer containing 20 mM sodium phosphate buffer and 0.5 M NaCl (pH 7.4).

To identify the fraction which contains the required target proteins, 10 μ l of each dialyzed samples were loaded onto a 16% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gel. To the identified fraction(s), 20 mM of imidazole was added and the mixture was loaded onto a 1 ml HisTrapTM HP affinity column (General Electric Healthcare, UK) that was pre-equilibrated with 5 column volumes of binding buffer containing 20 mM imidazole (pH 7.4). HisTrapTM HP is a pre-packed column containing precharged Ni SepharoseTM High Performance which consists of 34 μ m highly cross-linked agarose beads with an immobilized chelating group. The column was then washed sequentially with 15 column volumes of binding buffer. Bound proteins were eluted from the column using a linear gradient with 5 - 10 column volumes of elution buffer (20 mM sodium phosphate buffer, 0.5 M NaCl and 500 mM imidazole, pH 7.4). The whole process was carried out using a liquid chromatography system (ÄKTA designTM, General Electric Healthcare, UK). The fractions containing the recombinant (His)₆-PezA and (His)₆-PezAT were determined by SDS-PAGE analysis.

2.5.2 Purification of YefM_{Spn} and YefM-YoeB_{Spn} protein complex

To obtain adequate amounts of protein for this study, 8 ml of an overnight culture of *E. coli* BL21-CodonPlus(DE3)-RIL harboring the pET28a_HisYefMHis or pET28a_HisYefMYoeB recombinants (Table 5), were diluted into 800 ml of fresh

Luria-Bertani broth supplemented with 50 µg/ml Kanamycin and 34 µg/ml Chloramphenicol, and allowed to grow until $OD_{600} \sim 0.4$ with shaking at 250 rpm and at 37 °C. The cell cultures were cooled to 30°C before adding IPTG to a final concentration of 0.25 mM. Rifampicin was added (to a final concentration of 200 µg/ml) 15 min after addition of IPTG. After 2 h, the cells were harvested at $6,500 \times g$ at 4°C for 20 min. The cell pellets were then resuspended in 40 ml of 1 × His buffer (10 mM Tris, pH7.6, 1 M NaCl, 5 mM β-mercaptoethanol and 5% glycerol) containing 10 mM imidazole and 10 µl/ml protease inhibitor mix (General Electric Healthcare, UK) .

All purification steps were performed at 4°C. The cell suspension was then subjected to French Press (Constant System, UK) twice and the supernatant was separated from the cell debris and unbroken cells by centrifugation at $17,000 \times g$ for 20 min. The crude lysate was then loaded into a chromatography column packed with HIS-Select Nickel Affinity Gel (Sigma, USA) with the flow rate of 45 ml/h. HIS-Select Nickel Affinity Gel is an immobilized metal-ion affinity chromatography product which is a quadridentate chelate on 6% beaded agarose charged with nickel and is selective for recombinant proteins with histidine tags and exhibits low non-specific binding of other proteins. The column was then extensively washed with 100 ml of 1 × buffer A (10 mM Tris, pH 7.6, 0.3 M NaCl, 5 mM β-mercaptoethanol and 5% glycerol) containing 10 mM imidazole with a flow rate of 45 ml/h. The histidine-tagged YefM_{S_{pn}} and YefM-YoeB_{S_{pn}} proteins were eluted from the column using 1 × His buffer with 250 mM imidazole at a flow rate of 45 ml/h. The protein fractions were collected and analysed using SDS-PAGE.

2.5.3 Tris-Tricine SDS-PAGE

The protein fractions (Section 2.5.1 and 2.5.2) were analysed with SDS-PAGE under denaturing conditions. The separation of proteins was carried out using Mini-

PROTEAN Tetra Electrophoresis System (Bio-Rad Laboratories, USA). Polyacrylamide gels that were used comprised of a lower separating gel and an upper stacking gel. Separating gels, containing 16% polyacrylamide, were prepared by mixing the following solutions: 0.68 ml of sterile distilled water, 2.34 ml of acrylamide:bis-acrylamide (40:1), 2 ml of gel buffer (3 M Tris, pH 8.45, 0.3% SDS) and 0.92 ml of 87% Glycerol. Then, 60 μ l of 10% ammonium persulphate and 6 μ l of TEMED were added lastly and the mixture was then stirred gently and poured immediately into the gel caster to a level approximately 1 cm from the top of the smaller glass plate. The surface of the gel mix was over-layered with deionized water and the gel allowed to polymerize for about 45 min. The deionized water was then poured off and the gel mould filled with a 4% polyacrylamide stacking gel mix which contained 1.97 ml of sterile distilled water, 0.29 ml of acrylamide:bis-acrylamide (40:1), 0.74 ml of gel buffer (3 M Tris, pH 8.45 and 0.3% SDS) as well as 37.5 μ l of 10% ammonium persulfate and 6 μ l of TEMED which were added last. The stacking gel mix was filled to the top of the gel caster and a Teflon comb was then inserted into place.

After the stacking gel had polymerized, the comb was removed and the wells of the gel were rinsed thoroughly with 1 \times cathode buffer (0.1% SDS, 0.1 M Tricine and 0.1 M Tris, pH 8.25). The gel with the caster and a blank plate were then placed into the vertical electrophoresis tank where the inner space (space between the gel with the caster and the blank plate) was filled with 1 \times cathode buffer and the outer space filled with 1 \times anode buffer (0.2 M Tris, pH 8.9).

The protein fractions (from Section 2.5.1 and 2.5.2), with 10 μ l each, were mixed with 2 μ l of sample buffer (0.225 M Tris-Cl, pH 6.8, 50% glycerol, 5% SDS, 0.5% bromophenol blue and 0.25 M DTT) and heated at 95°C for 3 min before loading onto the gel. The protein marker was run along with the protein samples. Electrophoresis was performed at 60 V until the dye migrated down to the bottom of the

stacking gel, whereupon the voltage was then increased to 180 V until the dye migrated to the bottom of the separating gel.

The gel was then detached from the caster and stained with Coomassie Brilliant Blue (1 tablet of PhastBlue R (General Electric Healthcare, USA) in 1.6 L of 10% acetic acid) for 1 h by gentle shaking, followed by destaining in a 10% (v/v) acetic acid solution for 10 min. The image of the SDS-PAGE gel was photographed using the ChemiImager™ System (Alpha Innotech, USA) under white light.

2.5.4 Protein blotting for N-terminal sequencing

The protein samples were separated on SDS-PAGE along with The SeeBlue® Pre-Stained Protein Standard (Invitrogen™, USA) as a molecular weight marker, which facilitates the monitoring of the progress of the gel run. When the dye had migrated to the bottom of the gel, the gel was detached from the cast and sandwiched together with a Immun-Blot® PVDF membrane (Bio-Rad, USA), which was previously immersed in 100% methanol for a few seconds followed by transfer buffer (25 mM Tris, 192 M glycine and 20% methanol) for 3 min, as well as blotting paper and two sponges with a cassette. This assembly was then placed into a vertical electrophoresis tank and run with transfer buffer at 60 V for 1 h 30 min. The membrane was then detached from the assembly and rinsed three times with deionized water for 5 min before staining with 0.1% CBB R-250 in 1% acetic acid and 40% methanol for 20 min by gently shaking. After that, the membrane was de-stained with 50% methanol in deionized water until the protein band was observed. The membrane was dried at room temperature and the band excised for N-terminal sequencing which was carried out via a commercial service at the Centro de Investigaciones Biológicas, Spain.

2.5.5 Dialysis of purified proteins and determination of protein concentration

Dialysis tubing (1 ml/cm) was cut depending on the volume of protein samples and soaked in sterile distilled water for 30 min. Protein samples were added into the dialysis tubes, which were then sealed with clips and placed into a beaker with 1 L of chilled (4°C) 1 × dialysis buffer (20 mM Tris, pH 7.6, 1 mM EDTA, 1 mM DTT, 1% glycerol and 500 mM NaCl), and then gently stirred using an electromagnetic stirrer in a cold room. The buffer was changed twice every 3 h and then left overnight. The protein samples were then centrifuged at $6,500 \times g$ for 10 min at 4°C and the supernatant and any precipitate formed were separated on SDS-PAGE to confirm if the proteins were in the supernatant. The proteins in the supernatant were then concentrated using MicrosepTM Centrifugal Devices (Pall Corporation, USA) until the desired concentration was reached. The concentration of the protein was measured using a Nanodrop spectrophotometer (NanoDrop Technologies, USA).

2.6 Determination if the Toxin and Antitoxin Genes are Co-transcribed

To investigate the possibility that the *pezAT* and *yefM-yoeB_{Spn}* TA loci are co-transcribed, RT-PCR was utilized to detect for the presence of transcripts containing both toxin and antitoxin genes by using primer pairs that span both the toxin and antitoxin reading frames (i.e., the forward primer was designed to anneal within the antitoxin gene whereas the reverse primer was designed to anneal within the toxin gene). The primers used are listed in Table 6. Total RNA extracted from *E. coli* DH5 α harboring the pGEMT_PpezApezT (for *pezAT*) and pGEMT_P1P2yefMyoeB (for *yefM-yoeB_{Spn}*) recombinants was used as template for RT-PCR. For positive controls, primers to amplify the toxin and antitoxin genes separately were utilized whereas for negative controls, reverse transcriptase was not included in the reaction mixture to rule out DNA contamination.

RT-PCR was carried out using the AccessQuick™ RT-PCR System (Promega, USA). The AccessQuick™ Master Mix contains the *Tfl* DNA polymerase, dNTPs, MgSO₂ and reaction buffer whereas the AMV reverse transcriptase enzyme was provided in a separate tube. The 50 μ l of reverse transcription reaction mix contained 25 μ l of 2 \times AccessQuick™ Master Mix, 1 μ l of 10 μ M of each forward and reverse primer, 10 pg of total RNA template and 1 μ l (5 U) of AMV reverse transcriptase added as the final component. To initiate reverse transcription, the tubes were incubated at 45°C for 45 min, followed by the PCR thermal cycling (Section 2.3.3). The resulting products were then separated by electrophoresis on 1% agarose gels. The band was excised from the gel, purified as described in section 2.3.5 and verified by DNA sequencing.

2.7 Determination of transcriptional start sites

2.7.1 5'-RACE

The 5'-RACE system, Version 2.0 (InvitrogenTM, USA) was used to determine the transcriptional start sites for both the *pezAT* and *yefM-yoeB_{Spm}* TA genes in this study. RACE is a procedure used to amplify sequences from an mRNA template between a defined internal site and unknown sequences at either the 3' or 5'-end of the mRNA (Frohman *et al.*, 1988).

Total RNA extracted from the *E. coli* DH5 α cells harboring the pGEMT_PpezApezT (for *pezAT*) and pGEMT_P1P2yefMyoeB (for *yefM-yoeB_{Spm}*) recombinants was subjected to reverse transcription using gene-specific primer 1 (Table 6), which anneals at least 300 bp downstream from the antitoxin transcriptional start site. A homopolymeric tail was added using terminal transferase and dCTP to the 3' end of the synthesized cDNA. The dC-tailed cDNA was then PCR-amplified using the abridged anchor primer, which was complementary to the homopolymeric tail and another nested primer (gene-specific primer 2) (Table 6) which was complementary to a few codons upstream of the 5' end of the cDNA. The amplified product was then separated by electrophoresis on 1% agarose gels. The band was excised from the gel, purified as described in section 2.3.5 and verified by DNA sequencing. An overview of the procedure is schematically shown in Figure 6.

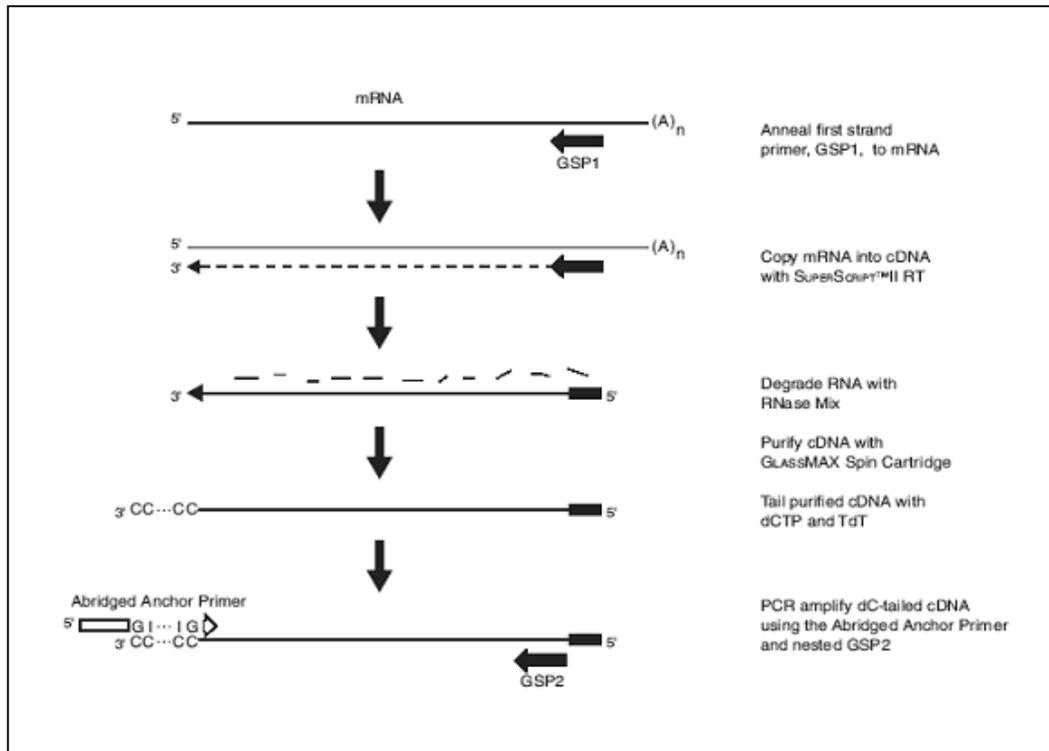


Figure 6: An overview of 5'-RACE system (Invitrogen™, USA). GSP1 denotes gene-specific primer 1 and GSP2 denotes gene-specific primer 2 (Figure adapted and modified from: 5'-RACE system, Version 2.0 instruction manual, Invitrogen™, USA).

To synthesize the first strand cDNA, 10–25 ng of gene-specific 1 and 1–5 µg of total RNA was added to a thin-walled PCR tube and brought up to a final volume of 15.5 µl with DEPC-treated water. The mixture was incubated at 70°C for 10 min to denature the RNA. The mixture was then chilled on ice for 1 min and centrifuged briefly to collect the contents. After that, 2.5 µl of 10 × PCR buffer, 2.5 µl of 25 mM MgCl₂, 1 µl of 10 mM dNTP mix and 2.5 µl of 0.1 M DTT were added to the mixture. The resulting mixture was suspended gently, centrifuged briefly to collect the reaction at the bottom of the tube, and incubated for 1 min at 42°C. To start the reaction, 1 µl of SuperScript™ II reverse transcriptase was added and synthesis of the first strand cDNA was allowed to proceed for 50 min at 42°C. The reaction was then terminated by incubating at 70°C for 15 min. To digest the RNA remaining in the reaction tube, 1 µl

of RNase mix was added, mixed gently and incubated for 30 min at 37°C. The reaction was then collected by brief centrifugation and placed on ice.

The synthesized cDNA was purified using S.N.A.P column purification (Invitrogen™, USA) by addition of 120 µl of binding solution (6 M NaI) to the first strand reaction and transferring the cDNA/NaI solution to a S.N.A.P column which was then subjected to centrifugation at $13,000 \times g$ for 20 s. The flow-through was discarded and the cartridge was placed back into the collection tube. The column was then washed using 0.4 ml of ice-cold $1 \times$ wash buffer which was added to the spin cartridge and subjected to centrifugation at $13,000 \times g$ for 20 s. The resulting flow-through was discarded. This wash step was repeated for three additional times. The cartridge was then washed twice using 400 µl of 70% ethanol at 4°C as in the above steps. After removing the final 70% ethanol wash from the tube, the column was centrifuged for another 1 min at $13,000 \times g$. The spin cartridge was then transferred to a fresh sample recovery tube and 50 µl of sterile distilled water (that was preheated to 65°C) was added to the column which was subsequently centrifuged at $13,000 \times g$ for 20 s to elute the cDNA.

The cDNA was then extended by adding dCTP at the 3' termini using terminal deoxytransferase. Into a thin-walled PCR tube containing 10 µl of the S.N.A.P-purified cDNA sample, 6.5 µl of DEPC-treated water, 5 µl of $5 \times$ tailing buffer, and 2.5 µl of 2 mM dCTP was added. The reaction was mixed gently, incubated for 2 min at 94°C, and immediately chilled on ice for 1 min prior to the addition of 1 µl of terminal deoxytransferase. The tailing reaction was allowed to proceed for 10 min at 37°C before terminating the reaction by heating the mixture containing terminal deoxytransferase for 10 min at 65°C. The contents of the reaction was collected by brief centrifugation of the tube and then placed on ice.

To amplify the dC-tailed cDNA, 31.5 μ l of sterilized distilled water, 5 μ l of 10 \times PCR buffer (containing 200 mM Tris-Cl, pH 8.4 and 500 mM KCl), 3 μ l of 25 mM MgCl₂, 1 μ l 10 mM dNTP mix, 2 μ l of 10 μ M nested gene-specific primer 2, 2 μ l of 10 μ M abridged anchor primer and the 5 μ l dC-tailed cDNA were added into a thin-walled PCR tube. After 0.5 μ l of *Taq* DNA polymerase (5 U/ μ l) was added, the reaction was mixed and subjected to PCR. The PCR reaction conditions were set as follows: Initial denaturation at 94°C for 2 min, denaturation at 94°C for 1 min, annealing of primers at 55°C for 30 s, and primer extension at 72°C for 45 s for 35 cycles. The final extension was set at 72°C for 7 min.

An aliquot of the amplified product was separated by electrophoresis on 1% agarose gel to estimate the sizes of the amplified products obtained. DNA fragments of the expected sizes were purified and cloned into pGEM-T Easy cloning vector and the resulting recombinants sequenced to determine the possible transcript start sites obtained by 5'-RACE. The experiment was repeated three times from three different preparations of RNA samples.

2.7.2 Primer extension analyses

Total RNA was isolated from *S. pneumoniae* R6 using the Aurum total RNA minikit (Bio-Rad Laboratories, USA). The primer extension assays were performed as described (del Solar *et al*, 1995), by annealing the total RNA with two different specific primers (located 8 nucleotides (yefM-far) or 90 nucleotides (yefM-near) downstream of the *yefM_{Spn}* ATG start codon, respectively at 65°C for 5 min. A mixture of dATP, dGTP, dTTP (100 μ M each) and 10 μ M [α -³²P]-dCTP were then added. The primers were then extended with ThermoScript™ reverse transcriptase (Invitrogen™, USA) at 50°C or at 55°C for 30 min. After that, 80 μ l of TE was added, followed by 100 μ l of phenol:chloroform (1:1). The mixture was vortexed for 5 min, and centrifuged at 13,000

$\times g$ for 5 min at room temperature. The upper aqueous phase was then transferred to a new microfuge tube. Later on, 1/10 volume of 3 M NaAc (pH 7.0) and 2.5 volume of ethanol were added and mixed. The mixture was then placed at -70°C for 20 min and then subjected to centrifugation at $13,000 \times g$ for 15 min at 4°C . The resulting supernatant was discarded. The pellet was dissolved in 8 μl of stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol). Dideoxy-mediated chain-termination DNA sequencing reactions (Section 2.10.2.4) was obtained by annealing the yefM-near or the yefM-far primers to the pMBH13 (Nieto *et al*, 2007) denatured plasmid DNA, respectively. The samples and the DNA sequencing reactions were separated on 8% sequencing gels (Section 2.10.2.2) at 1800 V.

2.8 Assays for Promoter Activity

Various fragments of *pezAT*, *yefM-yoeB_{Spn}* and their upstream regions were assayed for promoter activity using the promoter probe vector, pQF52 (McLean *et al.*, 1997), in which these fragments were cloned upstream of a promoter-less *lacZ* gene and the resulting recombinants assayed for β -galactosidase activity in an *E. coli* host. For the *pezAT* TA locus, the results obtained were validated using quantitative real-time RT-PCR to measure the transcript levels of *lacZ* in *E. coli* DH5 α cells harboring the pQF52-derived recombinants. For the *yefM-yoeB_{Spn}* TA locus, besides pQF52-derived recombinants, various fragments of *yefM-yoeB_{Spn}* were also cloned downstream of a P_{BAD} promoter in pLNBAD vector, and co-transformed into *E. coli* TOP10 together with pQF52-derived recombinants which harbour various fragments of *yefM-yoeB_{Spn}* promoter region upstream of the promoter-less *lacZ* gene. This was to provide the *yefM_{Spn}* and the *yefM-yoeB_{Spn}* reading frames in *trans* to their promoters and upstream regulatory regions. The pLNBAD vector harbours an inducible P_{BAD} promoter which is positively and also negatively regulated by the products of the *araC* gene (Ogden *et al.*, 1980; Schleif, 1992). The AraC dimer contacts with the I₁ site and O₂ site, which is further upstream of I₁, forming a DNA loop and therefore prevents transcription from the P_{BAD} promoter. In the presence of L-arabinose, AraC will form a complex with L-arabinose, causing AraC to detach from the O₂ site and bind to an I₂ site, which is adjacent and downstream of the I₁ site, and thus releasing the DNA loop and allowing transcription to begin (Figure 7). In addition, the cAMP activator protein (CAP)-cAMP complex will bind to the CAP site, which is located adjacent and upstream of I₁ site, and stimulates binding of AraC to I₁ and I₂ (Figure 7). Therefore, addition of glucose can repress basal expression or avoid leakage of the P_{BAD} promoter as glucose reduces the levels of 3',5'-cyclic AMP, which in turn decreases the binding of CAP (Miyada *et al.*, 1984). *E. coli* TOP10 was used in this assay because it is capable of transporting L-

arabinose, but not metabolizing it, and therefore the level of L-arabinose will be constant inside the cell and not decrease over time. Moreover, this strain is *araBADC*⁻ and *araEFGH*⁺. It has deletions for both *araBA* and *araC*, and the gene for *araD* has a point mutation in it, making it inactive (Instruction manual, InvitrogenTM, USA).

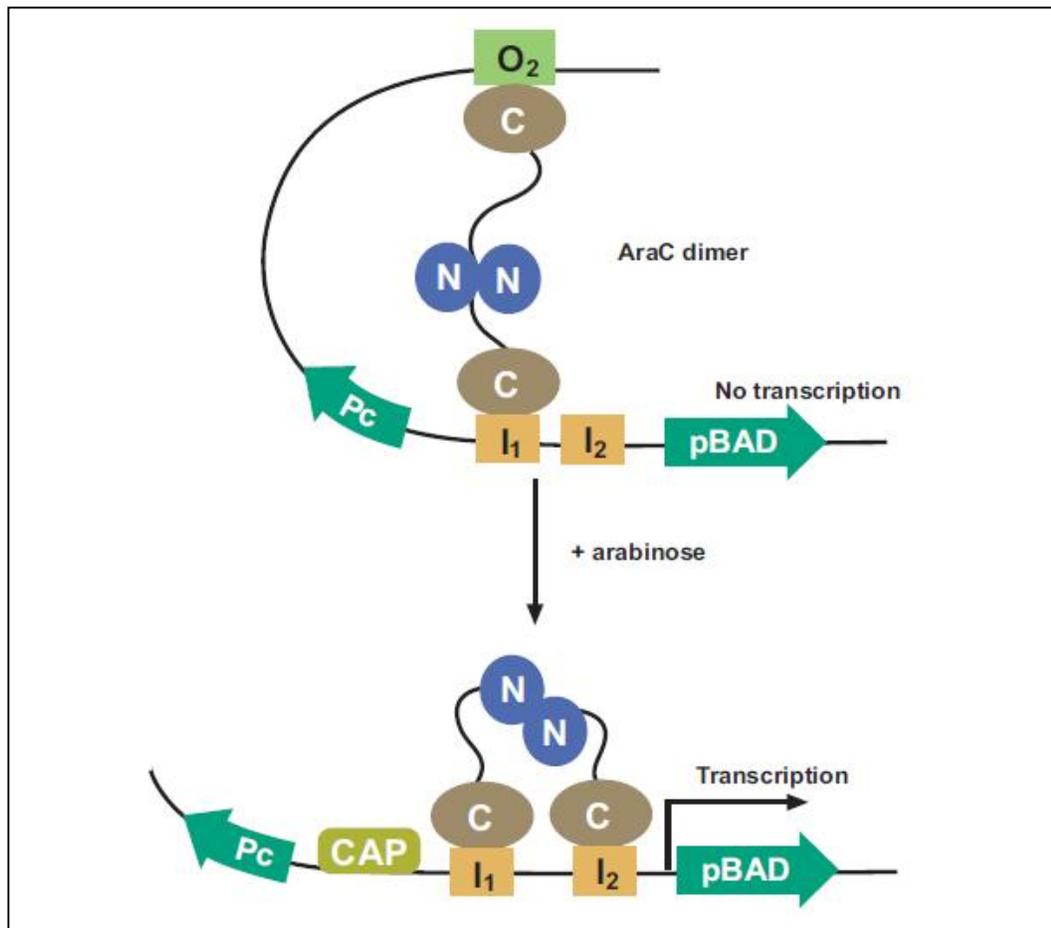


Figure 7: Regulation of the P_{BAD} promoter by L-Arabinose (Figure taken from: pBAD TOPO TA Expression Kit instruction manual, InvitrogenTM, USA).

2.8.1 Determination of promoter activities using β -galactosidase assays

The recombinant *E. coli* clones containing pQF52 and its derivatives, as well as pLNBAD-derived recombinants were cultured in 10 ml Luria-Bertani broth supplemented with the appropriate antibiotics and grown overnight at 37°C (for constructs that harbour the *pezAT* TA locus) or at 30°C (for constructs that harbour the *yefM-yoeB_{Spm}* TA locus) in a shaker incubator at 250 rpm. An aliquot of the overnight

bacterial cultures were inoculated into 10 ml of sterile fresh Luria-Bertani broth supplemented with appropriate antibiotics with starting $OD_{600} \sim 0.02$. For cells that harboured pLNBAD-derived recombinant plasmids, two sets of cultures were prepared where one set was induced by adding 1 mM of L-Arabinose and another set was grown without addition of any inducer. All cultures were allowed to grow under the same conditions until mid-log phase ($OD_{600} \sim 0.4 - 0.6$) where they were then placed on ice. A 2 ml aliquot of the bacterial culture was harvested in a 2 ml sterile microfuge tube by centrifugation at $3,500 \times g$ for 10 min and the resulting cell pellet resuspended in 2 ml chilled Z buffer (containing 0.06 M $Na_2HPO_4 \cdot 7H_2O$, 0.04 M $NaH_2PO_4 \cdot H_2O$, 0.01 M KCl, 0.001 M $MgSO_4$, 0.05 M β -mercapthoethanol, pH 7). The OD_{600} of the resuspended cells was measured spectrophotometrically with Z buffer as a blank. To permeabilize the cells, 1 ml cells in Z buffer was transferred into a 2 ml microfuge tube into which 100 μ l chloroform and 50 μ l 0.1% SDS were added. The mixture was then vortexed and equilibrated for 5 min in a 28°C heat block.

β -galactosidase, encoded by *lacZ*, hydrolyzes β -D-galactosides, allowing the bacteria to grow on carbon sources such as lactose by cleaving it into glucose and galactose for the cells to utilize as both carbon and energy sources. The β -galactosidase assay carried out in this study utilizes o-nitrophenyl β -D-galactopyraniside (ONPG) in place of lactose as the substrate. Cleavage of ONPG by β -galactosidase releases o-nitrophenol which has a yellow color and absorbs at 420 nm (Miller, 1972). Thus, the increase in absorbance at 420 nm would be a reflection of β -galactosidase activity.

The β -galactosidase assay is initiated by the addition of 0.2 ml ONPG (4 mg/ml) as the substrate into the lysate. The tubes were vortexed before incubation in the 28°C heat block and the colour changes was observed. After sufficient yellow color (as the colour of Luria-Bertani broth) had developed, the reaction was stopped by the addition of 0.5 ml 1 M Na_2CO_3 and mixed by vortexing. The addition of Na_2CO_3 will raise the

pH of the solution to 11 and thus stop the enzymatic reaction. The time taken from the addition of ONPG to the stopping of the reaction with Na₂CO₃ was precisely recorded. The mixture was transferred to a microfuge tube and centrifuged at maximum speed for 5 min to remove cellular debris and chloroform. The OD at 420 nm and at 550 nm for each tube was recorded (blanked against the same mixture but without cells). The reading at 420 nm is a combination of absorbance by *o*-nitrophenol and light scattering by cell debris. The absorbance at 550 nm corrects for light scattering as there is no absorbance from *o*-nitrophenol at this wavelength. The light scattering at 420 nm is proportional to that at 550 nm. The units of enzyme activity, expressed as Miller units, were calculated using the following equation (Miller, 1972):

$$\text{Miller Units} = 1000 \times [(\text{OD}_{420} - 1.75 \times \text{OD}_{550})] / (T \times V \times \text{OD}_{600})$$

where:

- OD₄₂₀ and OD₅₅₀ were read from the reaction mixture
- OD₆₀₀ reflected cell density in the washed cell suspension
- *T* = time of the reaction, in min
- *V* = volume of culture used in the assay, in ml.

2.8.2 Determination of transcript levels using quantitative real-time RT-PCR

The results obtained from the β-galactosidase assays were validated by measuring the *lacZ* transcript levels using quantitative real-time RT-PCR. Total RNA was extracted from *E. coli* DH5α cells harboring the recombinant pQF52 plasmids and subjected to quantitative real-time RT-PCR using primers designed based on the *lacZ* gene sequence (NCBI accession number: AAC73447). The gene encoding 16S rRNA, which is present at seven copies (GeneID: 944897, 947071, 947777, 748270, 948332, 948466 and 948511) in the *E. coli* K12 genome (NCBI accession number: NC_000913)

(Blattner *et al.*, 1997) was used as the reference housekeeping gene. Quantitative real-time RT-PCR was carried out using the LightCycler 2.0 (Roche, Switzerland).

2.8.2.1 Total RNA extraction

A 100 µl aliquot of the overnight cultures of *E. coli* DH5α cells carrying the recombinant pQF52 plasmids was inoculated into 10 ml sterile Luria-Bertani broth and grown with shaking at 250 rpm and at 37°C until mid-log phase (OD₆₀₀ ~0.4–0.6). RNA extraction was carried out using the RNeasy Mini Kit (Qiagen, Germany). A 1 ml aliquot of the mid-log phase *E. coli* cell culture was transferred into a sterile 1.5 ml microfuge tube and the cells harvested by centrifugation at 2,500 × *g* for 5 min at 4°C. The resulting cell pellet was resuspended thoroughly in 100 µl of 1 × TE (pH 8.0) buffer containing 400 µg/ml lysozyme by vortexing and the mixture incubated at room temperature for 5 min. Following that, 350 µl Buffer RLT was added, thoroughly mixed by vortexing, and the resulting cell lysate was then thoroughly mixed with 250 µl ethanol 99.5% (w/w). The sample was transferred onto an RNeasy mini column placed in a 2 ml collection tube and centrifuged for 15 s at ≥8,000 × *g*.

The flow-through was discarded and 350 µl buffer RW1 was added to the RNeasy column and centrifuged for 15 s at ≥8,000 × *g* to wash the column. The flow-through was again discarded and 80 µl DNaseI (27.27 Kunitz units in Buffer RDD) was added directly onto the RNeasy silica-gel membrane and incubated for 15 min at room temperature to digest any remaining DNA. The column was then washed with 350 µl buffer RW1 by centrifugation for 15 s at ≥8,000 × *g*.

The RNeasy column then was transferred to a new collection tube and washed with 500 µl Buffer RPE by centrifugation for 15 s at ≥8,000 × *g*. Washing with Buffer RPE was repeated twice but for the final wash, centrifugation was carried out for 2 min to ensure that no ethanol was carried over during elution of RNA. The RNeasy column

was then transferred to a new 1.5 ml collection tube for elution which was carried out by applying 50 μ l RNase-free water directly onto the RNeasy silica-gel membrane, incubated at room temperature for 1 min, and centrifuged for 1 min at $\geq 8,000 \times g$. The eluted RNA was electrophoresed on a 1% agarose gel to check its integrity. To estimate the purity of the RNA, 5 μ l of sample was diluted 10-fold with 10 mM of Tris-Cl, pH7.5 and then measured spectrophotometrically at 260 nm and 280 nm. Pure RNA has an A_{260}/A_{280} ratio of 1.9-2.1 in 10 mM of Tris-Cl, pH7.5 (Instruction manual, Qiagen, Germany).

2.8.2.2 Quantitative real-time RT-PCR

The LightCycler RNA master SYBR Green I kit was used to carry out quantitative real-time RT-PCR. SYBR Green I is a double-stranded DNA-specific dye, and its fluorescence is greatly enhanced by its binding to double-stranded DNA. During each phase of DNA synthesis, the SYBR Green I dye that is already included in the reaction mix binds to the amplified PCR products and the amplicons can be detected by its fluorescence. To synthesize the first strand of cDNA, reverse transcription was carried out at 61°C for 20 min. The cDNA/RNA hybrid was then denatured at 95°C for 30 s. After that, the value of the quantitative real-time RT-PCR protocol was set as follows: The template DNA was denatured at 95°C for 1 s, followed by annealing at 5°C below the melting temperature of primer pairs for 5 s and elongation at 72°C for 13 s. The amplification process was repeated for 45 cycles.

The volume of the LightCycler Capillaries used in this study was 20 μ l and the capillaries were chilled at -20°C before use. The final solution consisted of 3.25 mM MgCl₂, 0.3 μ M of each forward and reverse primers, 7.5 μ l of 1 \times LightCycler RNA master SYBR Green I and 500 ng of total RNA. The final reaction was brought up to 20 μ l with PCR grade water. As negative control, reverse transcriptase was not included in

the reaction mixture to preclude DNA contamination. The results were recorded and this experiment was repeated another two times.

The 16S rRNA gene located in the *E. coli* genome was used as the endogenous control and its expression was assumed to be constant for all the pQF52 clones (Lee *et al.*, 2008; Hays, 2009). The relative expression ratio of a target gene, *lacZ*, is calculated based on E and CP deviation of unknown sample versus a control, and expressed in comparison to a reference gene, 16S rRNA (Pfaffl, 2001) as follow:

$$\text{Ratio} = \frac{(E_{\text{target}})^{\Delta\text{CP}_{\text{target}}(\text{control-sample})}}{(E_{\text{ref}})^{\Delta\text{CP}_{\text{ref}}(\text{control-sample})}}$$

where:

- E_{target} = real-time RT-PCR efficiency of target gene, *lacZ* transcript
- E_{ref} = real-time RT-PCR efficiency of reference gene, 16S rRNA transcript
- $\Delta\text{CP}_{\text{target}}$ = CP deviation of control minus sample of the target gene transcript
- $\Delta\text{CP}_{\text{ref}}$ = CP deviation of control minus sample of the reference gene transcript

2.9 Determination of the Growth Curves

E. coli TOP10 cells harbouring the pLNBAD-derived recombinant plasmids pLN_yM, pLN_yMyB and pLN_CyMyB were cultured in 10 ml Luria-Bertani broth supplemented with the appropriate antibiotics and grown overnight at 30°C with shaking at 250 rpm. The overnight cultures were inoculated into fresh Luria-Bertani broth supplemented with the appropriate antibiotics with a starting OD₆₀₀ of ~0.02. Two sets of each culture were prepared and 1 mM of L-arabinose was added to one set of the culture at 0 h. The cultures were grown under the same conditions for 4 h. The OD₆₀₀ of each culture was measured at hourly intervals.

2.10 Determination of the DNA Binding Sites of the TA Loci

2.10.1 EMSA

2.10.1.1 The *pezAT* TA locus

DNA fragments containing the putative regulatory protein binding site(s) were amplified by PCR using primer pairs PS-F and PS-R (Table 6). The purified amplified fragments were labeled with biotin using the Biotin 3'-end DNA Labeling kit (Pierce, USA). The reaction mix was as follows: 1 × TdT Reaction Buffer, 5 pmol of sample DNA, 0.5 μM of Biotin-11-UTP, 0.2 U/μl of diluted TdT and ultrapure water to a final volume of 50 μl. The reaction was incubated at 37°C for 30 min before it was stopped by adding 2.5 μl of 0.2 M EDTA. Following that, 50 μl chloroform:isoamyl alcohol was added to the reaction to extract the TdT. The mixture was vortexed briefly followed by centrifugation at maximum speed for 2 min to separate the aqueous from the organic phase. The aqueous phase was saved and subjected to EMSA.

EMSA was carried out using the LightShift Chemiluminescent EMSA kit (Pierce, USA). Biotin end-labeled DNA (20 fmol) was incubated at room temperature for 20 min with 50 ng/μl of poly (dI•dC) and different amounts of purified proteins (PezA or the PezAT protein complex) in binding buffer (10 mM Tris, pH 7.5, 50 mM KCl and 1 mM DTT). The binding reaction was then loaded onto a 5% polyacrylamide gel (1.66 ml of acrylamide:bis-acrylamide (30:0.8), 1 ml of 5 × TBE, 7.06 ml of sterile distilled water, 5 μl of TEMED and 260 μl of 10% ammonium persulphate) and electrophoresis was conducted in 0.5 × TBE buffer on ice. Nucleoprotein complexes were separated from free DNA by applying 200 V until the bromophenol blue dye front had migrated approximately 2/3 the length of the gel and then transferred onto a Biorad® B Nylon Membrane (Pierce, USA) at 380 mA for 30 min. After that, the membrane was cross-linked by exposure to Ultraviolet light at 254 nm for 10 min.

The biotin end-labeled DNA was detected by streptavidin-horseradish peroxidase conjugate and chemiluminescent substrate supplied in the Chemiluminescent Nucleic Acid Detection Module (Pierce, USA). The membrane was blocked by adding 20 ml of blocking buffer (pre-warmed at 37 - 50°C) and incubated for 15 min with gentle shaking. The blocking buffer was decanted and replaced with the conjugate/blocking solution (66.7 µl of Stabilized Streptavidin-Horseradish Peroxidase Conjugate in 20 ml of blocking buffer) and incubated for another 15 min with gentle shaking. The membrane was then transferred to a new container and rinsed briefly with 20 ml of 1 × wash solution. The membrane was washed four times for 5 min each in 20 ml of 1 × wash solution with gentle shaking. The membrane was again transferred to a new container and 30 ml of Substrate Equilibrium Buffer was added and incubated for 5 min with gentle shaking. The Substrate Equilibrium Buffer was then removed and the edge of the membrane was blotted with a paper towel to remove the excess buffer. The membrane was then placed onto a clean sheet of plastic wrap on a flat surface. After that, 12 ml of Substrate Working Solution (Luminol/Enhancer Solution:Stable Peroxide Solution in a ratio of 1:1) was poured directly onto the membrane so that the surface of the membrane was completely covered and incubated for 5 min without shaking. The Substrate Working Solution was removed and the edge of the membrane was blotted with a paper towel to remove the excess buffer. The resulting chemiluminescence was captured using a ChemiImager™ System (Alpha Innotech, USA) and the exposure time was adjusted to obtain the desired result (~30 min).

2.10.1.2 The *yefM-yoeB_{Spn}* TA locus

Oligonucleotides PS1-F, PS2-F and PS2-R were first labeled with [γ -³²P]ATP with the following reaction mix: 20 pmol of oligonucleotides, 5 μ l of 10 \times Kinase Buffer, 12.5 μ l of 3000 Ci/mmol, 10 mCi/ml [γ -³²P]ATP, 1.5 μ l of 20 U/ μ l T4 polynucleotide kinase and sterile distilled water was added to a final volume of 50 μ l. The reaction mix was incubated in 37°C for 30 min and then heated at 65°C for 20 min to inactivate the enzyme. The non-incorporated nucleotides were eliminated using Illustra MicroSpin™ G-25 Columns (General Electric Healthcare, UK). The labeled oligonucleotides were then precipitated with 3 M NaAc, pH 4.8-5.2 and three volume of absolute ethanol in -20°C for 30 min. The oligonucleotide pellet was obtained by centrifugation at 13,000 \times g for 10 min, washed with 70% ethanol, air-dried and resuspended in 42 μ l of sterile distilled water.

DNA fragments containing both palindrome sequence 1 and palindrome sequence 2 were amplified using each of the labeled oligonucleotides and non-labeled oligonucleotides: Labeled PS1-F and non-labeled PS2-R as well as non-labeled PS1-F and labeled PS2-R, whereas DNA fragment containing PS2 was amplified using labeled PS2-F and non-labeled PS2-R as well as non-labeled PS2-F and labeled PS2-R. PCR amplification was carried out in a final volume of 50 μ l with the following reaction mix: 20 pmol of each labeled and non-labeled oligonucleotides (forward or reverse), 10 μ l of 5 \times HF buffer, 1 μ l of 10 mM dNTP mix, 1 ng DNA template (pQF_P1P2yMyB) and 0.5 μ l of Phusion™ DNA polymerase (Finnzymes, Finland) added lastly. The thermal cycling conditions were set as follows: Initial denaturation at 98°C for 30 s, denaturation at 98°C for 10 s, annealing of the primers to the DNA template at 5°C lower than the melting temperature of primer-template DNA duplex for 15 s, followed by an extension time of 15 s/kb at 72°C. This set of conditions (except for initial

denaturation of the DNA template) was repeated for 30 cycles, followed by a final extension at 72°C for 10 min.

A 5% polyacrylamide gel was prepared and pre-run for 30 min at 100 V prior to loading of the PCR products with one strand labeled with [γ -³²P]ATP. The entire labeled PCR-amplified DNA fragments were mixed with 1 × loading dye and then loaded onto the gel and electrophoresis was carried out at 100 V until the bromophenol blue dye front migrated 2/3 the length of the gel. The gel was then transferred to a film cassette and exposed to an X-ray film for 2 min. The film was developed and the band on the gel was excised corresponding to the band on the film. The gel containing the labeled DNA fragments was cut into smaller pieces and dissolved in elution buffer (0.2 M NaCl, 0.02 M Tris, pH 7.6 and 0.002 M EDTA, pH 8.0) overnight on a shaking incubator at 42°C. The polyacrylamide debris was removed using Costar® Spin-X® Centrifuge Tube Filters (Corning, USA) by centrifugation at 13,000 × *g* for 1 min. The labeled DNA fragments were then ethanol precipitated and dissolved in 20 μl of sterile distilled water.

EMSA was carried out in the following mix to a final volume of 5 μl: 1 × binding buffer (100 mM Tris, pH 7.6, 5 mM EDTA, 5 mM DTT and 5% glycerol), 3000 cpm of labeled DNA, 10 ng/μl of heparin and increasing concentration of proteins (YefM_{S_{pn}} or YefM-YoeB_{S_{pn}}). The reaction mix was incubated at room temperature for 20 min before mixing with 0.5 μl of 5 × loading dye and loaded into a 5% polyacrylamide gel. The gel was run with 0.5 × TBE buffer at 100 V on ice until the bromophenol blue dye front reached 2/3 down the length of the gel. The gel was then transferred to a film cassette and exposed to an X-ray film overnight at -70°C. The film was then developed according to the manufacturer's instructions.

2.10.2 DNaseI footprinting assay

2.10.2.1 DNaseI titration

DNaseI titration was carried out to determine the optimum amount of DNaseI to be used for DNaseI footprinting. The DNaseI was diluted to 1/100, 1/250 and 1/500 using DNaseI dilution buffer (25 mM Tris-Cl pH 7.6 and 50% glycerol). The reaction mix contained 1 × binding buffer (400 mM Tris-Cl pH 8, 60 mM MgCl₂, 10 mM CaCl₂ and 5% glycerol), 30,000 cpm labeled DNA, and sterile distilled water to a final volume of 50 µl. The reaction was initiated by adding 2 µl of the DNaseI dilutions into each reaction and incubated at room temperature for 5 min. The reaction was stopped by adding 25 µl of stop buffer (2 M ammonium acetate, 0.12 M EDTA, 0.8 M NaAc, pH 7.0 and 400 µl/ml tRNA) and 187 µl absolute ethanol, followed by freezing at -70°C for 30 min. Samples were then centrifuged at 13,000 × *g* at 4°C for 35 min, and the resulting pellet was then washed with 400 µl of 70% chilled ethanol. The pellet was air dried for 10 min and then dissolved in 3 µl of loading buffer (80% deionized formamide, 10 mM NaOH, 0.1% bromophenol blue, 0.1% xylene cyanol and 1 mM of EDTA, pH 8). The samples were heated at 80°C for 3 min and placed on ice immediately before loading onto an 8% urea polyacrylamide gel (7 M) which was preheated to 50°C and run at 45 W in 1 × TBE buffer at 50°C. The electrophoresis was stopped when the xylene cyanol dye had migrated 3/4 the length of the gel. The gel was then transferred to a blotting paper, covered with plastic wrap and vacuum dried for 45 min at 80°C. The gel was then placed in a film cassette and exposed to X-ray film at -70°C until the desired image was obtained. The X-ray film was then developed according to the manufacturer's instructions. The DNaseI dilution with the optimal result was used for the subsequent DNaseI footprinting assay.

2.10.2.2 Preparation of 8% urea polyacrylamide gel (7 M)

An 8% urea polyacrylamide gel (7 M) was prepared by dissolving 25.3 g of urea powder together with 12 ml of acrylamide:bis-acrylamide (38:2) and 12 ml of $5 \times$ TBE buffer at 50°C. When all the urea powder was completely dissolved, sterile distilled water was added to a final volume of 60 ml. The gel mixture was cooled on ice for a few seconds, followed by the addition of 20 μ l of TEMED and 360 μ l of 10% ammonium persulphate. The gel mix was stirred gently and then loaded using a syringe into the vertical gel caster followed by inserting the Teflon comb. When the gel had polymerized, the comb was removed and the wells were rinsed thoroughly with $1 \times$ TBE buffer before use.

2.10.2.3 DNaseI footprinting assay

DNaseI footprinting assays were carried out using the following reaction mix: $1 \times$ binding buffer, 30,000 cpm labeled DNA, 10 ng/ μ l of heparin, increasing concentrations of protein and sterile distilled water to a final volume of 50 μ l. The reaction mix was incubated at room temperature for 20 min before adding 2 μ l of 1/500 DNaseI and incubated at room temperature for 5 min. The following protocols were as described in DNaseI titration (Section 2.10.2.1). The sequencing ladder prepared using a dideoxy sequencing reaction was run along with the reaction samples.

2.10.2.4 Dideoxy sequencing reaction

A dideoxy sequencing DNA ladder was prepared using Sequenase Version 2.0 DNA Polymerase (United States Biochemical, USA) and pQF_P1P2yMyB was used as the DNA template. To denature the plasmid DNA, 2 μ l of 2 M NaOH was added to 8 μ l of DNA plasmid (2 μ g) and incubated at room temperature for 10 min. The denatured plasmid DNA was ethanol precipitated and resuspended in 10 μ l of sterile distilled

water. Annealing of the labeled oligonucleotides to the denatured DNA templates required the following reaction mix: 5 μ l of denatured plasmid DNA (1 μ g), 2 μ l of 5 \times Sequenase Reaction Buffer, 1.5 μ l of [γ - 32 P]ATP labeled PS2-F or PS2-R (1.5 pmol) and 1.5 μ l of sterile distilled water. The reaction mix was incubated at 65°C for 5 min, followed by cooling down slowly to 35°C over a period of 30 min and then chilled on ice.

The labeling reaction contained the following mixture in a final volume of 15.5 μ l: 10 μ l of annealing reaction mixture, 1 μ l of 0.1 M DTT, 1 \times labeling buffer, 1.5 pmol [γ - 32 P]ATP and 3.25 U of Sequenase DNA polymerase. The reaction mix was incubated at room temperature for 5 min. After that, 3.5 μ l of the reaction mixture was then added to each of the tubes containing 2.5 μ l of ddGTP, ddATP, ddTTP and ddCTP termination mix respectively, which were pre-warmed at 37°C, and incubation continued at 37 °C for a further 5 min after which 4 μ l of Stop Solution was added to each tube to terminate the reaction. Then, 4 μ l of each reaction was heated at 80°C for 2 min and loaded along with the reaction samples resulting from the DNaseI footprinting assay (Section 2.10.2.3) onto an 8% urea polyacrylamide gel (7 M) (Section 2.10.2.4).

2.10.3 Hydroxyl radical footprinting assay

2.10.3.1 Hydroxyl radical footprinting assay

The hydroxyl radical footprinting reagent was prepared by mixing equal volumes of 0.6% H₂O₂, 20 mM sodium ascorbate and Fe²⁺-EDTA (equal volumes of 0.4 mM Fe²⁺ in sterile distilled water and 0.8 mM of EDTA). First, 30,000 cpm of labeled DNA (Section 2.10.1.2) was incubated with increasing amount of proteins (YefM_{S_{pn}} or YefM-YoeB_{S_{pn}}) in 1 \times binding buffer, 10 ng/ μ l of heparin, and sterile distilled water, which was added to a final volume of 50 μ l. Then, the reaction mix was incubated at room temperature for 20 min to allow the binding of the proteins to the DNA fragments. The reaction was started by adding 9 μ l of the hydroxyl radical

footprinting reagent and incubated at room temperature for 2 min. The reaction was stopped by adding 14.7 μl of stop solution (0.041 M thiourea, 1.5 M NaAc, pH 6 and 0.68 mg/ml tRNA) and 187 μl of absolute ethanol followed by freezing at -70°C for 30 min. The mixture was then centrifuged at $13,000 \times g$ for 35 min at 4°C and the pellet was washed with chilled 70% ethanol. The pellet was then air dried and resuspended in 18 μl of sample loading dye (80% formamide, 10 mM NaOH, 0.1% bromophenol blue, 0.1% xylene cyanol and 1 mM EDTA). After that, 3 μl of the samples were preheated at 80°C for 3 min and placed on ice immediately before loading onto an 8% urea polyacrylamide gel (7 M) (pre-heated to 50°C). The sequencing ladder prepared using Maxam and Gilbert Reaction for (G+A) (Section 2.10.3.2) was run along with the reaction samples.

2.10.3.2 Preparation of sequencing ladder by the Maxam-Gilbert DNA sequencing method for (G+A)

A 2 μl aliquot of labeled DNA (30,000 cpm) was mixed with 1 μl of 1 mg/ml tRNA and 1 μl of 0.2 M formic acid and incubated at 65°C for 30 s. The reaction was stopped by adding 30 μl of 1.5 M NaAc, pH 7 and 150 μl of absolute ethanol, followed by freezing at -70°C for 10 min. The mixture was centrifuged at 4°C for 10 min at $12,000 \times g$ and the pellet was resuspended with 30 μl of sterile distilled water and 150 μl of absolute ethanol. The mixture was again centrifuged at 4°C for 10 min at $12,000 \times g$ and then washed with 80% ethanol. The pellet was dissolved in 50 μl of sterile distilled water and 5 μl of 10 M Piperidine. The mixture was heated at 100°C for 10 min, followed by extraction with 1 ml of 1-butanol. The mixture was again centrifuged at room temperature for 10 min and the supernatant was removed. The pellet was resuspended in 50 μl of 1% SDS and 500 μl of 1-butanol prior to centrifugation at room temperature for 10 min. The pellet was washed twice with 80% ethanol and air dried.

The pellet was then resuspended in 4 μ l of loading buffer as in DNaseI titration (Section 2.10.2.1), and heated at 80°C for 2 min before loading along with the reaction samples resulting from the hydroxyl radical footprinting assay (Section 2.10.3.1) onto an 8% urea polyacrylamide gel (7 M).

3 Results

3.1 The *pezAT* TA Locus of *S. pneumoniae*

3.1.1 Sequence analysis

The discovery of the *pezAT* TA locus in the chromosome of *S. pneumoniae* was due to the sequence similarity of the *pezT*-encoded protein with the ζ toxin of the ϵ - ζ TA locus encoded on plasmid pSM19035 of *S. pyogenes* (Meinhart *et al.*, 2003). The *S. pneumoniae* TIGR4 (NCBI accession number: NC_003028) (Tettelin *et al.*, 2001) SP_1051 protein (which was renamed as PezT; Khoo *et al.*, 2007) shares 42% amino acid sequence identity with the ζ protein. The sequence comparison analysis conducted in this project was an extension to the comparison study done by Khoo *et al.* (2007). The predicted PezT protein and ζ protein showed sequence conservation (Figure 8) particularly in the sequence at the phosphotransferase active site that was determined from the crystal structure of the ϵ - ζ TA complex (Meinhart *et al.*, 2003). The ζ phosphotransferase active site is characterized by a Walker A or P-loop motif (GXXGXGK[T/S]) that is typical of ATP/GTP-binding proteins (Lutkenhaus and Sundaramoorthy, 2003).

The C-terminal portion (residues 70 to 158) of the SP_1050 protein (designated PezA; Khoo *et al.*, 2007) that was encoded by the reading frame immediately upstream of *sp1051/pezT* shares only 21% sequence identity with the ϵ antitoxin (Figure 9) whereas helix-turn-helix motifs were found within the N-terminal portion of SP_1050 by InterProScan. The helix-turn-helix motif was not found in ϵ which functions solely as an antitoxin whereas another protein, ω , functions as the transcriptional repressor for the ϵ - ζ locus in pSM19035 (Meinhart *et al.*, 2003). No homologue of ω could be located within the *S. pneumoniae* genome. Therefore it is likely that PezA also functions as a transcriptional regulator like other TA loci but unlike ϵ - ζ where ϵ functions solely as the antitoxin with the regulatory role played by a third component, ω .

S.pneumoniaeTIGR4_PezT	-MEIQDYTDSEFKHALARNLRSRLTRGKKSSKQPIAILLGGQSGAGKTTI 49
S.pyogenespSM19035_zeta	MANIVNFTDKQFENRLNDNLEELIQGKKAVERPTAFLLGGQPGSGKTSIR 50
	:* :***:***: * **..* :***: :* * :***: .*:***:..
S.pneumoniaeTIGR4_PezT	RIKQKEFQGNVIVIDGDSFRSQHPHYLELQQEYKDSVEYTKDFAGKME 99
S.pyogenespSM19035_zeta	SAIFEETQGNVIVIDNDTFKQHPNFDELVKLYEKDVVKHVTTPYSNRMTE 100
	:* ***:***:***:***: * * : * * * :.. :..:..*
S.pneumoniaeTIGR4_PezT	SLVTKLSSLRYNLLIEGTLRTVDVPKKAQLLKNKGYEVQLALIAATKPEL 149
S.pyogenespSM19035_zeta	AIISRLSDQGYNLVIEGTGRITDVPITATMLQAKGYETKMYVMAVPKIN 150
	::::**.. ***:*** **..** :** :* : ***:.. :..*
S.pneumoniaeTIGR4_PezT	SYLSTLIRYEELYIINPNQARATPKHHDFIVNHLVDNTRKLEELAI 199
S.pyogenespSM19035_zeta	SYLGTIEREYETMYADDPMTARATPKQAHDIVVKNLPTNLETLHKTGLFSD 200
	.*: *** :* :* **: ***:***: * ..* : ..*
S.pneumoniaeTIGR4_PezT	IQIYQRDRSCVYDSKEN-TTSAADVLELLFGEWS-----QVEKEM 239
S.pyogenespSM19035_zeta	IRLYNREGVKLYSSLETPSISPKETLEKELNRKVSQKEIQPTLERIEQKM 250
	*:***: :*. * * : * :. :. :. : * : * :. :. :. *
S.pneumoniaeTIGR4_PezT	LQVGE-----KRLNELLEK----- 253
S.pyogenespSM19035_zeta	VLNKHQETPEFKAIQQKLESLOPPTPIPKTPKPLGI 287
	: . * :. :. *

Figure 8: Pairwise alignment of amino acid sequences of the *S. pyogenes* plasmid pSM19035 ζ toxin and PezT toxin encoded in the *S. pneumoniae* chromosome. The crystal structure of ζ revealed an active phosphotransferase site made up of residues GXXGXGKT (Meinhart *et al.*, 2003) which are highly conserved in PezT and indicated in a box. “*” indicates identical residues; “:” indicates conserved substitution; and “.” indicates semi-conserved substitution.

S.pneumoniaeTIGR4_PezA	MIGKNIKSLRKTDLTQLEFARIVGISRNSLSRYENGTSSVSTELIDIIC 50
S.pyogenespSM19035_epsilon	-----
S.pneumoniaeTIGR4_PezA	QKFNVSIVDIVGEDKMLNPVEDYELTLKIEIVKGERGANLLSRLRYQDSQ 100
S.pyogenespSM19035_epsilon	-----MAVTYEKTFEIEIINELASVYNRVNLVNLNH 32
	** * :***:*** :* . : . * : . *
S.pneumoniaeTIGR4_PezA	GISIDDESNPWILMSDDLSDLIHTNIYLVETFDEIERYSGYLDGIER-ML 149
S.pyogenespSM19035_epsilon	ELNKNSQLLEVNLLNQLKAKRVNLFDIS-LEELQAVHEYWRSMNRYSK 81
	. : . * . : : : . * . : . * : . : . * : . * : . *
S.pneumoniaeTIGR4_PezA	EISEKRMVA 158
S.pyogenespSM19035_epsilon	QVLNKEKVA 90
	: : . * . *

Figure 9: Pairwise amino acid sequence alignment of the *S. pyogenes* plasmid pSM19035 ϵ antitoxin and PezA found in the chromosome of *S. pneumoniae*. Only the C-terminal region of PezT shared 21% sequence identity with ϵ . “*” indicates identical residues; “:” indicates conserved substitution; and “.” indicates semi-conserved substitution.

Besides the *S. pneumoniae* TIGR4 strain, one copy of *pezAT* was also identified in the chromosomes of other annotated *S. pneumoniae* strains except for *S. pneumoniae* CGSP14 (NCBI accession number: NC_010582), where two copies of the *pezAT* locus were identified (Table 7). Multiple sequence alignment showed slight variation within both the PezA (Figure 11) and PezT (Figure 10) sequences with strict conservation of

residues within the phosphotransferase active site of PezT. Variations were mostly found in *S. pneumoniae* CGSP14 and *S. pneumoniae* Hungary 19A-6 (NCBI accession number: NC_010380), which suggests that the *pezAT* locus in these two *S. pneumoniae* strains are more divergent when compared to the other strains in the NCBI databases. The start codons for the various *pezA* loci were found to be differently annotated for several *S. pneumoniae* strains in the NCBI database leading to variations in the lengths of the PezA proteins (Figure 11). However, from experiments conducted by Khoo *et al.* (2007) and in this study, it is proposed that the *pezA* start codons in all *S. pneumoniae* strains that harbour the *pezAT* locus should be the same as that of *S. pneumoniae* TIGR4 (Figure 12).

Table 7: The PezAT homologues in the chromosomes of annotated *S. pneumoniae* strains.

<i>S. pneumoniae</i> strains	Accession numbers	ε homologues		ζ homologues	
		Locus tags	GI numbers	Locus tags	GI numbers
TIGR4	NC_003028	SP_1050	15900920	SP_1051	15900921
R6	NC_003098	Spr_0951	15902995	Spr_0952	15902996
D39	NC_008533	SPD_0930	116515960	SPD_0931	116515566
Hungary19A-6	NC_010380	SPH_1262	169833796	SPH_1263	169833311
CGSP14	NC_010582	SPCG_1029	182683999	SPCG_1030	182684000
CGSP14	NC_010582	SPCG_1296	182684266	SPCG_1295	182684265


```

S.pneumoniaeCGSP14_SPCG1029      MVHYAFLCYNGIIKIKKEFAMIGKNIKSLRKTHDLTQHEFARIVGISRNSL 50
S.pneumoniaeHungary19A-6_SPH1262  -----
S.pneumoniaeR6_spr0951          MVRYAFLCYNGFIKIKKEFAMIGKNIKSLRKTHDLTQPEFARIIGISRNSL 50
S.pneumoniaeD39_SPD0930        -----MIGKNIKSLRKTHDLTQPEFARIIGISRNSL 31
S.pneumoniaeTIGR4_SP1050       -----MIGKNIKSLRKTHDLTQLEFARIVGISRNSL 31
S.pneumoniaeCGSP14_SPCG1296    -----MIGKNIKSLRKTHDLTQLEFARIVGISRNSL 31

S.pneumoniaeCGSP14_SPCG1029      SRYENGTSSVSTELIDIICQKFNVSIVDIVGENKMLNPVEDYELTLKIEI 100
S.pneumoniaeHungary19A-6_SPH1262 -----MLNPVEDYELTLKIEI 16
S.pneumoniaeR6_spr0951          SRYENGTSSVSTELIDIICQKFNVSIVDIVGEDKMLNPVEDYELTLKIEI 100
S.pneumoniaeD39_SPD0930        SRYENGTSSVSTELIDIICQKFNVSIVDIVGEDKMLNPVEDYELTLKIEI 81
S.pneumoniaeTIGR4_SP1050       SRYENGTSSVSTELIDIICQKFNVSIVDIVGEDKMLNPVEDYELTLKIEI 81
S.pneumoniaeCGSP14_SPCG1296    SRYENGTSSVSTELIDIICQKFNVSIVDIVGEDKMLNPVEDYELTLKIEI 81
                                  *****

S.pneumoniaeCGSP14_SPCG1029      VKERGANLLSRLRYQDSQGISIDDESNPWILMSDDLSDLIHTNIYLVET 150
S.pneumoniaeHungary19A-6_SPH1262 VKERGANLLSRLRYQDSQGISIDDESNPWILMSDDLSDLIHTNIYLVET 66
S.pneumoniaeR6_spr0951          VKERGANLLSRLRYQDSQGISIDDESNPWILMSDDLSDLIHTNIYLVET 150
S.pneumoniaeD39_SPD0930        VKERGANLLSRLRYQDSQGISIDDESNPWILMSDDLSDLIHTNIYLVET 131
S.pneumoniaeTIGR4_SP1050       VKERGANLLSRLRYQDSQGISIDDESNPWILMSDDLSDLIHTNIYLVET 131
S.pneumoniaeCGSP14_SPCG1296    VKERGANLLSRLRYQDSQGISIDDESNPWILMSDDLSDLIHTNIYLVET 131
                                  *****

S.pneumoniaeCGSP14_SPCG1029      FDEIERYSYGLDGIERMLEISEKRMVA 177
S.pneumoniaeHungary19A-6_SPH1262 FDEIERYSYGLDGIERMLEISEKRMVA 93
S.pneumoniaeR6_spr0951          FDEIERYSYGLDGIERMLEISEKRMVA 177
S.pneumoniaeD39_SPD0930        FDEIERYSYGLDGIERMLEISEKRMVA 158
S.pneumoniaeTIGR4_SP1050       FDEIERYSYGLDGIERMLEISEKRMVA 158
S.pneumoniaeCGSP14_SPCG1296    FDEIERYSYGLDGIERMLEIFEKRMVA 158
                                  *****

```

Figure 11: Multiple amino acid sequence alignment of the PezA homologues found in annotated *S. pneumoniae* strains. “*” indicates identical residues; “.” indicates conserved substitution; and “:” indicates semi-conserved substitution. Differences in the length of the PezA homologues are due to differently annotated start sites for the various *pezA* loci.

S. pneumoniaeCGSP14_SPCG1029	MIGKNIKSLRKTHTDLTQHEFARIVGISRNSLSRYENGTSSVSTELIDIIC	50
S. pneumoniaeHungary19A-6_SPH1262	MIGKNIKSLRKTHTDLTQDDFARIVGISRNSLSRYENGTSSVSTELIDIIC	50
S. pneumoniaeR6_spr0951	MIGKNIKSLRKTHTDLTQPEFARIIGISRNSLSRYENGTSSVSTELIDIIC	50
S. pneumoniaeD39_SPD0930	MIGKNIKSLRKTHTDLTQPEFARIIGISRNSLSRYENGTSSVSTELIDIIC	50
S. pneumoniaeTIGR4_SP1050	MIGKNIKSLRKTHTDLTQLEFARIVGISRNSLSRYENGTSSVSTELIDIIC	50
S. pneumoniaeCGSP14_SPCG1296	MIGKNIKSLRKTHTDLTQLEFARIVGISRNSLSRYENGTSSVSTELIDIIC	50
	***** :****:*****	
S. pneumoniaeCGSP14_SPCG1029	QKFNVSYSVDIVGENKMLNPVEDYELTLKIEIVKGERGANLLSRLRYQDSQ	100
S. pneumoniaeHungary19A-6_SPH1262	QKFNVSYSVDIVGEDKMLNPVEDYELTLKIEIVKGERGANLLSRLRYQDSQ	100
S. pneumoniaeR6_spr0951	QKFNVSYSVDIVGEDKMLNPVEDYELTLKIEIVKGERGANLLSRLRYQDSQ	100
S. pneumoniaeD39_SPD0930	QKFNVSYSVDIVGEDKMLNPVEDYELTLKIEIVKGERGANLLSRLRYQDSQ	100
S. pneumoniaeTIGR4_SP1050	QKFNVSYSVDIVGEDKMLNPVEDYELTLKIEIVKGERGANLLSRLRYQDSQ	100
S. pneumoniaeCGSP14_SPCG1296	QKFNVSYSVDIVGEDKMLNPVEDYELTLKIEIVKGERGANLLSRLRYQDSQ	100
	***** :*****	
S. pneumoniaeCGSP14_SPCG1029	GISIDDESNPWILMSDDLSDLIHTNIYLVETTFDEIERYSGYLDGIERMLE	150
S. pneumoniaeHungary19A-6_SPH1262	GISIDDESNPWILMSDDLSDLIHTNIYLVETTFDEIERYSGYLDGIERMLE	150
S. pneumoniaeR6_spr0951	GISIDDESNPWILMSDDLSDLIHTNIYLVETTFDEIERYSGYLDGIERMLE	150
S. pneumoniaeD39_SPD0930	GISIDDESNPWILMSDDLSDLIHTNIYLVETTFDEIERYSGYLDGIERMLE	150
S. pneumoniaeTIGR4_SP1050	GISIDDESNPWILMSDDLSDLIHTNIYLVETTFDEIERYSGYLDGIERMLE	150
S. pneumoniaeCGSP14_SPCG1296	GISIDDESNPWILMSDDLSDLIHTNIYLVETTFDEIERYSGYLDGIERMLE	150

S. pneumoniaeCGSP14_SPCG1029	ISEKRMVA	158
S. pneumoniaeHungary19A-6_SPH1262	ISEKRMVA	158
S. pneumoniaeR6_spr0951	ISEKRMVA	158
S. pneumoniaeD39_SPD0930	ISEKRMVA	158
S. pneumoniaeTIGR4_SP1050	ISEKRMVA	158
S. pneumoniaeCGSP14_SPCG1296	IFEKRMVA	158
	* **** *	

Figure 12: Multiple amino acid sequence alignment of the PezA homologues found in *S. pneumoniae* strains following re-designation of the *pezA* start codon as in *S. pneumoniae* TIGR4 *pezA*. “*” indicates identical residues; “:” indicates conserved substitution; and “.” indicates semi-conserved substitution.

No homologues of the ω regulator were found encoded in the chromosomes of these *S. pneumoniae* strains except for the chromosome of *S. pneumoniae* CGSP14, where two ω homologues were discovered using a BLAST search. Both homologues displayed high sequence identity with the pSM19035-encoded ω (80% identity for SPCG_1326; GI number: 182684296 and 78% identity for SPCG_1322; GI number: 182684292) (Figure 13) but intriguingly, their gene sequences were not located upstream of, or anywhere near a *pezAT* locus. Likewise, no ϵ - ζ -like locus could be found in the vicinity of these ω -homologue-encoded reading frames. Whether the ω homologues encoded in the CGSP14 chromosome play any role in the regulation of the *pezAT* locus has yet to be determined.

Analysis of nucleotide sequences immediately upstream of *pezA* from the *S. pneumoniae* TIGR genome showed the presence of a canonical σ^{70} promoter sequence, designated P_{pezA} : the putative -10 region (5'-TATAAT-3') is located 33 nucleotides

upstream of the *pezA* start codon and is 18 nucleotides apart from the 5' end of the putative imperfect -35 region (5'-TTGgtg-3') (the lowercase letters indicate the nucleotides which are different from the σ^{70} -type promoter consensus sequence). Besides, a putative ribosomal binding site (5'-AGGAG-3') was also identified six nucleotides upstream of the *pezA* start codon. Moreover, a 56 bp palindrome sequence which overlaps with the -35 and -10 promoter regions was identified, and is centered 58 bp upstream of the *pezA* start codon (Figure 14). The upstream regions of the *pezAT* locus in other *S. pneumoniae* strains were found to be similar especially the putative ribosomal binding site, the -10 region and the putative imperfect -35 region (Figure 15).

The *pezA* and *pezT* reading frames overlap by one nucleotide (Figure 16) and thus would appear to constitute a bicistronic operon, as in the case of other TA loci (Hayes, 2003; Gerdes *et al.*, 2005). Although *pezAT* genes of *S. pneumoniae* Hungary 19A-6 have a different stop codon (TAG) for *pezA* and start codon (GTG) for *pezT*, both genes also overlap by one nucleotide 'G' (Figure 16). Analysis suggests that the *pezA* reading frame encodes a 158 amino acid protein with an estimated molecular weight of 18 kDa and a theoretical pI of 4.68 whereas *pezT* encodes a 253 amino acid protein with an estimated molecular weight of 29 kDa and a theoretical pI of 7.01 (ProtParam).

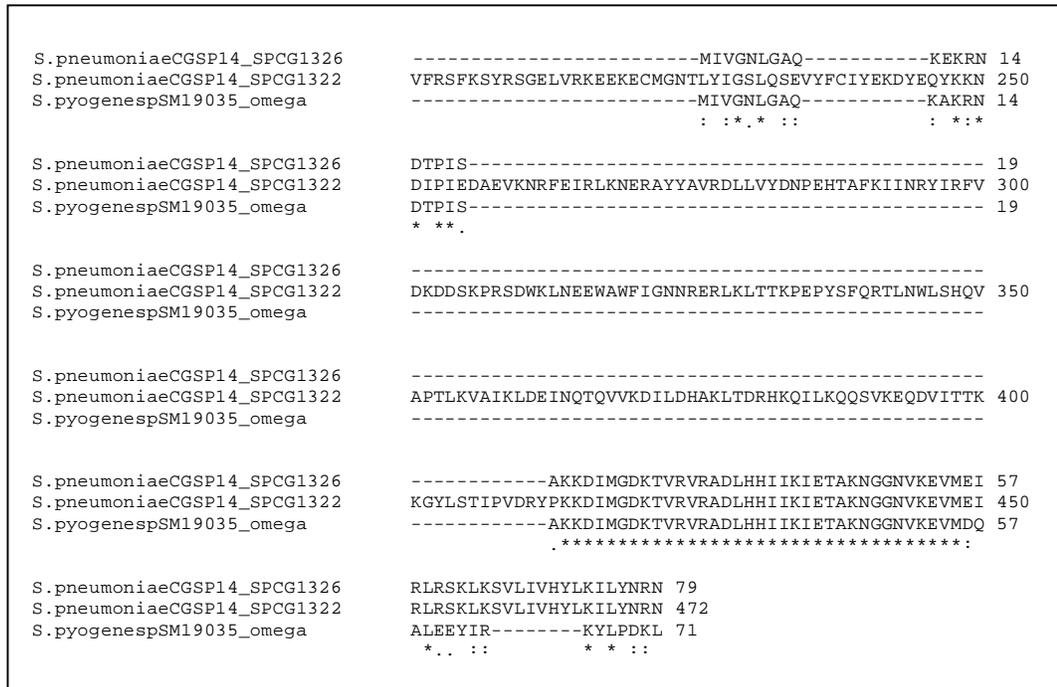


Figure 13: Multiple sequence alignment of the *S. pyogenes* plasmid pSM19035-encoded ω regulator with homologues from the chromosome of *S. pneumoniae* CGSP14. “*” indicates identical residues; “:” indicates conserved substitution; and “.” indicates semi-conserved substitution.

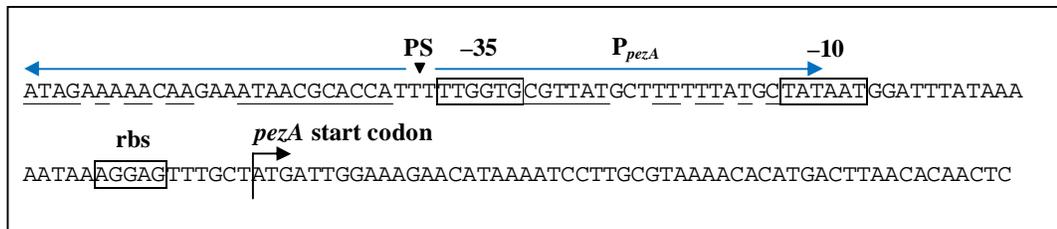


Figure 14: Nucleotide sequence of the upstream region of the *pezA* reading frame in the chromosome of *S. pneumoniae* TIGR4. A σ^{70} -type promoter sequence was located and designated P_{pezA} . The -10 putative promoter sequence (5'-TATAAT-3') and the -35 putative imperfect promoter sequence (5'-TTGgtg-3') as well as the putative ribosome binding site (rbs) are depicted within boxes. The divergent arrows (in blue) indicate a 56 nucleotide imperfect palindromic sequence (PS) which overlaps the -35 and -10 regions and is centered 58 nucleotides upstream of the *pezA* start codon. The inverted triangle (\blacktriangledown) represents the centre of the palindromic sequence and the nucleotides that are underlined are complementary to each arm of the palindromic sequence.

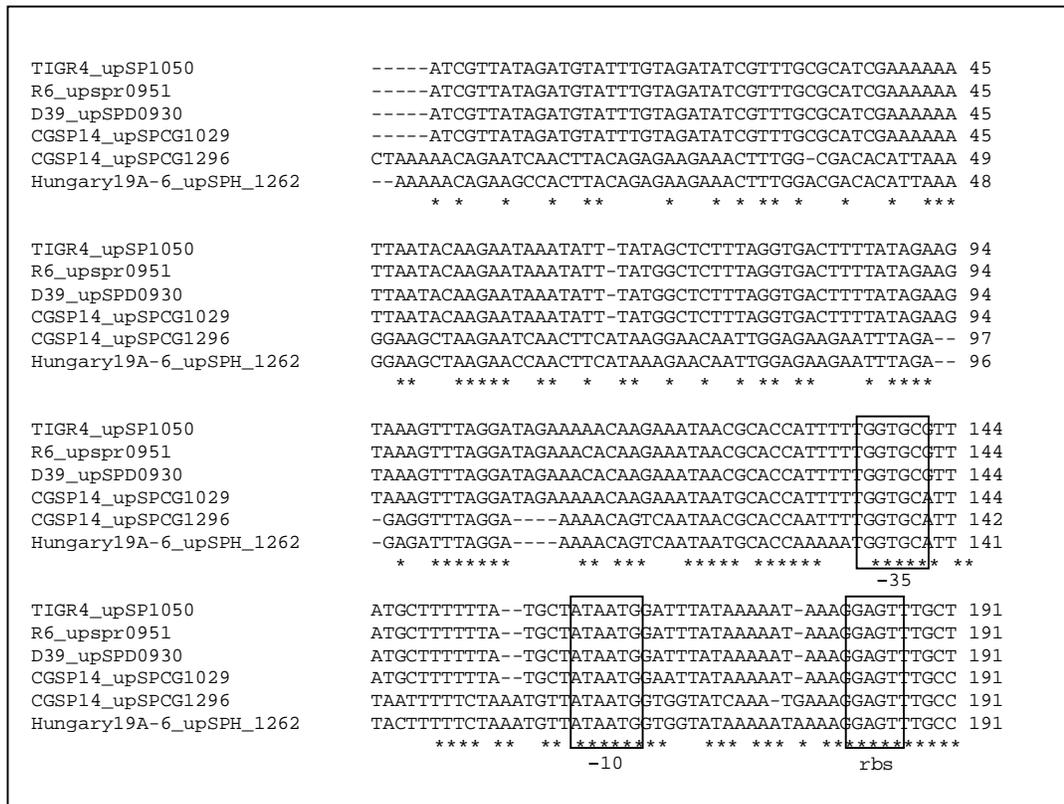


Figure 15: Multiple sequence alignment of nucleotide sequences upstream of the *pezA* reading frame in *S. pneumoniae* strains. The -10 putative promoter sequence and the -35 putative imperfect promoter sequence as well as the putative ribosome binding site (rbs) are depicted within boxes. “*” indicates identical residues; “.” indicates conserved substitution; and “.” indicates semi-conserved substitution.

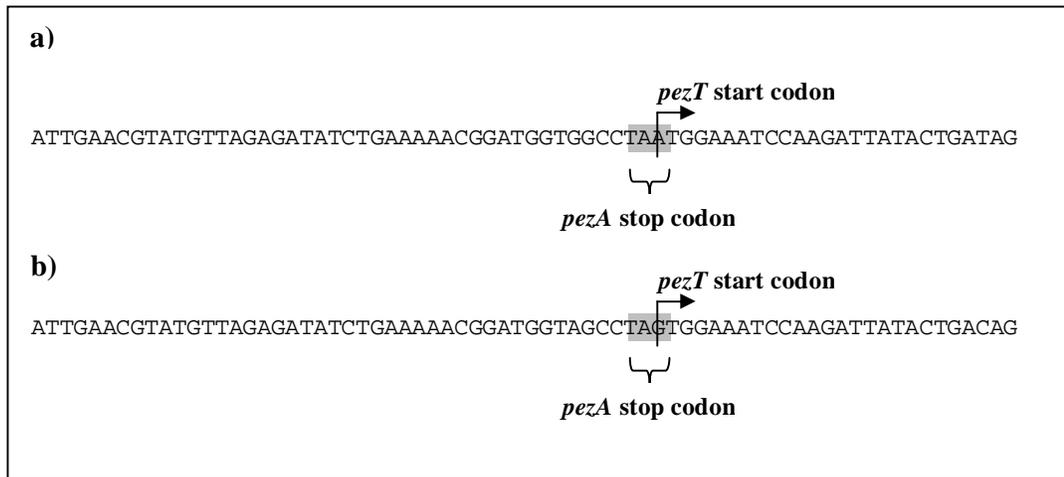


Figure 16: The *pezA* and *pezT* reading frames are overlapped. (a) Nucleotide sequence showing the *pezA* stop codon (TAA) overlapping by one nucleotide ‘A’ with the *pezT* start codon (ATG) in the *pezAT* TA locus in the chromosome of *S. pneumoniae* TIGR4. The arrangement of genes is identical in the other annotated chromosomes of *S. pneumoniae*. (b) Nucleotide sequence showing the *pezA* stop codon (TAG) overlapping by one nucleotide ‘G’ with the *pezT* start codon (GTG) in the *pezAT* TA locus in the chromosome of *S. pneumoniae* Hungary 19A-6.

3.1.2 Construction of recombinant plasmids

A 1,441 bp DNA fragment encompassing the *pezA* and *pezT* reading frames along with 191 bp of sequences immediately upstream of *pezA* was amplified from the chromosomes of several clinical isolates of *S. pneumoniae* and cloned into pGEM-T Easy vector. The resulting recombinant pGEM-T plasmids recovered from *E. coli* DH5 α transformants were sequenced and when compared to the NCBI database sequence (for *S. pneumoniae* TIGR4), a number of variations were found for all the isolates. The recombinant plasmid that contained the least number of variations when compared with the TIGR4 database sequence was chosen for subsequent experiments. The plasmid, designated pGEMT_PpezApezT, contained 31 nucleotide differences when compared with the TIGR4 database sequence: one variation ('A' \rightarrow 'G') was located 123 nucleotides upstream of the *pezA* start codon, 23 variations were within the *pezA* reading frame, which resulted in differences in four amino acid residues (L18P, V24I, H123Q and T131N) and seven variations were within the *pezT* reading frame which resulted in differences in two amino acid residues (K26R and K190Q). On the other hand, when comparing the sequence with the *S. pneumoniae* D39 genome (NCBI accession number: NC_003028), one variation ('C' \rightarrow 'A') was encountered 79 nucleotides upstream of the *pezA* start codon, six variations within the *pezA* reading frame, which resulted in differences in two amino acid residues (H123Q and T131N) and 15 variations within the *pezT* reading frame which resulted in differences in four amino acid residues of PezT (K26R, G109R, I121V and K190Q). As expected, when the nucleotide sequences of the *pezA* and *pezT* homologues from the various *S. pneumoniae* genomes were compared (Appendix Figures B1 and B2), a number of variations were observed. Therefore, nucleotide sequence variations within *pezA* and *pezT* appeared to be common among *S. pneumoniae* strains. The pGEMT_PpezApezT recombinant plasmid was used for subsequent cloning into the pQF52 promoter probe

vector and the pET28a expression vector. No mutations were found in all the pQF52 and pET28a derivative plasmids obtained.

3.1.3 Purification of PezA protein and the PezAT protein complex

The *E. coli* BL21-CodonPlus(DE3)-RIL cells harboring pET28a_HisPezA and pET28a_HisPezAPezT were used to overexpress the PezA protein and the PezAT protein complex for analysis. The fractions of crude lysate, in which the PezA protein and the PezAT protein complex were precipitated at 40-70% of ammonium sulfate, were used for subsequent purification under native conditions, as described in the Materials and Methods (Section 2.5.1). For the pET28a_HisPezAPezT construct, two distinct protein bands were observed on 16% SDS-PAGE following elution with 500 mM imidazole, corresponding to the expected sizes of the (His)₆-PezA fusion protein (20 kDa) and PezT (29 kDa), indicating that under native conditions, both PezA and PezT were co-purified (Figure 17a). This, in turn, suggests that the (His)₆-PezA protein possibly formed a tight complex with PezT. For the pET28a_HisPezA construct, one distinct protein band was observed corresponding to the expected size of the (His)₆-PezA fusion protein (20 kDa) (Figure 17b). Purified proteins were concentrated and subjected to EMSA.

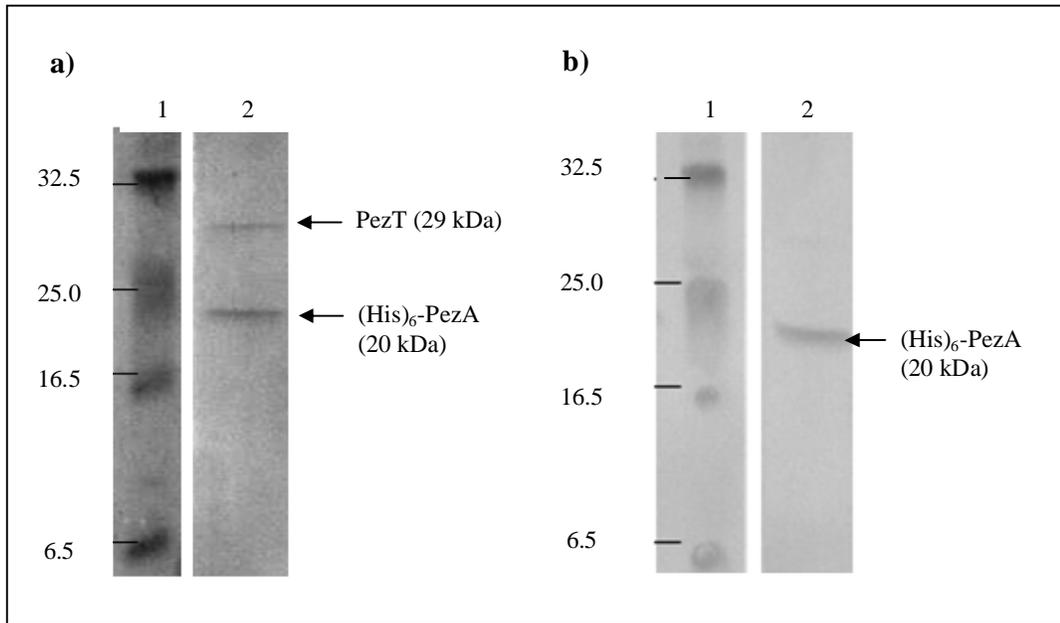


Figure 17: CBB-stained 16% SDS-PAGE gel showing (His)₆-PezA and PezT from *E. coli* BL21-CodonPlus(DE3)-RIL purified under native conditions. Lane 1, protein marker with sizes as indicated in kDa; lane 2, sample of purified (a) (His)₆-PezAT; and (b) (His)₆-PezA eluted from the nickel affinity column (as indicated by the arrows).

3.1.4 Investigation of co-transcription of *pezA* and *pezT*

The *pezA* stop codon and *pezT* start codon overlap by one nucleotide (Figure 16) and therefore are likely to be co-transcribed as reported for other TA loci (Hayes, 2003; Gerdes *et al.*, 2005). To verify this hypothesis, RT-PCR was carried out on total RNA extracted from *E. coli* DH5 α harboring pGEMT_PpezApezT, which contained the entire *pezA* and *pezT* reading frames, together with the 191 bp upstream sequence encompassing the P_{*pezA*} promoter. Reverse transcriptase was used to synthesize cDNAs that were complementary to *pezAT* mRNA and the resulting cDNAs were subjected to subsequent PCR amplification using RTpezA-F forward primer, which annealed within the *pezA* reading frame and RTpezT-R reverse primer, which annealed to the complementary sequence within the *pezT* reading frame (Figure 18a). For positive controls, the *pezA* reading frame was amplified using RTpezA-F forward primer and RTpezA-R reverse primer whereas the *pezT* reading frame was amplified using the RTpezT-F forward and RTpezT-R reverse primer pair (Figure 18a).

An amplified product was obtained using the RTpezA-F and RTpezT-R primer pair and when the reaction product was separated by 1% agarose gel electrophoresis, a single band corresponding to the expected size of 569 bp, which encompassed *pezA* and *pezT* in a single continuous transcript was present (Figure 18b). This band was excised and purified from the gel and sequencing of this fragment confirmed the identity of the amplified product. For the positive controls, a single band was detected for *pezA* and for *pezT*, corresponding to their expected sizes of 302 bp and 233 bp, respectively (Figure 18b). Possible DNA contamination was excluded as no PCR products were observed in the negative control where reverse transcriptase was not added in the reaction mixture (Figure 18b).

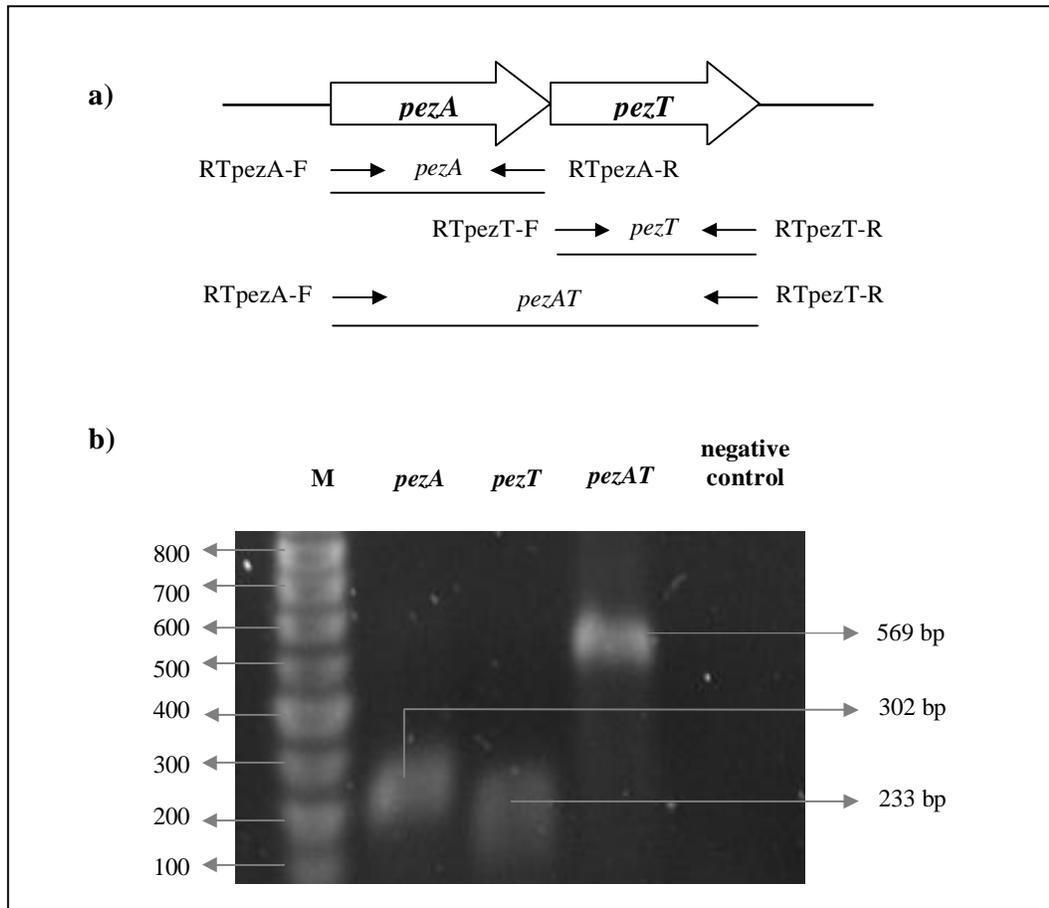


Figure 18: Genetic organization of the *pezAT* operon and the result from RT-PCR. (a) Genetic organization of the *pezAT* operon. Arrows depict the primers used for the RT-PCR analysis to amplify *pezA*, *pezT* and *pezA-pezT* cDNA fragments (b) Agarose gel electrophoresis of the amplified products obtained from RT-PCR. cDNA fragments spanning the *pezA*, *pezT* or *pezA-pezT* genes were synthesized using total RNA prepared from *E. coli* DH5 α harbouring recombinant plasmid pGEMT_PpezApezT. Subsequent PCR reactions were carried out using primer pairs as indicated in (a) Lane M, DNA molecular weight standard with sizes as indicated in bp; RT-PCR products for *pezA*, *pezT* and *pezA-pezT* are indicated by lane *pezA*, *pezT* and *pezAT*, respectively; negative control contained RTpezA-F and RTpezT-R primers but without reverse transcriptase added to preclude DNA contamination of the samples.

3.1.5 Determination of the *pezAT* transcriptional start site using 5'-RACE

To determine the transcriptional start site for the *pezAT* locus, total RNA was isolated from *E. coli* DH5 α harboring pGEMT_PpezApezT (which contained the entire *pezA* and *pezT* reading frames, together with the 191 bp upstream sequence encompassing the P_{*pezA*} promoter) and was subjected to 5'-RACE. Direct sequencing of the product obtained from the 5'-RACE reaction showed an abrupt loss of sequencing signal following the primer sequence despite several attempts. The product of the 5'-RACE reaction was then cloned directly into pGEM-T Easy vector and the recombinants obtained were then sequenced. Sequencing of the resulting recombinants obtained from three different preparations of RNA samples showed that the *pezAT* transcript was initiated at the 'A' residue located 21 nucleotides upstream of the *pezA* start codon (Figure 19 and Appendix Figure C1).

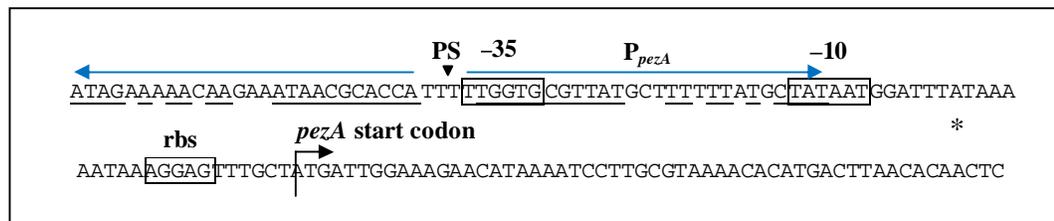


Figure 19: Nucleotide sequence of the region upstream of the *pezAT* locus in the chromosome of *S. pneumoniae* TIGR4. The transcriptional start site of the *pezAT* operon, which was determined using 5'-RACE, was an 'A' residue indicated by an asterisk "*". The start codon of *pezA* is indicated. The -10 and the -35 regions of P_{*pezA*} promoter as well as the putative ribosome binding site (rbs) are depicted within boxes. The divergent arrows (in blue) indicate the imperfect palindrome sequence (PS). The inverted triangle (▼) represents the centre of the palindrome sequence and the nucleotides that are underlined are complementary to each arm of the palindrome sequence.

3.1.6 Determination of promoter activity

3.1.6.1 β -galactosidase assay

The β -galactosidase assays included nine replicates for each clone and the mean values as well as the standard deviation were calculated. The functionality of the P_{pezA} promoter was validated when β -galactosidase activities of 130 ± 25 Miller units were detected for cells harbouring pQF_P (which encompassed just the 191 bp sequence upstream of the *pezA* start codon cloned upstream of the promoter-less *lacZ*) recombinant plasmid, as compared to the cells that contained pQF52 plasmid alone, which showed activities of less than 5 Miller units (Figure 20). For comparison with other recombinant pQF52 clones, the promoter activity of pQF_P was taken as 1.0. When the *pezA* reading frame was included along with the 191 bp upstream sequence in the pQF_P*pezA* recombinant, β -galactosidase activities were reduced by ~33% to 88 ± 9 Miller units. However, when both the *pezA* and *pezT* reading frames were included along with the 191 bp upstream sequence, β -galactosidase activities were further reduced by ~80% to only 26 ± 6 Miller units (Figure 20). The results thus suggest that PezA by itself might act as a weak repressor of its own transcription but this repression was much more significant in the presence of PezT. On the other hand, no promoter activity was detected for the 193 bp sequence upstream of *pezT* (i.e., the pQF_C*pezA* construct) (Figure 20), indicating that promoter activity for the *pezAT* locus is confined to the 191 bp sequence upstream of *pezA*.

Plasmid Constructs	β-galactosidase Activity	
	Miller units	Ratio
pQF52	<5	0.00
	130 ±25	1.00
	88 ±9	0.67
	26 ±6	0.20
	<5	0.00

Figure 20: β-galactosidase activity in *E. coli* DH5α cells carrying the recombinant plasmids pQF_P, pQF_PpezA, pQF_PpezApezT, pQF_CpezA and the parental pQF52 plasmid, as the negative control. β-galactosidase activities were expressed in Miller units and indicated as mean values and standard deviation of nine separate independent experiments. The ratio was calculated by normalization of the β-galactosidase activity of each clone against the β-galactosidase activity of the cells that harbour the pQF_P recombinant plasmid.

3.1.6.2 Quantitative real-time RT-PCR

Quantitative real-time RT-PCR was carried out to corroborate the results of the β-galactosidase assay. Total RNA was extracted from *E. coli* DH5α harboring the various pQF52-derived recombinants (i.e., pQF_P, pQF_PpezA, pQF_PpezApezT, pQF_CpezA and pQF52, as negative control) and quantitative real-time RT-PCR was carried out using primers against *lacZ* and 16S rRNA to determine their respective transcript levels. The values obtained were presented as ratios of *lacZ* over 16S rRNA, and the ratios were normalized with the ratio obtained from cells that harbour pQF_P (Figure 21). The results obtained validated the results from β-galactosidase assays: in the presence of the *pezA* reading frame, a 29% reduction in *lacZ* transcripts was seen when compared to the levels of just the P_{pezA} promoter alone. This was further decreased by 74% when both *pezA* and *pezT* were included along with the P_{pezA} promoter (Figure 21). No prominent *lacZ* transcripts were observed in cells that harbouring pQF_CpezA.

Thus, the PezA antitoxin likely functions as a transcriptional repressor to repress the transcription of the *pezAT* operon and this repression is even more pronounced in the presence of the PezT toxin.

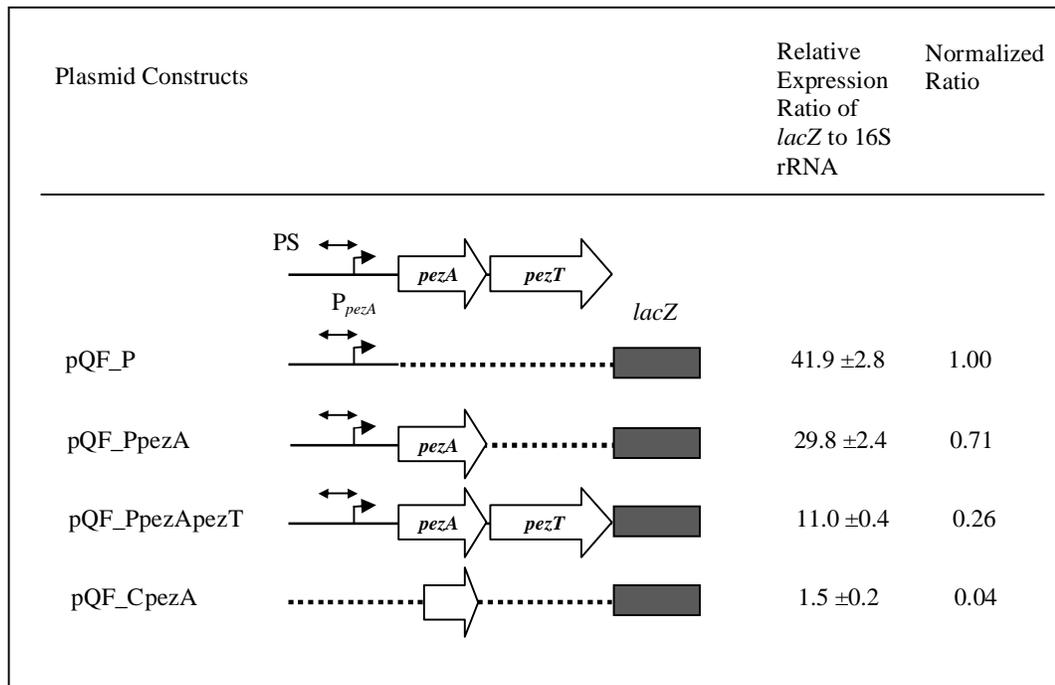


Figure 21: Relative expression ratio of *lacZ* to 16S rRNA transcript levels in *E. coli* DH5 α carrying plasmids pQF_P, pQF_PpezA, pQF_PpezApezT and pQF_CpezA, as determined by quantitative real-time RT-PCR.

3.1.7 Determination of the DNA binding site of PezA protein and PezAT protein complex using EMSA

EMSA was used to determine if the 56 bp imperfect palindrome sequence upstream of the *pezA* reading frame is the binding site (i.e., the operator) for the PezA protein and the PezAT protein complex. Increasing amounts (i.e., 0, 50, 150, 300, 500, 750, 850 and 1000 ng) of the purified PezA protein and the purified PezAT protein complex were incubated separately with a 203 bp biotin end-labeled DNA encompassing the imperfect palindrome sequence. The binding reactions were conducted in the presence of 50 ng of poly(dI-dC) as non-specific competitor DNA. Results showed that retardation of labeled DNA was detected when 150 ng of PezA protein was included into the binding reaction (Figure 22). Moreover, a majority of the labeled DNA was retarded in the presence of 750 ng of PezA protein (Figure 22). On the other hand, retardation of the labeled DNA occurred when only 50 ng of the PezAT protein complex was included in the binding reaction. Similar to the result for the PezA protein, a majority of the labeled DNA was retarded in the presence of 750 ng of the PezAT protein complex (Figure 22).

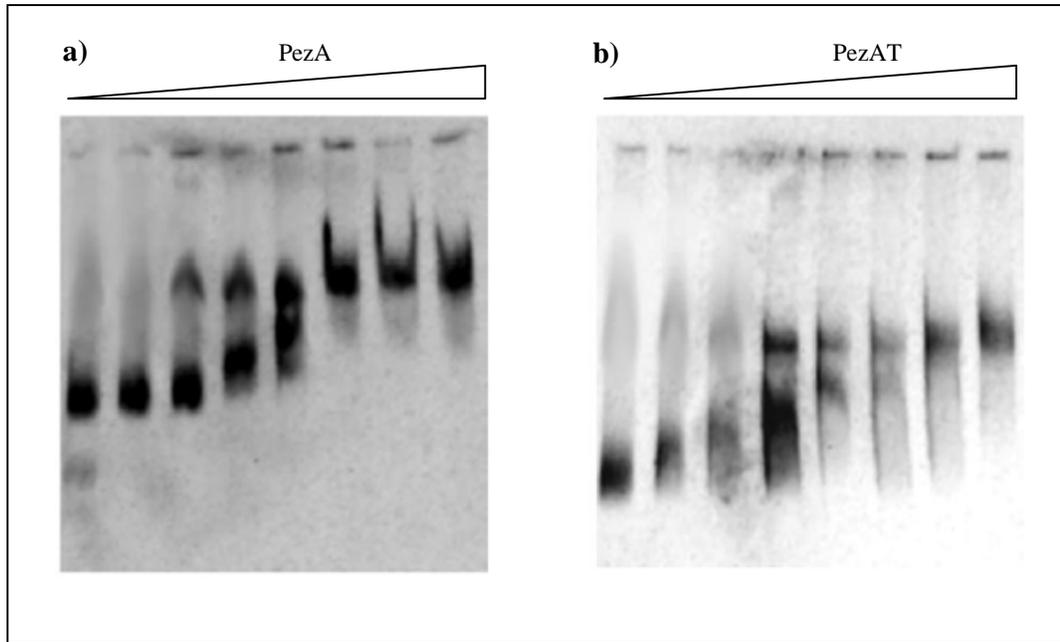


Figure 22: EMSA showing *in vitro* binding of purified PezA protein and PezAT protein complex with 20 fmol of a 203 bp DNA fragment encompassing the imperfect palindrome sequence. Increasing amounts (from left to right: 0, 50, 150, 300, 500, 750, 850 and 1000 ng) of the purified PezA protein (a) PezAT protein complex (b) were incubated with the 203 bp DNA fragment which was labeled with biotin. The unbound DNA fragments and the nucleo-protein complexes were separated by electrophoresis in 5% polyacrylamide gels.

3.2 The *yefM-yoeB*_{S_{pn}} TA Locus of *S. pneumoniae*

3.2.1 Sequence analysis

The pneumococcal *yefM-yoeB* (*yefM-yoeB*_{S_{pn}}) TA locus was discovered due to its sequence similarity with the *E. faecium* plasmid pRUM-encoded *axe-txe* TA locus (Grady and Hayes, 2003) as well as to the *E. coli* *yefM-yoeB* locus (*yefM-yoeB*_{E_{co}}). The YoeB_{S_{pn}} toxin exhibited higher sequence similarities than the YefM_{S_{pn}} antitoxin when compared to their respective homologues (Figure 23 and Figure 24). The YoeB_{S_{pn}} toxin shared 51% and 55% sequence identity with YoeB_{E_{co}} and Txe, respectively, whereas YefM_{S_{pn}} only shared 24% and 26% sequence identity with YefM_{E_{co}} and Axe, respectively. The *yefM-yoeB*_{S_{pn}} TA locus is present as a single copy in the genomes of a few annotated *S. pneumoniae* strains (Table 8). When comparing the *yoeB*_{S_{pn}} sequences of these *S. pneumoniae* strains, there were some nucleotide differences observed but the translated amino acid sequences were 100% identical except for *S. pneumoniae* strain JJA where differences were found in three of the 84 amino acid residues of YoeB_{S_{pn}} (Figure 25). The start codon of *yoeB*_{S_{pn}} among all the strains was identically annotated and the first 10 residues (MLLKFTEDAW) had been confirmed by N-terminal sequencing.

However, for the antitoxin encoded by *yefM*_{S_{pn}}, the gene was annotated to initiate from different start codons for different *S. pneumoniae* strains. In the genome sequences of *S. pneumoniae* strains R6 (NCBI accession number: NC_003098) (Hoskins *et al.*, 2001) and CGSP14 (NCBI accession number: NC_010582) (Ding *et al.*, 2009), *yefM*_{S_{pn}} was annotated to initiate from a GTG start codon 39 nucleotides upstream of the ATG start codon of strains TIGR4 (NCBI accession number: NC_003028) (Tettelin *et al.*, 2001), Hungary19A-6 (NCBI accession number: NC_010380) and JJA (NCBI accession number: NC_012466). Thus, the predicted YefM_{S_{pn}} protein from strains R6 and CGSP14 had an additional 13 amino acids at the

S.pneumoniaeR6_spr1586	MYNSGKELKRGVMVEAVLYSTFRNHLKDYMKKVNDEFEPLTVVNKNPDED	50
S.pneumoniaeTIGR4_SP1741	-----MEAVLYSTFRNHLKDYMKKVNDEFEPLTVVNKNPDED	37
S.pneumoniaeJJA_SPJ1637	-----MEAVLYSTFRNHLKDYMKKVNDEFEPLTVVNKNPDED	37
S.pneumoniaeD39_SPD1551	-----MVMEAVLYSTFRNHLKDYMKKVNDEFEPLTVVNKNPDED	39
S.pneumoniaeCGSP14_SPCG1715	MYNSGKELKRGVMVEAVLYSTFRNHLKDYMKKVNDEFEPLTVVNKNPDED	50
S.pneumoniaeHungary19A-6_SPH1850	-----MEAVLYSTFRNHLKDYMKKVNDEFEPLTVVNKNPDED	37

S.pneumoniaeR6_spr1586	IVVLSKSEWDSIQETLRIAQNKELSDKVLRGMAQVRAGSTQVHVIEE	97
S.pneumoniaeTIGR4_SP1741	IVVLSKSEWDSIQETLRIAQNKELSDKVLRGMAQVRAGSTQVHVIEE	84
S.pneumoniaeJJA_SPJ1637	IVVLSKSEWDSIQETLRIAQNKELSDKVLRGMAQVRAGSTQVHVIEE	84
S.pneumoniaeD39_SPD1551	IVVLSKSEWDSIQETLRIAQNKELSDKVLRGMAQVRAGSTQVHVIEE	86
S.pneumoniaeCGSP14_SPCG1715	IVVLSKSEWDSIQETLRIAQNKELSDKVLRGMAQVRAGSTQVHVIEE	97
S.pneumoniaeHungary19A-6_SPH1850	IVVLSKSEWDSIQETLRIAQNKELSDKVLRGMAQVRAGSTQVHVIEE	84

Figure 26: Multiple sequence alignment of the YefM_{pn} homologues from the annotated genomes of *S. pneumoniae* strains in the NCBI databases. Note that the additional amino acid residues at the N-terminus of YefM_{pn} from strains R6, CGSP14 and D39 are due to the different predicted start sites (see text for details). “*” indicates identical residues; “:” indicates conserved substitution; and “.” indicates semi-conserved substitution.

Analysis of sequences upstream of *yefM_{Spn}* indicated the presence of two σ^{70} -type promoter sequences, designated P_{*yefM1*} and P_{*yefM2*}. For P_{*yefM2*}, the putative -10 region (5'-TATAAT-3') is located 36 nucleotides upstream of the *yefM_{Spn}* start codon and is 17 nucleotides apart from the putative -35 region (5'-cTGACA-3') (Figure 27). P_{*yefM1*} is further upstream of P_{*yefM2*} with its -10 sequence (5'-TATAAA-3') located 95 nucleotides upstream of the *yefM_{Spn}* start codon, and 16 nucleotides apart from the putative -35 sequence (5'-TTGctc-3') (Figure 27). Both promoters are separated by 59 nucleotides (the distance between the centre of the -10 region of both promoters). A possible ribosome binding site (5'-AcGAGG-3') was identified seven nucleotides upstream of the *yefM_{Spn}* ATG start codon (Figure 27). In addition, an incomplete palindrome sequence, designated PS, was identified within the *yefM_{Spn}* upstream region. The palindrome sequence is 44 bp and is centered 62 bp upstream of the *yefM_{Spn}* ATG start codon and overlaps the -35 promoter sequence of P_{*yefM2*} (Figure 27). Interestingly, a BOX element, termed boxA-C element in this study, was identified upstream of P_{*yefM2*} and downstream of the putative transcription terminator of the gene upstream of *yefM_{Spn}*. BOX elements are mobile sequences typical of pneumococci and related species, and are probably involved in the regulation of virulence and genetic

competence (Knutsen *et al.*, 2006). Transcription of this boxA-C element was predicted to form a stem-loop structure, which is in agreement with the study by Martin *et al.* (1992). Moreover, the -35 region and 5 nucleotides (5'-TATAA-3') of the -10 region of P_{yefM1} are located at the 3' end of the boxC element (Figure 27 and Figure 28).

The upstream regions of $yefM_{S_{pn}}$, which include the boxA-C element, P_{yefM1} , P_{yefM2} as well as the ribosomal binding site, are highly conserved among the various *S. pneumoniae* strains in the database (Figure 29). The stop codon of $yefM_{S_{pn}}$ is followed by another three nucleotides prior to the $yoeB_{S_{pn}}$ initiation codon. Both the $YefM_{S_{pn}}$ and $YoeB_{S_{pn}}$ proteins are 84 amino acid residues with $YefM_{S_{pn}}$ having an estimated molecular weight of 9.7 kDa and a theoretical pI of 5.14 whereas the predicted molecular weight of $YoeB_{S_{pn}}$ is 10.1 kDa with a theoretical pI of 7.80 (as determined using ProtParam).

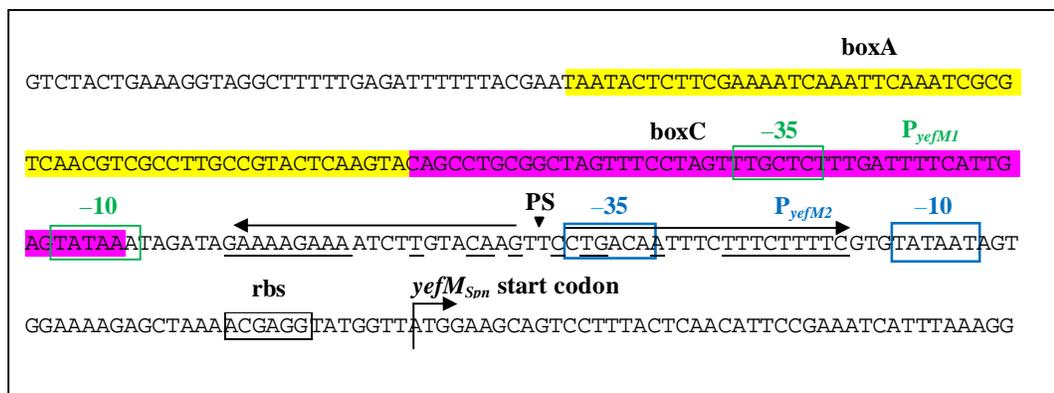


Figure 27: Nucleotide sequence of the $yefM_{S_{pn}}$ upstream regulatory region in the genome of *S. pneumoniae* R6. Two putative σ^{70} -type promoters were identified, depicted as P_{yefM1} and P_{yefM2} . The respective -10 and -35 regions of P_{yefM1} and P_{yefM2} are indicated within boxes along with the putative ribosome binding site (AcGAGG) designated 'rbs'. An incomplete palindrome sequences (PS) is shown with the centre of the palindrome marked by an inverted triangle (). The boxA element is highlighted in yellow whereas the boxC element is highlighted in pink.

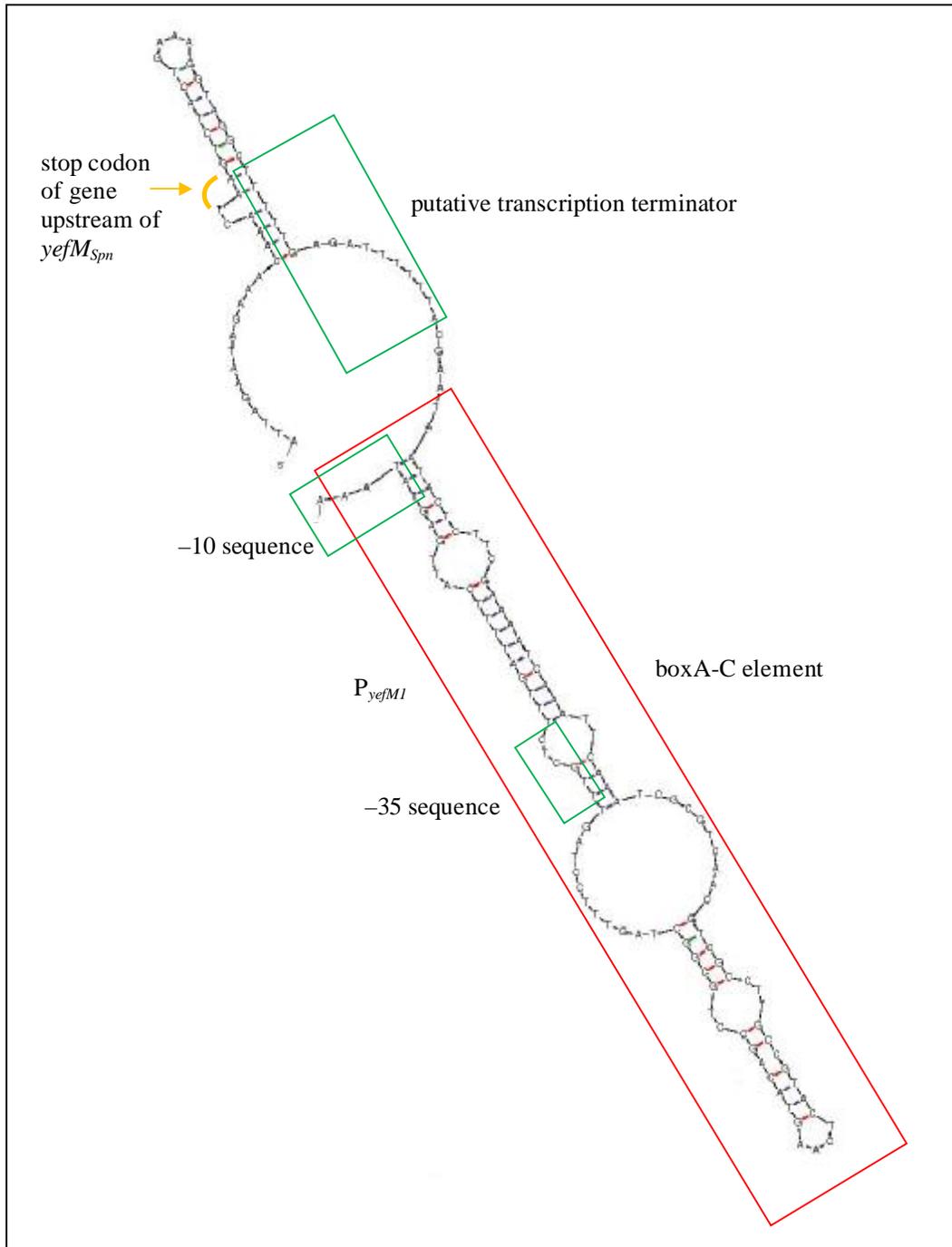


Figure 28: Predicted stem-loop structure of boxA-C element identified in region upstream of P_{yefM2} and downstream of the putative transcription terminator of gene upstream of *yefM_{Spn}*. Part of P_{yefM1} (5'-TATAA-3') is identified within the 3' proximal of boxC element.

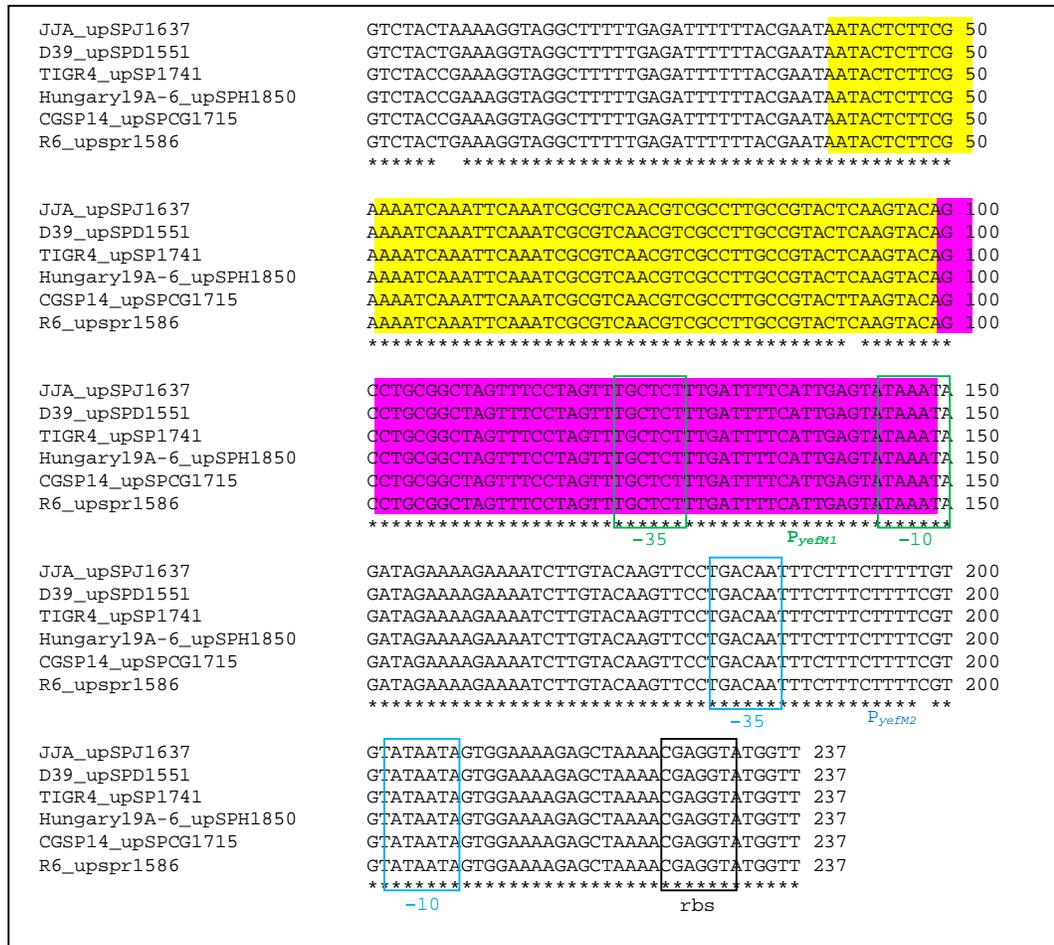


Figure 29: Multiple sequence alignment of nucleotide sequences upstream of *theyefM_{Spn}* reading frame in several *S. pneumoniae* strains available in the database. The -10 and the -35 putative promoter sequences of *P_{yefM1}* and *P_{yefM2}* as well as the putative ribosome binding site (rbs) are depicted within boxes. “*” indicates identical residues; “.” indicates conserved substitution; and “:” indicates semi-conserved substitution. The boxA element is highlighted in yellow whereas boxC element is highlighted in pink.

3.2.2 Construction of recombinant plasmids

The pGEM-T Easy recombinant plasmid that had the same sequence as the *yefM-yoeB_{Spn}* locus of *S. pneumoniae* R6, pGEMT_P1P2yefMYoeB, was used for the subsequent cloning. No mutations were found in the resulting recombinant plasmids (pQF52 and pLNBAD derivative recombinant plasmids) obtained. For pET28a_HisYefMHis and pET28a_HisYefMYoeB, the recombinant plasmids contained the extra 39 bp upstream of the *yefM_{Spn}* ATG start codon, as the construct was designed according to the annotated *S. pneumoniae* R6 genome sequence where the predicted *yefM_{Spn}* start codon, GTG, was located 39 bp upstream of the actual ATG start codon of *yefM_{Spn}* (Table 5) as described in section 3.2.1 and the Materials and Methods. The pET28a recombinant plasmid was constructed such that *yefM-yoeB_{Spn}* was in-frame with the (His)₆ coding sequence of the pET28a vector, thus resulting in the eventual expression of N-terminal (His)₆-fusion YefM-YoeB_{Spn} proteins. Similar to *yefM-yoeB_{Spn}*, *yefM_{Spn}* was also constructed in-frame with the (His)₆ which would result in the (His)₆ fused with YefM_{Spn} at its N-terminus. However, this protein fusion had a tendency to form aggregates during purification and the protein obtained was not sufficient for subsequent assays. Nevertheless, another construct was obtained during cloning with a 'G' deletion 3 nucleotides upstream of the *yefM_{Spn}* stop codon. This deletion led to an extension of 16 amino acids (EDPNSSSVDKLAAALE), followed by a (His)₆-tag at the C-terminus of YefM_{Spn} (Figure 30). Moreover, due to the additional 39 bp DNA sequence which contained the P_{yefM2} promoter as well as the native ribosome binding site, YefM_{Spn} was likely translated from this ribosome binding site in addition to the translation from the ribosome binding site of the pET28a plasmid. Therefore, a combination of two YefM_{Spn} fusion proteins were obtained and purified: one fusion protein with (His)₆-tags at both C- and N-terminus of YefM_{Spn} and another fusion protein with the (His)₆-tag at the C-terminus of YefM_{Spn}. Surprisingly, these

fusion proteins were soluble and easy to purify. Therefore, this variant construct, designated pET28a_HisYefMHis, was subsequently used for the overexpression and purification of recombinant YefM_{S_{pn}}.

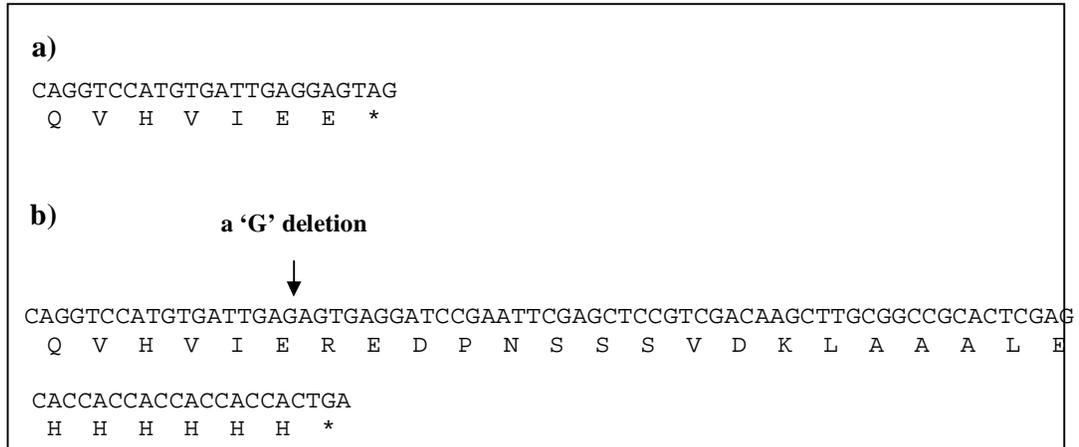


Figure 30: Nucleotide and amino acid sequences of the C-terminus of *yefM_{S_{pn}}* of the genome of *S. pneumoniae* R6 and the pET28a_HisYefMHis recombinant plasmid. (a) Nucleotide and amino acid sequence with the native TGA stop codon at the C-terminus end of YefM_{S_{pn}} and; (b) a 'G' deletion 3 nucleotides upstream of the *yefM_{S_{pn}}* TAG stop codon which led to an extension of amino acids, followed by a (His)₆-fusion at the C-terminus of YefM_{S_{pn}} in the pET28a_HisYefMHis recombinant plasmid.

3.2.3 Purification of YefM_{Spn} protein and YefM-YoeB_{Spn} protein complex

For the (His)₆-YefM-YoeB_{Spn} construct, two distinct protein bands were observed on 16% SDS-PAGE gel, which corresponded to the expected sizes of the (His)₆-YefM_{Spn} fusion protein (13.5 kDa) and YoeB_{Spn} (10.1 kDa), indicating that under native conditions, both YefM_{Spn} and YoeB_{Spn} were co-purified (Figure 31a). For the pET28a_HisYefMHis construct, two distinct protein bands were observed in the SDS-PAGE gel of the purified proteins: a band which corresponded to the size of (His)₆-YefM_{Spn}-(His)₆ (15.8 kDa) and a band which corresponded to the size of YefM_{Spn}-(His)₆ (12.1 kDa) (Figure 31b). N-terminal sequencing of these two protein bands validated their identities and showed that two fusion proteins were expressed: YefM_{Spn} with (His)₆-tagged at both N- and C-terminal regions, as well as a C-terminal (His)₆-tagged YefM_{Spn} resulting from translation from the ATG start codon and likely utilizing the native ribosome binding site located within the additional 39 bp.

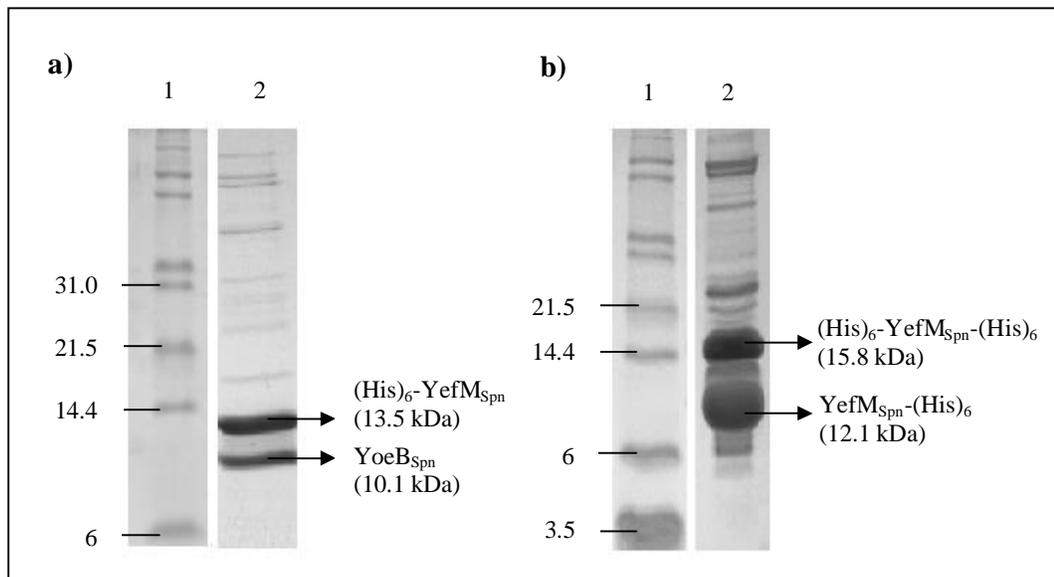


Figure 31: CBB-stained 16% SDS-PAGE gel showing purified YefM-YoeB_{Spn} and YefM_{Spn} fusion proteins from *E. coli* BL21-CodonPlus(DE3)-RIL under native conditions. Lane 1, Protein Marker with sizes shown in kDa; lane 2, sample of (a) YefM-YoeB_{Spn} and, (b) YefM_{Spn} fusion proteins eluted from the nickel affinity column as indicated by the arrows.

3.2.4 Investigation of co-transcription of *yefM_{Spn}* and *yoeB_{Spn}*

The *yefM_{Spn}* stop codon and *yoeB_{Spn}* start codon were separated by three nucleotides. To find out if the *yefM-yoeB_{Spn}* genes are co-transcribed as reported for other TA loci (Gerdes *et al.*, 2005), RT-PCR was carried out on total RNA extracted from *E. coli* DH5 α harboring pGEMT_P1P2yefMyoeB, which encompassed the *yefM-yoeB_{Spn}* reading frames along with the 237 upstream promoter sequences. cDNA that was complementary to the *yefM-yoeB_{Spn}* mRNA was synthesized using reverse transcriptase and subsequently amplified by PCR using the RTyefM-F forward primer and the RTyoeB-R reverse primer (Figure 32a). At the same time, *yefM_{Spn}* was amplified using RTyefM-F forward primer and RTyefM-R reverse primer whereas *yoeB_{Spn}* was amplified using RTyoeB-F forward primer and RTyoeB-R reverse primer as positive controls. As for the negative control, the RTyefM-F and RTyoeB-R primer pair with same reaction mixture was used but reverse transcriptase was not included to preclude the possibility of DNA contamination.

When the RT-PCR products were separated by agarose gel electrophoresis, a single band corresponding to the expected size of 475 bp, which encompassed both *yefM_{Spn}* and *yoeB_{Spn}* in a single continuous transcript was observed (Figure 32b). This band was excised and purified from the agarose gel and subsequent sequencing of this fragment confirmed the identity of the amplified product. For the positive controls, a single band with 201 bp and 220 bp were observed, which corresponded to the expected sizes for *yefM_{Spn}* and *yoeB_{Spn}*, respectively (Figure 32b). No PCR product was detected in the negative control thus ruling out any DNA contamination in the RNA sample used (Figure 3.2.10b).

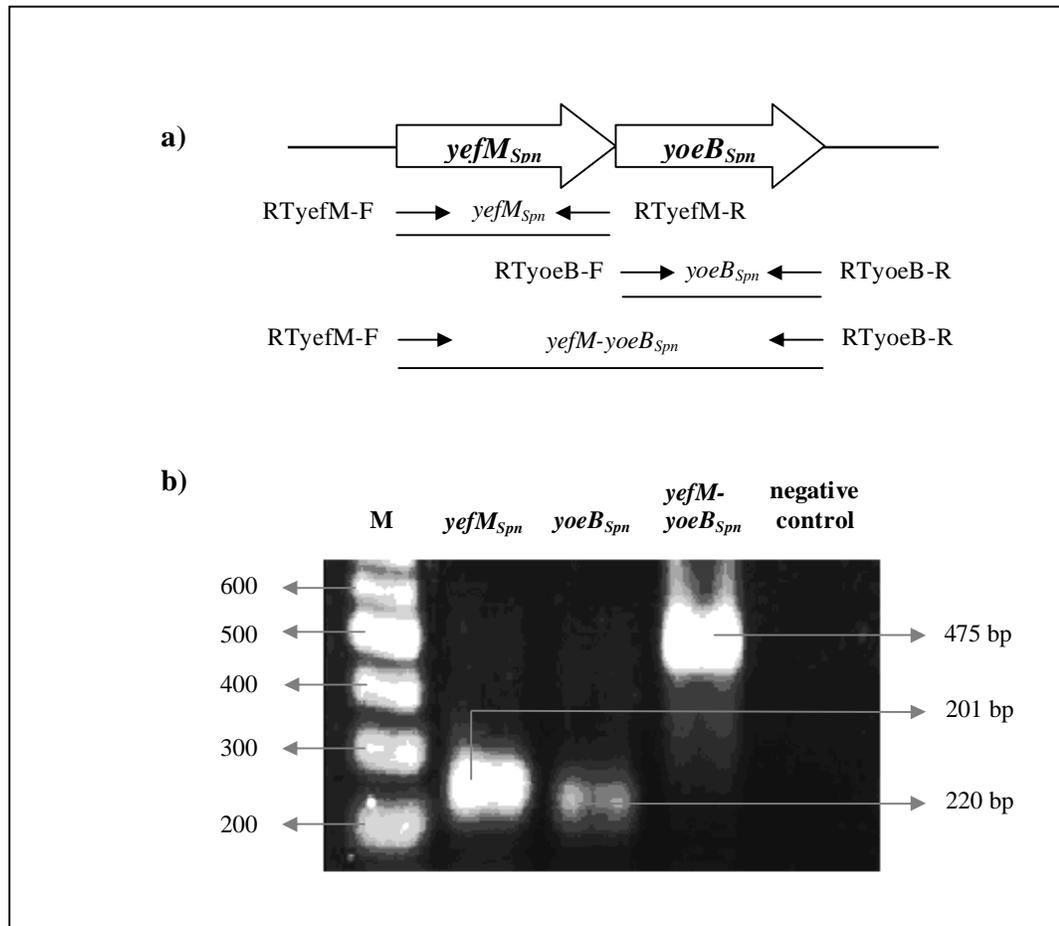


Figure 32: Genetic organization of the *yefM-yoeB_{Spn}* operon and the results from RT-PCR. (a) Genetic organization of the *yefM-yoeB_{Spn}* operon. Arrows depict the primers used for the RT-PCR analysis to amplify *yefM_{Spn}*, *yoeB_{Spn}* and *yefM-yoeB_{Spn}* cDNA fragments. (b) Agarose gel electrophoresis of the amplified products obtained from RT-PCR. cDNA fragments spanning the *yefM_{Spn}*, *yoeB_{Spn}* and *yefM-yoeB_{Spn}* genes were synthesized using total RNA prepared from *E. coli* DH5 α harbouring recombinant plasmid pGEMT_P1P2yefMyoeB. Subsequent PCR reaction was carried out using primer pairs as indicated in (a). Lane M, DNA molecular weight standard with sizes as indicated in bp; RT-PCR products for *yefM_{Spn}*, *yoeB_{Spn}* and *yefM-yoeB_{Spn}* were also indicated; negative control contained RTyefM-F and RTyoeB-R primers and the same reaction mixture but without reverse transcriptase.

3.2.5 Determination of *yefM-yoeB_{Spn}* transcriptional start sites

3.2.5.1 5'-RACE

To determine the transcriptional start site of the *yefM-yoeB_{Spn}* locus, total RNA isolated from *E. coli* DH5 α harboring pGEMT_P1P2yefMyoeB, which encompassed the *yefM_{Spn}* and *yoeB_{Spn}* reading frames along with the 237 upstream promoter and regulatory sequences were subjected to 5'-RACE. The amplified product from the 5'-RACE reaction was cloned directly into pGEM-T Easy vector and the recombinants obtained were sequenced. The sequencing results from the resulting recombinants indicated that the *yefM-yoeB_{Spn}* transcript initiated from an 'A' residue located 25 nucleotides upstream of the *yefM_{Spn}* start codon (Figure 33). Inspection of the promoter sequence revealed the distance between the -10 element of the P_{yefM2} promoter and the transcriptional start site is 5 nucleotides (Figure 33). This indicated the resulting 5'-RACE products would be the mRNAs transcribed from the P_{yefM2} promoter.

3.2.5.2 Primer extension

To verify the result obtained from 5'-RACE, a primer extension assay was conducted using total RNA extracted from *S. pneumoniae* R6. A *yefM*-near or *yefM*-far specific primer was used to anneal to the RNA samples. The primer was extended, yielding transcripts of 92 nucleotides or 115 nucleotides, at 50°C or at 55°C respectively. The higher temperature, (i.e., 55°C), was used to eliminate the possibility of false negative results in the primer extension conducted at 50°C which could be due to the secondary structure of the mRNA that was not completely disrupted. Two transcriptional start sites were detected, which were the 'A' residue located 25 nucleotides (i.e., identical to the result obtained from 5'-RACE) and another 'A' residue located 84 nucleotides upstream of the *yefM_{Spn}* ATG start codon (Figure 34). The primer extension product that was initiated from 84 nucleotides upstream of the *yefM_{Spn}* ATG

start codon is likely transcribed from the P_{yefM1} promoter as the distance between the –10 element of the P_{yefM1} promoter and the transcriptional start site is five nucleotides (Figure 33). However, this result was not evident in 5'-RACE. The results from primer extension were in agreement with the DNA sequence analysis where two canonical σ^{70} promoters could be identified, denoted as P_{yefM1} and P_{yefM2} . Transcription from P_{yefM1} would lead to the transcript initiating 84 nucleotides upstream of the $yefM_{S_{pn}}$ start codon whereas transcription from P_{yefM2} would result in a transcript that initiates 25 nucleotides from the $yefM_{S_{pn}}$ start codon.

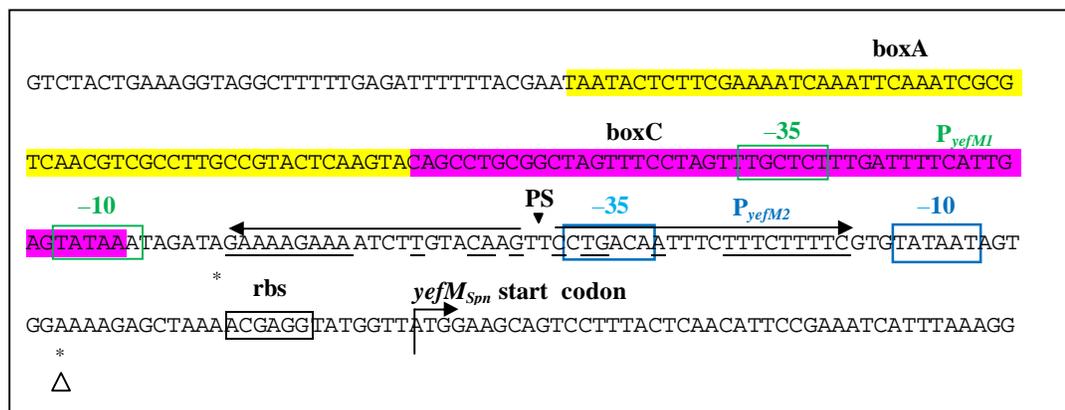


Figure 33: Nucleotide sequence of the region upstream of the $yefM$ - $yoeB_{S_{pn}}$ TA locus in the genome of *S. pneumoniae* R6. The transcriptional start site of $yefM$ - $yoeB_{S_{pn}}$ TA locus as determined by 5'-RACE was an 'A' residue indicated by "Δ". Asterisks "*" denote the two transcriptional start sites of $yefM$ - $yoeB_{S_{pn}}$ determined by primer extension analyses. The two putative σ^{70} -type promoters are depicted as P_{yefM1} and P_{yefM2} . The respective –10 and –35 regions of P_{yefM1} and P_{yefM2} are indicated within boxes along with the putative ribosome binding site (designated 'rbs'). The palindrome sequences (PS) is shown with the centre of the palindrome marked by an inverted triangle (▼). The boxA element is highlighted in yellow whereas the boxC element is highlighted in pink.

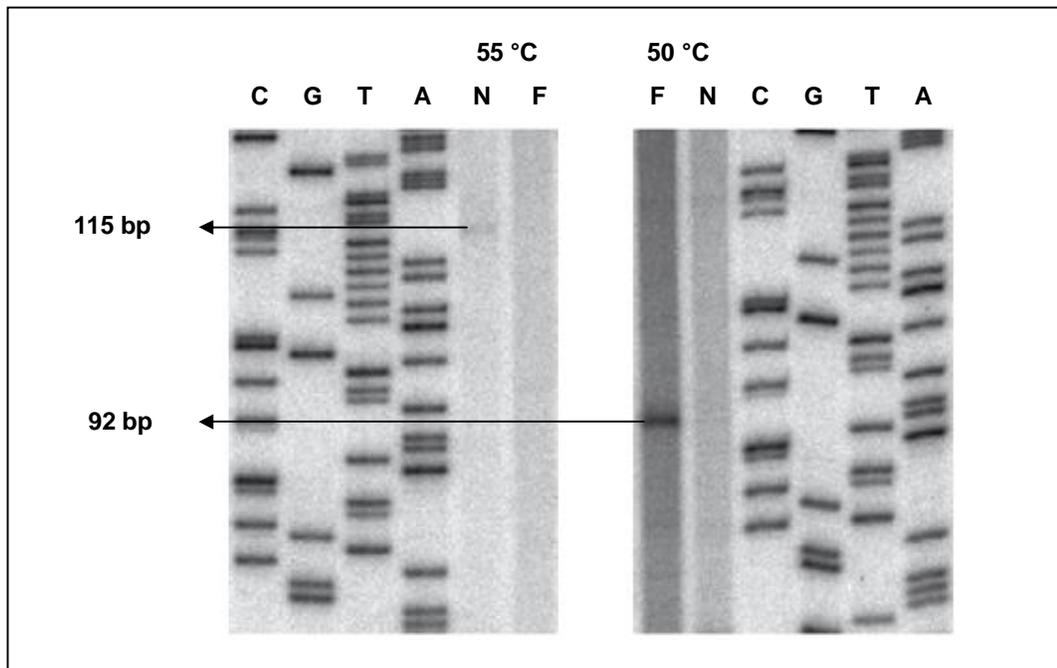


Figure 34: Transcriptional start sites of the *yefM-yoeB_{Spn}* operon determined by primer extension analyses. A *yefM*-near (N) or *yefM*-far (F) specific primer was used to anneal to the RNA samples prepared from *S. pneumoniae* R6. The primer was extended, yielding transcripts of 92 nucleotides or 115 nucleotides, at 50°C or at 55°C respectively. The C, G, T and A DNA sequencing ladder were obtained using the dideoxy-mediated chain-termination sequencing method, by annealing the *yefM*-near or *yefM*-far primers to pEMB13 (Niet α *et al.*, 2007), respectively.

3.2.6 Determination of promoter activity

3.2.6.1 β -galactosidase assays

The β -galactosidase assay for each pQF52 construct was repeated at least eight times and the mean value was calculated. P_{yefM2} was shown to be a functional promoter when β -galactosidase activities of 122 ± 35.7 Miller units were detected for cells that harboured pQF_P2, as compared to the cells that harboured the pQF52 plasmid alone, which had activities of <5 Miller units (Figure 35). Interestingly, β -galactosidase activities were reduced by 43% to 70 ± 9.5 Miller units in cells harbouring pQF_P1P2, which encompassed both P_{yefM1} and P_{yefM2} (Figure 35). To determine the contribution of the P_{yefM1} promoter to the overall promoter activity in pQF_P1P2, site-directed mutagenesis was carried out to knockout the P_{yefM2} promoter by changing its -10 sequence from $5'$ -TATAAT- $3'$ to $5'$ -ctgcAg- $3'$ resulting in the pQF_P1 plasmid. β -Galactosidase activities of only 8 ± 0.9 Miller units were detected in cells that harboured pQF_P1 (Figure 35) suggesting that P_{yefM1} is a much weaker promoter compared to P_{yefM2} . To validate that the $5'$ -TATAAT- $3'$ to $5'$ -ctgcAg- $3'$ mutation truly abolished P_{yefM2} promoter activity, the mutation was carried out on pQF_P2. Cells that harboured the resulting pQF_nP recombinant showed β -galactosidase activities of <5 Miller units (Figure 35), thus confirming that the mutation resulted in the abolishment of P_{yefM2} activity.

Plasmid Constructs	β-galactosidase activities	
	Miller units	Ratio
pQF52	<5	-
pQF_P1P2	70 ±9.5	0.57
pQF_P2	122 ±35.7	1.00
pQF_P1	8 ±0.9	0.07
pQF_nP	<5	-

Figure 35: β-galactosidase activities of *E. coli* TOP10 cells carrying the recombinant plasmids pQF_P1P2, pQF_P2, pQF_P1, pQF_nP and the parental pQF52 plasmid as the negative control. The ratios were calculated by normalization of the mean β-galactosidase activity of each clone against the β-galactosidase activity of the cells that harbour the pQF_P2 recombinant plasmid.

To investigate the effects of YefM_{Spn} and YoeB_{Spn} on P_{yefM2}, pQF_P2 plasmids were co-transformed with either pLN_yM (for the arabinose-induced expression of YefM_{Spn}) or pLN_yMyB (for the expression of both YefM_{Spn} and YoeB_{Spn}) into *E. coli* TOP10 cells. When the YefM_{Spn} antitoxin was expressed in *trans* (in cells that harboured both the pLN_yM and pQF_P2 plasmids), the β-galactosidase activity was slightly reduced by about 7% to 43 ±10.2 Miller units, when compared to the uninduced cells, which showed activity levels of 46 ±7.1 Miller units (Figure 36). Moreover, when both the YefM_{Spn} antitoxin and the YoeB_{Spn} toxin were co-expressed in *trans* (in cells that harboured both the pLN_yMyB and pQF_P2 plasmids), the β-galactosidase activity was further reduced by 87% to 11 ±2.1 Miller units when compared to the uninduced cells which showed β-galactosidase activity levels of 87 ±26.3 Miller units (Figure 37). From these results, it can be inferred that the YefM_{Spn} antitoxin by itself acts as a weak

repressor and the YoeB_{Spn} toxin acts as a co-repressor to further repress transcription from the P_{yefM2} promoter, as had been reported for other TA loci.

When P_{yefM1} and boxA-C element were included in the construct along with P_{yefM2} and palindrome sequence, expression of YefM_{Spn} in *trans* (in arabinose-induced cells harbouring pLN_{yM} and pQF_P1P2) led to a 12% reduction in β -galactosidase activity (14 ± 1.2 Miller units) when compared to uninduced cells which was 16 ± 4.0 Miller units (Figure 36). However, further repression was not observed when YefM-YoeB_{Spn} was expressed in *trans* (in induced cells harbouring pLN_{yMyB} and pQF_P1P2), where a similar level of reduction in β -galactosidase activity (i.e., 16% reduction) was observed when compared to uninduced cells (Figure 37).

The β -galactosidase activity levels in cells that harboured the pQF52-derived recombinant with only functional P_{yefM1} and box A-C element (i.e., pQF_P1 with the –10 region of P_{yefM2} mutated from 5'-TATAAT-3' to 5'-ctgcAg-3') were low (7 ± 2.0 Miller units in Figure 36 and 5 ± 2.7 Miller units in Figure 37) and merely slight repression was observed when either YefM_{Spn} or YefM-YoeB_{Spn} were expressed in *trans*, respectively (<5 Miller units for both constructs) (Figure 36 and Figure 37). Cells that harboured the pQF_{nP} recombinant (i.e., with only P_{yefM2} but with its –10 region mutated from 5'-TATAAT-3' to 5'-ctgcAg-3') showed no detectable β -galactosidase activities regardless of whether YefM_{Spn} or YefM-YoeB_{Spn} were expressed in *trans* (in cells that harboured pQF_{nP} and pLN_{yM} or pQF_{nP} and pLN_{yMyB}, respectively; Figure 36 and Figure 37).

Plasmid Constructs		β-galactosidase activities	
		Miller units	Ratio
pQF_P1P2 &		UI 16 ±4.0	1.00
pLN_yM		I 14 ±1.2	0.88
pQF_P2 &		UI 46 ±7.1	1.00
pLN_yM		I 43 ±10.2	0.93
pQF_P1 &		UI 7 ±2.0	-
pLN_yM		I <5	-
pQF_np &		UI <5	-
pLN_yM		I <5	-

Figure 36: β-Galactosidase activities of *E. coli* TOP10 cells carrying the pLN_yM-derived recombinant plasmids harboring *yefM_{Spn}* under the control of the arabinose-inducible P_{BAD} promoter and pQF52-derived recombinant plasmids pQF_P1P2, pQF_P2, pQF_P1 and pQF_nP containing various upstream regulatory fragments as depicted. β-Galactosidase activities were measured when cells were either uninduced (UI) or induced (I) with 1 mM L-arabinose for the expression of *yefM_{Spn}*. Ratios were obtained by normalization of the mean β-galactosidase activity of induced cells against uninduced cells harbouring the various recombinant plasmids, respectively.

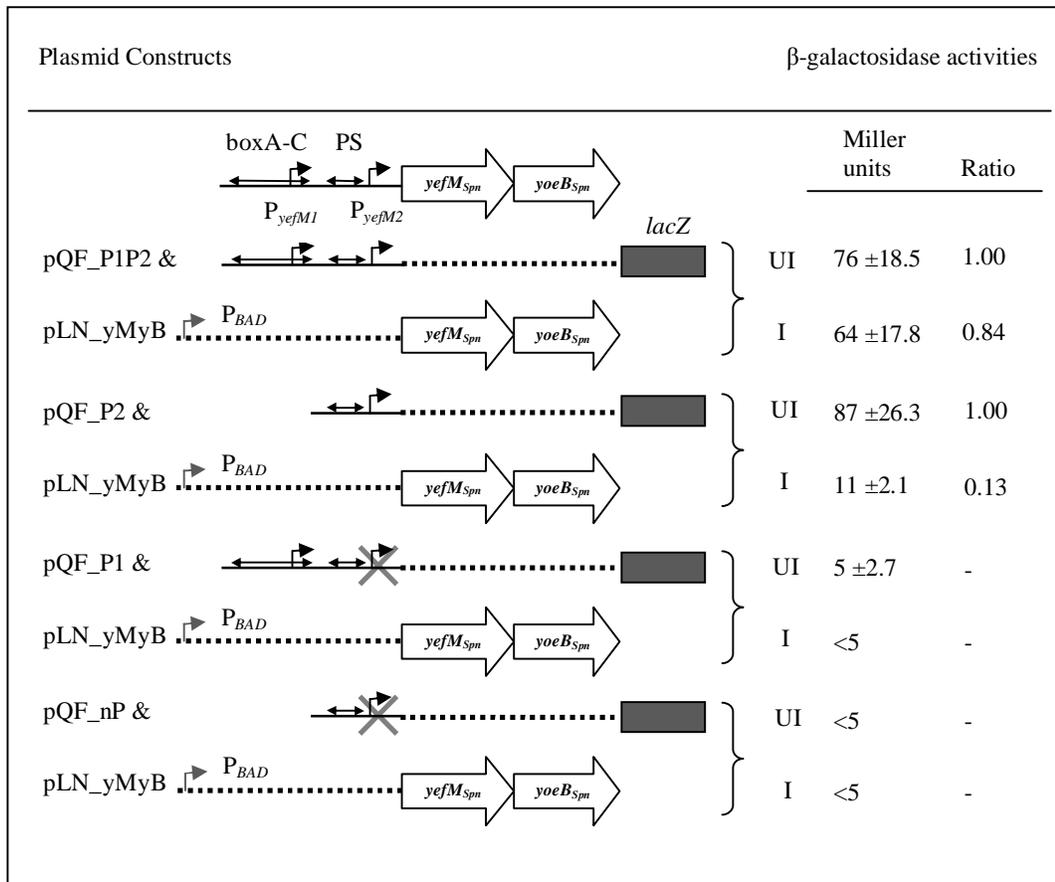


Figure 37: β-Galactosidase activities of *E. coli* TOP10 cells carrying the pLN_yM-derived recombinant plasmids with *yefM-yoeB_{Spn}* under the control of the arabinose-inducible P_{BAD} promoter and pQF52-derived recombinant plasmids pQF_P1P2, pQF_P2, pQF_P1 and pQF_nP containing various upstream regulatory fragments as depicted. β-Galactosidase activities were measured when cells were either uninduced (UI) or induced (I) with 1 mM L-arabinose for the expression of YefM-YoeB_{Spn}. Ratios were obtained by normalization of the mean β-galactosidase activities of the induced cells against the uninduced cells harbouring the various recombinant plasmids, respectively.

Similar to the β-galactosidase activity profiles observed when YefM_{Spn} was expressed in *trans*, repression was also observed when YefM_{Spn} was expressed in *cis* with the P_{yefM2} promoter. β-Galactosidase activity levels were reduced 60% to 49 ± 14.8 Miller units in cells that harboured pQF_P2yM when compared to cells that harboured pQF_P2 (pQF52 containing P_{yefM2} alone; Figure 38). The repression was more pronounced when YefM-YoeB_{Spn} was co-expressed in *cis* along with the P_{yefM2} promoter as observed in cells that harboured pQF_P2yMyB where the β-galactosidase

activity levels were 87% lower (16 ± 8.6 Miller units) when compared to cells that harboured pQF_P2 (Figure 38).

Intriguingly, when P_{yefM1} and boxA-C element were included along with P_{yefM2} and the $yefM_{Spn}$ reading frame, an apparent activation was observed instead of repression as cells that harboured pQF_P1P2yM had 171% higher β -galactosidase activity levels (190 ± 13.9 Miller units) when compared to cells that harboured pQF_P1P2 which encompassed P_{yefM1} and P_{yefM2} (70 ± 9.5 Miller units; Figure 38). When both the $yefM_{Spn}$ and $yoeB_{Spn}$ reading frames were included along with the P_{yefM1} and P_{yefM2} promoters (in the pQF_P1P2yMyB construct), lower β -galactosidase activity levels were observed (138 ± 9.9 Miller units) when compared with pQF_P1P2yM. Nevertheless, the activity levels were still 97% higher than in cells harbouring pQF_P1P2 (Figure 38).

To investigate if the YefM_{Spn} and YefM-YoeB_{Spn} proteins could actually serve as activators of their own promoters as suggested by the β -galactosidase results, an amber stop codon was introduced at the 7th codon of YefM_{Spn} (Figure 39) to abolish the translation of full length YefM_{Spn}. Surprisingly, the activation was even more prominent for cells that harboured pQF_M1yM and pQF_M1yMyB recombinants which had the $yefM_{Spn}$ amber mutation. In cells that contained pQF_M1yM, a 230% increase in β -galactosidase activity (231 ± 50.6 Miller units) was observed when compared to cells that harboured pQF_P1P2 (Figure 38). In the presence of the $yoeB_{Spn}$ reading frame in *cis*, the activation was further increased by 254% (248 ± 62.8 Miller units) for cells harbouring pQF_M1yMyB, when compared to cells that harboured pQF_P1P2 (Figure 38). To rule out the possibility that in pQF_M1yM and pQF_M1yMyB, translation of a truncated YefM_{Spn} could still occur from the Met-17 residue and thus causes the apparent activation, three stop codons were introduced further downstream – i.e., at residues 27, 32 and 33 of YefM_{Spn} (Figure 39). The β -galactosidase activity levels of

the resulting pQF_M2yM and pQF_M2yMyB recombinants were similar when compared with the pQF_M1yM and pQF_M1yMyB recombinants (i.e., with the single amber stop codon in *yefM_{Spn}*), respectively. When compared with cells harbouring pQF_P1P2, β -galactosidase activities of pQF_M2yM cells increased by 226% (228 \pm 19.0 Miller units) and by 260% (252 \pm 24.7 Miller units) for pQF_M2yMyB cells (Figure 38).

To investigate whether there are any functional promoters present within the *yefM_{Spn}* reading frame that could have contributed to the activation observed in pQF_P1P2yM, two pQF52-derived recombinant plasmids were constructed: pQF_yM which contained the entire *yefM_{Spn}* reading frame cloned upstream of the promoter-less *lacZ* gene of pQF52, and pQF_CyM which contained only the C-terminal region of YefM_{Spn} (195 bp upstream of the *yoeb_{Spn}* start codon). β -Galactosidase activity was not detected in cells harbouring either pQF_yM or pQF_CyM (Figure 38), ruling out the presence of an internal promoter within the *yefM_{Spn}* reading frame. Taken together, these results appeared to suggest that the activation observed in cells that harboured pQF_P1P2yM was not due to the presence of an internal promoter within the *yefM_{Spn}* reading frame or was activated by the YefM_{Spn} protein but rather caused by an unknown *cis*-acting element. It has to be emphasized that this activation could only be observed in the presence of the entire upstream regulatory region (which encompassed the boxA-C element, P_{yefM1}, P_{yefM2} and palindrome sequence) along with the *yefM_{Spn}* reading frame.

Plasmid Constructs	β -galactosidase activities	
	Miller units	Ratio
pQF_P1P2	70 \pm 9.5	1.00
pQF_P1P2yM	190 \pm 13.9	2.71
pQF_P1P2yMyB	138 \pm 9.9	1.97
pQF_P2	122 \pm 35.7	1.00
pQF_P2yM	49 \pm 14.8	0.40
pQF_P2yMyB	16 \pm 8.6	0.13
pQF_M1yM	231 \pm 50.6	3.30
pQF_M1yMyB	248 \pm 62.8	3.54
pQF_M2yM	228 \pm 19.0	3.26
pQF_M2yMyB	252 \pm 24.7	3.60
pQF_yM	< 5	-
pQF_CyM	< 5	-

Figure 38: β -Galactosidase activities of *E. coli* TOP10 cells carrying the following pQF52-derived recombinant plasmids: pQF_P1P2, pQF_P1P2yM, pQF_P1P2yMyB, pQF_P2, pQF_P2yM, pQF_P2yMyB, pQF_M1yM, pQF_M1yMyB, pQF_M2yM, pQF_M2yMyB, pQF_yM and pQF_CyM. For cells that harbour pQF_P1P2, pQF_P1P2yM, pQF_P1P2yMyB, pQF_M1yM, pQF_M1yMyB, pQF_M2yM and pQF_M2yMyB, ratios were calculated by normalization of the β -galactosidase activity levels with that of pQF_P1P2. For cells that harbour pQF_P2, pQF_P2yM and pQF_P2yMyB, ratios were calculated by normalization of the β -galactosidase activity levels with that of pQF_P2. Stars (\star) indicate stop codons that were introduced into the *yefM_{Spn}* reading frame as detailed in Figure 39.

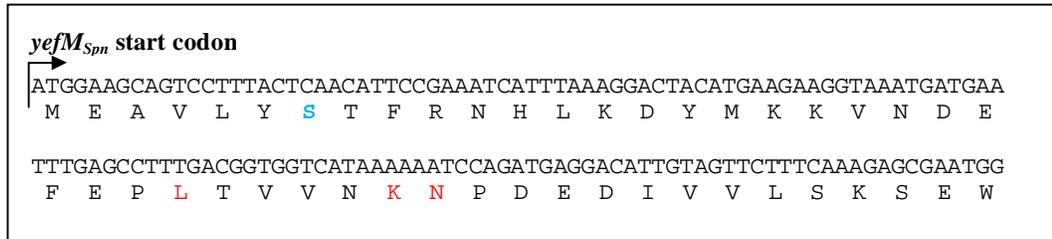


Figure 39: Sequence of the first 48 codons of *yefM_{Spn}* indicating the locations in which stop codons were introduced within the *yefM_{Spn}* reading frame using site-directed mutagenesis. In pQF_M1yM and pQF_M1yMyB, the 7th codon, S (indicated in blue) was replaced by an amber stop codon. Three other codons, L-27, K-32 and N-33 (indicated in red), were replaced by amber, ochre and amber stop codons, respectively, in the recombinant plasmids pQF_M2yM and pQF_M2yMyB.

To investigate the effects of *yefM_{Spn}* and *yefM-yoeB_{Spn}* on the P_{yefM1} promoter, the -10 region of P_{yefM2} was mutated from 5'-TATAAT-3' to 5'-ctgcAg-3'. This mutation was shown to inactivate P_{yefM2} as β -galactosidase activities of *E. coli* cells harbouring pQF_nP was <5 Miller units as had been previously shown in Figure 35. The presence of *yefM_{Spn}* in *cis* (pQF_P1yM) led to a slight increment (13%) in β -galactosidase activity when compared to cells harbouring just the active P_{yefM1} alone (pQF_P1) (Figure 40). However, the presence of *yoeB_{Spn}* led to a massive increase in β -galactosidase activity levels where 150 \pm 10.2 Miller units was measured in cells that harboured pQF_P1yMyB as compared to only 8 \pm 0.9 Miller units in cells that harboured pQF_P1 (Figure 40). Similar results were obtained in constructs containing just the mutated inactive P_{yefM2} without the presence of P_{yefM1} (i.e., pQF_nP, pQF_nPyM and pQF_nPyMyB; Figure 40). No detectable β -galactosidase activities (<5 Miller units) were observed in cells harbouring the inactive P_{yefM2} with *yefM_{Spn}* (i.e., pQF_nPyM) but the additional presence of the *yoeB_{Spn}* reading frame (in pQF_nPyMyB) led to β -galactosidase activity levels of 132 \pm 24.4 Miller units (Figure 40). These results appeared to suggest the possible presence of a functional promoter within the *yoeB_{Spn}* reading frame.

To determine if this was the case, two pQF52-derived recombinant plasmids were constructed: pQF_yB which contained the entire *yoeB_{Spn}* reading frame cloned

upstream of the promoter-less *lacZ* reporter gene, and pQF_CyB which contained DNA encoding just the C-terminal portion of YoeB_{Spn} (159 bp upstream of the *yoeB_{Spn}* stop codon). *E. coli* cells harbouring pQF_yB and pQF_CyB displayed β -galactosidase activities of 119 ± 0.1 and 156 ± 1.7 Miller units, respectively (Figure 40), indicating a strong likelihood for the presence of a functional promoter within the *yoeB_{Spn}* reading frame, particularly within the last 159 nucleotides. Sequence analysis of the *yoeB_{Spn}* reading frame led to the identification of a putative σ^{70} -type promoter within this section: an imperfect -35 region (5'-TaGgaA-3') which is 17 nucleotides apart from the putative -10 region (5'-TATgAT-3') (Figure 41a). Furthermore, a putative ribosome binding site (5'-AGGAGc-3') was also located 5 nucleotides downstream of the -10 sequence along with an alternative start codon GTA (encoding valine; Kim *et al.*, 2007) 11 nucleotides downstream of the ribosome binding site. If this sequence is translated, the resulting peptide is only seven amino acid residues, as a TGA stop codon is identified six amino acids downstream of this alternative start codon (Figure 41a Frame 1). No homologous sequence was obtained in the NCBI database for this short peptide. Another two possible start codons (both were ATG that encodes methionine) were located 40 and 43 nucleotides downstream of the putative ribosome binding site and were in frame with *yoeB_{Spn}*, and could therefore possibly lead to the translation of the last 16 or 15 amino acids of the C-terminal of YoeB_{Spn}, respectively (Figure 41a Frame 2). The optimal spacing between the ribosome binding site and the AUG initiation codon varies from 5 to 13 nucleotides (Chen *et al.*, 1994). However, high levels of protein expression have been observed in constructs with the spacer between the ribosome binding site and the AUG initiation codon of approximately 40 nucleotides (Berwal *et al.*, 2010). Hence, whether any of the short peptides is actually translated, and from which possible start codons it is initiated (if any), remains to be determined. Nevertheless, no other possible start codons could be identified in the other reading

frame. The possibility of this putative promoter to regulate the gene downstream of *yoeB_{Spn}* was also examined. The reading frame (*spr1584* from R6 strain) downstream of *yoeB_{Spn}* encodes a hypothetical protein which has high similarity to phosphodiesterase/phosphohydrolase. However, this gene is located ~200 nucleotides apart from the putative promoter and therefore it is not likely to be transcribed by the internal promoter within the C-terminal region of the *yoeB_{Spn}* gene. In addition, no positive BLASTN or BLASTX result was obtained for the sequences between the *yoeB_{Spn}* stop codon and the start codon of *spr1584*, which indicates that these nucleotides may not encode for any known functional gene. Even if these sequences are translated, a few stop codons are encountered within the 3 possible reading frames (Figure 41c). Moreover, an RNA hairpin followed by a stretch of poly-U (five consecutive of 'U' residues) was predicted immediately after the *yoeB_{Spn}* stop codon, which indicates the presence of a rho-independent terminator right after the *yoeB_{Spn}* transcript (Figure 41b). Formation of an RNA hairpin loop pauses the transcribing RNA polymerase and the low stability between the DNA template and transcribed RNA at the rU-dA hybrid allows the dissociation of the RNA polymerase from the DNA strand (Farnham and Platt, 1980; Martin and Tinico, 1980; Arndt and Chamberlin, 1988). The possibility of this RNA serving as a regulatory antisense RNA was also considered. However, no gene was annotated on the opposite strand of this RNA transcript, and no sequence homology was detected with any other portion of *yefM-yoeB_{Spn}* or any other sequence in the NCBI database. Although the *yoeB_{Spn}* internal promoter is indeed functional in *E. coli*, whether it is similarly functional in its native *S. pneumoniae* host and whether the transcribed sequences are ultimately translated has yet to be determined.

If this internal promoter of *yoeB_{Spn}* is functional in *E. coli*, its promoter activity should be reflected in the constructs that harboured the *yoeB_{Spn}* reading frame, as in

pQF_yB, pQF_CyB, pQF_P1yMyB and pQF_nPyMyB (Figure 38). However, this internal promoter did not seem to affect the promoter activity of pQF_P2yMyB (Figure 35), as the promoter activity of P_{yefM2} was reduced by 87% in the presence of *yefM-yoeB_{Spn}* in both *trans* (Figure 37) and *cis* (Figure 38). In other words, if the internal promoter of *yoeB_{Spn}* was functional constantly, higher promoter activity should be detected when the *yoeB_{Spn}* reading frame was present in *cis*, which was not observed in pQF_P2yMyB. Furthermore, in cells that harboured plasmid constructs in which the translation of *yefM_{Spn}* was abolished, the presence of the *yoeB_{Spn}* reading frame (pQF_M1yMyB, 248 ±62.8 Miller units; pQF_M2yMyB, 252 ±24.7 Miller units) had only mere differences in the β-galactosidase activities when compared with the constructs without the *yoeB_{Spn}* reading frame (pQF_M1yM, 231 ±50.6 Miller units; pQF_M2yM, 228 ±19.0 Miller units) (Figure 38). Taken together, the results appeared to indicate that when the P_{yefM2} promoter is functioning, transcription from the internal promoter within *yoeB_{Spn}* is not apparent.

Plasmid Constructs	β-galactosidase activities	
	Miller units	Ratio
pQF_P1	8 ±0.9	1.00
	9 ±1.4	1.13
	150 ±10.2	18.75
pQF_nP	<5	-
	<5	-
	132 ±24.4	-
pQF_yB	119 ±0.2	-
pQF_CyB	156 ±1.7	-

Figure 40: β-Galactosidase activities of *E. coli* TOP 10 cells carrying the following pQF52-derived recombinant plasmids: pQF_P1, pQF_P1yM, pQF_P1yMyB, pQF_nP, pQF_nPyM, pQF_nPyMyB, pQF_yB and pQF_CyB. Ratios were calculated by normalization of the β-galactosidase activity levels of cells that harboured pQF_P1yM and pQF_P1yMyB against activity levels of cells that harboured pQF_P1 (taken as 1.00). Grey crosses indicate mutation of the -10 region of P_{yefM2} from 5'-TATAAT-3' to 5'-ctgcAg-3' thereby rendering P_{yefM2} inactive, as indicated by the pQF_nP β-galactosidase activity levels of < 5 Miller units.

While *E. coli* cells harbouring the various recombinant plasmids were grown for β -galactosidase assays, growth retardation was observed in uninduced cells that harboured the pLNBAD-derived recombinant plasmid pLN_yMyB which encompassed *yefM_{Spn}* along with *yoeB_{Spn}* but this retardation was not apparent in arabinose-induced cells (Figure 42). The uninduced cells harbouring pLN_yM which contained just the *yefM_{Spn}* reading frame in pLNBAD showed a similar non-retarded growth curve as the arabinose-induced cells (Figure 43). It would thus appear that in uninduced cells harbouring pLN_yMyB, some *yoeB_{Spn}* toxin gene may have been expressed. However, this could not have been due to a leaky P_{BAD} promoter since if that was the case, the *yefM_{Spn}* antitoxin gene would also be co-expressed (as P_{BAD} was located immediately upstream of *yefM-yoeB_{Spn}*) and this would have counteracted the toxic effect of YoeB_{Spn}. Could this growth retardation be perhaps due to the presence of the internal promoter within the *yoeB_{Spn}* reading frame which leads to the expression of the C-terminal portion of YoeB_{Spn}? This is not likely as cells that harboured pQF52 plasmids encompassing *yoeB_{Spn}* (pQF_yB) and DNA encoding the C-terminal half of YoeB_{Spn} (pQF_CyB) showed no growth retardation. Thus, even if the C-terminal region of YoeB_{Spn} was expressed due to the presence of the internal promoter, the C-terminal half of YoeB_{Spn} may not be toxic to the host cells.

The growth retardation of the cells harbouring pLN_yMyB suggested that the YoeB_{Spn} toxin was perhaps expressed when the cells were not induced by arabinose. Since this expression was not likely driven by the internal promoter within the *yoeB_{Spn}* reading frame, it can be inferred that there may be another promoter upstream of *yoeB_{Spn}*, i.e., within the *yefM_{Spn}* reading frame. However, β -galactosidase assay results were not indicative of the presence of a functional promoter within the *yefM_{Spn}* reading frame as non-detectable β -galactosidase activities (i.e., <5 Miller units) were observed

for cells that harboured pQF_yM which encompassed the entire *yefM_{Spn}* reading frame as well as pQF_CyM which spanned only the C-terminal portion of *yefM_{Spn}* (Figure 38).

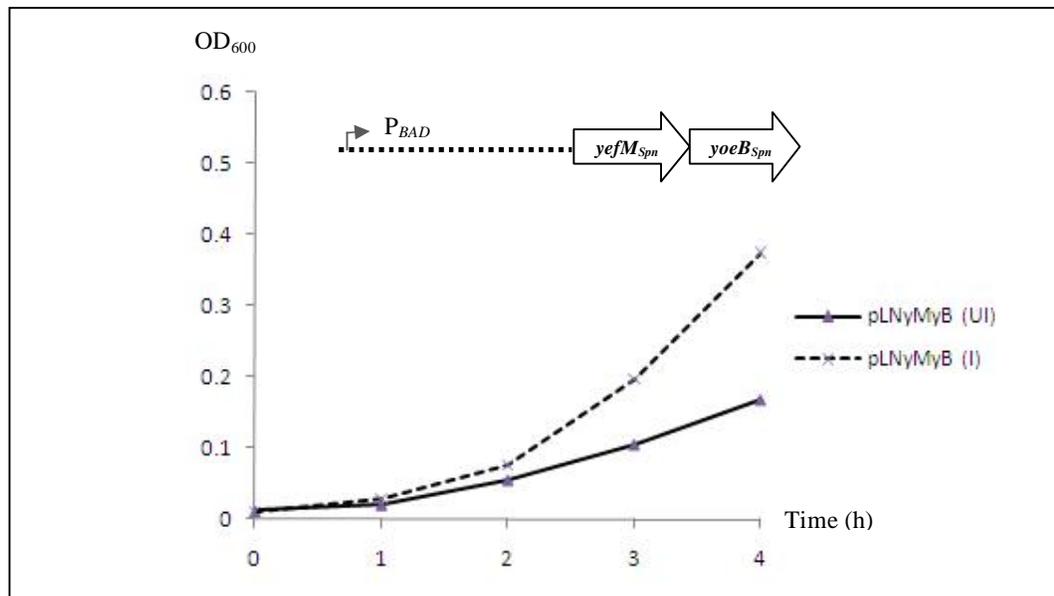


Figure 42: Growth curve of *E. coli* TOP10 cells harbouring the pLNBAD-derived recombinant plasmid pLNyMyB encompassing the *yefM-yoeB_{Spn}* reading frames cloned downstream of the arabinose-inducible P_{BAD} promoter of the vector. The cells were grown with starting OD₆₀₀ ~0.02. I: cells induced with L-arabinose at a final concentration of 1 mM at 0 h; UI: uninduced cells.

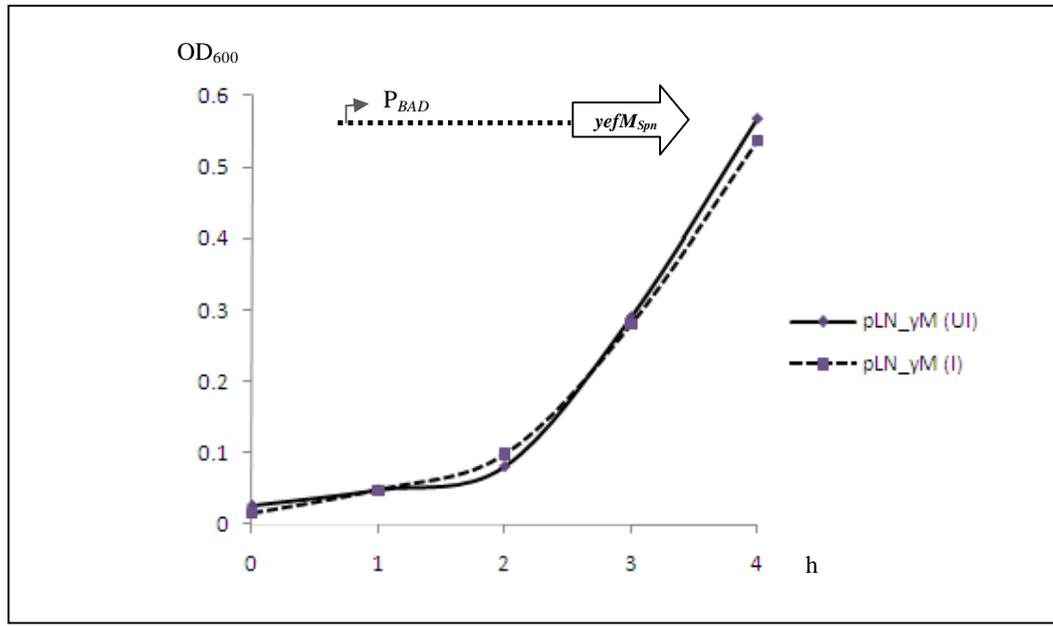


Figure 43: Growth curve of *E. coli* TOP10 cells harbouring the pLNBAD-derived recombinant plasmid pLN_yM encompassing the *yefM_{Spn}* reading frame inserted downstream of the arabinose-inducible P_{BAD} promoter of the vector. The cells were grown with a starting OD_{600} ~0.02. I: cells induced with L-arabinose at a final concentration of 1 mM at 0 h; UI: uninduced cells.

Intriguingly, when cells that harboured pQF_CyM were grown on Luria-Bertani agar supplemented with X-gal at a final concentration of 80 $\mu\text{g/ml}$, light blue colonies were observed when compared to cells that carried the pQF52 negative control plasmid (Figure 44). The blue colouration was certainly not as intense as that of cells that harboured the functional P_{yefM2} promoter in the pQF_P2 construct. Subsequent sequence analysis indicated the presence of a plausible σ^{70} -like promoter sequence within the reading frame that encodes the C-terminal half of YefM_{Spn}: The putative -10 region (5'-aATAAg-3') was located 81 bp upstream of the *yoeB_{Spn}* start codon whereas the putative -35 region (5'-agGAaA-3') was 17 nucleotides apart from the -10 region. A putative ribosome-binding site (5'-AGGAGt-3') was also identified six nucleotides upstream of the *yoeB_{Spn}* start codon (designated 'rbs2' in Figure 45). These observations seemed to suggest the possibility of a very weak promoter within the *yefM_{Spn}* reading frame which was not detectable by β -galactosidase assays but nevertheless was

sufficient to express YoeB_{Spn} in quantities that was able to moderately inhibit growth of cells harbouring pLN_yMyB.

If the region of the *yefM_{Spn}* reading frame encoding the C-terminal portion of YefM_{Spn} truly harbours a functional, albeit very weak promoter, some growth retardation would be observed as well in cells that harboured the *yefM-yoeB_{Spn}* reading frames along with the upstream regulatory region but with stop codons introduced into the *yefM_{Spn}* reading frame (i.e., pQF_M1yMyB and pQF_M2yMyB; see Figure 38). This is because no functional YefM_{Spn} antitoxin would be expressed from these constructs to counteract the YoeB_{Spn} toxin expressed from the putative internal promoter within *yefM_{Spn}*. However, growth retardation was not observed in cells that harboured either pQF_M1yMyB or pQF_M2yMyB. In pQF_M1yMyB, the 7th codon of YefM_{Spn} was replaced with a stop codon whereas in pQF_M2yMyB, stop codons were introduced at the 27th, 32nd and 33rd codons of YefM_{Spn}. Interestingly, a putative ribosome binding site, 5'-AtGAGG-3' (designated "rbs1" in Figure 45) could be located four nucleotides upstream of an alternative start codon, GTA, encoding valine which is the 39th codon of YefM_{Spn}. It is thus possible that despite the presence of three stop codons in the *yefM_{Spn}* reading frame in pQF_M2yMyB, a truncated C-terminal YefM_{Spn} may still be expressed from the GTA alternative start codon due to the presence of the internal ribosome binding site.

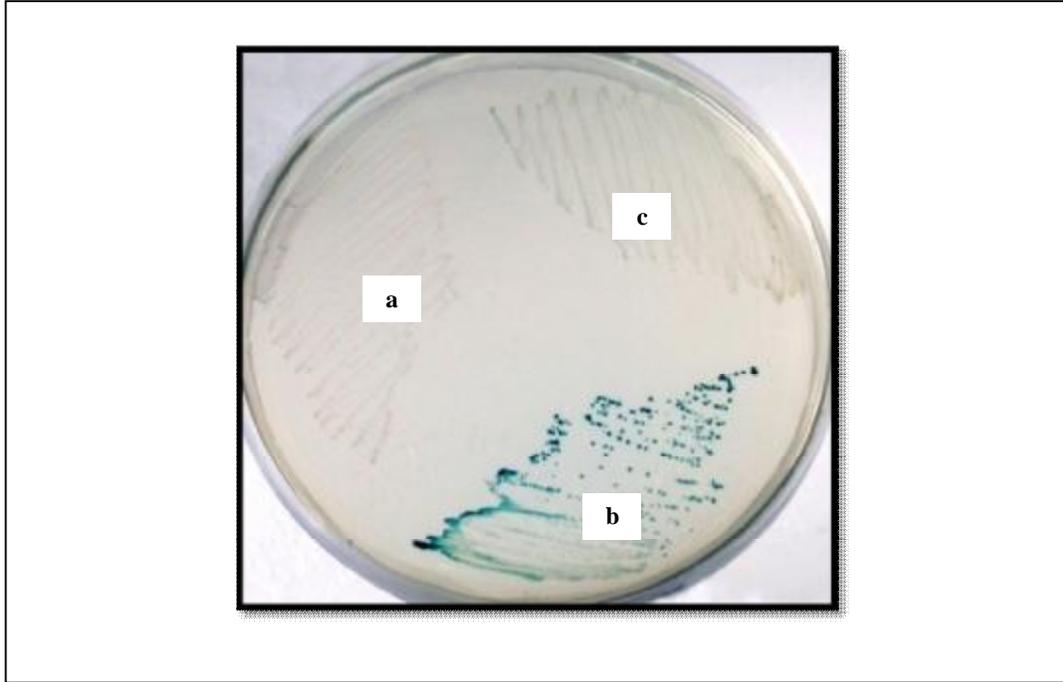


Figure 44: *E. coli* cells on Luria-Bertani agar supplemented with 100 µg/ml of Ampicillin and 80 µg/ml of X-gal. (a) Cells harbouring the pQF52 plasmid as a negative control; (b) pQF_P2 plasmid containing the functional P_{yefM2} promoter as a positive control; and (c) pQF_CyM. Note the light blue colouration of cells in (c) indicative of low amounts of β -galactosidase produced.

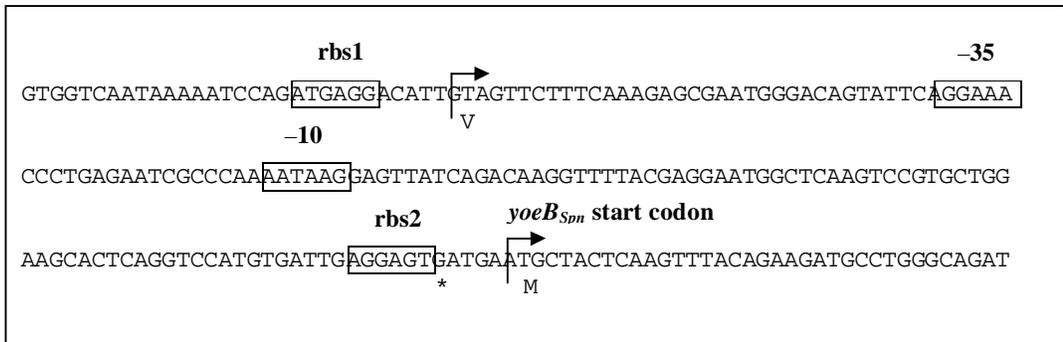


Figure 45: Nucleotide sequence of the *yefM_{Spn}* reading frame encoding the C-terminal portion of YefM_{Spn}. The putative -35 and -10 regions of the σ^{70} -like promoter sequence are indicated along with two putative ribosome-binding sites, designated 'rbs1' and 'rbs2'. The stop codon for *yefM_{Spn}* is indicated by an asterisk, '*'. A GTA alternative start codon (encoding Valine) within the *yefM_{Spn}* reading frame is indicated as 'V', as is the 'ATG' start codon for *yoeB_{Spn}*.

To ascertain if this C-terminal portion of YefM_{Spn} is sufficient to counteract the toxicity of YoeB_{Spn}, a pLNBAD-derived recombinant plasmid encompassing the last 159 bp of the *yefM_{Spn}* reading frame (i.e., from the 34th amino acid residue of YefM_{Spn} and thus containing the putative ribosome-binding site rbs1 as well as the GTA alternative start codon) along with the *yoeB_{Spn}* reading frame, was constructed. Under

non-induced conditions, cells that harboured the resulting pLN_CyMyB plasmid showed slight growth retardation whereas cells induced with arabinose displayed a normal growth profile (Figure 46) as was earlier observed for pLN_yM (Figure 43). This indicated that the C-terminal portion of YefM_{Spn} is sufficient to counteract the toxicity of the YoeB_{Spn} toxin.

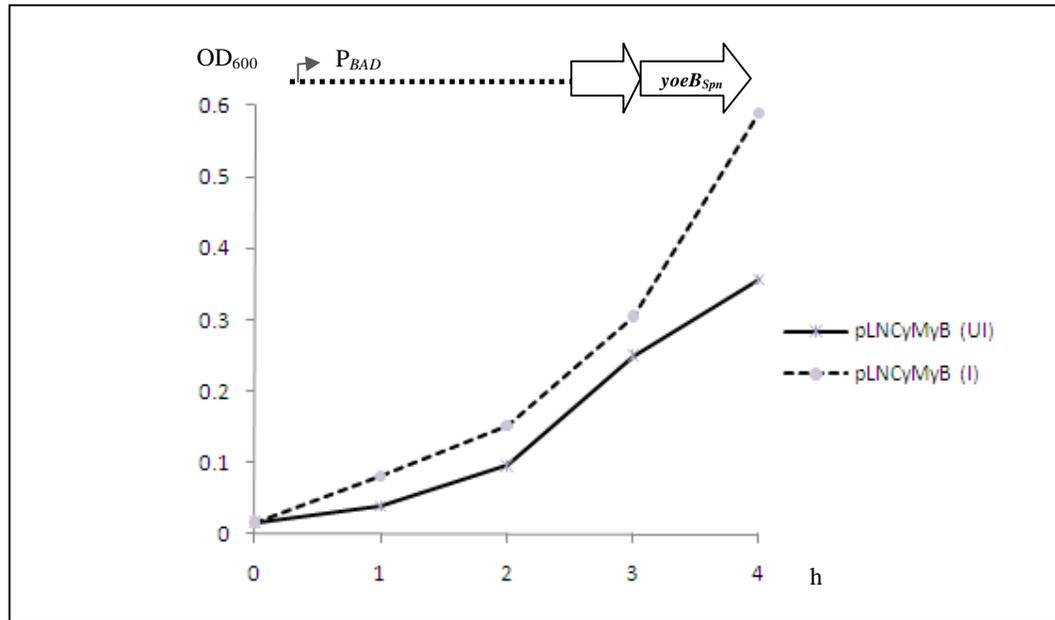


Figure 46: Growth curve of the *E. coli* TOP10 cells harbouring pLNBAD-derived recombinant plasmid containing the DNA fragment encoding amino acid residues 34 – 84 of YefM_{Spn} and the entire *yoeB_{Spn}* reading frame inserted downstream of the arabinose-inducible *P_{BAD}* promoter (pLN_CyMyB). The cells were grown with a starting OD₆₀₀ = 0.02 at h = 0. I: cells with L-Arabinose at a final concentration of 1 mM added at h = 0; UI: cells not induced with L-arabinose.

To further investigate if the C-terminal of YefM_{Spn} is also sufficient to autoregulate transcription of the *yefM-yoeB_{Spn}* operon, the pLN_CyMyB recombinant plasmid was co-transformed together with each of the pQF52-derived recombinant plasmids harbouring the various upstream promoter fragments, i.e., pQF_P1P2, pQF_P2, pQF_P1 and pQF_nP, respectively, into *E. coli* TOP10 cells. Results of β-galactosidase assays indicated no prominent repression for all these constructs when the cells were induced with arabinose when compared to the uninduced cells (Figure 47). This is in stark contrast to the repression observed when full length YefM_{Spn} and

YoeB_{Spn} were provided in *trans* to the upstream regulatory region (Figure 37). Hence, although the C-terminal residues 34 – 84 of YefM_{Spn} were sufficient to counteract the toxicity of YoeB_{Spn}, they were unable to autoregulate the transcription of the *yefM-yoeB_{Spn}* operon.

Plasmid constructs	β-galactosidase activities	
	Miller units	Ratio
pQF_P1P2 & pLN_CyM yB	UI 38 ±4.4	1.00
pLN_CyM yB	I 36 ±4.4	0.95
pQF_P2 & pLN_CyM yB	UI 54 ±9.3	1.00
pLN_CyM yB	I 55 ±16.9	1.02
pQF_P1 & pLN_CyM yB	UI 5 ±1.4	-
pLN_CyM yB	I <5	-
pQF_nP & pLN_CyM yB	UI <5	-
pLN_CyM yB	I <5	-

Figure 47: β-Galactosidase activities of *E. coli* TOP 10 cells carrying the pLNBAD-derived recombinant pLN_CyMyB plasmid and each of the following pQF52-derived recombinant plasmids harbouring various upstream regulatory fragments: pQF_P1P2, pQF_P2, pQF_P1 and pQF_nP. β-galactosidase activities were measured when cells harbouring pLN_CyMyB and the relevant pQF-derived recombinant were either induced with 1 mM L-arabinose (I) or uninduced (UI). Ratios were calculated by normalization of the β-galactosidase activity of the induced cells against the uninduced cells for the respective pair of plasmids. Grey crosses indicate mutation of the -10 region of P_{yefM2} from 5'-TATAAT-3' to 5'-ctgcAg-3' thereby rendering P_{yefM2} inactive.

3.2.7 Determination of the DNA binding site of the YefM_{Spn} antitoxin and the YefM-YoeB_{Spn} TA complex

3.2.7.1 Optimization of EMSA condition and the length of DNA fragment for footprinting assays

A 322 bp [γ -³²P]ATP labeled PCR-amplified DNA fragment encompassing both the boxA-C element and palindrome sequence upstream of *yefM-yoeB_{Spn}* was incubated with increasing amounts of purified (His)₆-YefM-YoeB_{Spn} protein complex with or without 10 ng/ μ l heparin (as non-specific competitor). The unbound DNA fragments and the nucleoprotein complexes were separated by electrophoresis in 5% polyacrylamide gels. Results showed that 10 ng/ μ l heparin had significantly improved the binding of the (His)₆-YefM-YoeB_{Spn} protein complex to the labeled DNA (Figure 48). Retardation of labeled DNA was observed when 0.01 μ g of protein complex was added (Figure 48). This suggested that the (His)₆-YefM-YoeB_{Spn} protein complex did bind to the DNA fragment containing boxA-C element and palindrome sequence. Nevertheless, the binding of the protein was likely at one site of the DNA as only one band of nucleoprotein complexes was observed when the protein (0.10 μ g) was added until a majority of the DNA fragment had been shifted (Figure 48). The same EMSA reaction mixture was then prepared, but was scaled up to a final volume of 50 μ l and subjected to DNaseI footprinting assay. Preliminary observations indicated that the binding site for the (His)₆-YefM-YoeB_{Spn} protein complex was at the palindrome sequence overlapping the -35 region of the P_{*yefM2*} promoter with no binding detected at other nucleotides including within the boxA-C element. Similar results were obtained with (His)₆-YefM_{Spn}. Consequently, another set of primers (PS2-F and PS2-R) were designed to amplify the palindrome sequence flanked by 80 nucleotides to produce the optimal length DNA fragment for footprinting assays.

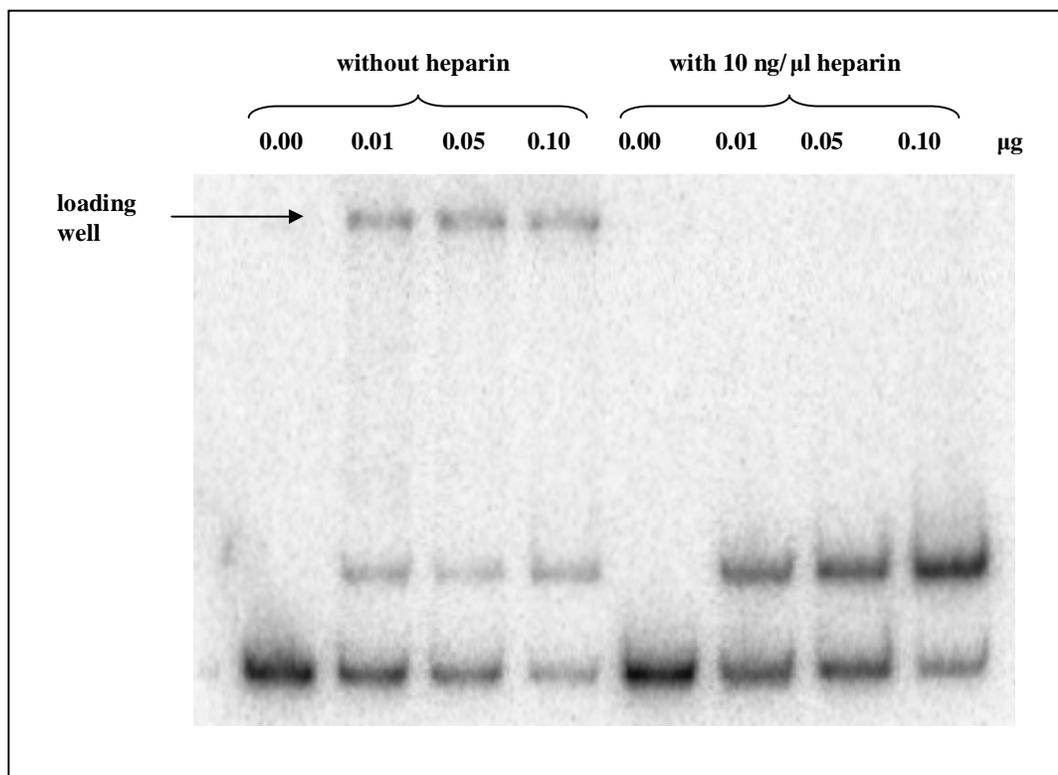


Figure 48: EMSA showing *in vitro* binding of increasing amounts (from left to right: 0.00, 0.01, 0.05 and 0.10 μg) of purified YefM-YoeB_{Spn} protein complex with 322 bp of [γ -³²P]ATP labeled DNA fragment (3000 cpm per lane) encompassing the boxA-C element and palindrome sequence, with (lanes 5 to 8) and without (lanes 1 to 4) 10 ng/ μl of heparin. The unbound DNA fragments and the nucleoprotein complexes were separated by electrophoresis in 5% polyacrylamide gels.

3.2.7.2 EMSA

A 284 bp of [γ -³²P]ATP-labeled DNA fragment containing the palindrome sequence was PCR-amplified using the PS2-F and PS2-R primer pair. The palindrome sequence-containing DNA fragment (3,000 cpm) was incubated with increasing amounts of either purified YefM_{Spn} fusion protein or the YefM-YoeB_{Spn} fusion protein complex in the EMSA reaction mixture containing 10 ng/ μl of heparin. With YefM_{Spn}, retardation of the DNA fragment was detected when 3.00 μg of YefM_{Spn} protein was added into the reaction and most of the DNA fragment was retarded or shifted when 5.00 μg of the YefM_{Spn} protein was included. However for the YefM-YoeB_{Spn} protein complex, even as little as 0.05 μg of the protein complex resulted in retardation of

>70% of the palindrome sequence-containing DNA fragment and most of the DNA fragment was shifted when 0.25 μg of the protein complex was added (Figure 49).

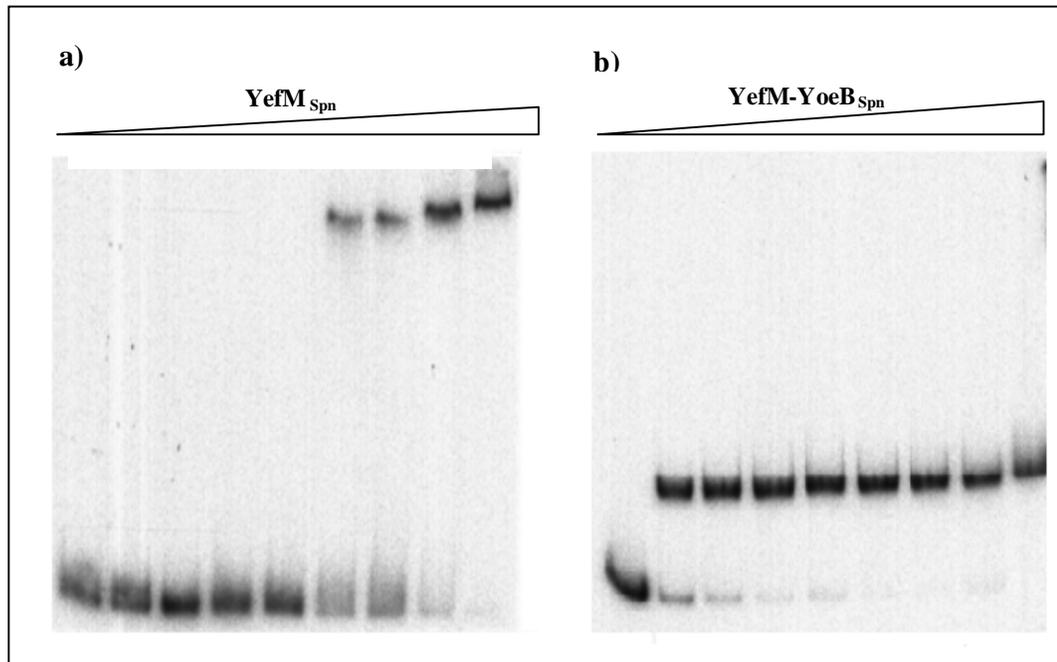


Figure 49: EMSA showing *in vitro* binding of purified YefM_{Spn} protein and the YefM-YoeB_{Spn} protein complex with the 284 bp of [γ -³²P]ATP labeled DNA fragment (3000 cpm per lane) encompassing the PS palindrome sequence upstream of *yefM-yoeB_{Spn}* with 10 ng/ μl of heparin. Increasing amounts of purified (a) YefM_{Spn} protein was added from left to right: 0.00, 1.00, 1.50, 2.00, 2.50, 3.00, 3.50, 4.00, 5.00 and 10.00 μg ; (b) YefM-YoeB_{Spn} protein complex was added from left to right: 0.00, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.40, 0.50 and 1.00 μg . The unbound DNA fragments and the nucleoprotein complexes were separated by electrophoresis in 5% polyacrylamide gels.

3.2.7.3 DNase I footprinting assays

The 284 bp DNA fragment encompassing the PS palindrome sequence was labeled with [γ -³²P]ATP (30,000 cpm) at either the coding or non-coding strand and subsequently incubated with 0, 10, 20 and 50 μg of purified YefM_{Spn} protein or 0, 1, 2 and 5 μg of purified YefM-YoeB_{Spn} protein complex, respectively, and with 10 ng/ μl of heparin. The reactions were subjected to DNase I digestion before loading onto an 8% polyacrylamide gel containing 7 M urea along with the DNA sequencing ladder prepared using the dideoxy sequencing reaction as mentioned in the Materials and Methods (Section 2.10.2). For YefM_{Spn}, one region of the coding-strand was protected

by the DNaseI digestion, which was 20 nucleotides in length and located within the PS palindrome sequence that overlapped the whole -35 element of the P_{yefM2} promoter (Figure 50a and Figure 51a). A similar protected region was observed for the non-coding strand, but was located two nucleotides downstream and was extended for another two nucleotides (Figure 50b and Figure 51a). Hypersensitivity was observed for three nucleotides located between the -35 and -10 elements on the coding stand (Figure 50a and Figure 51a) when $YefM_{Spn}$ was added from which it was inferred that DNA bending may occur at this point (based on the similar prediction for the DNase I hypersensitivity observed around the region of the transcriptional start site induced by the TATA-binding protein by Coulombe and Burtun, 1999). For $YefM$ - $YoeB_{Spn}$, two shorter protected regions were identified instead on the coding strand, which was also located within the PS palindrome sequence and covered the entire -35 element of the P_{yefM2} promoter (Figure 50c and Figure 51b). Hypersensitivity was also detected exactly the same positions on the coding stand as observed in footprinting where $YefM_{Spn}$ alone was added (Figure 50c and Figure 51b). For the non-coding strand, one region comprising 23 nucleotides was protected further downstream of the PS palindrome sequence (Figure 50d and Figure 51b). Thus, the PS palindrome sequence likely constitutes the operator site for the autoregulation of the $yefM$ - $yoeB_{Spn}$ locus. Nevertheless, the protected regions were slightly more extended in the presence of the $YefM$ - $YoeB_{Spn}$ complex as compared to the $YefM_{Spn}$ antitoxin alone. These, along with the EMSA results, indicated that the $YefM$ - $YoeB_{Spn}$ TA complex had a higher binding affinity than the $YefM_{Spn}$ antitoxin to the operator site of $yefM$ - $yoeB_{Spn}$ locus.

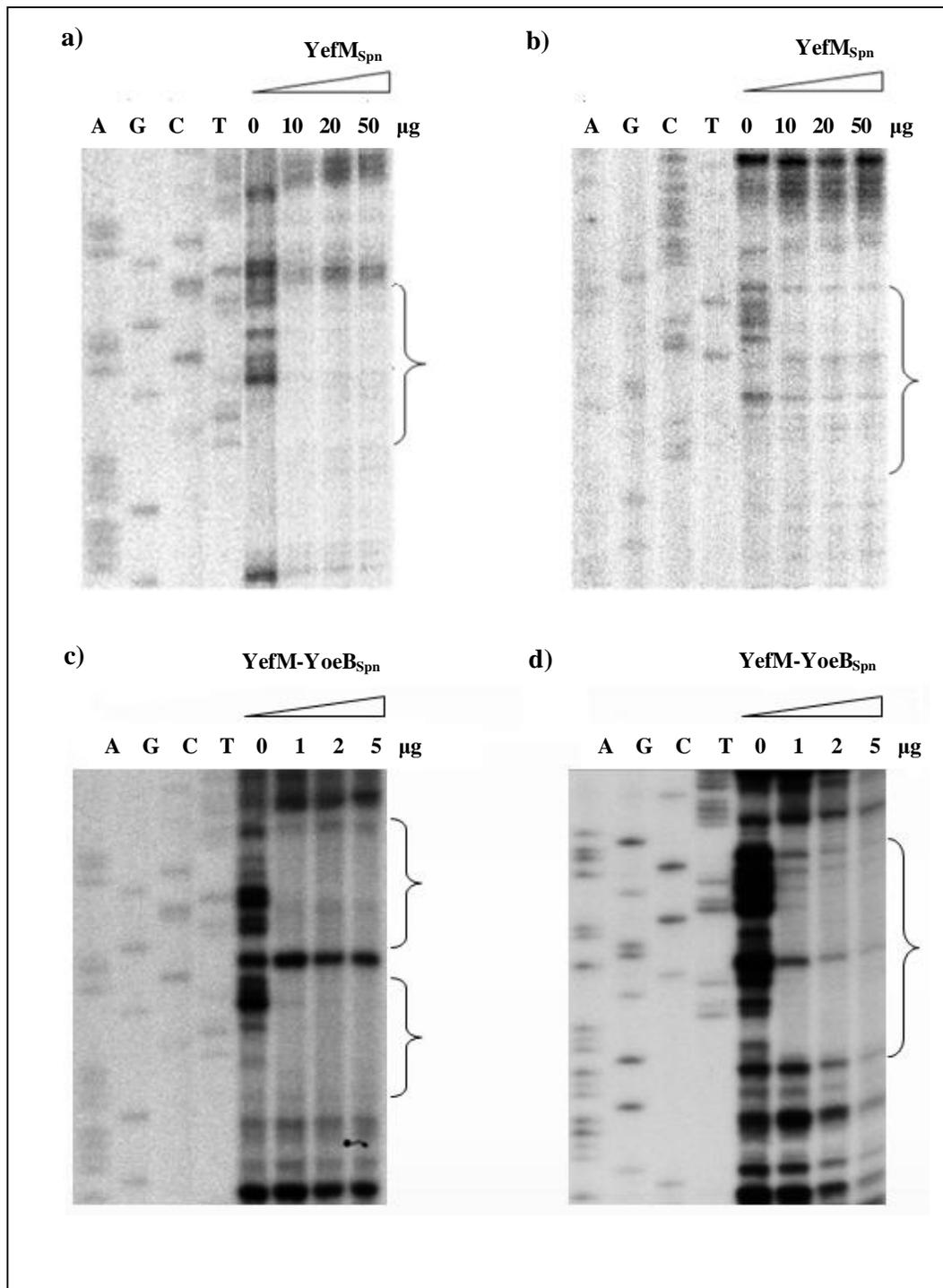


Figure 50: DNaseI footprinting assays for the purified YefM_{Spn} and YefM-YoeB_{Spn} proteins on the DNA fragment containing the PS palindrome sequence. Both coding strands (a and c) and non-coding strands (b and d) of [γ -³²P]ATP labeled DNA fragments containing the PS palindrome (30,000 cpm) were incubated with increasing amounts of (a and b) YefM_{Spn} protein at 0, 10, 20 and 50 µg or (c and d) YefM-YoeB_{Spn} protein complex at 0, 1, 2 and 5 µg, prior to DNaseI digestion. The reaction was then separated on a 8% polyacrylamide gel containing 7 M urea along with the DNA sequencing ladder prepared using the dideoxy sequencing reaction. Brackets () indicated the nucleotide sequences protected by YefM_{Spn} or YefM-YoeB_{Spn} proteins from DNase I digestion.

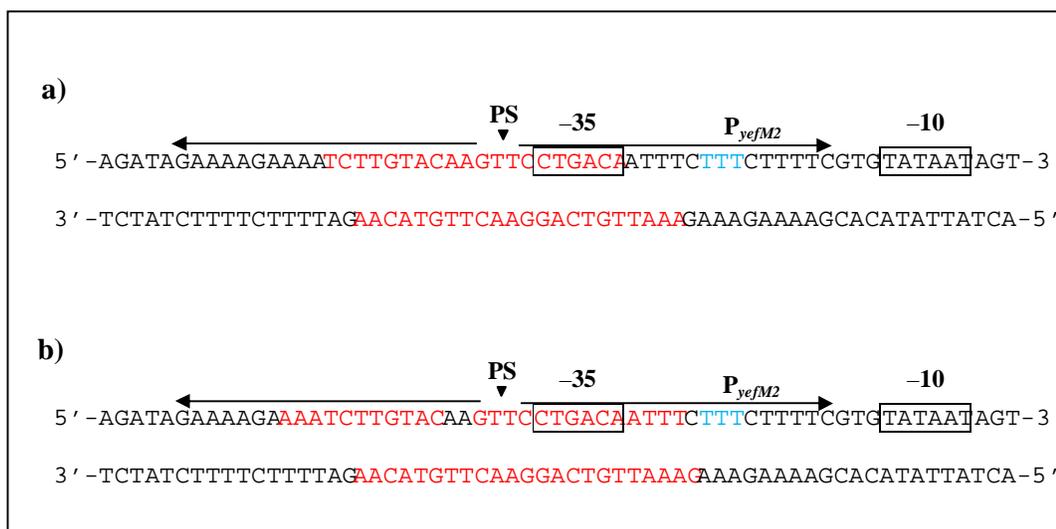


Figure 51: Nucleotide sequence of the coding and non-coding DNA strands of the *yefM_{Spn}* upstream region containing the PS palindromic sequence. The -10 and -35 regions of the P_{yefM2} promoter are indicated within boxes. The inverted triangle (\blacktriangledown) indicates the centre of the palindrome. The nucleotides protected from DNase I digestion by YefM_{Spn} (a) and YefM-YoeB_{Spn} (b) proteins are highlighted in red. The nucleotides with hypersensitivity are indicated in blue.

3.2.7.4 Hydroxyl radical footprinting assays

To further examine the DNA binding sites for YefM_{Spn} and the YefM-YoeB_{Spn} proteins, 30,000 cpm of the 284 bp [γ -32P]ATP labeled DNA fragment containing the PS palindromic sequence for both coding and non-coding strands were incubated with or without 50 μ g of the purified YefM_{Spn} protein or 5 μ g of the purified YefM-YoeB_{Spn} protein complex. The reaction was then subjected to hydroxyl radical reagent mixture as detailed in the Materials and Methods (Section 2.10.3). Two areas within the palindromic sequence (PS) of the coding strand were protected from the hydroxyl radical attack by YefM_{Spn}, and the distance between the first nucleotides of both areas was 8 nucleotides (Figure 52a and Figure 53a). Similarly, two areas with a distance of 8 nucleotides between the first nucleotides of both areas were detected for the non-coding strand, but were located 9 nucleotides downstream of the PS palindromic sequence when compared to the coding strand (Figure 52a and Figure 53a). For YefM-YoeB_{Spn}, the two areas protected from hydroxyl radical attack were separated by 10 nucleotides (between

the first nucleotides of each area) for both coding and non-coding strands (Figure 52b and Figure 53b). For the coding strand, the protected area included the -35 element of PyefM2 promoter, which was not evident in the presence of YefM_{S_{pn}}. Nonetheless, both YefM_{S_{pn}} and YefM-YoeB_{S_{pn}} proteins were found to bind along one face of the DNA helix, thus leaving the DNA backbone on the opposite face exposed to hydroxyl radical attack (Hertzberg and Dervan, 1984) (Figure 54).

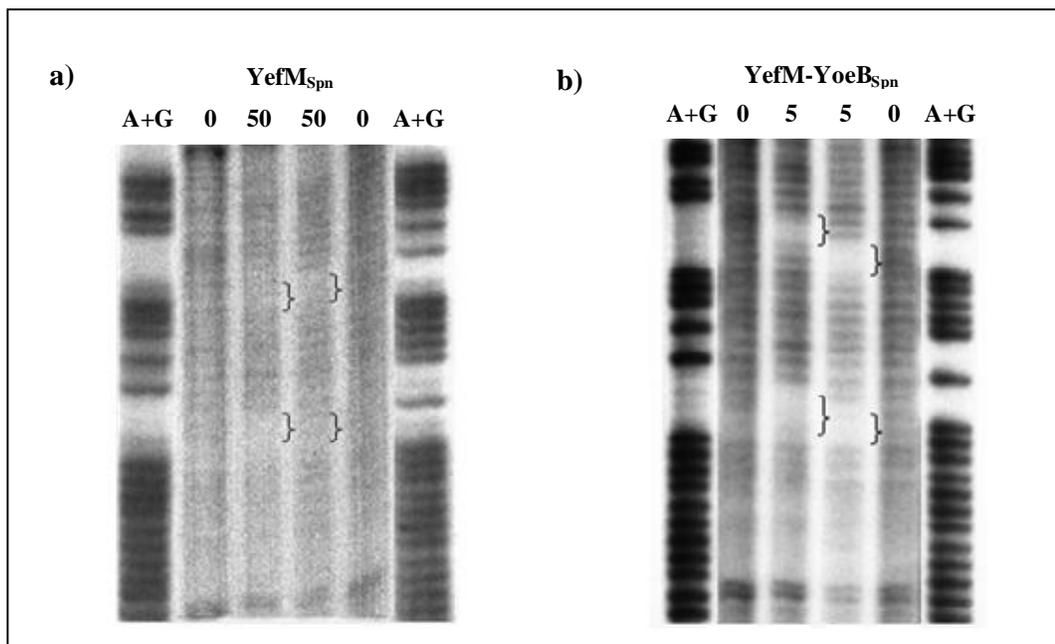


Figure 52: Hydroxyl radical footprinting assay for YefM_{S_{pn}} and YefM-YoeB_{S_{pn}} on the PS palindrome sequence. 30,000 cpm of both coding and non-coding strands of γ -³²P]ATP labeled DNA fragment containing the PS palindrome sequence were incubated with or without 50 μ g of purified YefM_{S_{pn}} protein (a) and with or without 5 μ g of purified YefM-YoeB_{S_{pn}} protein complex (b) prior to hydroxyl radical treatment. The reaction was then separated on 8% polyacrylamide gel containing 7 M urea along with a DNA sequencing ladder (A+G) prepared by the Maxam-Gilbert sequencing method. Brackets (}) indicated the nucleotide sequences that were protected by either the YefM_{S_{pn}} protein or the YefM-YoeB_{S_{pn}} protein complex from hydroxyl radical attack.

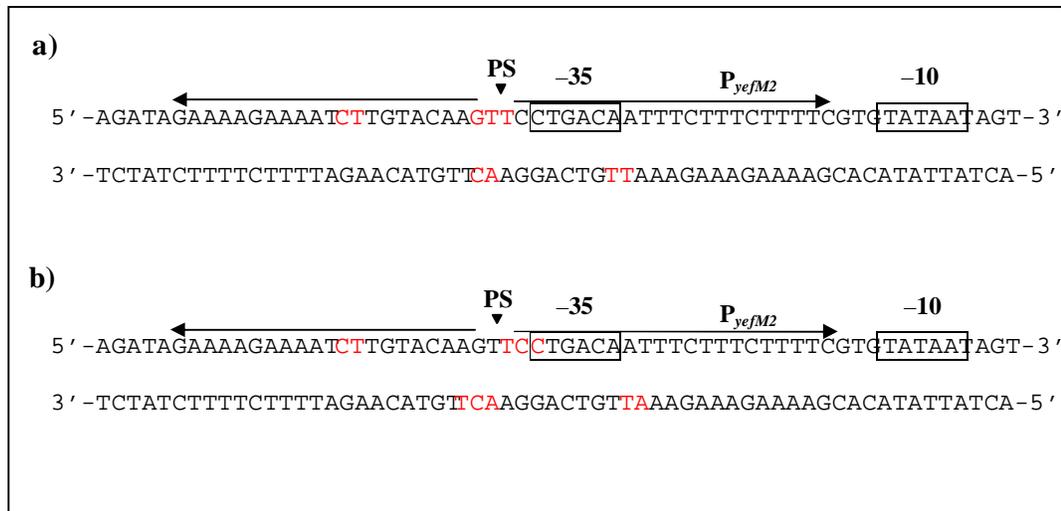


Figure 53: Nucleotide sequence of the coding and non-coding DNA strands of the *yefM_{Spn}* upstream region containing the PS palindrome sequence. The inverted triangle (▼) indicates the centre of the palindrome. The -10 and -35 regions of the P_{yefM2} promoter are indicated within boxes. The nucleotides protected by the YefM_{Spn} protein (a) and the YefM-YoeB_{Spn} protein complex (b) from hydroxyl radical attack are highlighted in red.

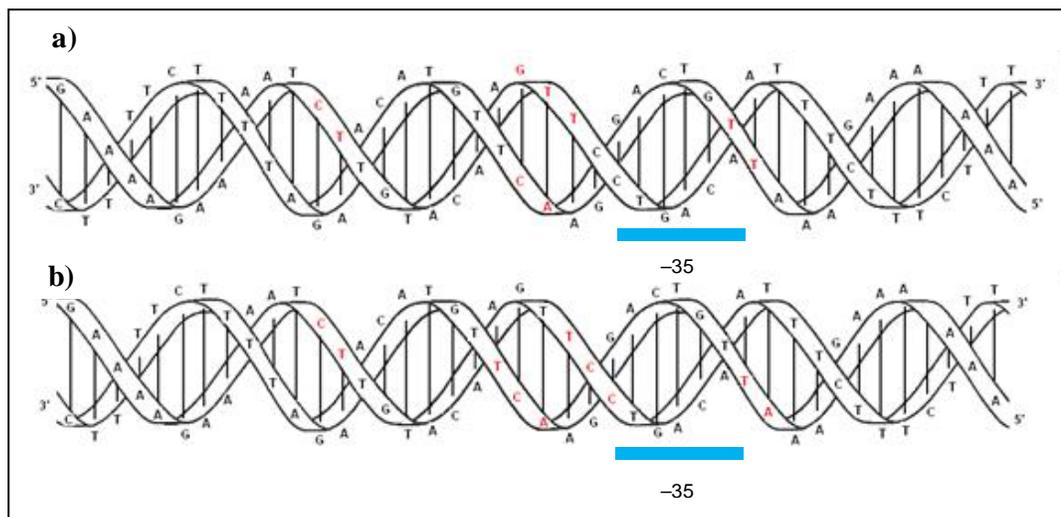


Figure 54: Schematic drawing of the DNA double helix of the *yefM_{Spn}* upstream region containing the PS palindrome sequence. Nucleotides protected by the YefM_{Spn} protein (a) and the YefM-YoeB_{Spn} protein complex (b) from hydroxyl radical attack are indicated in red. The -35 region of the P_{yefM2} promoter is indicated. The schematic drawing of the DNA double helix template was provided by Gloria del Solar at the Department of Molecular Microbiology and Infection Biology, Centro de Investigaciones Biológicas, Spain (Hernández-Arriaguet *et al.*, 2009).

4 Discussion

4.1 The *pezAT* TA locus of *S. pneumoniae*

The *pezAT* TA locus constitutes a bicistronic operon which is co-transcribed from a canonical σ^{70} promoter located upstream of *pezA* designated P_{pezA} . Like most proteic TA loci, *pezAT* negatively autoregulates its own promoter at the transcriptional level. The PezA antitoxin protein serves as a weak repressor to repress its own transcription and consequently its translation as shown in the β -galactosidase and real-time RT-PCR assays. Repression is more pronounced in the presence of the PezAT TA protein complex. From the EMSA (Section 3.1.7, Figure 22), the PezA antitoxin protein and the PezAT TA protein complex were demonstrated to bind to a DNA fragment that encompassed the P_{pezA} promoter and an incomplete palindrome sequence which overlaps the -10 and -35 elements of P_{pezA} . It is therefore likely that binding of the PezA protein and the PezAT protein complex occur at this palindromic region, thus hindering RNA polymerase from binding to the P_{pezA} promoter and resulting in repression of transcription. This would, however, require validation through footprinting assays to identify the exact location to which the PezA protein and the PezAT complex bind. Similar findings have been reported for other proteic TA loci (Gerdes *et al.*, 2005) such as the *higBA* locus where the HigA antitoxin protein acted as a repressor by binding and covering the -35 and -10 regions of the P_{hig} promoter (Tian *et al.*, 2001) thereby preventing RNA polymerase from accessing the promoter.

Sequence variation of the *pezAT* locus was observed in different *S. pneumoniae* strains (Section 3.1.2), however, no variation was observed within the phosphotransferase active site (GQSGAGKT) at the N-terminus of the PezT toxin. The highly conserved nature of this motif implied an important role in the functionality of PezT. The N-terminus of ζ , which contains the Walker A motif, is responsible for the toxicity of ζ in *E. coli* cells (Meinhart *et al.*, 2003; Zielenkiewicz *et al.*, 2009). Even

though this motif is present in many protein families with various functions (Zielenkiewicz and Ceglowski, 2005), the functional mechanism and the target substrate for either PezT or ζ remain to be elucidated. Although PezT is found to be very similar to ζ , its cognate antitoxin, PezA, only shares similarity with it at its C-terminus. Unlike other TA loci, the ε - ζ TA locus is regulated by a third component, the ω protein that is encoded upstream of ε . Although no ω -encoded sequence was found immediately upstream of *pezA*, there are extended amino acid sequences found at the N-terminus of PezA (158 amino acid residues) which is not found in ε (90 amino acid residues). These extended amino acid sequences comprise a Cro-like helix-turn-helix motif. Helix-turn-helix motifs are commonly found on prokaryotic DNA-binding proteins that function to regulate gene expression (Brennan and Matthews, 1989). The motifs found in Cro, CAP and X-repressor are very similar in sequence and structure, and are characterized by two α -helices. One helix contacts with the operator DNA via hydrogen bonds and hydrophobic interactions between the side chain of the protein and the exposed bases within the major groove of the DNA, whereas the other helix helps in stabilizing the DNA-protein structure (Brennan and Matthews, 1989). Not all the helix-turn-helix motifs interact in exactly the same way with the DNA, however, they adopt similar but distinctly different binding geometries with the DNA (Brennan and Matthews, 1989). Results from the *lacZ* transcriptional fusion experiments (Section 3.1.6.1) indicated that PezA likely functions as a transcriptional regulator like other TA loci but unlike ε - ζ where ε functions solely as the antitoxin with the regulatory role played by a third component, ω .

TA loci have been known to function in plasmid maintenance. It has also been postulated that TA loci contribute to stabilization of mobile elements or genomic islands. A thorough bioinformatics search showed that large SIs harboured TA loci but smaller SIs did not (Szekeres *et al.*, 2007). Five functional TA loci were identified in a

153 kb superintegron of *Vibrio vulnificus* CMCP6 (Szekeres *et al.*, 2007). A recombinant plasmid designated pWEBRNC::VvuSI carried a 36 kb fragment of the *V. vulnificus* 75.4T SI which was deleted of its TA loci. When the *intI1* integron-encoded integrase expression was induced, DNA losses of up to 30 kb were observed, whereas cassette losses were suppressed when the *relBE1* and *parDE1* TA loci were incorporated into the pWEBRNC::VvuSI plasmid. This indicated that the TA loci inhibited large-scale deletion of genes due to integrase-mediated recombination (Szekeres *et al.*, 2007). However, the TA loci did not preclude integrase-mediated microevolution from occurring as SI cassette relocation was observed in another study (Szekeres *et al.*, 2007). In addition, the *relBE1* and *parDE1* TA loci were also shown to stabilize a 165 kb dispensable region of *E. coli* (Szekeres *et al.*, 2007). The *pezAT* locus was identified within a 27 kb pathogenicity island of *S. pneumoniae*, termed pneumococcal pathogenicity island I (PPI1), which was thought to be acquired horizontally (Brown *et al.*, 2004). Whether *pezAT* could function to stabilize the PPI1 as the TA loci that were found in large SIs remains to be determined.

A global PCR analysis of the chromosomal structures of clinical *S. pneumoniae* isolates (21 Spanish isolates and 58 Polish isolates) as well as bioinformatic analysis of the genome sequences of 31 strains that were available at the NCBI Genome and at the Sanger Institute databases, showed that *relBE2Spn* was conserved in all isolates (Nieto *et al.*, 2010). However, analysis of the neighbouring genes led to the identification of three categories of gene arrangements: Type I had the genetic organization found as in TIGR4, D39 and R6 strains, type II was where the SP1222 and SP1221 ORFs were replaced by a cation channel protein-like family, whereas type III was similar to type II but contained an IS1167 transposon sequence insertion upstream of the promoter of the *relBE2Spn* operon (Nieto *et al.*, 2010). Sequencing analyses of several clinical isolates belonging to the three pneumococcal *relBE* types showed no variation in the antitoxin-

encoding *relB2Spn* gene. However, several nucleotide changes which led to silent mutations were identified in the toxin-encoding *relE2Spn* gene. In addition, two polymorphisms in which minor amino acid changes occurred were detected (T34I, and D39G). However, cells that harboured a recombinant plasmid encompassing this mutated *relE2Spn* still led to growth arrest, although the toxicity was lower than in the wild type. Changes were also observed in the region spanning the –35 and –10 region of the *relBE2Spn* promoter (A to G at position –28), a region which is probably involved in the transcriptional self-regulation of the operon (Nieto *et al.*, 2010). Although polymorphisms were found at the vicinity of this TA region, all the strains contained a functional *relBE2Spn* locus and the type of its structure correlated with the multilocus sequence type (Nieto *et al.*, 2010). Variation in gene organization (Brown *et al.*, 2004) and sequence variation (Khoo *et al.*, 2007) were also observed for the *pezAT* loci. A 11.3 kb block of genes (Sp1047 to Sp1063), which included Sp1051 (PezT), at the 3' region of PPI1 that was present in the TIGR serotype 4 strain, were missing in a serotype 17 strain (strain 142). This block of genes was flanked by direct repeats and therefore could be acquired or lost via horizontal gene transfer (Brown *et al.*, 2004). On the other hand, the unencapsulated R6 strain and the capsular serotype 19F, G54 strain, have unique complement of genes within the 3' end region of PPI1, which was designated the PPI1 variable region (Brown *et al.*, 2004). These findings suggest that PPI1 is a mosaic pathogenicity island with sets of conserved genes and Sp1051 (PezT) to Sp1053 were acquired at different times (Brown *et al.*, 2004). PCR of six *S. pneumoniae* clinical isolates also revealed that the *pezAT* sequences differed by 4-6% from the TIGR4 strain sequence (Khoo *et al.*, 2007). Most of these differences led to silent mutations; however, the translated PezA and PezT proteins differed by up to 3% compared to the TIGR4 strain (Khoo *et al.*, 2007). Due to a similar degree of variation found when comparing the TIGR4 strain with two other strains (R6 and D39), it was

thus suggested that the variation in sequence was due to natural inheritance (Khoo *et al.*, 2007).

The PPI1 pathogenicity island also contains *apiaABCD* locus which encodes an ABC transporter for iron uptake that is required for full virulence in mice (Brown *et al.*, 2001; 2002). Disruptions in Sp1051 (PezT) did not lead to growth retardation in broth, serum or blood but resulted in impaired virulence in mouse and inability to compete with a wild-type strain under both systemic and respiratory infection. This indicated that the reduced virulence is not due to the reduced growth rate (Brown *et al.*, 2004). Moreover, PCR and Southern hybridization showed that 33% of Sp1051 was undetected out of 26 strains tested and this indicates Sp1051 is not likely essential for virulence, but rather modulates the virulence of those strains which carry it (Brown *et al.*, 2004). Although *pezAT* may have a role in guarding against the loss of DNA, as in other TA systems found on large mobile elements, the relatively large percentage of strains (33%) in which *pezAT* was absent and the detection of an 11.3 kb deletion within PPI1 (Brown *et al.*, 2004) indicates that this is not the case for all virulence factors. The ability of *S. pneumoniae* to cause disease varies among strains within a species and one major cause of this variation is differences in gene content. The striking variation of gene content and structure of the PPI1 variable region among the *S. pneumoniae* strains suggests that this could be one of the factors that modulate different virulence among strains (Brown *et al.*, 2004). Perhaps the relative inefficiency in maintaining stability of PPI1 plays a key role in virulence modulation of this bacterium. Nevertheless, the function of *pezAT* in the virulence of *S. pneumoniae* may be clearer once the target of the PezT toxin is finally determined.

4.2 The *yefM-yoeB_{Spn}* TA locus of *S. pneumoniae*

The *yefM_{Spn}* antitoxin gene is located upstream of the *yoeB_{Spn}* toxin gene and three nucleotides are identified between the stop codon of *yefM_{Spn}* and the start codon of *yoeB_{Spn}*. Although the genes do not overlap, they have been demonstrated to be co-transcribed (Section 3.2.4) as a single RNA transcript and thus indicated translation coupling. A σ^{70} -type promoter, termed P_{*yefM2*}, was identified upstream of the *yefM_{Spn}* antitoxin gene. Interestingly, a BOX element was identified upstream of P_{*yefM2*} as well. In general, the BOX elements found in *S. pneumoniae* are short repeats (Martin *et al.*, 1992) and they are likely derived from mobile genetic elements. This is due to single-nucleotide duplication with identical nucleotide pairs that has been observed at the base of the predicted stem-loop structure of the BOX element as well as significant homology between BOX elements and ISS_{*Spn2*}/ISS_{*Sts01*} transposable elements (Knutsen *et al.*, 2006). They are distributed randomly and are numerous in intergenic spaces in the genomes of the TIGR4 (127 copies reported by Mrazek *et al.*, 2002; Tettelin *et al.*, 2001) and R6 strains (115 copies reported by Hoskin *et al.*, 2001). The BOX elements consist of three different modules; boxA (59 bp), boxB (45 bp) and boxC (50 bp) (Martin *et al.*, 1992). BoxB is located between boxA and boxC and is present at between zero to eight copies. BOX elements which contain a boxA and boxC module have the potential to form a stable stem-loop structure which could affect the expression of the neighbouring genes (Knutsen *et al.*, 2006). The secondary structure is more important than the base changes within the consensus sequence, as in certain cases, the BOX elements differ from the consensus, but the secondary structure remains the same (Martin *et al.*, 1992). The BOX elements might enhance gene expression by either stabilizing mRNAs (e.g. increasing half life of the mRNA) or serving as binding sites for regulatory proteins (Knutsen *et al.*, 2006; Croucher *et al.*, 2009).

The BOX element upstream of the *yefM-yoeB_{Spn}* locus contains only boxA and boxC elements, which is therefore categorised as a boxA-C element, and is flanked by an ‘A’ base at both the proximal ends of the boxA-C element (Figure 27). This could be the single nucleotide-duplication that resulted from the insertion of the boxA-C element upstream of P_{*yefM2*}. The insertion of the boxA-C element led to the addition of a σ^{70} -type promoter designated P_{*yefM1*} to the *yefM-yoeB_{Spn}* locus. The functionality of P_{*yefM2*} as well as P_{*yefM1*} was validated by the β -galactosidase assay results (Figure 35) as well as primer extension analysis (Figure 34) with P_{*yefM1*} being 15-fold weaker than P_{*yefM2*}. This was also reflected in the results of the 5’-RACE used to initially identify the transcription start site(s) for the *yefM-yoeB_{Spn}* locus. 5’-RACE only identified one transcriptional start site (Section 3.2.5.1) i.e. the ‘A’ residue 25 nucleotides upstream of the *yefM_{Spn}* ATG start codon and initiating from the P_{*yefM2*} promoter. Transcription from the P_{*yefM1*} promoter that initiated from the ‘A’ residue located 84 nucleotides upstream of the *yefM_{Spn}* start codon was only detected using primer extension and not by 5’-RACE. 5’-RACE may therefore be not sensitive enough to detect the presence of the transcript from the much weaker P_{*yefM1*} promoter.

As the boxA-C element along with the P_{*yefM1*} and P_{*yefM2*} promoters are highly conserved among the *S. pneumoniae* strains, the promoters could have been optimized to regulate the *yefM-yoeB_{Spn}* operon concomitantly. Tandem promoters were also observed in another chromosomal TA locus, *mazEF*, in *E. coli*. The two promoters P₂ and P₃ located upstream of *mazEF* are 13 nucleotides apart. The P₂ promoter is about 10-fold stronger than the P₃ promoter (Marianovsky *et al*, 2001). Besides two promoters, there is an “alternating palindrome” located upstream of *mazE*, which functions as the operator of the *mazEF* operon. This alternating palindrome could exist in one of two alternative states: its middle part, designated “a”, complemented with either of the outer parts, designated “b” or “c” (Marianovsky *et al*, 2001). Numerous

mutations that were introduced into the “alternating palindrome” did not at all affect the binding efficiency of the MazEF complex, suggesting that the secondary structures of the regulating region was more important than its DNA sequence *per se*. The role of the additional d-e palindrome within the “alternating palindrome” is unclear. The duplication of the promoters or binding sites for auto-regulation was proposed as ensuring that the *mazEF* regulation will be adequate even in the case that one of these elements may be destroyed (Marianovsky *et al.*, 2001). Perhaps for the *yefM-yoeB_{Spn}* locus, the presence of the two promoters, P_{*yefM1*} and P_{*yefM2*}, also enable a safer and more dynamic regulation to the TA operon. Tandem promoters are not uncommon in prokaryotes, especially for genes that encode enzymatic proteins such as *glnA* (Reitzer and Magasanik, 1985), *carA* (Piette *et al.*, 1984), and *gal* (Musso *et al.*, 1977). The tandem promoter pair of a gene can be regulated differently to serve different purposes in cellular functions (Queen and Rosenberg, 1981). In certain bacterial genes, developmental control is affected by differential use of tandem promoters (Johnson *et al.*, 1983; Wong *et al.*, 1983). As the boxA-C elements are highly conserved among all the *S. pneumoniae* strains that harbour a *yefM-yoeB_{Spn}* locus, perhaps these elements have given the host a selective advantage and increased fitness for the cells.

Besides the two tandem promoters, an incomplete palindrome sequence was identified upstream of the *yefM_{Spn}* start codon (Figure 27). The 44 bp palindrome sequence is centered 62 bp upstream of the *yefM_{Spn}* ATG start codon and overlaps the -35 promoter sequence of P_{*yefM2*}. Purified YefM_{Spn} protein and YefM-YoeB_{Spn} protein complex were found to bind to this palindrome sequence (Section 3.2.7) and regulate P_{*yefM2*}. On the other hand, no binding of these proteins was observed on other nucleotide sequences nor was any signal regulating P_{*yefM1*} observed for these proteins. It seems likely that P_{*yefM1*} is expressed independent of the YefM_{Spn} and YefM-YoeB_{Spn} proteins. Combining the results from β -galactosidase assays, it was thus deduced that P_{*yefM1*} is

expressed constitutively, providing basal transcription (based on its weak expression detected in β -galactosidase assays as indicated in Figure 35) to the TA genes so that the host cells could have a 'quick response' to encounter any sudden environmental or physical changes.

Tandem promoters or a combination of multiple promoters are usually stronger than the individual promoters (Wei *et al.*, 2004; Mirjana *et al.*, 1988). However, the overall promoter activity of both P_{yefM1} and P_{yefM2} promoters was observed to be lower when compared to the P_{yefM2} promoter alone (Figure 3.2.13). This could be due to the competition between the two adjacent promoters to recruit RNA polymerase. This scenario has been observed in the gene for ATPase subunit 9 of yeast mitochondria (Oli 1) which contained tandem promoters (Op_1 and Op_2) separated by 78 nucleotides (Biswas and Getz, 1988). The upstream promoter (Op_1) is 12-15 times stronger than the downstream promoter (Op_2). Removal of the upstream strong promoter enhanced the weak promoter activity by seven-fold. The strong promoter competed effectively with the weak promoter for limited RNA polymerase. Moreover, when the distance between the two promoters was increased, the inhibitory interaction between them decreased. In addition, the weak promoter was more sensitive to inhibition by KCl salt concentration in transcription than the strong promoter (Biswas and Getz, 1988). Therefore, the close proximity (30 nucleotides) of P_{yefM1} and P_{yefM2} may influence the binding affinity of each other for RNA polymerase and thus reduce the overall promoter activity. As in the case of Op_1 and Op_2 towards sensitivity to salt, perhaps P_{yefM1} and P_{yefM2} may have differential transcriptional activities which could be attributed to different regulation of transcription under different environments.

When only the P_{yefM2} promoter is present in the construct, the expression of the YefM_{S_{pn}} antitoxin protein in *trans* slightly represses the P_{yefM2} promoter (Figure 36) and YoeB_{S_{pn}} toxin protein serves as a co-repressor to further repress the P_{yefM2} promoter

activity in *trans* (Figure 37). This observation was similar to that seen for most other TA loci, including its homologue in *E. coli*, the *yefM-yoeB_{Eco}* locus. When YefM_{Eco} was expressed in *trans* from an arabinose-inducible promoter, the promoter activity was reduced by 5.5-fold. Coexpression of YefM-YoeB_{Eco} in *trans* further reduced the promoter activity to background level (Kedzierska *et al.*, 2007). A similar pattern was observed when the genes were constructed in *cis* (Figure 38). EMSA (Figure 49) and footprinting assays (Figure 50 and Figure 52) showed that both the YefM_{Spn} protein and the YefM-YoeB_{Spn} protein complex bind to the palindrome sequence which overlaps the -35 promoter sequence of P_{yefM2}, and thus possibly competing with the RNA polymerase reducing binding to the promoter thereby repressing transcription, as was described for the *pezAT* TA locus. The amount needed for binding of the YefM-YoeB_{Spn} protein complex to the DNA target is a lot lower compared to the YefM_{Spn} protein (Figure 49) indicating that YoeB_{Spn} likely enhances the binding affinity of YefM_{Spn} to the palindrome sequence and therefore represses the transcription more prominently. A similar observation was reported for the *E. coli yefM-yoeB_{Eco}* TA loci where less YefM-YoeB_{Eco} complex was needed compared to YefM_{Eco} protein for binding to the DNA target upstream of *yefM_{Eco}*, implying that YoeB_{Eco} improves DNA binding by YefM_{Eco} either by enhancing the stability of YefM_{Eco} or altering the YefM_{Eco} conformation to one which is more favorable for DNA binding (Kedzierska *et al.*, 2007).

On the other hand, when P_{yefM1} and the boxA-C element were included in the construct along with P_{yefM2} and the palindrome sequence, the overall promoter activity was slightly reduced when YefM_{Spn} protein was expressed in *trans* (Figure 36). However, no further repression was observed when the YefM-YoeB_{Spn} protein complex was co-expressed in *trans* (Figure 37). Intriguingly, activation instead of repression was observed when YefM_{Spn} was expressed in *cis* (Figure 38). Although lower promoter

activity was observed when YefM-YoeB_{Spn} was expressed in *cis*, the promoter activity was still higher than in cells that harbour a construct with the two promoters, the palindrome sequence and boxA-C element. Even when stop codons were introduced within the *yefM*_{Spn} reading frame to abolish its translation, activation was still observed. In addition, no prominent promoter activity was detected within the *yefM*_{Spn} reading frame (as shown in the β -galactosidase assays in Figure 38) that could contribute to this activation. It is worth noting that the activation could only be observed in the presence of the entire upstream regulatory region (which encompasses P_{yefM1}, boxA-C element, P_{yefM2} and the palindrome sequence) along with the *yefM*_{Spn} reading frame. Thus, the activation is likely due to *cis*-acting elements instead of the YefM_{Spn} protein serving as an activator, as no activation was seen when the protein was expressed in *trans*, nor any internal promoter detected within the *yefM*_{Spn} reading frame.

For the *mazEF* TA locus, besides autoregulation by its own proteins, the expression from the *mazEF* promoters was activated 1.6-fold by a host factor, FIS protein, where the FIS binding site is located upstream of the “alternating palindrome” and the two promoters (Marianovsky *et al*, 2001). It had been suggested that by binding to the DNA target, the FIS homodimer (Koch and Kahmann, 1986) causes the DNA to bend, thus increasing the binding efficiency of the RNA polymerase (Panet *et al.*, 1996). The *yefM-yoeB*_{Spn} TA locus may have a similar modulation but different mechanism as besides sequences upstream of P_{yefM2}, the activation was only detected in the presence of the *yefM*_{Spn} reading frame. It is thus speculated that the sequences upstream of P_{yefM2}, which includes the boxA-C element and sequences downstream of P_{yefM2} harbour possible activation binding sites of the host factor, which causes the activation seen in the β -galactosidase assays. Generally, very few prokaryotic activators bind downstream of the promoter -10 element. Most prokaryotic regulatory proteins with downstream binding sites reduce the level of transcription, and therefore the region downstream of

the -10 element seems to be exclusive for repressors (Gralla and Collado-Vides, 1996). Rns, a member of the AraC family, is a transcriptional activator for the pili genes in enterotoxigenic *E. coli* (Caron *et al.*, 1989). Rns also positively autoregulates its own expression (Froehlich *et al.*, 1994). Instead of binding to the normal binding region of the activator, an engineered maltose-binding protein:Rns fusion protein was observed to bind to a region centred at -227, which is far upstream of its transcriptional start site. In addition, the maltose-binding protein:Rns fusion protein has another two binding sites centred at +43 and +82, which are atypical locations for prokaryotic activators (Munson and Scott, 2000). *In vivo*, Rns-dependent activation required the binding of the Rns to the distal upstream binding site and at least one site downstream of the transcriptional start site. *In vitro*, the binding of a maltose-binding protein:Rns fusion protein to each of these sites prevented RNA polymerase from forming open complexes at two distinct sites and thus facilitating the binding of RNA polymerase to the *rns* promoter and the subsequent formation of an open complex at a location consistent with the transcription start site (Munson and Scott, 2000). Rns is thus an example in which binding of the prokaryotic factor downstream of the target promoter is needed for activation and this arrangement is extensively used in eukaryotes. In addition, many activator-dependent promoters are also co-regulated either by a repressor or by a second activator (or by both) (Barnard *et al.*, 2004). Promoters can be complex and the complexity arises because of the cells' need to couple the expression of individual genes to different environmental signals, interpreted by different transcription factors (Barnard *et al.*, 2004). The regulation of the *yefM-yoeB_{Spn}* operon is conceivably more complex when compared to most reported TA loci due to the insertion of the boxA-C mobile element, which led to the presence of an additional promoter, P_{*yefM1*}, and the possible involvement of host factor(s) possibly acting as an activator to the *yefM-yoeB_{Spn}* operon.

The promoter region of *yefM-yoeB_{Spm}* seems to share some similarity and dissimilarity with the promoter region of the *E. coli*-encoded *yefM-yoeB_{Eco}*. The *yefM-yoeB_{Eco}* locus harbours only one promoter instead of two as in *yefM-yoeB_{Spm}*. This is not unexpected as the boxA-C element which contributed the additional promoter in *yefM-yoeB_{Spm}* is not found in *yefM-yoeB_{Eco}*. Although the -10 (5'-TAcAAT-3') and the -35 sequences (5'-TaGttA-3') of the *yefM-yoeB_{Eco}* promoter (Kedzierska *et al.*, 2007) diverged from the P_{*yefMI*} promoter of *yefM-yoeB_{Spm}*, the spacer between the -10 and -35 regions as well as the spacer between the -10 region and the transcriptional start site (as determined by primer extension) were the same, i.e. 17 bp and 5 bp, respectively. The operator site for *yefM-yoeB_{Eco}* comprises adjacent long and short palindromes with core 5'-TGTACA-3' motifs (Kedzierska *et al.*, 2007). DNaseI footprinting results showed that YefM_{Eco} binds to the long palindrome, followed sequentially by short palindrome recognition when the amount of YefM_{Eco} added was increased (Kedzierska *et al.*, 2007). The preferential binding of YefM_{Eco} to the long palindrome might be facilitated by the additional palindromic nucleotides that flanked the hexameric core sequence, which was absent in the short repeat. The YefM-YoeB_{Eco} protein complex had been shown to bind to both of the palindromes more avidly than the YefM_{Eco} protein alone (Kedzierska *et al.*, 2007). Multiple substitution mutagenesis demonstrated that the short repeat was crucial for correct interaction of YefM-YoeB_{Eco} with the operator site *in vivo* and *in vitro* (Kedzierska *et al.*, 2007). Moreover, altering the relative positions of the two palindromes perturbed the cooperative interactions within the YefM-YoeB_{Eco} protein complex with the repeats: such that the binding of the complex to the long repeat was retained but was abrogated to the short repeat (Bailey *et al.*, 2009). Bioinformatics analyses showed that the long and short palindromes were conserved in a number of diverse bacteria. The paired hexamer repeats were consistently separated by 6 bp, which supported the study by Bailey *et al.* (2009), but the sequence of the spacers, as well as

the sequence between the short palindrome and the start codon, and the 5' end of the long palindrome, were varied (Kedzierska *et al.*, 2007). The hexamer repeat along with two extended nucleotides (5'-CTTGTACAAG-3') that could possibly form a long repeat was identified 5 bp upstream of the -35 sequence of P_{yefM2}, however, the sequence located 6 bp (the consistent spacer between the long and short repeats) downstream of the long repeat (5'-TGacaA-3') did not share much consensus with the short repeat which was known to be very conserved among the bacteria studied. The 5'-TGTACA-3' motif was not seen in many genomes of bacteria that harbour *yefM-yoeB* homologues, which suggests that the YefM-YoeB homologues of these bacteria recognize different regulatory motifs (Kedzierska *et al.*, 2007). The binding sites of the YefM_{Eco} antitoxin and the YefM-YoeB_{Eco} protein complex covered both the long and short palindromes, which overlapped the -10 element (which is located within the long palindrome) and the transcriptional start site (which is located within the short repeat) (Kedzierska *et al.*, 2007). The long repeat upstream of *yefM*_{Spn} did not overlap with the promoter sequence of P_{yefM2}. DNaseI footprinting assays (Figure 50) showed that YefM_{Spn} and YefM-YoeB_{Spn} covered the long repeat and the -35 element instead of the -10 element and transcription start site. It seems that YefM_{Spn} and YefM-YoeB_{Spn} recognize a similar motif as YefM_{Eco} and YefM-YoeB_{Eco}, but the regulation mechanism is likely different due to the location of the motif and the lack of consensus sequence of the short repeat in the *yefM-yoeB*_{Spn} promoter, as well as the likely involvement of the boxA-C element and sequences downstream of the *yefM*_{Spn} promoters. Circular dichroism spectra indicated that YefM_{Eco} and the YefM-YoeB_{Eco} complex undergo structural transitions when bound to DNA (Kedzierska *et al.*, 2007). In contrast, the operator region DNA does not undergo major structural changes when bound by its cognate proteins. However, the presence of a DNaseI hypersensitive cleavage site in the YefM- and YefM-YoeB-operator complexes suggests that the operator within the

nucleoprotein complex may undergo deformation as DNA conformational changes such as DNA bending, major groove opening and kinking are not uncommon in repressor-operator interactions (Kedzierska *et al.*, 2007). Hypersensitivity was also evident at three nucleotides located between the –35 and –10 elements of P_{yefM2} on the coding strand in the presence of the YefM_{S_{pn}} and YefM-YoeB_{S_{pn}} complex, which indicated that the DNA structure was possibly perturbed at this point.

Intriguingly, a functional promoter was detected within the C-terminal coding region of the *yoeB_{S_{pn}}* gene. The functionality of this internal promoter was scrutinized *in silico*. The possibility of this promoter regulating any gene downstream of *yoeB_{S_{pn}}* is not high as the gene downstream of *yoeB_{S_{pn}}* is located a distance away (~200 nucleotides) from this promoter. Furthermore, any transcription from this internal promoter will likely be terminated after the *yoeB_{S_{pn}}* stop codon due to the presence of a typical transcriptional terminator immediately after it (Figure 41b). In addition, no homologues of any known genes were found in the region between the *yoeB_{S_{pn}}* stop codon and the start codon of the gene downstream of *yoeB_{S_{pn}}*. The likelihood that this promoter results in transcription of an antisense regulatory RNA was also investigated. Antisense RNA is usually a small, untranslated transcript (70–110 nucleotides), which pairs with target RNA that is complementary in sequence and thereby controls the expression of the gene target by transcriptional attenuation, facilitation of mRNA decay and translational inhibition (Simons and Kleckner, 1988). Usually, the antisense and target RNAs are transcribed from opposite strands of the same piece of DNA template by opposing promoters, therefore antisense RNAs are completely complementary to their targets (Thisted *et al.*, 1994). For the RNA transcript transcribed from the internal promoter of *yoeB_{S_{pn}}* (Section 3.2.6.1, Figure 41), no reading frame was annotated on the opposite strand of this sequence, or adjacent to this sequence, based on the BLASTN results. This RNA transcript is also not homologous to any other part of the *yefM-yoeB_{S_{pn}}* locus

along with its promoter as well as to any other prokaryotic sequence in the databases. Therefore, it seems unlikely that this RNA transcript could serve as an antisense regulatory RNA to *yefM-yoeB_{Spm}*, or even for other genes. The possibility that this RNA transcript could be translated and the function of the translated peptide (if any) was also inspected. Kolodkin-Gal and colleagues (2007) had related *mazEF*-mediated cell death to quorum sensing as *mazEF*-mediated cell death was found to require a novel quorum-sensing peptide called the EDF. Structural analysis revealed that EDF was a linear pentapeptide, Asn-Asn-Trp-Asn-Asn (NNWNN), and each of the five amino acids of EDF was important for its activity, with the N- and C-terminal residues being the most crucial. It was suggested that the *zwf* gene product (glucose-6-phosphate-dehydrogenase), containing the amino acid sequence NNWDN may be the precursor of EDF (Kolodkin-Gal, *et al.* 2007). For *S. pneumoniae*, the identified quorum-sensing signal molecule is called the competence stimulating peptide (CSP) (Håvarstein *et al.* 1995). CSP in *S. pneumoniae* is derived from a 44-amino-acid-peptide precursor that contains a typical Gly-Gly cleavage site and an N-terminal leader sequence that is cleaved via proteolysis to produce a 17-amino-acid-peptide that is biologically active (Håvarstein *et al.* 1995). The peptide that could be translated from the putative start codon within the *yoeB_{Spm}* reading frame has only seven amino acids, which is comparatively a lot shorter than the *S. pneumoniae* and *mazEF*-mediated quorum-sensing peptide precursors. Furthermore, it does not share any consensus sequence with the EDF, the Gly-Gly motif of CSP or even any peptide sequences that are available in the NCBI databases. Therefore, this putative peptide is not likely to serve as a quorum-sensing peptide. Additionally, if this peptide is translated, it is not toxic to the *E. coli* cells. Intriguingly, based on the β -galactosidase assay results (Figure 38), the internal promoter of *yoeB_{Spm}* appeared to be nullified when transcription from P_{*yefM2*} is active. It is speculated that the binding affinity of RNA polymerase to P_{*yefM2*} promoter is stronger

than the internal promoter within *yoeB_{Spn}* and when transcription from P_{yefM2} is active, transcription from the internal promoter within *yoeB_{Spn}* is disrupted. Although this internal promoter is functional in *E. coli*, whether the transcript from this promoter is translated or if this promoter is indeed functional in its native host, *S. pneumoniae*, is yet to be elucidated.

In addition, a putative weak promoter and ribosomal binding sites were subsequently identified within the C-terminal region of *yefM_{Spn}* even though β -galactosidase assays were negative for the presence of a promoter within the *yefM_{Spn}* reading frame. Are these promoters (including the internal promoter within the C-terminal coding sequence of *yoeB_{Spn}*) indeed functional in their native *S. pneumoniae* host? The genome of *S. pneumoniae* is A+T-rich (61%) (Cook, 1999) and therefore the likelihood of the sequence resembling an *E. coli* σ^{70} promoter sequence is high (Samul and Leng, 2007). However, *S. pneumoniae* RNA polymerase recognizes more stringent consensus sequences than *E. coli* RNA polymerase (Lim *et al.*, 2003; Opel and Hatfield, 2001). *S. pneumoniae* sequences which are non-promoters but that function as promoters in *E. coli* have been observed (Rostoks *et al.*, 2000). As the promoter activity assay (β -galactosidase assay) in this study was done in an *E. coli* host, it is possible that these promoters found within the *yefM_{Spn}* or *yoeB_{Spn}* reading frame are recognized by RNA polymerase in *E. coli* but not in *S. pneumoniae*. The functionality of these promoters would thus require verification by primer extension analysis with RNA extracts from *S. pneumoniae*.

5 Conclusion

S. pneumoniae chromosomal *pezAT* and *yefM-yoeB_{Spn}* loci share the characteristics of the common TA loci described so far. Although the *pezA* and *pezT* genes overlap by one nucleotide whereas the *yefM_{Spn}* and *yoeB_{Spn}* genes are separated by three nucleotides, the antitoxin and the toxin genes of both TA loci were shown to be co-transcribed as a single transcript in RT-PCR analyses. Northern blots could also be performed to detect if the mRNA is further processed or truncated. Both TA operons are negatively autoregulated by their own proteins at the transcriptional level. The antitoxin protein (PezA or YefM_{Spn}) serves as a repressor that binds to the DNA operator site upstream of the antitoxin gene, and the repression is augmented by the toxin protein (PezT or YoeB_{Spn}). EMSA conducted for both TA loci showed that the amount of the TA protein complex needed to retard the DNA target was much lower than the corresponding level of antitoxin protein. This indicated that the TA complex binds more avidly to the DNA target compared to antitoxin alone. Moreover, from the footprinting assays, the protected region of the YefM-YoeB_{Spn} complex was more extended when compared to YefM_{Spn} alone. Therefore, it is postulated that the toxin protein improves the binding of the antitoxin protein to the DNA operator sites. In the DNaseI footprinting assays, the binding sites of YefM_{Spn} and the YefM-YoeB_{Spn} complex were within a palindrome sequence which overlapped with the -35 element of the P_{yefM2} promoter. It is suggested that for the *yefM-yoeB_{Spn}* TA locus, the binding of the antitoxin or TA protein complex to the palindrome, which overlaps the -35 element, hinders the binding of RNA polymerase or interfere with the RNA polymerase activity to the promoter and thus thwarts transcription. In the hydroxyl radical footprinting assays, the binding sites for both YefM_{Spn} and YefM-YoeB_{Spn} proteins were also within the palindrome sequence, but the -35 element of P_{yefM2} was protected only in the presence of the YefM-YoeB_{Spn} protein complex and not by YefM_{Spn}. This further

explained that the more prominent repression observed with the YefM-YoeB_{Spn} protein complex compared to YefM_{Spn} was due to the improved binding of YefM_{Spn} by the YoeB_{Spn} toxin on the operator sites which covered the -35 element of P_{yefM2}. For the *pezAT* locus, the DNA fragment bound to the PezA and PezAT proteins in the EMSA assays encompassed a palindrome sequence which overlapped with the -35 and -10 elements of the P_{pezA} promoter. It is therefore likely that for *pezAT*, the negative autoregulation of the locus is similarly due to hindrance of either the antitoxin or the TA complex binding on the operator site preventing RNA polymerase from accessing the P_{pezA} promoter or interfering with the RNA polymerase activity. Future studies with footprinting assays will enable a more precise determination of the binding sites for the PezA antitoxin and the PezAT TA complex.

The regulation of *yefM-yoeB_{Spn}* seems to be more complicated than that of *pezAT* or other TA loci as the insertion of a boxA-C element upstream of the locus contributed the additional promoter, P_{yefM1}, upstream of the P_{yefM2} promoter. The P_{yefM1} promoter is 15-fold weaker than the P_{yefM2} promoter and is not likely regulated by the YefM_{Spn} or YefM-YoeB_{Spn} proteins as no binding of these proteins was observed in this region. The P_{yefM1} promoter which lies at the 3'-end of the boxC element is located 30 bp upstream of P_{yefM2}. When only the P_{yefM2} promoter is present, YefM_{Spn} represses the promoter activity both in *cis* as well as in *trans*, and the repression is augmented by YoeB_{Spn}. The presence of the boxA-C element and P_{yefM1} caused a drastic difference in the regulation of this operon. When YefM_{Spn} was expressed in *trans*, a slight repression could still be observed, however, no further repression was detected with the co-expression of YoeB_{Spn} in *trans*. When the TA locus was provided in *cis*, activation instead of repression was observed in the presence of the *yefM_{Spn}* reading frame. Activation was still observed in the presence of *yoeB_{Spn}* although the promoter activity was lower compared to that with the *yefM_{Spn}* reading frame alone with the two tandem

promoters. The *yefM-yoeB_{Spn}* locus may be regulated by multiple factors and therefore its regulation appears to have been adapted to ensure its expression is at an optimal level to the cells. It is postulated that the activation observed when the *yefM-yoeB_{Spn}* locus was provided in *cis* is due to *cis*-acting elements which include the boxA-C element, the tandem promoters, the *yefM_{Spn}* reading frame and/or other *E. coli* host factor(s). The activation was not likely due to YefM_{Spn} acting as an activator as no activation was observed when YefM_{Spn} was expressed in *trans*. It appears that P_{*yefM1*} is a weak constitutive promoter which could provide a basal level of transcription for *yefM-yoeB_{Spn}* so that the cells may have a prompt response to counter any sudden changes in the environment. It has to be highlighted that the promoter activity assays were carried out in *E. coli* and have yet to be validated in the native *S. pneumoniae* host (such as a *yefM-yoeB_{Spn}* deletion mutant). Pull-down assays could also be carried out to identify the host factor(s) (if any) that could be involved in activation of promoter activity that was observed in the promoter activity assays of the *yefM-yoeB_{Spn}* TA operon.

Intriguingly, functional promoters have been discovered within the *yefM_{Spn}* and *yoeB_{Spn}* reading frames in assays carried out in *E. coli*. However, the functionality of these promoters in the *S. pneumoniae* host is questionable as the recognition of the *S. pneumoniae* RNA polymerase to its consensus sequence is more stringent compared to the *E. coli* RNA polymerase. Nevertheless, although no known functions could be ascribed to these internal promoters, their functionality in *S. pneumoniae* could not be ruled out, and these putative promoters could be verified using primer extension analyses using RNA extracts from *S. pneumoniae*.

Variation of gene organization and polymorphisms has been reported for the *relBE2Spn* TA locus of *S. pneumoniae* (Nieto *et al*, 2010). A similar observation was found for the *pezAT* locus, which is located in the PPI1 pathogenicity island of *S. pneumoniae* and direct repeats were observed flanking a set of genes which includes the

pezAT locus (Khoo *et al.*, 2007; Brown *et al.*, 2004). This indicated that horizontal gene transfer could happen at these direct repeat sites therefore causing variation in gene organization. However, this is not the case for *yefM-yoeB_{Spn}* as the arrangement of neighboring genes of the *yefM-yoeB_{Spn}* locus seems to be more homogeneous among *S. pneumoniae* strains and no variation of translated amino acids were found for the strains analyzed in this study with the exception of the JJA strain (where 3 out of 84 amino acids of the YoeB_{Spn} toxin differed from those of other strains in the database). Nonetheless, the YefM_{Spn} antitoxin sequence was conserved for all the *S. pneumoniae* strains. This is similar to the *pezAT* and the *relBE2Spn* TA loci where the PezA and the RelB2Spn antitoxins are more conserved than the PezT and RelE2Spn toxins among the strains. The basis behind this observation is not yet understood as when comparing the YefM-YoeB_{Spn} and the PezAT homologues from different organisms, the toxins appeared to have higher conservation than the antitoxins.

Understanding the mechanism of regulation of a TA locus is important as inhibition of the transcription of the TA locus and triggering the release of the toxin protein from the inactive TA protein complex could lead to bacteriostasis or possibly be bacteriocidal when exceeding a certain period of time. This could then serve as the basis for alternative antimicrobial therapy in the near future to counter the increased resistance to antibiotics by *S. pneumoniae* and similar pathogens.