Use of High-Density Transposon Libraries to Analyze Oxidative Protein Folding and Protein Translocation in *Escherichia coli*

By

Chen Jiang

A thesis submitted to The University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

School of Biosciences

College of Life and Environmental Sciences

University of Birmingham

June 2024

Abstract

Transposon mutagenesis is a widely used method to introduce random mutations into bacterial genomes. Transposon directed insertion-site sequencing (TraDIS) combines transposon mutagenesis with next-generation sequencing to identify genes required for viability under experimental conditions, which provides linkage of phenotype to genotype. In this study, the Tn5-based transposon mutant libraries were used to investigate the oxidative protein folding in *Escherichia coli* (*E. coli)* cytoplasm and the azide resistance in *E. coli* K-12 strain MG1655. SHuffle strains are engineered to produce disulfide bonded proteins to high yields within the cytoplasm. TraDIS was used to analyze the physiology differences between SHuffle strains and their wild type strains by comparing the essential genes in each strain. SHuffle B was identified to have nearly two-fold unique essential genes that are mainly involved in cell response to oxidative stress, suggesting that the SHuffle mutations might cause more oxidative stress in *E. coli* B. The transposon libraries were further used to probe the genomic requirement to suppress the temperature sensitivity of SHuffle B. Insertion disrupting genes involved in aerobic respiration decreased susceptibility of SHuffle B to higher temperatures. TraDIS was also used to investigate the possible function of trigger factor (TF) and DsbC by determine the essential genes of *E. coli* strains containing Δ*tig* and Δ*dsbC* mutations, which increase susceptibility to sodium azide. Since the mechanism of azide was yet unknown, six transposon mutations conferred azide-resistance were isolated from the transposon library. All of affected genes were involved in ribosomal biogenesis. SecA migrated as two isoforms in SDS-PAGE, the large isoform was full-length SecA and the small isoform was determined to be C-terminal tail truncated SecA. Expression of both isoforms increased in the isolated azide-resistant mutants in the presence of azide. However, only the small isoform accumulated in the azide-resistant *secA* mutants in the presence of azide, suggesting that the expression of the small isoform could affect the susceptibility to azide and two isoforms might have distinct physiological roles in vivo. The results presented highlight the potential applications of transposon mutagenesis in dealing with complex biological questions.

Thank you to everyone who loves me.

I dedicate this to you all.

Acknowledgements

Firstly, I would like to thank my family, my mom and dad, for not only supporting me for the expense of studying and living abroad, but also for the continuing love and encouragement. Without your efforts, I cannot devote myself to study and work. Even you were not with me in the UK, I can feel family is my solid backing.

To my supervisor Damon Huber, who provided me with excellent supervision and gave me fantastic lab experience. I would like to say thank you for all the help and guidance in the lab, your support in language and as a proofreader to help with my writing skills and grammar, which gave me confidence and helped me make progress constantly. The information about science and upcoming events you provided me enriched my PhD life. I would specially thank your patience and concern during the COVID-19 lockdown period and when I was suffering from anxiety, make me feel the "warmth in winter".

To the PIs and professors, thank you Pete Lund, Steve Busby, it's my pleasure to learn from you. Thanks for your help to provide me convenience by sharing the lab materials and techniques. Also, your advice and ideas for my project helped me improve my experimental design.

To Mat, thank you for all your help in my PhD life, I could not imagine without you teaching me TraDIS and answering my questions. Also, thanks for your coffee and drinks.

I would also like to thank everyone in T101 and students worked in our group on this project, Manpreet, Dana, Santoch, Gabby, Dena and Harriet in T101, Max, Georgia, Sravanthi, Evie and Micheal in our group. Thanks for giving me a convivial lab atmosphere and joyful experience. To Manpreet, for your help with TraDIS and interesting topics in everyday life. To Max and Georgia, for laughs, coffee, drinks, especially for the soup and mixed nuts prepared by Georgia's dad.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Genetic screening

Genetic screening, also known as mutagenesis screen, is a widely used experimental tool to link the phenotype and genotype. The basic screening is divided into two approaches, which are forward genetic and reverse genetics. Forward genetics aims to identify particular genes that are responsible for a phenotype of interest and further investigate the possible mechanisms. In general, forward genetics seeks to select and identify individual mutants in a population which contribute to a certain phenotype, for example, mutants that survive under specific environmental conditions, such as abnormal temperatures and antibiotic treatment. Whereas reverse genetics is used to identify phenotypes caused by disruption of a known gene, usually based on inactivating methods such as single-gene knockout. It has been widely used to determine the function of uncharacterized genes. Even for *E. coli* strain, which is a versatile cell factory that has been studied over many decades, there are approximately 35% genes within its genome that are missing their functional annotations (Ghatak et al., 2019).

1.1.1 Random mutagenesis and targeted mutagenesis

Generation of genetic variation is normally based on random mutagenesis and targeted mutagenesis. Random mutagenesis aims to create libraries containing randomly distributed mutations induced by mutagens or transposons. The classic approaches using mutagens include environmental mutagens such as ionizing radiation dating back to the 1920s and chemical mutagens such as N-methyl-N'-nitro-N-nitrosoguanidine and ethylmethane sulfonate to generate microbes with desired phenotypes. The random mutations in specific proteins provide molecular basis for protein function identification and are developed to screen for mutants with desired properties. The identification of the mutational profiles of mutagens such as UV radiation is necessary for the understanding of the evolutionary process (Shibai et al., 2017). However, classic mutagens conferred low rates of mutagenesis that can only generate a small fraction of mutants in the population. The greater understanding of antibiotics in the 1950s provided a broader screen for particular genetic system in the given instance and yield new insight into therapeutic development (Hutching et al., 2019). The DNA libraries generated by random mutagenesis using error-prone PCR or transposon mutagenesis were later developed.

Transposons, also known as transposable elements or jumping genes, are DNA elements that can be transferred to a random location within the chromosome. Transposition was first discovered by McClintock during the 1940s-1950s, the existence of transposable elements was discovered to alter the unusual phenotypic futures in maize (McClintock, 1950). The transposable elements were finally cloned and isolated in the early 1980s (Fedoroff et al., 1983). Tn*5* and Tn*10* were found in the 1970s as transposons containing the antibiotic-resistance genes, which conferred resistance to kanamycin and tetracycline respectively (Berg et al., 1975; Kleckner et al., 1975). The Tn*5* transposon was used in this study to generate the mutant library. It is composed of a DNA sequence containing an antibiotic resistance gene flanked by 19-base pair Mosaic Ends (ME), which can be recognized by transposase. The transposase is activated by Mg^{2+} in the host's cellular environment and facilitates the insertion of transposon into the host strain by electroporation. The mutants containing transposon insertions are selective on Luria bertani (LB) plates containing appropriate antibiotics. Transposon mutagenesis is widely used to produce random gene mutations within bacterial genomes in genetic studies. The transposon insertion sites can be identified by whole genome sequencing or genetic mapping. Compared to chemical mutagenesis, transposon mutagenesis is more efficient due to higher mutation frequency, lower chance to kill the organism and easier single hit mutation and mutant cell recovery (Seifert et al., 1986). The single hit mutations are able to introduce selectable markers in the targeted strains and can be recovered after mutagenesis. The disadvantages are the inaccuracy of the transposition system in living cells.

Unlike Random mutagenesis that cannot target specific region or genes within the genome, targeted (site-directed) mutagenesis is more straightforward to introduce controlled mutations in a known gene or genetic region using methods such as error-prone polymerase chain reaction (PCR), RNA-mediated gene interference (RNAi) and CRISPR/Cas9. Error-prone PCR is applied for direct evolution, especially for engineering of catalytic enzymes with a high specificity (Beaudry and Joyce, 1992) owing to the adjustable mutation rate and easy manipulation. RNAi is a method based on the expression of a sequence-specific silencing antisense either activating the degradation process and resulting in a genetic null mutant or an allelic series of mutants (Xu et al., 2019). CRISPR/Cas9 is more efficient in editing and mutagenesis of a genome in vivo (Wang et al., 2018). It has been widely used for reverse genetic screening to identify the phenotypic change from a defined mutation and to proceed with the analysis of unknown gene functions. One of the most famous targeted mutagenesis libraries is the Keio library (Baba et al., 2006), which contains a set of single-gene knock out mutants containing mutations that inactive all nonessential genes within the genome of *E. coli* K-12. Each of the nonessential genes has two independent deletion strains which the genes were replaced by a kanamycin resistance cassette flanked by FLP recognition target site combined with specific primers to replace the in-frame coding regions. Apart from the identification of essential genes, the individual mutants can be used separately for phenotypic studies and mapping of genetic networks. By contrast, targeted mutagenesis can introduce mutations at desired locations in the genome sequence with high precision and is typically used to generate non-functional, out-of-frame or gain-of-functional mutations. Random mutagenesis is appropriate for generating a comprehensive library and can be used to investigate the intragenic region which are normally overlooked in the targeted mutagenesis.

1.1.2 High-throughput sequencing technologies

Sequencing technology has been developed since the 1950s by Sanger and was first used to sequence the amino groups of insulin (Sanger et al., 1965). The first complete DNA was sequenced by electrophoresis-based sequencing in 1977 (Sanger et al., 1977).

Since then, 'Shotgun' sequencing (Staden, 1979), florescence-based Sanger sequencing (Smith et al., 1986; Mitra and Church, 1999), as well as the next-generation sequencing, such as Illumina sequencing (Canard and Sarfai, 1994; Slatko et al., 2018) have been developed and used to generate sequence data stored in repositories such as GenBank (https://www.ncbi.nlm.nih.gov/genbank/statistics/). Combined with a mutagenesis screen, sequences can provide a new and promising way to study the linkage between genetics and phenotypic differences, especially for the transposon mutagenesis.

Traditionally, studies have identified transposon insertion sites individually by PCR using specific PCR primers. However, sequencing all of the transposons in a library using Sanger sequencing is not feasible if there are a large number of mutations that cause a particular phenotype. Furthermore, studies that rely on the isolation and identification of individual mutations can result in selection bias. For example, they may miss rare but informative mutations at the expense of common mutations. In addition, because positive selection is required to isolate mutations from a library, these studies miss genes that are required for growth (*i.e.* essential for viability). Later approaches combine transposon mutagenesis with next-generation sequencing, which provide a high-throughput method for sequencing of the transposon junctions in a pooled library. Four insertion sequencing methods were published in 2009, INseq, Tn-seq, TraDIS and HIST (Goodman et al., 2009; van Opojnen et al., 2009; Langridge et al., 2009; Gawronski et al., 2009). Despite the similarity in protocols, all of these methods have been used to identify the transposon insertion sites and contribute to the analysis of essential and conditional essential genes under specific laboratory conditions. Compared to the traditional methods, transposon sequencing is more efficient and accurate based on short experimental period and development of sequencing and statistical analysis methods. The saturated library provides a wide coverage of the organism genome including the non-coding regions.

1.1.3 Transposon directed insertion-site sequencing

TraDIS is a high-throughput experimental technique based on TIS workflow, which allows the identification of the transposon insertion sites within a high-density library. Tn*5* transposon without insertion site bias is generally used in TraDIS, although it was considered to have preference in GC rich region (Green et al., 2012). The genomic DNA is fragmented by sonicating and the method uses adaptor ligation and a series of PCR based approaches to amplify the transposon junctions. By sequencing and mapping the junctions to the genome, the location and frequency of the transposon insertion sites can be identified, which can be used to determine the gene essentiality (Figure 1.6). Since the insertions of transposon resulting in disruption of the gene function, genes that are essential for the organism's growth rate or survival can be identified as ones that do not tolerate any or a few insertions. Under different laboratory conditions, TraDIS can be used to identify conditional essential genes, as well as genes with decreased or increased insertion frequency compared to the control sequencing library. Genes in which insertion frequency does not change under given conditions means that they have no effect on the phenotype. As the sensitivity of this method is boosted by sufficient saturated mutagenesis libraries for even small fitness differences, apart from genes, TraDIS can be used to assay essentiality of smaller components such as promoter regions and intergenic non-coding regions (van Opijnen and Camilli, 2013).

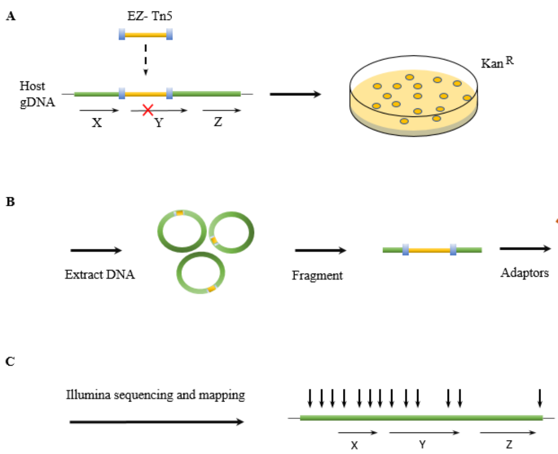


Figure 1.1 Workflow of transposon-directed insertion site sequencing (TraDIS). (A) The Tn5 transposon is inserted to random locations within the host's genome

leading to a loss of function mutation in gene Y represented by a red cross. The transformants are selected on the LB plates containing Kanamycin to pool the library. (B) The genomic DNA is extracted from the mutants in the library and fragmented by sonicating. The adaptors (brown rectangles) are ligated to the end of sheared genomic DNA, followed by PCR amplification and addition of P5 primer (P5), barcodes (BC) and P7 primer (P7) for Illumina sequencing and library distinguish. (C) Sequences out from the transposon junction are mapped onto the reference genome, the vertical black arrow represents a single transposon insertion site. Genes that do not tolerate any insertion or contain a few insertions are considered to be essential (gene Z).

1.2 Disulfide bonds

Protein disulfide bonds form between two cysteine residues (Cys) within a protein when the thiol groups of the cysteines undergo oxidation to form a covalent sulfur-sulfur bond. The conversion of SH groups to disulfide bonds was discovered to be a rate-limiting factor in the RNase A folding experiment (Anfinsen and Haber, 1961). This experiment studies the reduction and reformation of disulfide bonds in reduced ribonuclease. The conversion of SH groups to disulfide bonds was extremely slow during the regeneration of the RNase, and the disulfide bonds formation reform and stabilize the secondary and tertiary structures to ensure the high yield enzymatically activity (Anfinsen and Haber, 1961). The introduction of disulfide bond in proteins, also known as oxidative protein folding, is a post-translational modification essential for many proteins in folding and structural stability (Darby and Creighton, 1995). Disulfide bonds also play a prominent role in cellular processes such as cell division, molecular transportation, and cell responses to experimental stress (Kadokura et al., 2004). Understanding the role of disulfide bond enables advances in antibacterial innovation, drug development and genetic engineering (Dyotima et al., 2024).

1.2.1 Reducing pathways in the cytoplasm

E. coli strain is not capable of producing disulfide bonds in the cytoplasm due to the presence of two highly conserved reducing pathways, the thioredoxin and the glutathione/glutaredoxin pathways (Prinz et al., 1997). Ribonucleotide reductase is one of the infrequent proteins containing disulfide bond formed in the cytoplasm. These proteins are normally enzymes involved in the catalytic cycles and ultimately reduced by reducing pathways, or transient disulfide bonds in proteins generated under oxidative stress, which are regulated by the transcription factor OxyR. The change in thio-disulfide redox status in the cytoplasm stimulates the activation of OxyR to maintain cell homeostasis. The oxidized OxyR induces transcription of *trxC*, *grxA* and *gorA* (Ritz et al., 2000; Storz et al., 1990; Zheng et al., 1998), which encode for thioredoxin 2, glutaredoxin 1 and glutathione reductase, and reduced by thioredoxin and glutaredoxin systems.

Both thioredoxin and glutathione pathways rely on the reducing potential of NADPH and multiple thiol-disulfide oxidoreductases to reduce disulfide bonds in the cytoplasm. The thioredoxin pathway is composed of thioredoxin 1 (Trx1), thioredoxin 2 (Trx2) and thioredoxin reductase (TrxR), which are encoded by *trxA*, *trxC* and *trxB* (Laurent et al., 1964). TrxR utilizes the electrons provided by NADPH to maintain Trx1 and Trx2 in their reduced states and in turn catalyzes the reduction of substrates such as ribonucleotide reductase. The functions of Trx1 and Trx2 are overlapped, the growth defect of mutations in *trxA* was complemented by Trx2 (Ritz et al., 2000). However, *trxA* gene is normally more highly expressed than *trxC* gene and superior in the reducing ability (Ritz et al., 2000; Miranda-Vizuete et al., 1997). Both two thioredoxins were found to be involved in redox regulation by interacting with proteins either directly or associated with the regulation of cell response to oxidative stress (Kumar et al., 2004). Mutations in both thioredoxins were shown to be more resistant to hydrogen peroxide due to the activation of catalase in an oxidized environment. However, Trx2 was found to be significantly increased under oxidative stress, which could be explained by the regulatory function of OxyR on the expression of *trxC* (Ritz et al., 2000; Prieto-Alamo et al., 2000), indicating a specific role of Trx2 under oxidative stress distinct from Trx1.

The glutathione pathway is composed of three glutaredoxins, glutaredoxin 1 (Grx1), glutaredoxin 2 (Grx2) and glutaredoxin 3 (Grx3), which are encoded by *grxA*, *grxB* and *grxC*. The glutaredoxins are reduced by glutathione (GSH) encoded by *gshA* and *gshB* after reducing the substrate. GSH is in turn reduced by NADPH and GSH reductase encoded by *gor* (Holmgren, 1989). NADPH is the necessary cofactor for glutathione reductase activity which is generated by the pentose phosphate pathways (PPP) to provide electrons. Among the glutaredoxins, Grx1 is an efficient substitute for Trx1 that is able to reduce the ribonucleotide in vivo (Holmgren, 1976). Grx3 was shown to reduce the ribonuclotide reductase in low activity when *trxA* and *grxA* were mutated in *E. coli* (Aslund et al., 1994). Grx2 is involved in the reduction of mixed disulfide bonds generated by arsenate reductase together with Grx1 and Grx3 (Liu and Rosen, 1997).

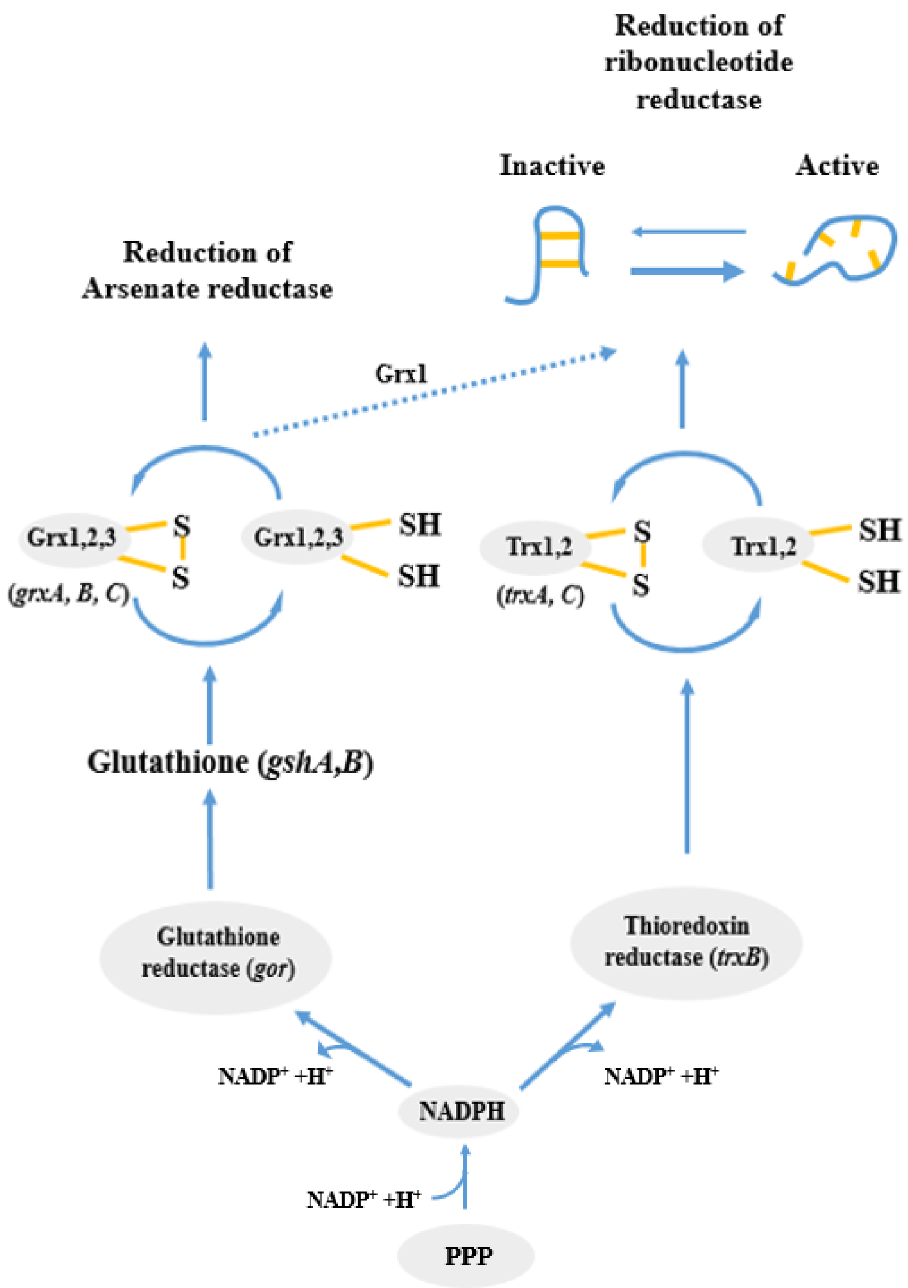


Figure 1.2. The *E. coli* cytoplasmic reducing pathways: thioredocxin pathway and glutathione pathway. The thioredoxin pathway is composed of thiroredoxin reductase encoded by *trxB* and thioredoxins (thioredoxin 1 encoded by *trxA* and thioredoxin 2 ecodede by *trxC*). Thioredoxins are responsible for the reduction of substrates such as ribonucleotide reductase. The glutathione pathway is composed of glutathione reductase encoded by *gor*, glutathione encoded by *gshA* and *gshB* and 3 glutaredoxins (glutaredoxin 1, 2, 3 encoded by *grxA*, *grxB* and *grxC*). Glutaredoxin 1 can reduce

ribonucleotide reductase mildly in a mutant inactivating the thioredoxins. Glutaredoxin 2 and 3 are involved in the reduction of substrates such as arsenate reductase. The electrons are provided by NADPH produced by the pentose phosphate pathway (PPP).

Both pathways are crucial for maintaining the reduced environment in the cytoplasm and protecting the cell from oxidative stress. The biosynthetic enzymes encoding by genes possess redundant effects, which lead to overlapping between thioredoxin and glutathione pathway in function and substrate specificity. Disruption of one of the pathways is defective in reducing cytoplasmic disulfide bonds without detriment in cell viability (Stewart et al., 1998). However, mutants lacking both systems confer lethal phenotype under aerobic conditions (Prinz et al., 1997).

The AhpCF system in *E. coli* is responsible for protecting the cells from oxidative damage, such as hydrogen peroxide. AhpC is a cysteine-dependent peroxidase. The cysteine in AhpC attacks the peroxide and reduces it by generating a sulfenic acid derivative and ultimately water and disulfide bond between two cysteines (Yamamoto et al., 2007). The disulfide bond is reduced by the Trx domain of AhpF using electrons provided by NADH and AhpC is returned to the active reduced state (Poole, 2005). It could be converted to a disulfide bond reductase AhpC* by a reversible triplet repeat expansion leading to the addition of single amino acid. AhpC* was shown to facilitate the reductase activity in reducing the mixed disulfides involved in the glutathione reductase pathway and alleviate the disulfide stress (Yamamoto et al., 2008). The mutated AhpC* still needs AhpF for its reductive activity (Ritz et al., 2001).

1.2.2 Disulfide bond formation in bacteria

In *E. coli* and other Gram-negative bacteria, disulfide bonds are usually only formed catalytically in the periplasm, which is an oxidizing cellular compartment between the cytoplasmic and bacterial outer membrane. The machinery of the disulfide bond formation, reduction and isomerization has been comprehensively studied. The formation of disulfide bonds is catalyzed by DSB proteins (Bardwell et al., 1991) and the oxidative pathway is catalyzed by DsbA and DsbB (Bardwell et al., 1993; Figure 1.3). The oxidative protein folding was shown to be deffected in cell absence of either DsbA or DsbB (Barder et al., 1999; Eckels et al., 2021). Proteins that include disulfide bonds in their mature structures are translocated across the cytoplasm to the periplasm, where the formation and isomerization of disulfide bond are processed.

DsbA is a soluble periplasmic protein composed of a thioredoxin (TRX) domain and an inserted helical domain (Martin et al., 1993). TRX domain is a common structure fold of thio-dusulfide oxidoreductase catalyzed thiol-disulfide exchange reactions. This structure is highly conserved and facilitates the formation of a functional active site typically containing a CXXC motif active site, which is two cysteine residues separated by two amino acids. The active site motif in DsbA, which is Cys-Pro-His-Cys, interacts with the reduced unfolded substrate and introduces disulfide bonds via disulfide exchange as they enter the periplasm via the Sec translocation pathway (Kadokura and Beckwith, 2009). The oxidation of the cysteine pair results in reduction of the active site, whereas the reduced DsbA is restored to the oxidized state via transfer of electrons to DsbB.

DsbB is an integral membrane protein composed of four transmembrane helices and 2 pairs of active-site Cys residues in the periplasm loops (Bardwell et al., 1993; Jander et al., 1994; Tang et al., 2013). The reoxidation of DsbA is achieved by rearrangement of the Cys residues in DsbB, the electrons extracted are transferred to quinone molecules and ultimately to the oxidase of the respiratory chain (Kobayashi et al., 1997; Bader et al., 1999). The oxidation and regeneration of DsbB relies on the oxidative activity of quinones, which were found to be ubiquinone under aerobic condition or menaquione under anaerobic condition (Kobayashi et al., 1997; Bader et al., 2000; Takahashi et al., 2004).

1.2.3 Disulfide bond isomerization in bacteria

Substrate proteins containing more than one pair of Cys residues are potentially forming non-native disulfide bonds, the formation of disulfide bonds between consecutive cysteine residues can also lead to misfolded disulfides (Berkmen et al., 2005; Manta et al., 2022). Therefore, it is necessary for cells to possess protein isomerase that facilitate the formation of correctly folded disulfide bonds. The isomerization pathways in bacteria are catalyzed by DsbC, DsbG and DsbD (Figure 1.3). DsbC and DsbG are responsible for the rearrangement of misfolded disulfide bonds, whereas DsbD is required to maintain these two proteins in their active reduce states.

DsbC is a soluble homodimeric protein function as protein disulfide isomerase in the periplasm to rescue incorrectly folded proteins. Each monomer is composed of an N-terminal dimerization domain and a C-terminal TRX fold catalytic domain containing CXXC motif linked by an α-helix (Sun and Wang, 2000; McCarthy et al., 2000). The thioredoxin fold domain contains four cysteines, two of them form disulfide which is essential to maintain the structural stability of DsbC, another two form a redox-active pair (Rietsh et al., 1997; Liu and Wang., 2001). The dimerization of the two monomers forms an uncharged groove, which is not only responsible for the interaction with the misfolded proteins but also involved in the chaperone activity in vivo (Darby et al., 1998; Chen et al., 1999).

The CXXC motif within the catalytic domain allows DsbC to perform thiol-disulfide exchange reactions. The active site cysteines attack incorrect disulfide bonds in substrate proteins, leading to the formation of transient mixed disulfide. There are two proposed mechanisms for DsbC, the reducing model and the isomerase model (Manta et al., 2022). In the reducing model, the substrate in the mixed disulfide complex is reduced then reoxidized by DsbA-DsbB system. The oxidized DsbC is returned to the reduced state by DsbD, which transfers electrons from the cytoplasm to the periplasmic space. DsbD itself is reduced by thioredoxin or other cytoplasmic reductants (Missiakas et al., 1995; Chung et al., 2000). Whereas in the isomerase model, DsbC binds to the substrate until it reaches the native state via thiol:disulfide exchange with no change of electrons in DsbC. Therefore, DsbC remains in the reduced states, and DsbD in this model is assumed to prevent DsbC from being oxidized in an oxidative environment (Walker and Gilbert, 1997). DsbC was also found to promote bacterial survival under oxidative stress by maintaining the protein stability and function under adverse conditions (Imlay, 2013).

DsbG is a newly discovered disulfide bond isomerase in *E. coli* (Heras et al., 2004). It is highly similar to its structural homolog DsbC, which is also composed of an N-terminal dimerization domain and C-terminal TRX catalytic domain. Similar to DsbC, the CXXC motif of DsbG is maintained in its active states by DsbD. In addition, it also functions as a molecular chaperone to prevent the aggregation of misfolded proteins in the periplasm. It was shown in precious studies that overexpression of DsbG as a reductase or isomerase, was able to complement the *dsbC* mutant in multiple disulfide-bonded proteins (Bessette et al., 1999). The main difference between DsbG and DsbC is the substrate specificity of DsbG is more limited than that of DsbC (Hiniker and Barwell. 2004). Additional acidic residues were found in the substrate-binding region of DsbG and form negative charged surface, distinct from the uncharged groove in DsbC, which further proved the difference in substrate specificity (McCathey et al., 2000; Heras et al., 2004). In addition, the feature of the binding sites makes DsbG prone to catalyze the disulfide rearrangement in proteins in the later folding process rather than in its unfolded states (Heras et al., 2004).

DsbD is an inner-membrane protein composed of three domains, an immunoglublin-like N-terminal domain (Goulding et al., 2002), a central transmembrane domain and a carboxy-terminal domain containing TRX fold and CXXC active site motif (Haebel et al., 2002; Porat et al., 2004). The N-terminal and C terminal domain of DsbD are located in the periplasm and connected by the transmembrane domain, each of them contains two cysteines. DsbD mediates the transfer of electrons from the cytosolic NAPDH pool via reduction of the disulfide pair in the transmembrane domain by cytosolic Trx1 (Katzen and Beckwith. 2000; Cho et al., 2007). The electrons are transferred to the C-terminal domain in the periplasm for maintaining DsbC and DsbG in their active reduced states via thiol:disulfide exchange reactions (Rietsch et al., 1997; Krupp et al., 2001). Cysteine pairs in the N-terminal are proposed to reduce the oxidized C-terminal dithiol (Chung et al., 2000).

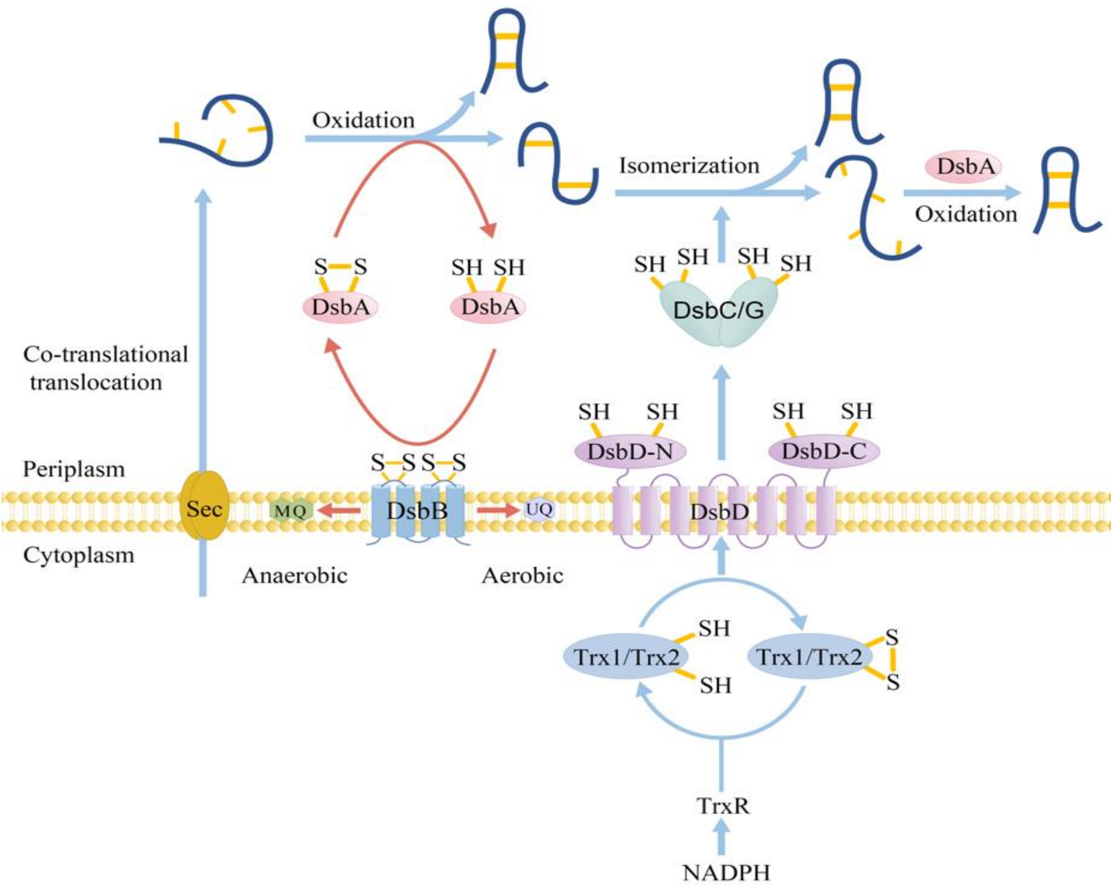


Figure 1.3 Disulfide bond formation and isomerization in the periplasm. The substrate protein that requires the disulfide bonds in its mature structure is exported to the periplasm by co-translational translocation via the sec machinery. DsbA interacts with the reduced substrate and introduces disulfide bonds via disulfide exchange immediately. The reduced DsbA is re-oxidized by DsbB, which is an inner membrane protein transferring the electrons obtained from DsbA either to ubiquinone (UQ) under aerobic conditions or menaquinone (MQ) under anaerobic conditions. The misfolded

disulfide bond is recognized by DsbC/DsbG, either reduced to the substrate then oxidized by DsbA or the native state or isomerize to the native state. DsbC/DsbG is maintained in its reduced states by DsbD, which is an innermembrane protein transferring electrons from the cytoplasmic thioredoxins (Trx1 and Trx2). The electrons are ultimately received from the NADPH pool catalyzed by thioredoxin reductase (TrxR).

1.3 Engineering of *E. coli* strains

E. coli strains are effective hosts for production of recombinant proteins with low cost and large quantity, which have been widely used in therapeutics and biological engineering. The well characterized genetics and ease of manipulation make *E. coli* an important tool in both laboratory and industrial use. According to the phylogeny, *E. coli* strains have been divided into five groups, including group A, B1, B2, E and D, each group contains numerous derivatives (Sims and Kim, 2011; Chaudhuri and Henderson, 2012). The majority of laboratory-used *E. coli* strains are *E. coli* K-12 strains and *E. coli* B strains. *E. coli* K-12 strain was isolated in 1922 and the derivatives such as MG1655, DH5α and DHB4 are the most frequently used. Whereas *E. coli* B is less well studied. It was named by Delbruk and Luria in 1943 (Luria and Delbruk, 1943) and originally used for phage study. BL21 is more extensively used *E. coli* B strain for basic research and biology applications.

1.3.1 *E. coli* K-12 and B strains

E. coli K-12 and B strains are derived from the same ancestor (Daegelen et al., 2009) and their genome sequences are highly similar. However, there are distinctive differences in functional genome, insertion sequence (IS) elements and complex deletions which contribute to particular physiological behaviors (Schneider et al., 2002; Studier et al., 2009). At present, both strains are used for producing recombinant proteins as well as metabolites such as amino acids. However, *E. coli* B was shown to be more suitable for producing recombinant proteins since enzymes involved in amino acid biosynthesis were synthesized at high levels in B strains but enzymes involved in degradation were not (Han, 2017). In addition, *E. coli* B was shown to decrease the acetate accumulation (Phue et al., 2010) and has an additional secretion system (Yone et al., 2009). The *E. coli* K-12 strains are favorable in producing membrane proteins associated with cell motility as *E. coli* B lacks flagella and has increased membrane permeability (Han, 2015; Han, 2017; Herrera et al., 2002). *E. coli* K-12 strain also confers more resistance to cellular stress therefore the expression of chaperones associated with stress response, such as DnaK and ClpB are expressed in higher levels in K-12 strain (Hyun Seo et al., 2013).

Despite both strains being versatile host in producing heterologous proteins, producing proteins requiring post-translational modifications such as disulfide bonds in appropriate conditions used to be a big challenge. Numerous approaches were applied to investigate the mechanisms of the formation of native disulfide bonds in the cytoplasm in various modified strains. The modifications include but are not limited to engineering the oxidative pathways by introduction of the de novo dithiol oxidants (Hatahet et al., 2010) or inactivating the reducing pathways by mutations inactivating reductases. The main difficulty is disulfide bonds are normally formed in a more oxidizing environment such as periplasm, but now has been overcome by engineering *E. coli* to diminish the cytoplasmic reducing pathways and rendering an oxidizing environment. These mutations were elucidated by Beckwith lab (Derman et al., 1993; Stewart et al., 1998) and optimized to construct the SHuffle strains (Lobstein et al., 2012).

1.3.2 SHuffle strains

E. coli SHuffle is a strain that has been engineered to produce disulfide bonds in proteins with high efficiency in the cytoplasm. Because proteins can be produced with high yields in the cytoplasm compared to when exported to the periplasm, these strains have become an important tool for the production of recombinant proteins. The SHuffle strains contain mutations inactivating genes encoding thioredoxin reductase (*trxB*) and glutathione reductase (*gor*). Thioredoxin reductase and glutathione reductase are required in the cytoplasmic reducing pathways which are responsible for maintaining the enzymes in their reduced active state to reduce any disulfide bonds in the cytoplasm. Normally, cells lacking both pathways are not viable because they cannot reduce ribonucleotide reductase, which is inactivated by thiol redox stress under aerobic conditions, which disrupts the deoxyribonucleotide synthesis. However, the lethal phenotype is suppressed in SHuffle by a mutation in a gene encoding the alkylhydroperoxidase AhpC, known as *ahpC**. This mutation disrupts the peroxicase function of AhpC but allows it to catalyze the reduction of Grx-1, which gains function to reduce essential proteins such as ribonucleotide reductase to enable the cell growth (Ritz et al., 2001; Figure 1.3). The thioredoxins accumulate their oxidized forms in this strain and consequently catalyze the formation of disulfide bonds (Bessette et al., 1999). In addition, SHuffle strains (New England Biolabs) have been engineered to express a cytoplasmic variant of the DsbC to generate an oxidative strain with isomerize activity, which mitigates some problems with protein misfolding and aggregation. Overexpression of DsbC increases the production of multiple recombinant proteins containing disulfide bonds (Zhang et al., 2002; Lobstein et al., 2012) and also protects SHuffle strains against the oxidative stress (Denocin et al., 2014).

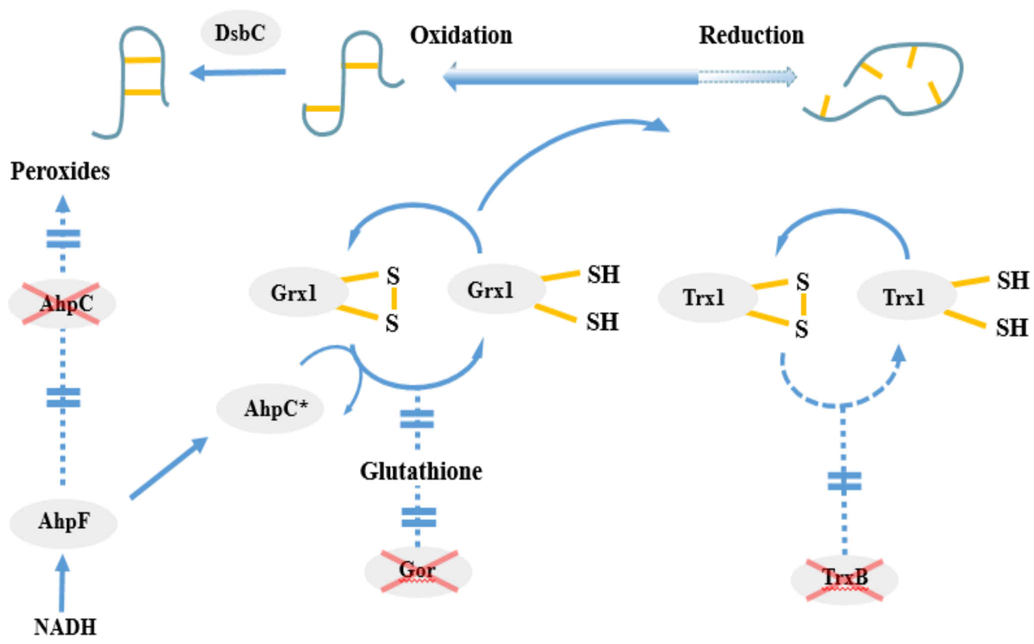


Figure 1.4 Modification in SHuffle strains to produce disulfide bonds in *E. coli* cytoplasm. The thioredoxin pathway and glutathione pathways is blocked in SHuffle strains by inactivating genes encoding for thioredoxin reductase and glutathione reductase (*trxB* and *gor*). The mutant AhpC* restores the reductive function of Gtx1, in turn relieves the lethal phenotype induced by double mutation of *trxB* and *gor*. AhpF is a flavoprotein transfer electrons from NADH to AhpC, but AhpC* in SHuffle strains. Overexpressed cytoplasmic DsbC facilitates aberrant disulfide bonds correction. Disabled protein interactions are shown by dotted lines. The reductase activity of glutaredoxin, Grx2, Grx3 and thioredoxin Trx2 are blocked as well but omitted in this figure.

The SHuffle modifications have been introduced into two parental lines of *E. coli*: a derivative of *E. coli* K-12 (SHuffle K-12), DHB4 and a derivative of *E. coli* B (SHuffle B), ER2566, with both SHuffle strains producing higher yields of recombinant proteins compared to the wild type, irrespective of whether the protein contains disulfide bonds or not (Anton et al., 2016). Since the production of SHuffle strains in 2009, various proteins with no disulfide bond such as green fluorescent protein (GFP) have been expressed in the SHuffle strains and show improved expression but the mechanism is not fully understood. The possible explanation is the increased oxidative stress in *E. coli* cells resulting the upregulation of the chaperones. Compared to SHuffle K-12, SHuffle B conferred higher enzyme activity and greater yield in expression of particular proteins, such as vtPA (Lobstein et al., 2012). The differences were proposed to be owing to the genetic differences between two *E. coli* strains.

SHuffle strains have been used to express various recombinant proteins containing disulfide bonds, including the full length biologically active immunoglobulin G (IgG) antibodies (Robinson et al., 2015). To improve the yield and activity of desired proteins, there is massive research dedicated to optimizing the expression conditions such as different temperature, growth media or co-expression with other proteins (Lobstein et al., 2012; Ren et al., 2016). Temperature was found to have a significant effect on protein expression in SHuffle cells, but was protein specific. Expression of proteins at high temperatures such as 37°C is harmful to the cell (Lobstein et al., 2012), which might be owing to the oxidative stress caused by SHuffle mutations.

1.4 Bacterial protein translocation pathways

Protein translocation across membranes is required for proteins located outside the cytoplasm, which is achieved by specialized mechanisms sharing several common features. There are two ubiquitous protein transportation pathways, Sec-dependent pathways and twin arginine translocation (Tat) translocation pathway (Palmer and Berks, 2012). The Sec machinery is responsible for translocation of polypeptides in an unfolded state, whereas the Tat pathway is responsible for the translocation of folded proteins.

1.4.1 Sec machinery

The Sec machinery is an evolutionary conserved system responsible for the localization of newly synthesized protein substrates into or across the cytoplasmic membrane. The translocation process relies on an evolutionarily conserved membrane-embedded heterotrimeric channel SecYEG in bacteria, which is formed by the integral cytoplasmic membrane proteins (IMPs) SecY, SecE and SecG (Blaauwen and Driessen, 1996). Sec-dependent protein translocation in *E. coli* is normally achieved by two different pathways, translationally coupled translocation (CT) along with the protein synthesis or afterwards, which refers to uncoupled (post-translational) translocation (UT) (Cranford-Smith and Huber, 2018). The insertion of IMPs is prone to be co-translational, whilst other soluble proteins for the periplasm or outer membrane tend to be post-translational (Osborne et al., 2005). In *E. coli*, Sec machinery is involved in the translocation of approximately 20% of secretory proteins, of which 7.5% of all newly synthesized proteins are translocated by CT and 13.5% of all newly synthesized proteins are translocated by UT (Cranford-Smith and Huber, 2018).

Newly synthesized protein substrates interact with the Sec machinery with the aid of amino-terminal signal sequence (SS). Proteins transported by the Sec machinery share common features in the 18-26 signal amino acids signal sequences in their amino-terminal region (von Heijne, 1986). The signal sequence contains three characteristic regions including a positively charged N-terminal amino acid, a hydrophobic region, and a C-terminal signal peptidase cleavage site in a polar region (Berks, 1996). The positively charged amino acids are involved in the interaction with lipid bilayer of the membrane (Van Voorst and De Krujiff, 2000) of Sec related proteins, such as signal recognition particle (SRP) and SecA (Batey et al., 2000; Gelis et al., 2007) and facilitate different types of secretory pathways (Owji et al., 2018). Mutations in the N-terminus region leading to reduced activity in translocation, but do not completely block the translocation (Vlasuk et al., 1983). The C-terminal region is responsible for cleavage of the SS after translocation (Paetzel et al., 2002).

The SSs of a subset (~10%) of the nascent chains are recognized by SRP as it emerges from the exit tunnel in CT (Huber et al, 2005). SRP is a ribonucleoprotein complex that mediates delivery of targeted nascent proteins to the SecYEG-bound membrane associate receptor FtsY (Halic et al., 2006; Figure 1.4). SRP in bacteria consists of an evolutionarily conserved protein, the fifty-four kDa homolog (Ffh) and a 4.5S RNA subunit (Eisner et al., 2003). Ffh and FtsY share conserved homologous α-helices N domain and GTPase domain (Kusters et al., 1995; Angelini et al., 2005). The GTPase is stimulated by their interaction, leading to the release of signal protein via conformation changing for both ribosome and translocation channel. SRP can also deliver the substrate proteins to the YidC insertase, which is a membrane protein that mediates insertion of polypeptide both dependent and independent with Sec translocon (Chen et al., 2002; Figure 1.4). YidC is able to export less complex membrane proteins. It is notable that SRP is prone to bind signal sequences which are hydrophobic or with strong helicity, otherwise, the signal sequence can be recognized predominantly by trigger factors (TF) or SecA, and undergoes UT instead (Lee and Bernstein, 2001; Schibich et al., 2016).

Proteins are delivered to the UT by SecA, and proteins delivered to this pathway are frequently fully synthesized when they start the process of translocation. In bacteria, the transportation of most soluble periplasmic proteins and outer membrane proteins across the cytoplasm membrane are dependent on SecA-mediated pathway through the SecYEG translocon (Oliver and Beckwith, 1981; Huber et al., 2005). SecA is an ATPase that drives the insertion of nascent proteins through the channel formed by SecYEG through ATP binding and hydrolysis. SecA interacts with the SS of the substrate protein either near the polypeptide exit channel of the ribosome or when it presents at the SecYEG translocon (Cranford-Smith and Damon, 2018; Figure 1.4). Binding of SecA undergoes ATP hydrolysis partially opens the SecYEG channel and facilitates the export (Allen et al., 2016). SecB interacts with SecA and maintains protein substrates in the unfolded state. TF can also interact with the nascent chain but not recognize the SS, with the help of SecB and SecYEG-bound SecA, mediating the UT (Oh et al., 2011; Geyter et al., 2020) (Figure 1.4).

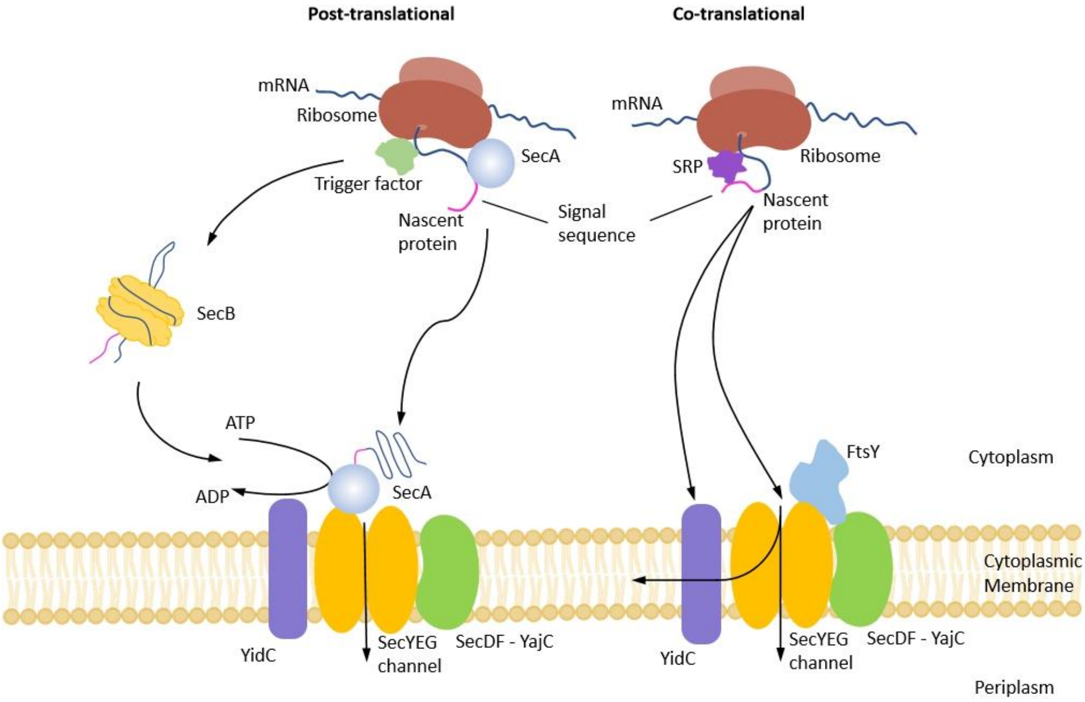


Figure 1.5 Post-translational and co-translational protein translocation pathways. The post-translational pathway also refers to SecA-dependent pathway, which is mediated by SecA and assisted by chaperones like trigger factor and SecB. SecA interacts with the signal sequence of nascent protein near the polypeptide exit channel of the ribosome and drives the insertion of protein through SecYEG translocon through ATP hydrolysis. Trigger factor can also bind to nascent proteins and ribosome without recognizing the signal sequence. SecB interacts with SecA or trigger factor to maintain the protein in its unfolded state. Co-translational pathway is dependent on SRP. SRP binds to translating ribosome and the signal sequence of the nascent protein as it emerges from the polypeptide exit channel, then deliver it to FtsY on the SecYEG translocon. SRP can also deliver nascent protein to YidC insertase to transfer less-complex proteins.

1.4.2 SecA

SecA as an ATP motor protein interacts dynamically with SecYEG to drive the protein translocation across the cytoplasm membrane though repeat cycles of ATP binding and hydrolysis. In *E. coli*, SecA is located half in the cytoplasm and half associated with inner membrane which is modulated by ATP (Cabelli et al., 1991). SecA is composed of two nucleotide-binding domains (NBD1 and 2), a pre-protein crosslinking domain (PPXD) and C-terminal domain (Zimmer et al., 2008). The C-terminal domain contains the helical scaffold domain (HSD), the helical wing domain (HWD) and the C-terminal tail (CTT) (Tsirigotaki et al., 2016).

Two NBDs are structurally similar to the Dead-box helicase where the ATP binding site is located. PPDX is associated with the interaction with nascent protein and the conformational change of SecA (Hunt et al., 2002). Substrate proteins bind to SecA in the groove formed by two NBDs and PPXD (Zimmer and Rapoport, 2009). PPDX and HSD are responsible for SecA interacting with SecY, where the PPDX interact with the cytoplasmic loops of SecY and HSD insert into the SecYEG translocon (Zimmer et al., 2008).

The CTT is proposed to alter the conformation of SecA as well. It is composed of a highly conserved metal binding domain and a flexible linker domain (FLD) (Jamshad et al., 2019). The metal binding domain (MBD) can be co-purified with metal ions, Zn^{2+} and Fe^{3+} (Cranford-Smith et al, 2019), and was shown to be involved in the interaction with SecB, and ribosome (Breukink et a., 1995). FLD is proposed to change the conformation of PPXD when SecA binds to the protein substrate and auto-inhibits the SecA activity (Jamshad et al., 2019).

1.4.3 SecB

SecB is a cytosolic molecular chaperone that interacts with fully translated preproteins and nascent Sec polypeptide sequence over 150 amino acids (Topping and Randall, 1994). It is highly conserved in α-,β- and γ- proteobacteria but not present in the Gram-positive bacteria (Van Der Sluis and Driessen, 2006). SecB is an acidic homotetrameric protein formed by a dimer of dimers, each monomer consists of 4 anti-parallel β-sheets and two α- helices (Muren et al., 1999; Bechtluft et al., 2009). The tetramer is connected by a long α- helices and the dimer is formed by interations between the first β-sheets and the firstα- helices, the structure is stabilized by the hydrogen bonds in between (Dekker et al., 2003).

SecB facilitates the protein translocation across the cytoplasmic membrane by preventing nascent proteins from folding and aggregation, and subsequently delivers them to the SecYEG-bound SecA. The chaperone activity associated with unfolded substrate proteins is similar to the general chaperones such as trigger factor, GroEL and DnaK (Fekkes and Driessen, 1999; Sakr et al., 2010; Jiang et al., 2021). The main difference is SecB can interact with the SecA MBD with high affinity and further contribute to the UT. SecB is also responsible for the translocation of proteins without SS (Driessen, 2001). Although SecB is not essential for viability in *E. coli*, inactivating SecB results in a mild translocation defect and severe protein aggregation (Baars et al., 2005). In addition, SecB null mutants have severe cold-sensitive (CS) phenotype (Ullers et al; 2007). Overexpression of SecB in turn relieves the CS phenotype and protein aggregation in a strain lacking TF or DnaK (Ullers et al., 2004).

1.4.4 Other Sec components

In addition to SecA and SecB, other auxiliary proteins including YidC, SecD/F, YajC and SecH (YecA) are also involved in the Sec-dependent protein translocation. SecD, SecF, YajC and YidC form a super-complex with SecYEG in the cytoplasmic membrane which assists the Sec-dependent protein translocation. SecD and SecF functions together to stimulate the Sec-dependent translocation. Inactivating of SecD and SecF exhibits CS phenotype and reduces the protein translocation (Pogliano and Bechwith, 1994). SecDF-YajC is associated with the membrane cycling of SecA and stabilizes SecG (Kato et al., 2003; Hand et al., 2006).

YidC is a membrane protein that involved in the SRP-dependent protein translocation

(Petriman et al., 2018) and functions as a chaperone assisting the folding of membrane proteins (Kumazaki et al., 2014; Wang and Dalbey, 2011). There are multiple studies proved the interaction between YidC insertase and the SecYEG translocon (Kumazaki et al., 2014; Sachelaru et al., 2017). However, YidC can promote the protein translocation either independently or by the Sec translocase (Celebi et al., 2009), the switch is dependent on the N-terminal length and change density of the substrate protein (Shanmugam and Dalbey, 2019).

SecH (YecA) is a novel accessory protein in the Sec machinery which assists the protein translocation. It is composed of an N-terminal domain with unknown function and a C-terminal domain which is highly identical to the SecA MBD (Cranford-Smith, 2017). The sequence homologue results in the functional similarity, SecH is shown to interact with client proteins, ribosome and SecB in vivo as SecA does (Cranford-Smith, 2017), but not modulate the ATP hydrolysis. SecH is also proved to interact with SecA in the co-purification experiments (Wayne, 2023).

1.4.5 Trigger factor

TF is a ribosomal-associated molecular chaperone, which assists the folding of nascent polypeptides and prevents protein aggregation. Monomeric TF is an ATP-independent chaperone consisting of three core domains. The N-terminal domain contains the TF signature motif which is responsible for modulating the interaction with ribosome and contribute to the chaperone activity (Kristensen and Gajhede, 2003). The central domain actives as peptidyl-prolyl cis/trans ismerise (PPIase) but not essential as a folding catalyst (Scholz et al., 1997; Zarnt et al., 1997). The C-terminal domain is the main chaperone module of TF which displays chaperone activity on its stabilized isolates (Merz et al., 2006).

TF was first isolated by its ability to stabilize the outer membrane protein A (proOmpA) in an assembly competent form (Crooke and Wickner, 1987), and it has been proved to interact with ribosomes, proOmpA and membranes, which lead to a cyclic action (Lill et al., 1988). TF interacts co-translationally with the nascent polypeptide chains as they emerge from the ribosome, either facilitates the spontaneously folding, or requires downstream chaperones, such as DnaK and GroEL to assist the subsequent protein assembly (Figure 1.5). TF cooperates with DnaK to facilitate protein folding but was shown to be functionally overlapped with DnaK. Inactivating TF was shown to increase the binding of DnaK to nascent polypeptides (Teter et al., 1999). However, cells lacking both TF and DnaK results in a lethal phenotype or limited growth under specific temperature but exhibit high-level protein aggregation (Deuerling et al., 1999; Genevaux et al., 2004). This growth defect can be suppressed by over-expression of GroEL/GroEL (Genevaux et al., 2004). TF was shown to interact with GroEl and its cofactor GroES functions in degradation of certain polypeptides (Kandror et al., 1995) and promotes the binding affinity of GroEL to unfolded proteins (Kandror et al., 1997). Unlike the DnaK and GroEL are heat-shock proteins, TF has higher expression level and affinity under lower temperatures (Kandror and Goldberg, 1997).

The expression of TF is two- or three-fold excess to the ribosomes (Crooke et al., 1988; Hoffmann et al., 2010), apart from facilitating the folding of nascent polypeptides, TF was shown to prevent unfolded proteins from aggregation (Saio et al., 2014; Wu et al., 2022). TF is also found associated with UT in which mediates transport of the fully synthesized protein across the cytoplasmic membrane. A recent study proposed that TF can interact with SecB and translocase and mediate the protein secretion (Geyter et al., 2020), even SecB was shown to be suppressed greatly by ribosome-bound TF in the previous study (Ullers et al., 2004). Both SecB and TF are dispensable for the cell, but the disruption of TF accelerates protein export by facilitating SecA and ribosome interacting to the SecYEG translocon, the mechanism is yet unknown (Lee and Bernstein, 2002; Ullers et al., 2006; Figure 1.5). In addition, mutations disrupting TF were shown to suppress the CS and protein aggregation caused by mutations in *secB*. The current consensus is that TF facilitates the correct folding and prevents polypeptide chains from mis-folding and aggregation but interferes the nascent chains targeting to the SecYEG translocon. Nonetheless *tig* mutants display a moderate outer membrane defect and lead to premature, cotranslational export of certain proteins (Oh et al., 2011), suggesting this delay in targeting may be important for OMP biogenesis. TF was also shown to be involved in cell division (Guthrie and Wickner, 1990).

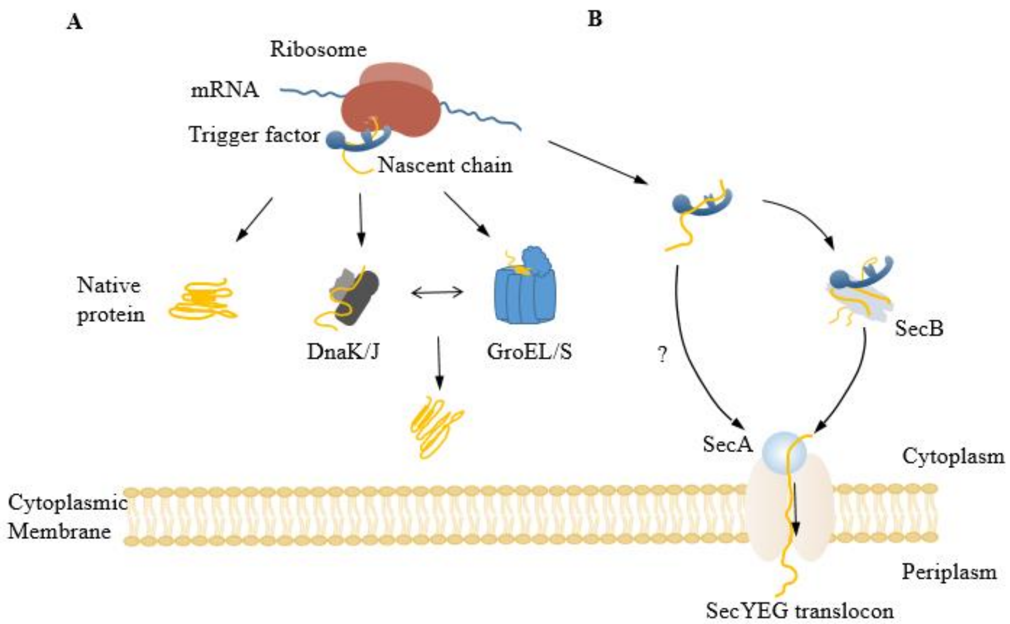


Figure 1.6 Mechanism of action of the trigger factor. (A) Trigger factor interacts with the nascent polypeptide chains as it emerges from the exist channel of ribosome. The nascent chains either fold immediately upon release from the trigger factor or requires folding assistant by interaction with the Hsp70 chaperone system which consists of the DnaK and cofactor DnaJ, and the Hsp60 chaperonin GroEL and its cofactor. There is shuttling between the chaperone systems to facilitate particular protein folding. (B) Trigger factor is associated with Sec dependent protein translocation. Nascent proteins interacting with TF binds to SecB and translocase to mediate the translocation across the cytoplasmic membrane. Trigger factor interferes nascent proteins and secA binding to SecYEG translocation by an unknown mechanism. Figure adapted from (Hoffmann et al., 2010).

1.5 Azide-mediated inhibition of SecA

Sodium azide is an inhibitor of SecA which inhibits SecA dependent protein tanslocation and the ATPase activities. Severe translocation defects were reported in *E.coli* in the presence of 1 mM sodium azide due to the accumulation of protein substrates and defect in ATP turnover (Oliver et al., 1990). Sodium azide was shown to block the ATP hydrolysis by interacting with the catalytic side chain in F_1-ATPase and preventing the release of ADP (Bowler et al., 2006). It is thought to inhibit the nucleotide exchange by SecA in a similar mechanism. All of the azide-resistance (*azi*^{R}) mutations were mapped to alleles within the *secA* gene (Oliver et al., 1990), resulting in amino acid substitution in one of the NBDs (Huie and Silhavy, 1995; Figure 1.6).

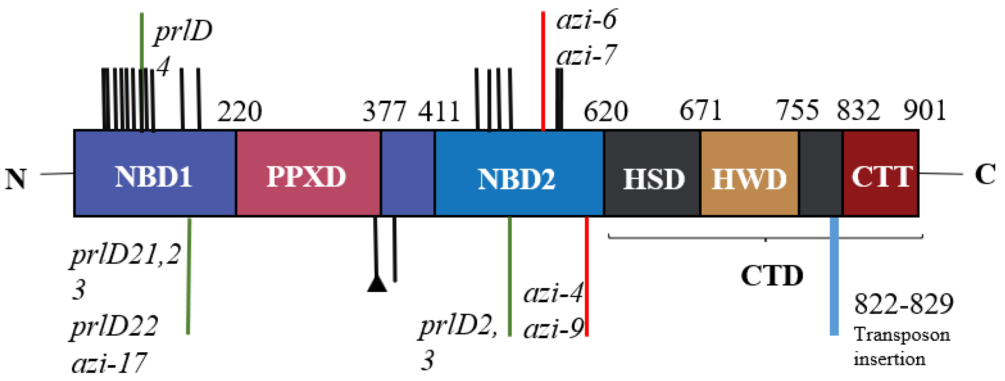


Figure 1.7 Distribution of the azide-resistant mutations in the amino acids of SecA. SecA in *E. coli* is composed of four domains: nucleotide binding domain 1 (NBD1; amino acids 1-220, 278-411), polypeptide crosslinking domain (PPXD; amino acids 221-377), nucleotide binding domain 2 (NBD2; aa 412-620) and C-terminal domain (CTD; aa 378-830). CTD includes the helical scaffold domain (HSD; aa 620-671, 755-832), helical wing domain (HWD; aa 671-755) and a C-terminal tail (Chandler et al., 2018). The vertical lines represent mutations resistant to sodium azide identified in the previous studies. The red lines represent four mutations located at two regions (Oliver et al., 1990). The green lines refer to *prlD* and *azi* mutations conferred azide resistance (Huie and Silhavy, 1995). The black lines are secA mutations isolated as suppressors against a cold-sensitive *secY* mutation (*secY*205) but increase the resistance to azide,

only mutations resistant to 1.0 mM sodium azide and above are presented (Matsumoto et al., 2001). The thick blue lines represent majority of enriched insertions in amino acids 822-829 of SecA in the presence of azide identified in a TraDIS experiment which result in truncation of the CTT (Chandler et al., 2018).

Translocation-coupled hydrolysis of ATP requires SecA interacting with substrate protein and SecYEG. Sodium azide was shown to selectively inhibit the protein-stimulated ATPase of SecA in the protein translocation (Wolk et al., 1997), most likely by inhibiting the membrane SecA cycling or preventing the release of the nucleotides. The membrane inserted SecA was stabilized in the presence of azide and the de-insertion reaction of SecA was blocked by inhibiting the release of ADP (Vasilyeva et al., 1982; Minkov and Strotmann, 1989; van der Wolk et al., 1997). SecA protein prepared from the *azi*^{R} mutants were found to reduce the SecA affinity for ADP (Schmidt and Kiser., 1999). SecB, which is a protein enhancing the membrane-associated ATPase activity of SecA by regulating the conformation of the nucleotide binding site to assemble an ATP-bound form. The enhancement is weakened by azide and the presence of SecB was shown to increase the SecA susceptibility to azide (Miller et al., 2002). However, the concentration of azide required for inhibiting the translocation is lower than required for partially inhibiting the ATP turnover in vivo (Oliver et al., 1990), suggesting the inhibition of azide on SecA needs to be further identified.

To this end, recent studies used TraDIS to screen a mutagenesis library for mutants that affect the susceptibility to azide (Chandler et al., 2018; Cranford-Smith et al., 2020). The transposon insertions in the Sec components, such as YajC, were shown to increase susceptibility to azide, however, insertions in the CTT of SecA were shown to be more resistant to azide (Chandler et al., 2018). The CTT of SecA is responsible for SecA interacting with ribosome, metal ions, SecB and phospholipids (Breukink et al., 1995), as well as altering the conformation of SecA. Sodium azide is proposed to disrupt the CTT, subsequently inhibiting MBD of the CTT binding to metal ions (Chandler et al., 2018). Further study proved that azide disrupts the iron binding to the MBD of SecA, without interfering with iron binding to other proteins, such as YecA and YchJ which containing SecA-like MBDs (Chandler et al., 2018; Cranford-Smith et al., 2020). Sodium azide is also shown to prevent SecA from interacting with the membrane phospholipids without inhibiting the nucleotide exchange (Chandler et al., 2018). Therefore, apart from inhibiting the rate of ATP turn over, sodium azide can also inhibit the SecA ATPase activity by disrupting the binding of SecA to the cytoplasmic membrane.

1.6 Project aims

The first aim of this work was to understand the physiological difference between SHuffle strains and their isogenic parental strains by comparison of the essential genes of each strain generated by TraDIS. The transposon libraries created were further used to analyze the temperature sensitivity of SHuffle strains, especially for SHuffle B. The second aim of this work was to identify the conditional essential genes in strains containing mutations in *tig* and *dsbC* compared to the model *E. coli* K-12 strain MG1655 to identify the possible function of TF and DsbC. In addition, the transposon mutations that conferred azide resistance were isolated to understand the molecular mechanisms of azide resistance, which was the final aim.

Chapter 2
Materials and Methods

2.1 Bacteria strains and plasmids

The bacteria strains and plasmids used in this study are listed in Table 2.1 and 2.2, respectively. SHuffle T7 express K12 strains, SHuffle T7 express B strains and its isogenic parent strains in this work were provided by New England Biolabs (NEB). DHB4 and MG1655 strains were lab stocks.

Lab stocks were made with 500 µl overnight culture and 250 µl sterilized 50% glycerol then stored at -80 °C. The reference genomes are obtained from NCBI database: NEB T7 express (C2566; CP014268), SHuffle-B (C3029; CP014269), DHB4 (CP014270), and its cognate F′ episome (CP014271), SHuffle K-12 (C3026; CP014272) and its cognate F′ episome (CP014273), and MG1655 (NC_000913.3) (Anton et al., 2016; Riley et al., 2006).

Table 2.1 Strains used in this study

Strain	**Description**	**References/Source**
SHuffle® T7 Express Competent *E. coli*	*fhuA2 lacZ::T7 gene1 [lon] ompT ahpC gal λatt::pNEB3-r1 cDsbC* (Spec^R, *lacIq*) *ΔtrxB sulA11 R(mcr-73::miniTn10*--Tet^S*)2 [dcm] R(zgb-210::Tn10* --Tet^S*) endA1 Δgor Δ(mcrC-mrr)114::IS10*	NEB # C3029
SHuffle® T7 Competent *E. coli*	F′ *lac, pro, lacIq / Δ(ara-leu)7697 araD139 fhuA2 lacZ::T7 gene1 Δ(phoA)PvuII phoR ahpC* galE (or U) galK λatt::pNEB3-r1-cDsbC* (Spec^R, *lacIq*) *ΔtrxB rpsL150*(Str^R) *Δgor Δ(malF)3*	NEB # C3026
T7 Express Competent *E. coli*	*fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10*--Tet^S*)2 [dcm] R(zgb-210::Tn10*--Tet^S*) endA1 Δ(mcrC-mrr)114::IS10*	NEB # C2566

DHB4	DHB4	Lab stock
E. coli DH5α	F– *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* φ80d*lac*ZΔM15 Δ(l*acZYA-argF*)U169, hsdR17(*rK*–*mK*+), λ–	Lab stock
DRH-503	MC4100 F- *araΔ139 DlacU169 rpsL150 thi rbsR*	Lab stock
DRH-545	BL21 DE3+pCA597(Strep-SUMO)-SecA(nt1-2700)	Lab stock
DRH-584	BL21 DE3 + pCA528-SecA-biotin	Lab stock
DRH-717	DH5a + Δ*secA*::KanR + pTrcSpc-SecA-biotin PhoAss-Met-FLAG-PhoA	Lab stock
DRH-731	DHB4 + pTrc99a	Lab stock
DRH-736	DHB4 +Δ*secA*::KanR + pTrcSpc-SecA-biotin Δ*phoAss*::*secMarrest* -*secM*	Lab stock
DRH-855	MG1655	Lab stock
DRH-856	MG1655 Δ*tig*::Kan	Lab stock
DRH-952	MG1655 Δ*dsbC*::Kan	Lab stock
DRH-997	MC4100 *azi-D1. azi*^{R}	
DRH-998	MC4100 *azi-D2. azi*^{R}	Lab stock
DRH-1027	BW25113	Lab stock
DRH-1064	MC4100 Ara+ *secA827*::IS*1*..*leuA*::Tn*10*	Lab stock
DRH-1071	DRH1064 Leu+..*secA827*::IS*1* (truncated SecA)	Lab stock
DRH-1103	BW25113 *secA207. azi*^{R}	Lab stock
DRH-1116	BW25113 *azi-D1. azi*^{R}	Lab stock
DRH-1122	BW25113 *secA219. azi*^{R}	Lab stock
DRH1127	BW25113 *azi-D2. azi*^{R}	Lab stock
DRH1292	MG1655 Δ*tig* Δ*dsbC*	
SHuffle B/DHB4 Δ*aroB*	SHuffle B/DHB4 Δ*aroB*::Kan	this study*
SHuffle B/DHB4 Δ*dgkA*	SHuffle B/DHB4 Δ*dgkA*::Kan	This study*

SHuffle B/DHB4 Δ*cyoA*	SHuffle B/DHB4 Δ*cyoA*::Kan	This study*
SHuffle B/DHB4 Δ*arcB*	SHuffle B/DHB4 Δ*arcB*::Kan	This study*
SHuffle B/DHB4 Δ*fis*	SHuffle B/DHB4 Δ*fis*::Kan	This study*
SHuffle B/DHB4 Δ*glpR*	SHuffle B/DHB4 Δ*glpR*::Kan	This study*
SHuffle B/DHB4 Δ*prlC*	SHuffle B/DHB4 Δ*prlC*::Kan	This study*
SHuffle B/DHB4 Δ*fre*	SHuffle B/DHB4 Δ*fre*::Kan	This study*
SHuffle B/DHB4 Δ*fur*	SHuffle B/DHB4 Δ*fur*::Kan	This study*
SHuffle B/DHB4 Δ*ubiI*	SHuffle B/DHB4 Δ*ubiI*::Kan	This study*
SHuffle B/DHB4 Δ*rbsR*	SHuffle B/DHB4 Δ*rbsR*::Kan	This study*
SHuffle B Δ*rbsD*	SHuffle B Δ*rbsD*::Kan	This study*
SHuffle B Δ*rbsC*	SHuffle B Δ*rbsC*::Kan	This study*
SHuffle B Δ*rbsA*	SHuffle B Δ*rbsA*::Kan	This study*
TR1	SHuffle B *prlC*::miniTn5 (--Kan^R)..*gor*	This study
TR2,5	SHuffle B *dmsA*::miniTn5 (--Kan^R)..*trxB* (insertions in different orientation)	This study
TR3	SHuffle B *ycaK*::miniTn5 (--Kan^R)..*gor*	This study
TR4	SHuffle B *hdeD*::miniTn5 (--Kan^R)..*gor*	This study
TR6	SHuffle B *rbsA*::miniTn5 (--Kan^R)	This study
AR1,2,4	MG1655 *rpsQ*::miniTn5 (--Kan^R)	This study
AR3	MG1655 *rpsA*::miniTn5 (--Kan^R)	This study
AR5	MG1655 *srmB*::miniTn5 (--Kan^R)	This study
AR6	MG1655 *nusB*::miniTn5 (--Kan^R)	This study
MC736	MC4100 + pTrcSpc-SecA-biotin Δ*phoAss*::*secMarrest* -*secM*	This study
MC717	MC4100 + pTrcSpc-SecA-biotin PhoAss-Met-FLAG-PhoA	This study

* Represents strain made by P1 transduction from mutants obtained from Keio library (Baba et al., 2006).

Table 2.2 Plasmids used in this study

Plasmid	Description	Reference
Pcp20	Contains FLP recombinase. Temperature sensitive plasmid cured at 37°C	(Cherepanov and Wachernagel, 1995)
pTrc99a	Amp^R	Invitrogen
pDSW204	Amp^R	(Weiss et al., 1995)
pDH811	Expresses a signal sequence-less alkaline phosphatase (ΔssphoA) under control of an IPTG-inducible promoter, Amp^R	Lab stock

2.2 Media and growth conditions

2.2.1 Culture media

Luria broth (LB) agar/broth (Milner, 1991) were supplied by T101 lab and all chemicals were provided by Huber's Lab. LB broth was composed of 10 g/L tryptone, 5g/L yeast extract, 10 g/L NaCl and adjusted to pH = 7. LB agar was made by addition of 15 g/L agar to LB. Super optimal broth with Catabolite repression medium (SOB) was composed of 20 g/L tryptone, 5 g/L yeast extract, 10 mM NaCl and 2.5 Mm KCl, the SOC used for recovery of competent cells after electroporation was made by addition of 10 mM $MgCl_2$ and 20 mM glucose and adjust the pH = 7. Top agar was prepared using 7 g/L Agar, 10 g/L tryptone, and 8 g/L NaCl. All of the media in this study were autoclaved (15 min at 120 °C and 15 psi) before use. The pH was measured using a Jenway 3510 pH meter with a Thermo Scientific Orion 815600 ROS combination probe and adjusted using 1 M HCl or 1 M NaOH solutions. For small volume Ph-Fix 0-14 pH test strips were used.

2.2.2 Growth conditions

Overnight bacteria cultures were set up in a 30 ml plastic universal by inoculating a single colony of bacteria into 5 ml LB media and incubated in a shaking incubator (180 rpm) at appropriate temperatures. Overnight growth of cells on the agar plates were incubated in the static incubator. Media was supplemented with 200 µg/ml ampicillin, 25 µg/ml chloramphenicol or 50 µg/ml kanamycin to select the antibiotic markers. Stock solution of the antibiotics used in this study are listed in Table 2.3. All of the stock solutions were filter sterilized by 0.22 µm syringe filters. Top agar was melted at 100 °C water bath and kept at 55-60 °C before used for P1 lysis.

Table 2.3 Stocks of antibiotics used in this study

Antibiotic	**Stock concentration**	**Solvent**	**Storage**
Kanamycin sulfate	50 mg/ml	Water	-4 °C
Chloramphenicol	25 mg/ml	Ethanol	-20 °C
Ampicillin	100 mg/ml	Water	-20 °C
Rifampicin	50 mg/ml	Water	-20 °C
Streptomycin	100 mg/ml	Water	-20 °C
Neamine	1 mg/ml	Water	-4 °C
Sodium azide	0.5 M	Water	RT

2.2.3 96-well plate growth curve

Greiner bio-one 96-well F-bottom plates were used to grow cells overnight with shaking at 37 °C in a CLARIOostar 96-well plate reader. 200 µl LB broth was inoculated by overnight cultures at a starting OD600 = 0.02. The optical density was recorded every half an hour for 18 hrs.

2.3 Buffers

The buffers used in this study are listed in Table 2.4.

Table 2.4 Buffers used in this study

Buffer	Components
Phosphate buffered Saline (PBS; 10X)	80 g/L NaCl, 2 g/L KCl, 14.4 g/L NA_2HPO_4, 2.4 g/L KH_2PO_4; pH 7.4
TAE Buffer (1X)	40 mM Tris, 20 mM glacial acetic acid, 1 mM EDTA
Laemmli Sample Buffer (5X)	0.02% (w/v) bromophenol-blue, 30% (v/v) glycerol, 10% (w/v) SDS and 250 mM Tris-HCl; pH 6.8
SDS Running Buffer (1X)	25 mM Tris, 192 mM glycine, 0.1 % (w/v) SDS
Transfer Buffer	25 mM Tris, 192 mM glycine, 20% methanol
DNA buffer for SPRI beads	48 ml Nuclease-free water, 0.5 ml 1M Tris base, 0.5 ml 0.1 M EDTA, 0.25 ml 10% (v/v) Tween 20, 0.186 ml 1N HCl
TBS	2.4 g/L Tris base, 8.8 g/L NaCl, pH 7.8

2.4 General microbiology techniques

2.4.1 Miniprep of plasmid DNA

Plasmid DNA was prepped from an overnight culture growing in the shaking incubator (180 rpm) at an appropriate temperature using the GeneJET plasmid Miniprep Kit (ThermoFisher) following the manufacturer's instructions. Overnight culture was pelleted by centrifugation at 13000 rpm and resuspended in 250 µl of the resuspension solution. The lysis solution and neutralization solution were then added to the mixture. After centrifuging to pellet cell debris and chromosomal DNA, the supernatant was transferred to the supplied GeneJET spin column and centrifuged. The flow-through was discarded and the column was put back into the collection tube. Wash solution with ethanol was added to the column twice, after centrifuging the column and removing the residual ethanol, the plasmid was eluted by elution buffer and stored at -20 °C.

2.4.2 DNA concentration determination

Qubit™ HS dsDNA assay kit (Invitrogen) was used to quantify the DNA concentration accurately. The assay was performed at room temperature (RT). The two standards were made by adding 189 µl of master mix of Qubit buffer, 1 µl of the Qubit dye and 10 µl prediluted DNA standard. Samples were prepared using 198 µl master mix of Qubit buffer, 1 µl of the Qubit dye and 1µl DNA. The Qubit assay tubes were mixed by vortex and incubated at RT for 2 min before reading. The DNA concentration was measured in µg/µl.

The concentration of DNA prepared for Sanger sequencing was determined by Nanodrop™ ND-1000 Spectrophotometer (ThermoFisher), measuring light absorbance at 260 nm by placing 1-2 µl of prepped DNA on the pedestal.

2.4.3 Restriction digestion of DNA

The PCR product and plasmids were digested using restriction enzymes, NcoI-HF, BamHI-HF, BspHI (NEB) in this study. The reaction is consisting of 20 µl DNA plasmid, 16 µl nuclease-free water, 4 µl 10X rCutSmart Buffer™ (NEB) and 1µl of each digestion enzymes. The mixture was incubated at 37 °C for 3 hours or overnight. The product was purified by PCR purification (Section 2.4.4) and kept at -20 °C for further use.

2.4.4 DNA ligation

Reaction was set up on ice and consisting of 2-4 µl vector (cut plasmid), 4-8 µl DNA insert, which is the amplified gene of interest using primers containing sequences of the enzyme restriction sites, 2 µl 10X T4 DNA ligase buffer 1 µl T4 DNA ligase (NEB) and nuclease-free water made up to 20 µl in total. The reaction was incubated at 20 °C overnight. The negative control contains only vectors without any insert.

2.4.5 DNA precipitation

DNA precipitation was used to remove salts in the DNA buffer from the ligation products before electroporation. It used 1:1 volume of 7.5 M NH_4OAc to the ligated

DNA and 2 volumes of isopropanol to wash the ligation product. The reaction was mixed and centrifuged for 15 min at RT. The supernatant was discarded, and the pellet was washed with 75 µl 70% ethanol and centrifuged again to discard the supernatant. The residual ethanol was evaporated using Concentrator 5301 (Eppendorf) at RT for 5-10 min. 10 µl nuclease-free water was used to resuspend the DNA pellet and incubated at 50 °C for 5-10 min to ensure the resuspension.

2.5 Polymerase Chain Reaction (PCR) and sequencing

2.5.1 Oligonucleotides

Oligonucleotides were synthesized by either ThermoFisher Scientific or Sigma-Aldrich (Merck). The primers were resuspended at a concentration of 100 µM using nuclease-free water (Invitrogen) and diluted to 10 µM before use. The primers were stored at -20 °C for further use. Oligos used in this study are listed in Table 2.5.

Table 2.5 Oligonucleotides used in this study

Primer	Sequence (5' – 3')	Description
TKK_F	ACCTGCAGGCATGCAAGCTTCAGG	Forward primer for PCR1 in the TraDIS library preparation
TKK_R	GACTGGAGTTCAGACGTGTGCTCTT CCGATC	Reverse primer for PCR1 in the TraDIS library preparation
TKK_6.1	AATGATACGGCGACCACCGAGATCT ACACTCTTTCCCTACACGACGCTCT TCCGATCTCGTACGAGCTTCAGGGT TGAGAT GTGTA	Inline barcode primers for PCR2 in the TraDIS library preparation, the barcode sequences are underlined
TKK_7.4	AATGATACGGCGACCACCGAGATCT ACACTCTTTCCCTACACGACGCTCT TCCGATCTTAGCTAGAGCTTCAGGG	

	TTGAGA TGTGTA	
TKK_8.3	AATGATACGGCGACCACCGAGATCT ACACTCTTTCCCTACACGACGCTCT TCCGATCTCATGCATGAGCTTCAGG GTTGAG ATGTGTA	
TKK_9.2	AATGATACGGCGACCACCGAGATCT ACACTCTTTCCCTACACGACGCTCT TCCGATCTATCGATCGAAGCTTCAG GGTTGA GATGTGTA	
*visC*_F	CGTTGAAGGCGGTTTTAAGGA	To confirm the mutants obtained from Keio library contain mutations in desired genes and confirm the mutations have been transduced into the recipient strain
*visC*_R	TGGGTATCAAATAAACAACAGAGG	
*fur*_F	TTAGTAACAGGACAGATTCCGC	
*fur*_R	TTGGCTTTTCTCGTTCAGGC	
*fre*_F	TTGCCCTATTTATCGATCCG	
*fre*_R	TTTAGTTGCCGTTCTTCCCG	
*cyoA*_F	TTAAACGCCACCAGATCCCG	
*cyoA*_R	ATCTTTATTCTTCCTCAACCCC	
*cyoE*_F	ATGATGATGCACTAAGAGCG	
*cyoE*_R	AAGAAAAAAGGTGCCGTAGC	
*aroB*_F	AGTAATTAAGGTGGATGTCGC	
*aroB*_R	ATAAGCGGCCTGACCTTTC	
*dgkA*_F	TGACGCTCATAATGTAAAAAGG	
*dgkA*_R	AGATTTATCGAATTCTGGAAGG	
*prlC*_F	TCTTTAACCAGGACTGCGC	
*prlC*_R	TGCCTGTTTCATCAATTAAGC	
*glpR*_F	TCGCTCAATGCGCGAAAACG	
*glpR*_R	TTTTGGTCTGACGTGGGAAGC	
*basS*_F	TATATGCTGGTCGCGAATGAGG	
*basS*_R	AACTACCGTGTTCAGCGTGC	
*fis*_F	TGCGTAAACAGAAATAAAGAGC	
*fis*_R	AATCAAGCATTTAGCTAACCTG	
*arcB*_F	TTATTGGGGCAGGTTGTCGTG	

*arcB*_R	TGTTGGCGCAGTATTCGCG	
*rbsR*_F	CGACGCATTTTTAGACAGG	
*rbsR*_R	TTCAGCAGCAATGTTCATGC	
*rbsA*_F_NcoI	CGCGGAACCATGGAAGCATTACTTCAGC	Plasmid cloning of genes in the *rbs* operon
*rbsA*_R_BamHI	GCGAGCGGATCCTTACTCCTAGATTCACGCG	
*rbsD*_F_BspHI	CGACTCATGAAAAAAGGCACCGTTCTTAATTC	
*rbsD*_R_BamHI	AAATCAGGATCCTCAGAACGTCACGCCAGCAC	
*rbs*_F_BspHI	CCAGGATCATGAAAAAAGGCACCGTTCTTAATTCTG	
*rbs*_R_BamHI	AAATAAGGATCCCTAAGCCGAACCGCGTTCCATC	
*rbsK*_F_NcoI	CACAAACCATGAAATGCAAAACGCAGGCAGCCTCG	
*rbsK*_R_BamHI	ACGCGAGGATTCTCACCTCTGCCTGTCTAAAAATGC	
KAN-2_F	ACCTACAACAAAGCTCTCATCAACC	To identify the transposon insertion sites
KAN-2_R	GCAATGTAACATCAGAGATTTTGAG	
HIB17	CGGAATTCCGGATNGAYKSNGGNTC	
*rpsA*_F	GATCAGAAGCGCAACAACG	To confirm the transposon insertions in *rpsA* and *rps*Q gene
*rpsA*_R	GCTTGATTACAGGACGAAACC	
*rpsQ*_F	CGAGAACAACGAATGCGG	
*rpsQ*_R	GCATTATAACACCGCTTCAAGG	

The primer names "_F" or "_R" refer to the forward and reverse primers.

2.5.2 Colony PCR

Linear DNA fragments for cloning were amplified using Q5® High-Fidelity DNA Polymerase combined with dNTP mix and 5% (v/v) DMSO. The colony PCR used 25 µl MyTaq Red Mix (Bioline), 2 µl forward and reverse primers, single colony of interest or extracted DNA and nuclease-free water (Invitrogen) up to total volume of 50 µl. The primers used in this study are listed in Table 2.3 and the PCR steps are listed in Table

2.6.

Table 2.6 Colony PCR steps

Step	Temperature	Cycle	Time
Initial denaturation	98 °C	1	30 s
Denaturation	98 °C	30-35	10 s
Anealing	50-60 °C		30 s
Extension	72 °C		30 s/kb
Final extension	72 °C	1	2-3 min
Hold	4 °C	-	Hold

2.5.3 Touchdown PCR

The preparation of the touchdown PCR followed (Del Aguila et al., 2005), which is used to determine the transposon insertion sites. The azide-resistant mutants (AR1-6) were re-streaked from the -80 °C stock and incubated overnight before use. The fresh colony was boiled in 100 µl nuclease-free water and centrifuged at 13000 rpm for 30 s to get the supernatant. The PCR used the degenerate primer HIB17 combined with either the KAN-2_F or KAN-2_R (Table 2.3) to amplify the transposon insertions. The PCR reaction included 25µl MyTaq Red Mix (Bioline), 10 µl HIB17 (10 µM), 2 µl of either KAN-2_F or KAN-2_R (10 µM), 5 µl boiled DNA preparation of the mutant colony and nuclease-free water to 50 µl. The touchdown PCR steps are listed in Table 2.8.

Table 2.7 Touchdown PCR steps

<table>
<tr><th>Temperature</th><th>Cycle</th><th>Time</th></tr>
<tr><td>95 °C</td><td>1</td><td>5 min</td></tr>
<tr><td>95 °C</td><td rowspan="3">25</td><td>45 s</td></tr>
<tr><td>60 °C (drop by 1 °C each cycle)</td><td>1 min</td></tr>
<tr><td>72 °C</td><td>2 min</td></tr>
<tr><td>95 °C</td><td rowspan="3">25</td><td>45 s</td></tr>
<tr><td>45 °C</td><td>45 s</td></tr>
<tr><td>72 °C</td><td>2 min</td></tr>
<tr><td>4 °C</td><td>-</td><td>Hold</td></tr>
</table>

2.5.4 PCR purification

The purification of the PCR products followed the instruction on the protocol in QIAquick® PCR purification kit. 5 volumes of the buffer PB (125 µl) were added to the PCR sample and the mixture was transferred to a QIAquick spin column in a 2 ml collection tube. The column was centrifuged for 60 seconds. The filtrate was discarded and the column was placed back to the collection tube. 750 µl buffer PE was added to the column and centrifuged for 60 seconds. The filtrate was discarded, and the column was centrifuged for another 1 min to remove the residual ethanol in buffer PE. 50 µl nuclease-free water was added to the column, the column was placed in a new 1.5 ml receiver tube and incubated for 1min. The column was centrifuged for 1 min to collect the eluted DNA.

2.5.5 Agarose gel electrophoresis

The DNA fragments were loaded onto a 1% (w/v) agarose gel (Bioline) in 1x Tris-acetate-EDTA (TAE) buffer with 2.5 µl Midori Green (Nippon Genetics) or 2 µl SYBR® Safe (Invitrogen). The concentration of the agarose gel depended on the expected size of the fragments. A 2% (w/v) agarose gel was used for higher resolution. The gel was submerged in the 1x TAE buffer before loading ladders and samples. 5-10 µl of the PCR product were loaded, and 5 µl 1 kb or 100 bp hyperladder (Bioline) were loaded onto the gel as size references of the DNA fragments. The gel was run at 110 V for 30-40 min depending on the size of the DNA fragment and the size of the agarose gel. Gels were then visualized in a gel documentation (Biorad Gel Doc XR+) system.

2.5.6 Sanger sequencing

The purified PCR product and plasmid DNA were sequence by Sanger sequencing when required using the service from Source Biosciences. The samples were prepared to contain 10 µl of 10 ng/µl PCR products and 5 µl of 3.2 pmol/µl primers following the company's instructions.

2.5.7 Whole genome sequencing

The whole genome sequencing in this study was performed by MicrobesNG (https://microbesng.com). The samples were prepared following the strain submission instructions. A single colony of strain for sequencing was resuspended in PBS and spread on the LB agar plates to get a lawn covering half or two third of the plate. The lawn was harvest and resuspended in PBS, 4 x 10^9 – 6 x 10^9 cells were required for sequencing. The cells were washed by PBS and resuspended in the inactivation buffer in the sample tube provided.

2.6 Bacteria transformation

2.6.1 TSS chemical transformation

TSS transformation was performed as described (Miller, 1992). 5 ml LB media was inoculated by a single colony of the recipient bacteria and grown in the 37 °C shaking incubator to early exponential bacterial growth phase (OD 0.2-0.3), the cell culture turned to be slightly cloudy. The culture was placed on ice to stop the incubation. 100 µL 2X TSS reagent (5 g 10% Polyethylene glycol, 2.5 ml 5% Dimethyl Sulfoxide, 1 ml 20mM MgCl2 in 50 ml sterile LB) and 4 µl of desired plasmid were added to 100 µl cell culture in 1.5 ml Eppendorf tube and mixed by pipetting. The mixture was incubated on ice for 10 min. Heat shock was performed in 42 °C (30 °C if the plasmid using pCP20) water bath for 1 min. 500 µl LB broth was added and the mixture was incubated at 37 °C (30 °C if the plasmid using pCP20) for 30-40 min. Cells were harvested by centrifugation at 13000 rpm for 1 min. Most of the supernatant was discarded, cells were resuspended with the residual supernatant (~200 µl). Cells were plated on the LB plates with appropriate antibiotic and grown at appropriate temperature (normally at 37 °C). Blank controls were set up as overnight culture only and plasmid only. Colonies grown on the selective plates were re-streaked on the selective plates to confirm the antibiotic resistance. The plasmid pCP20 was used for removal of the kanamycin cassette.

2.6.2 Preparation of electrocompetent cells

Cells were made competent for electroporation followed the protocol as described (Chung and Miller, 1993). Each step of the protocol was performed on ice. 5 ml of overnight culture was inoculated with a single colony and grown at appropriate temperature in the shaking incubator. 500 ml LB broth was inoculated with 5 ml overnight culture and grown to reach OD_{600} = 0.5-0.7. The bacterial culture was split equally into 50 ml pre-cooled falcons on ice and centrifuged at 4000 XG at 4 °C for 10 min. The pellets were washed step-by step with the same volume of ice-cold distilled water (50 ml), 1/3 volume of ice-cold distilled water (17 ml) and 1/50 volume of 10% glycerol (1 ml). The pellets were finally resuspended with 1/1000 volume of 10% glycerol (400 µl) and aliquoted with 100 µl each into pre-cooled sterilized 1.5 ml eppendorf tubes then stored at -80 °C.

2.6.3 Electroporation

Electroporation cuvettes and sterilized eppendorfs were pre-chilled on ice before use. The Eppendorf Thermomixer® C was pre-warmed to 30 °C (SHuffle strains) or 37 °C (other strains like MG1655). 30 µl of cells were transferred along the side of the cuvette without inducing bubbles. 0.1 µl of Tn5 transposon was added to the cells and mixed by flicking gently. The cells were observed deposited across the bottom of the cuvettes. The cuvettes containing mixture were then placed in the Eppendorf Eporator® and electroporated using 1750 V. 1 ml SOC recovery medium was added immediately to the cuvettes, cells were gently mixed up and down several times and transferred to 1.5 mL microcentrifuge tube. The ideal time constant is ~5.0-6.0 ms. Cell cultures were shaken vigorously (350 rpm) at an appropriate temperature for 1-2 hours. Cells were diluted as appropriate (1:5 in this experiment) then 200 µl cells were spread onto selective plates (kanamycin plates were used in this study). The plates were incubated overnight in the static incubator under appropriate temperature. Blank control was the electroporation of 30 µL of cells without inserted DNA recovered in 1 mL SOC media.

2.7 P1 transduction

2.7.1 P1 liquid lysate

The subcultures for lysis were inoculated by adding 50 µl of the overnight culture to 5 mL LB media. 125 µl of 1 M sterilized CaCl2 was added to facilitate the attachment of the bacteriophage. Different volumes of the P1 phage stock (1 µl, 2 µl and 5 µl) were addition to the subcultures. Cultures were incubated at proper temperature until the lysis occurred. Culture with no P1 phage was set up as a blank control which would undertake normal bacteria growth and turn to be cloudy. The P1 lysate was collected until the cell culture turned to be clear, 5 drops of chloroform were added and the stock was stored at 4 °C for P1 transduction.

2.7.2 P1 plate lysate

5 ml LB broth was inoculated with the donor strain and grown overnight in the shaking incubator at appropriate temperature. Inoculating 50 µl of the overnight culture to 5 mL LB media and the subculture was incubated to the late log phase (OD_{600} ~0.8). 50 µl of 0.5 M $CaCl_2$ was added to the cell culture and grown for 10 min. 1 ml of the cell culture was transferred to the 1.5 ml microcentrifuge tube and different volume of P1 phage (50 µl, 100 µl and 200 µl) was added to the mixture and incubated for 20 min. The cell

culture was mixed in the tubes containing 2 drops of 20% glucose, 2 drops of 0.5 M $CaCl_2$ and 2 ml H top agar and the mixture was poured onto a LB plate and grown overnight at appropriate temperature. To harvest the lysates, 2 ml LB was added to the plate. The top agar was scraped from each plate by sterile spreader and transferred to the falcons. The scraped top agar was centrifuged at 10K rpm and the debris was removed. Several drops of chloroform were added and the lysate was stored at 4°C.

2.7.3 P1 transduction

Overnight cultures of recipient strains were prepared and grown at an appropriate temperature in the previous day. 500 µl of the overnight was centrifuged and the pellet was resuspended in 500 µl of solution containing 100 mM CaCl2 and 10 mM MgCl2. 100 µl cells were mixed with 100 µl P1 lysate (1:10 dilution) and incubated for 20 min at appropriate temperature without shaking (Miller, 1972). The blank controls were cells with no P1 lysate and P1 lysate without cells. 200 µl of 1 M NaCitrate to stop the reaction and 0.5 ml LB broth were added for cell growth. The mixture was incubated in the thermocycler for 1 hour with continuous shaking at 350 rpm. The cell culture was centrifuged and the pellet was resuspended by 200 µl of 1 M NaCitrate then plated onto the antibiotic selective plate as well as the overnight culture and lysate alone as blank controls. No growth of colonies on the control plates indicated no contamination in the bacterial culture or the P1 lysate used.

2.8 TraDIS

The TraDIS protocol used in this study was optimized by Ian Henderson's lab and previously described in (Langridge et al., 2009; Goodall et al., 2018, Milner, 2022). The analysis scripts and the GFF3 files parsing were implemented by Mathew Milner.

2.8.1 Construction of the transposon library

The construction of the transposon mutagenesis library was achieved by transformation with transposome. 0.1 µl of EZ-Tn5TM <KAN-2>Tnp transposome was added to 30 µl of the electrocompetent cells (Section 2.5.2) and the mixture was transferred into the 1 mm electroporation cuvette. After incubation for 30 min on ice, the electroporation was performed (Section 2.5.3). The transformants were counted and pooled to get a desired population. 50% glycerol was added to the collected library to obtain a final concentration of 15% glycerol and stored at -80 °C.

2.8.2 Genomic DNA extraction

The genomic DNA extraction was performed following the protocol of the Stratec RTP Bacteria DNA Mini kit for genomic extraction of bacterial cells. The OD_{600} of the mutagenesis library was measured in 1:1000 dilution, the ideal OD_{600} of the cells is 1 (1 x 10^9 cells) to undertake the DNA extraction. The proper amount of library was transferred into 1.5 ml centrifuge tube and the supernatant was discarded. The cell pellet was resuspended in 400 µl of the resuspension buffer R and mixed by pipetting up and down. The mixture was transferred into the extraction tube containing preformulated solid lysis reagent, lytic enzyme, proteinase K, carrier nucleic acid and internal DNA extraction control. Samples were mixed by vortexing and incubated in the thermomixer at 65 °C for 10 min with continuous shaking at 300 rpm. The extraction tube was then incubated at 95 °C for 5-10 min with continuous shaking at 300 rpm. 400 µl binding buffer B6 was added to the sample and vortex shortly. The mixture was loaded onto the RTA spin filter set and incubated for 1 min then centrifuged at 11.000 rpm in a standard centrifuge for 2 min. The filtrate was discarded, and the spin filter was put back into the receiver tube. 500 µl wash buffer I was added to the filter set and centrifuged at 11.000 rpm for 1 min. The filtrate and the receiver tube were discarded. The RTA spin filter was placed into a new receiver tube. 600 µl of the wash buffer II was added and centrifuged at 11.000 rpm for 1 min. The filtrate was discarded. A maximum of 4 min centrifuge was performed to remove the ethanol remaining in the DNA sample completely. The RTA spin filter was placed into a new 1.5 mL receiver tube. 200 µl of low salt elution buffer was added into the DNA sample and incubated for 1 min at RT. The genomic DNA was collected by centrifugation at 11.000 rpm for 1 min.

2.8.3 DNA fragmentation

The proper volume of DNA was calculated to get the final concentration of 2 μg/μl DNA in 500 μl nuclease-free water. The mixture was spun to collect the liquid on the side of the falcon and kept on ice for approximately 10 min before shearing. A Diagenode Bioruptor sonication device was turned on 30 min before running the sample. The water level in the water bath was maintained above the minimum lane. The probes were cleaned by sterilized tissue with ethanol and put in the center of the falcon. Samples were loaded and balanced, blank 500 μL water was used if needed. The sonicator was adjusted with the instructions: 30 seconds on, 90 seconds off, low intensity and 13 cycles. Samples after shearing were collected by spinning down in a quick and transferred into the 1.5 ml Eppendorf tubes. The transferred DNA samples were condensed to 50 μl using a vacuum concentrator (Eppendorf). The final volume was made up to 55.5 μl with nuclease-free water and stored at -20 °C.

2.8.4 Construction of SPRI beads

The fragmented DNA was size selected and cleaned using solid-phase reversible immobilization (SPRI). SPRI beads were used to separate specific sizes of DNA from the samples relying on different concentration of beads added. The beads were lab-made from Sera-Mag beads following the protocol described by (Jolivet and Foley, 2015). 1-1.5 ml stock beads were mixed well and transferred to the 1.5 ml centrifuge tube, and the tube was placed on the magnet stand until the supernatant was clear. The supernatant was discarded, and the beads were mixed with a DNA buffer (Table 2.4) by vortexing. The mixture was placed on the magnetic stand and the supernatant was removed. This wash step was repeated 3 times and the last supernatant was kept to mix with 3.528 ml nuclease-free water, 25 ml 5 M NaCl, 0.168 ml 1 N HCl as well as 0.5 ml 1M Tris base and 0.5 ml 0.1 M EDTA. The mixture was placed on the magnet stand, the supernatant was removed. This step was repeated once again but keep the mixture and 20 ml of 50% PEG 8000 was added followed by 0.25 ml of 10% (v/v) Tween 20. The mixture was mixed until the color became homogeneous and stored at 4 °C. The constructed SPRI beads were confirmed for function using 1 kb hyperladder (Bioline) and compared to the AMPure XP beads (Beckman Coulter), which are commercially used beads.

2.8.5 NEBNext DNA library preparation for Illumina

NEBNext Ultra DNA Library Prep Kit for Illumina (E7370) was used in this study and followed steps shown in Figure 2.1. All of the reagents and enzymes were kept on ice. 55.5 µl fragmented DNA was added into a sterile nuclease-free PCR tube combined with 3.0 µl end prep enzyme and 6.5 µl 10X end repair reaction buffer then mixed by pipetting. The tube was placed in a thermocycler with the heated lid on and incubated at 20 °C for 30 min and 65 °C for 30 min then held on ice. 15 µl Blunt/TA ligase master mix, 2.5 µl NEBNext adaptor for illumina and 1 µl ligation enhancer was added into the end prep reaction mixture and mixed well. The mixture was incubated at 20 °C for 15 min in the thermocycler. 3 µl of USER enzyme was then added to the ligation mixture and mixed well. Sample was incubated at 37 °C for 15 min.

The SPRI beads (Section 2.7.4) were removed from fridge half an hour before the size selection and vortexed thoroughly before use. 13.5 µl nuclease-free water was added to the ligation reaction to make up 100 µl final volume. 44 µl of the resuspended beads were added to the sample and transferred into a 1.5 ml Eppendorf tube. The mixture was incubated at RT for 5 min. The tube was placed on an appropriate magnetic stand to separate the beads from the supernatant. After the solution is clear (2-5 min), the supernatant containing the desired DNA was carefully transferred to a new tube. The beads were discarded with unwanted large DNA fragments. 20 µl resuspended beads were added to the supernatant and mixed well by pipetting then incubated at RT for 5 min. The tube was placed on the magnetic stand to separate the beads from the supernatant. The supernatant was carefully removed with unwanted DNA after the solution turned out to be clear. 200 µl of 80% freshly made ethanol was used to wash the beads twice for 30 seconds and carefully discarded using a 200 µl pipette. The residual ethanol was removed by pipetting and air dried for maximum 5 min. The tube was removed from the magnetic stand and 17 µl elution buffer EB (Qiagen) was used to elute the DNA from the beads. The mixture was mixed by pipetting and incubated for 2 min at RT. The tube was placed on the magnetic stand and 15 µl DNA solution was transferred to a new PCR tube after the solution is clear.

The beads selected DNA (~250 kb) containing transposon and a piece of gDNA was amplified by the 1st PCR reaction. The reaction solution contains 25 µl of NEBNext Q5 Hot Start Hifi PCR Master Mix, 2.5 µl 10 µM TKK_F and TKK_R (Table 2.3), 15 µl beads selected DNA, and 5 µl nuclease-free water. The reaction followed the steps described in Table 2.8.

Table 2.8 1st PCR reaction steps for library preparation

Step	**Temperature**	**Cycle**	**Time**
Initial denaturation	98 °C	1	48 s
Denaturation	98 °C		15 s
Annealing	65 °C	10	30 s
Extension	72 °C		30 s
Final extension	72 °C	1	1 min
Hold	4 °C	-	Hold

The PCR product was purified using SPRI beads as described above and eluted with 17 µl buffer EB. The 2nd PCR was performed to add custom inline barcode, Illumina barcode (Table 2.3), P5 and P7 homology sequences. The 2nd PCR reaction included 25 µl of NEBNext Q5 Hot Start Hifi PCR Master Mix, 2.5 µl 10 µM TKK barcode and NEBnext Multiplex Oligos for Illumina (Table 2.3), 15 µl beads selected DNA, and 5 µl nuclease-free water. The reaction followed the steps described in Table 2.9.

Table 2.9 2nd PCR reaction steps for library preparation

Step	Temperature	Cycle	Time
Initial denaturation	98 °C	1	48 s
Denaturation	98 °C	20	15 s
Annealing	65 °C		30 s
Extension	72 °C		30 s
Final extension	72 °C	1	1 min
Hold	4 °C	-	Hold

The PCR product was purified and size selected by SPRI beads followed by the protocol stated earlier in this section, but with a final elution of 33 μl buffer EB. 32 μl of the purified DNA was transferred into a 1.5 ml Eppendorf and stored at -20 °C.

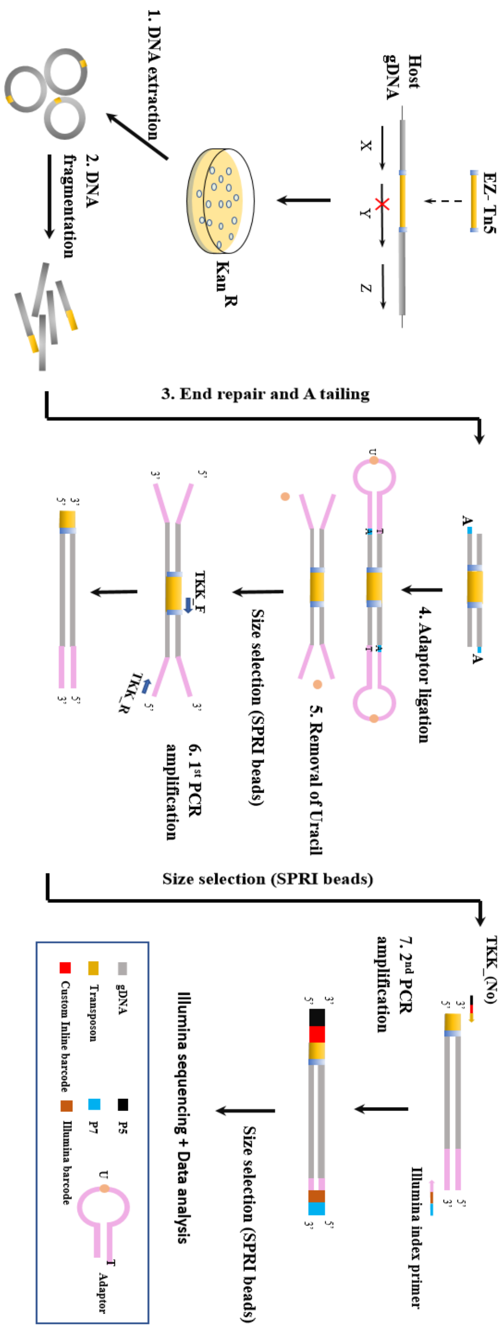


Figure 2.1 Overview the library preparation of TraDIS for Illumina sequencing. Genomic DNA is extracted from the transposon library collected from selective LB agar plates and fragmented by sonicating. The fragment ends are repaired and add A tailing then ligated by adaptor. The transposon junction is amplified by 1st PCR using primers specific to the transposon and adaptor. The 2nd PCR is performed to add P5, custom barcode, Illumina barcode and P7 for Illumina sequencing and subsequent data analysis.

2.8.6 Quantification of the prepped library by qPCR

Genomic libraries are quantified using KAPA DNA Library Quantification Kit (Illumina). Two dilutions were used for qPCR: 1:50000, 1:5000000 and 3 replicates were made to be read out. 10 mM Tris-HCl (pH 8) was used to dilute the DNA samples. The Semi-Skirtted PCR plate was kept on a cold block in the whole process. 6 µl SYBR dye with primer master mix, 2 µl nuclease-free water and 2 µl DNA sample was added to the Semi-Skirtted PCR plate for each sample and mixed by pipetting. 6 standards and at least 2 standard controls were loaded as the standards for quantification. The plate was sealed with 8-strip optical seal caps and vortexed on plate shaker then spun to collect. Samples were quantified using the Mx3005P qPCR machine (Agilent Technologies), a brown mat covered the plates and the holes was lined up. The qPCR was performed following the conditions described in Table 2.10.

Table 2.10 QPCR steps to quantify prepped library

Step	Temperature	Cycle	Time
Initial denaturation	95 °C	1	5 min
Denaturation	95 °C	35	30 s
Annealing, extension, read fluorescence	60 °C		30 s
Dissociation melt curve	60-95 °C	1	-

2.8.7 Illumina sequencing and data analysis

The Illumina sequencing was performed by Novogene Co, Ltd, Cambridge, United Kingdom. A series of custom scripts were used to assess and trim the raw data, the sequencing analysis and essential gene prediction were performed according to the methods described in the published paper (Goodall et al, 2018). The scripts were provided on a linux server by Mathew Milner. Briefly, the sequence reads were identified and trimmed by a custom inline barcode, then filtered based on the presence of sequences of the transposon. These steps ensured the sequence reads containing transposon were mapped to the genome sequence. The reference genomes were downloaded from NCBI (Tatusova et al, 2014; Section 2.1), the location and frequency of transposon insertion sites were inspected manually using the Artemis genome browser (Rutherford et al, 2000). The circular genome diagrams with insertion sites were generated by DNA plotter. Since the SHuffle strains and *E. coli* B are not well annotated, their generic feature format (.gff) files were edited by gff3 parse to ensure the comparability with the TraDIS files. The edition involved altering the ID annotation in the gff3 files which represent the position of the genome to the MG1655 genome annotation.

2.8.8 TraDIS pipelines

Log_2 likelyhood was used as described in (Goodall et al., 2018) to determine the essential genes, the scripts 'Langridge_R_Script_FD_v1.2.R' was gifted by Mathew Milner and was used in R studio (version 3.6.0; http://www.r-project.org). The scripts distributed genes to essential, nonessential and ambiguous based on the insertion index score which corresponds to the number of unique insertion sites normalized by gene length. The essential mode was fitted to the left exponential distribution, whereas the nonessential mode fitted to the right gamma distribution. The log-likelihood score is calculated by the probability of a gene belonging to each mode, the threshold is 12-fold. If a gene is 12 times more likely to be distributed in the essential mode, then that gene was determined as essential and vice versa. The unclear genes refer to genes which the log-likelihood scores is between $\log_2(12)$ and $-\log_2(12)$. BioTraDIS is a pipeline run on the CLIMB (Barquist et al., 2016; Connor et al., 2016) based on the $\log_2$ likelyhood method but a lower threshold.

2.8.9 Read depth analysis

To determine the number of reads required for sufficient coverage of the original SHuffle B transposon library to prepare sub-libraries for TraDIS analysis under different temperatures, an equation proposed by Goodall et al (Goodall et al., 2018) was used to predict the sequencing reads needed. The scripts used were gifted by Mathew Milner and provided in Goodall's paper. Considering that all of the transposon insertions in the original library have been identified, then the maximum calculated insertions should be the total amount of insertions in the original library. The scripts run the equation multiple times to create a curve show correlations between read count and estimated insertions, and the sequence depth (estimated insertions divided by total insertions in the original library) refers to the percentage saturation. The equation below was used to estimate the number of insertions in the sub-library. S corresponds to the total insertion sites in the original library, n refers to a number of reads used.

$$\mathrm{I} = \mathrm{S} - \left(\frac{S-1}{\mathrm{S}}\right)^{n}$$

2.8.10 Comparative analysis

Venny 2.1 was used for comparison of the essential genes identified between each library dataset (Oliveros, 2007-2015).

2.8.11 Gene ontology enrichment analysis

The enrichment analysis was performed on Ecocyc (http://www.ecocyc.org), only enrichment has a p value < 0.05 (Benjamini Hochberg) was considered to be significant.

2.9 Protein expression analysis

2.9.1 SDS-PAGE

The protein samples were prepared by growing the strain of interest to early exponential phase (OD_{600} = 0.5-0.7), 1 ml of cells were transferred to an Eppendorf and resuspend to get cell pellet. The pellet was resuspended in certain amount (normally OD_{600} x 10) of 1X Laemmli sample buffer and boiled for 5-10 min at 95 °C. The protein sample was stored at -20 °C and boiled for 5 min at 95 °C before use. Typically, 10 µl of each sample was loaded onto an SDS-PAGE gel. Proteins in this study were separated using a 12 % resolving gel, the recipes of resolving gel and stacking gel is described in Table 2.11.

Table 2.11 Formula of SDS-PAGE used in this study

Component	**15% resolving gel (ml)**	**6 mL stacking gel (ml)**
H_2O	2.3	4.1
30% acrylamide mix (Protogel)	5.0	1.0
1.5 M Tris-HCl (pH 8.8)	2.5	-
1.0 M Tris-HCl (pH 6.8)	-	0.75
10% SDS	0.1	0.06
10% ammonium persulfate (APS)	0.1	0.06
TEMED	0.004	0.006

2.9.2 Western blotting

Western blot was done as previously described (Towbin et al, 1979). Proteins separated by SDS gel were transferred to a nitrocellulose membrane (UltraCruz) using the Western Sandwich cassette in 1X western blot transfer running buffer (Table 2.4) at 50 V for 3 hours or 20 V overnight. The membrane was blocked 1 hour at room temperature in 5% casein (1.25 g skimmed milk powder in 25 ml TBS) or overnight at 4 °C on the orbital shaker. The membrane was then washed 3 times with 1X TBS (Table 2.4), each time for 15 minutes. Then the membrane was incubated with the anti-YecA primary antibody (1: 2000 diluted in 1X TBS) for 1 hour. The membrane was washed three times with TBST (0.1% Tween-20 in 1X TBS), each for 15 minutes. The membrane was then incubated with anti-rabbit antibody for an hour then washed with further 3 times with TBST. After being washed by TBS for 5 minutes, the membrane was incubated with 1:1 mixture of buffer A for peroxide and buffer B in ECL^{TM} Prime Western Blotting Detection Reagent (Cytiva Amersham^{TM}) and visualized using a Gel Doc Chemi XR^{+} and Image Lab software.

2.9.3 Alkaline phosphate (PhoA) assay

PhoA assay (Kreuzer et al., 1975) in this study was used to indicate the disulfide bond formation in the cytoplasm. Strains for this test were all transformed with pDH811 (Table 2.2). 50 µl of the overnight culture was subcultured in 5 ml LBB in the presence of 0.5 mM IPTG and grown at appropriate temperature to reach OD_{600} = 0.28-0.7. The culture was cooled on ice for 20 min to stop growth and the cell density was recorded by measuring the absorbance at 600 nm. The cell culture was aliquoted and added into 100 Mm Tris-HCl, pH = 8 addition with 2 drops of chloroform and 1 drop of 0.1% SDS then mixed by vortexing. The mixture was incubated at an appropriate temperature for 5 min. 0.2 ml 4 mg/ml PNPP was added and mixed by a shaking incubator. The time was recorded after the yellow color was developed, 0.5 ml of 1M K_2PO_4 was added to stop the reaction. The mixture was centrifuged for 1 min and the absorbance of the supernatant at 550 nm and 420 nm (OD_{550} and OD_{400}) were measured and recorded. The PhoA activity was calculated in Milner units by the equation below, in which t = reaction time in minutes and v = volume of culture used for this assay.

$$\text{Unit of PhoA activity} = \frac{1000\text{x } (OD_{420} - 1.75 \text{ x } OD_{550})}{\text{t x v x } OD_{650}}$$

2.10 Filter disc susceptibility test assay

Filter disc assays were performed following the protocol as described (Bauer et al., 1966). Overnight cultures were streaked evenly over the surface of the LB plates containing different antibiotics by using sterilized cotton swabs. A 6 mm antibiotic assay filter disc (Merck-Millipore) was then placed at the center of the plate. The antibiotics were pipetted on to the filter disc and the plates were grown overnight at appropriate temperature. A lawn of the bacteria and a clear zone would be observed if the bacteria strain were sensitive to the antibiotic. The diameter of inhibition zone which refers to the sensitivity including the 6 mm disc was measured and recorded.

Chapter 3
Identification of the physiological differences between SHuffle strains and their isogenic parental strains using TraDIS

3.1 Introduction

SHuffle strain is an *E. coli* strain capable of producing correctly folded disulfide bonded protein in the cytoplasm. This strain contains mutations inactivating the genes encoding thioredoxin reductase (*trxB*) and gluathione reductase (*gor*), which diminish the cytoplasmic reductive pathways. Deletion of *trxB* and *gor* genes is normally lethal. However, the lethal phenotype is suppressed in SHuffle by a mutation in the gene encoding the alkylhydroperoxidase AhpC, known as *ahpC**. Furthermore, over-expression of DsbC is applied to assist the formation of correctly folded multi-disulfide bonded proteins. At present, the SHuffle mutations have been introduced into two parental lines of *E. coli*: a derivative of *E. coli* K-12 (SHuffle K-12) and a derivative of *E. coli* B (SHuffle B).

However, the SHuffle mutations result in several unexplained "knock on" effect, for example, SHuffle strains produce higher yields of recombinant proteins compared to the wild type, irrespective of whether the protein contains disulfide bonds or not (Anton et al., 2016). The expression of proteins containing no disulfide bond such as wild type GFP has been increased in the SHuffle cells compared to the standard protein expression strain such as BL21 (DE3) (Ren et al., 2016). However, the biological mechanism is currently not understood. In addition, SHuffle B strain was proved to be superior to SHuffle K-12 in expression of majority proteins, even though *E. coli* B and K-12 share great similarity in the genome size and sequence identity (Lobstein et al., 2012; Jeong et al., 2009). The level of redox states of proteins and mutations in *ahpC* in two SHuffle strains are distinct as well. The sequence of the *ahpC* gene in SHuffle K-12 and SHuffle B compared to their wild type strains indicated that there was a triplet condon expression *ahpC** allele in SHuffle K-12 but a triplet condon contraction allele *aphC*$^{\Delta}$ in majority of SHuffle B strain (Lobstein et al., 2012). However, there was no significant differences in the mechanism of disulfide bond formation and the activities of the enzyme in two strains. Therefore, a deeper understanding of SHuffle strains is required. To this end, the physiological differences between SHuffle strains (SHuffle K-12 and SHuffle B) and their isogenic parental strains were identified by TraDIS.

TraDIS as a high-throughput method was used to generate the essential gene lists from the transposon mutagenesis libraries constructed for four strains. For essential genes, by definition, any disturbance in these genes will cause a lethal phenotype. Therefore, no transposon insertion site could be identified in the transposon library. The manual inspection of the statistical datasets was used to avoid the false positive identifications and essentiality affected by domain essential and translation orientations. The essential genes determined in the library datasets were compared to investigate the physiological differences between SHuffle strains and their wild type strains. Currently complete essential genome analysis was reported for an *E. coli* BW25113 transposon library constructed using a mini Tn5 transposon containing a chloramphenicol resistance cassettes (Goodall et al., 2018). Since the BW25113 and DHB4 are both *E. coli* K-12 strain, comparing these strains as well as the wild type B strain with regard to the essential genes may provide new insight into the physiology differences.

3.2 Results

3.2.1 Construction and sequencing of the transposon insertion libraries

To generate the high-density transposon libraries, SHuffle K-12, SHuffle B and their isogenic parental strains were mutagenized by Tn*5*-based transposon containing a kanamycin resistance gene flanked by 19 bp inverted repeats. The inverted repeats were recognized by transposase which could be transformed into the target genomic DNA at random places. Since the transposons do not contain the transposase gene, once it has been inserted, it cannot be moved. Therefore, each individual colony selected on the LB plates in the presence of 50 μg/ml kanamycin should only contain one transposon insertion.

The colonies were collected and pooled to construct the transposon mutagenesis library following the protocol described in Section 2.7.1, each library in this study was created to consist of approximately 500,000 individual colonies. Since SHuffle strain are temperature sensitive, unlike the libraries of DHB4 and T7 Express were grown at 37℃, SHuffle K-12 and SHuffle B libraries were grown at 30℃. Once being pooled, the genomic DNA in each library was extracted and prepared for Illumina sequencing. The sequencing data was then used to map the transposon insertion sites to the genome, the overall protocol for TraDIS was described in Section 2.7.

To ensure the reproducibility of the results, two technical replicates, S1 and S2, for the DHB4 and SHuffle K-12 libraries were sequenced and the results were compared, the reads associated with transposon junctions were identified and aligned to the corresponding genome. To ensure the correlation between two replicates in DHB4 and SHuffle K-12, the metrics for insertion index score (IIS) which is the unique insertion sites normalized for gene length showed the high correlation between two replicates, with the correlation coefficient of 0.94 and 0.96 (Figure 3.1), indicating the two samples were high technical reproducible. 9033850 total reads were mapped to DHB4 genome (CP014270) and its cognate F' episome (CP014271), for each replicates corresponded to 142137 and 434642 unique insertion sites, respectively. 6635884 total reads were mapped to the genome of SHuffle K-12 (CP014272) and its cognate F' episome (CP014272), which corresponded to 218115 and 353632 unique insertion sites for sample 1 and sample 2, respectively. There were overlapped unique insertion sites between each replicate therefore could not be added up together (Table 3.1). No repeats were applied for libraries of wild-type B and SHuffle B strains, 4037580 sequence reads and 428843 unique insertion sites were mapped to T7 express (CP014268), 6391671 sequence reads and 381263 unique insertion sites were mapped to SHuffle B (CP014269). The detailed results of each library were displayed in Table 3.1.

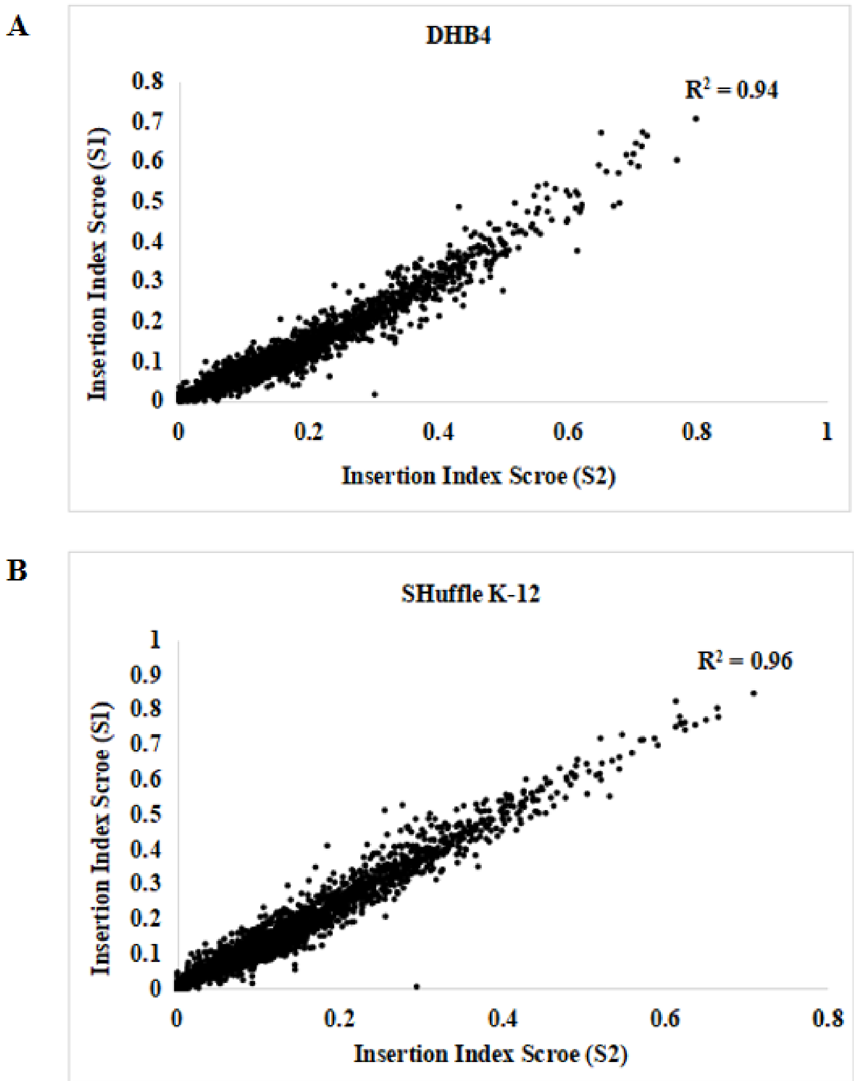


Figure 3.1: The correlation of insertion index scores to show similarity of two technical repeats taken from the transposon libraries of DHB4 (A) and SHuffle K-12 (B). The correlation coefficient shows the similarity was generated for each library.

Table 3.1 Parameters for analysis of the TraDIS data derived from four transposon libraries

TraDIS Library	No. of transformants	No. of mapped sequence reads	No. of insertion sites*
DHB4 (S1)	518682	2695654	142137
DHB4 (S2)		6338196	434642
DHB4 combined		9033850	-
SHuffle K-12 (S1)		2558374	218115
SHuffle K-12 (S2)	579135	4077510	353632
SHuffle K-12 combined		6635884	-
T7 express	503621	4037580	428843
SHuffle B	835354	6391671	381263

*No. of insertion sites in each replicate may overlap and therefore cannot be added up.

3.2.2 Identification and comparison of essential genes in four libraries

An uneven distribution of sequencing reads and insertion sites was observed throughout the genome of each strain, and there was no particular region that lacked insertion sites (Figure 3.2). In previous studies, the frequency of insertions was likely to be increased around origin of replication (Milner, 2022), enrichment insertions close to the origin site were observed in all libraries. The origin of replication is where the DNA replication initiated, more copies of region close to the origin of replication result in an increasing distribution in read. For DHB4 and SHuffle K-12, there were significant enriched insertion sites mapped between 210000 and 317600 on chromosome, suggesting this region could have distinct growth disadvantages in DHB4 compared to *E. coli* B strains but the reason was not clear.

The essentiality of genes was quantified by IIS, absence of insertions within the coding sequence (CDS) suggesting that the gene could be essential. The inserted transposon could partially or completely disturb the function of the gene. Essential genes are usually defined as genes required for cell viability, therefore, when function is disrupted, will lead to cell death and will not be included in the population. Correspondingly, insertions in the nonessential genes will not affect the organism's ability to survive and could have multiple insertion sites throughout the CDS. Genes can be classed as essential in one strain but nonessential in another, suggesting the gene function is required for one strain but not in the other. In addition, there are domain essential genes, only a part of the gene is essential, which need to be manually inspected on Artemis browser (Figure 3.3).

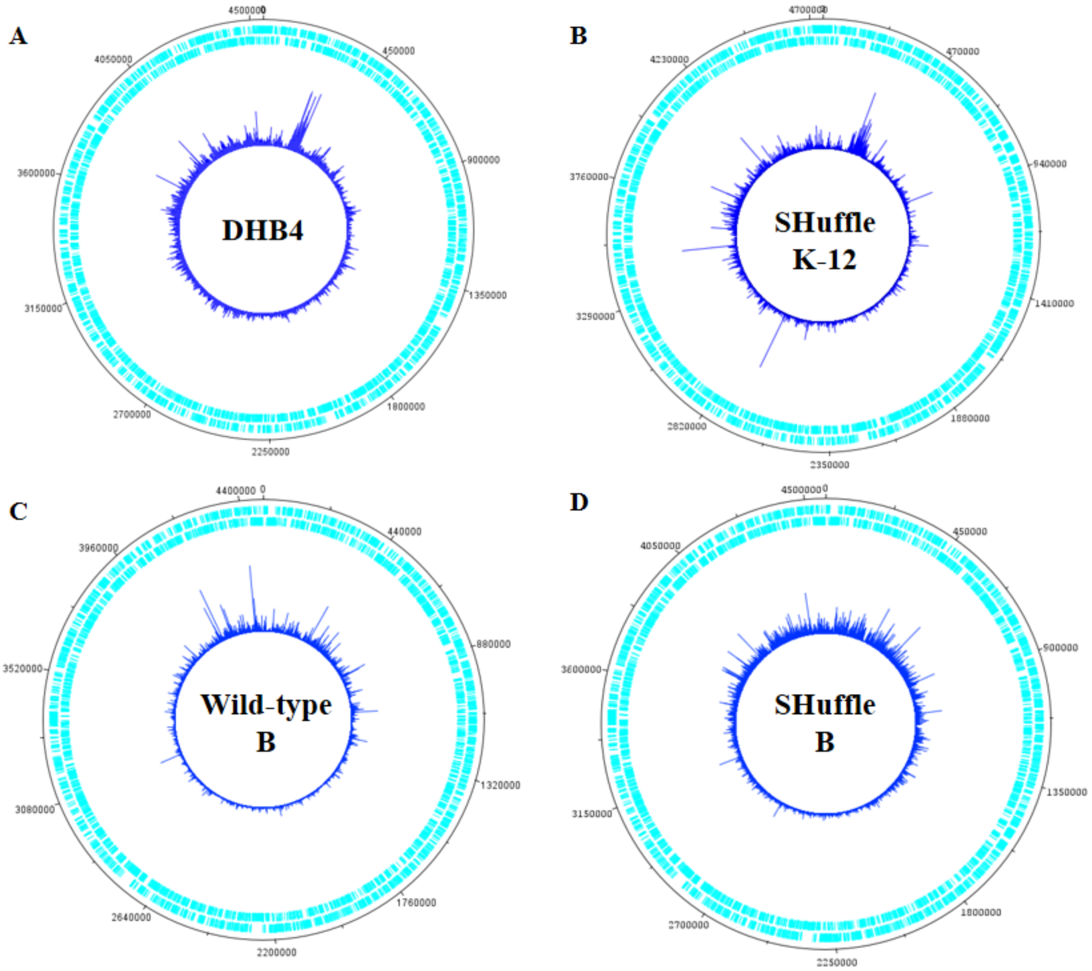


Figure 3.2 The distribution of Tn5 transposon insertion sites mapped to genomes of DHB4 (A), SHuffle K-12 (B), wild-type B (C) and SHuffle B (D). The dark blue inner circle represents the relative number and location of mapped transposon insertion sites. The outer black and light blue circles represent genome in bp and corresponding forward and reverse annotated CDS respectively.

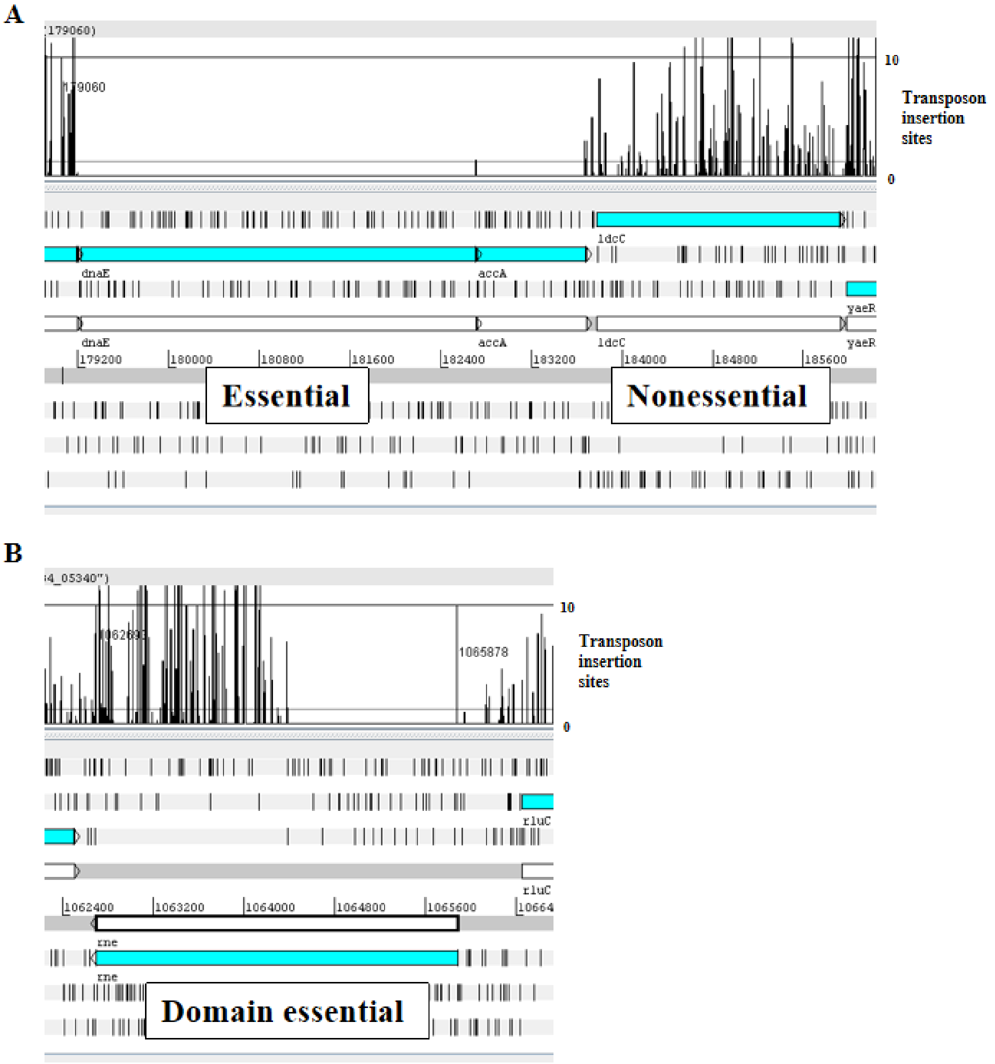


Figure 3.3 Visualization of example genes identified as essential, nonessential (A) and domain essential (B) on the Artemis browser. Each vertical black line represents a transposon insertion site mapped to the DHB4 genome, the length of it represents the read count. The scale was set to 0-10 for better inspecting the presence or absence of insertions.

The essential genes were determined by analyzing the sequencing data using several computational methods combined with manual inspection. Two statistical methods used in this study were log_2 likelihood and Biotradis pipeline, which were described in Section 2.7.8. The log_2 likelihood score quantified IIS and calculated the probability of belonging to each mode based on the bimodal distribution (Figure 3.4). This method was previously described (Goodall et al., 2018). The frequency of IIS was plotted using R MASS library (version: 7.3-60.0.1), the exponential distribution (red) was fitted to essential mode with no insertions or low insertion score, a gamma distribution (blue) was fitted to the nonessential mode. Essential and nonessential genes were determined by calculating the ratio of $\log_2$ likelihood of a gene, the threshold was 12-fold. Genes were classified as essential when 12-fold more likely to be in the essential mode (lower than -3.6), respectively, genes were classified as nonessential if 12-fold more likely to be in the nonessential mode (greater than 3.6), ambiguous if with scores in between. However, there could be false positive identifications caused by computational analysis on its own. Genes such as *rne* only contain an essential domain (Figure 3.3B) or some of the genes are only essential in one transcriptional orientation, which could not be identified by statistical analysis based on the IIS. Therefore, manual inspection was applied for each library to reveal features that have not been detected.

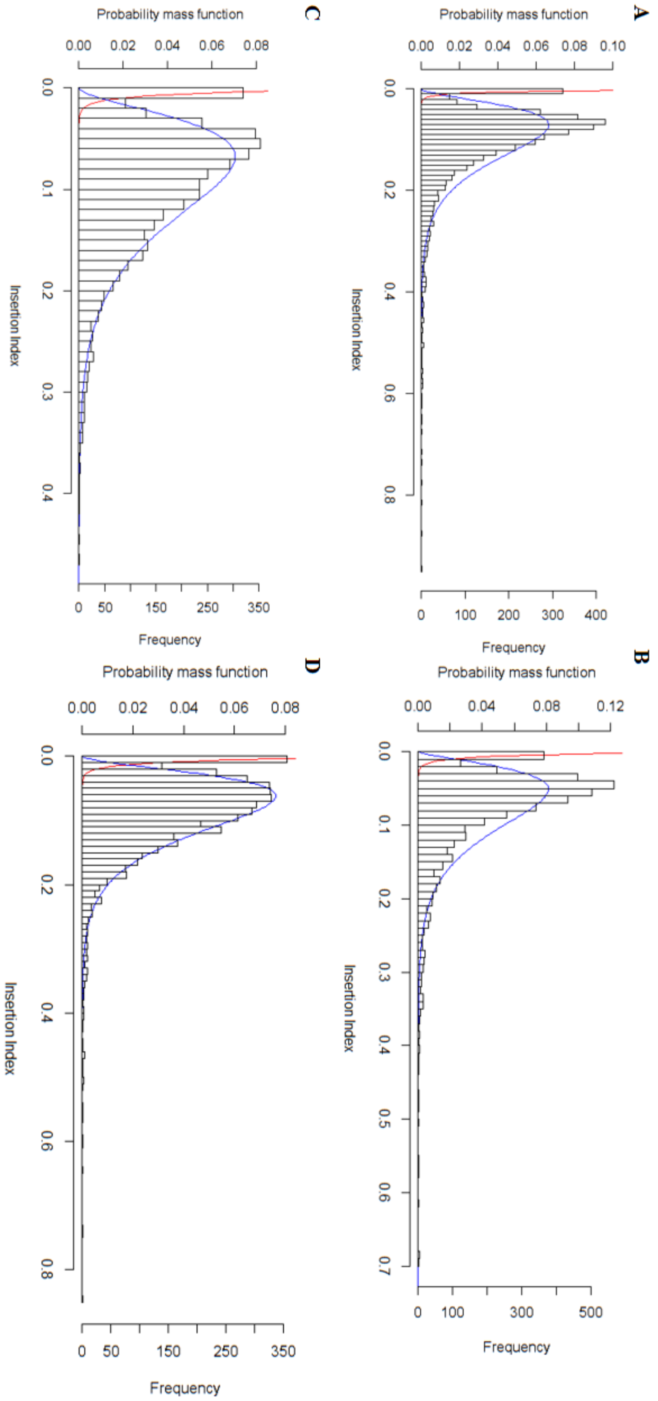


Figure 3.4 Bimodal distribution of insertion index scores in the transposon libraries of DHB4 (A), SHuffle K-12 (B), wild-type B (C) and SHuffle B genome (D). The exponential distribution (red line) fitted to essential mode and the gamma distribution (blue line) fitted to nonessential mode. The threshold is 12-fold, essential genes were determined with log-likelihood scores less than $\log_2(12)$, suggesting the genes were 12 times more likely to be in the essential mode.

Log_2 likelihood and the Biotradis pipeline used similar methods but a lower threshold (cutoff of ±2) for the Biotradis pipeline to determine the essential and ambiguous genes. Since the threshold has been cutoff, there were less genes classified as ambiguous, but more false positive results for each library. In addition, the annotations reported in Biotradis showed a slight difference in reporting the genome features in addition to genes encoding proteins. To generate the most accurate essential gene lists, manual inspection was applied to the unique essential genes determined only in one method, and the false positives were removed. The results of genes within four libraries were listed in Table 4.2 and the final determined essential genes were compared between four library datasets in Figure 3.5. Since T7 Express and SHuffle strains were not well annotated in the previous studies, the GFF3 files of each strain were edited by GFF3 passer scripts to ensure the compatibility of the TraDIS files. It altered the ID annotation in the GFF3 files, changing it from what it was in the original genome representing the position of the genome to corresponding ID in MG1655. The sequence ID in four strains were all converted, and gene names were identified with reference to genes in MG1655. However, not all of the annotations could be recognized according to the data frame. Therefore, the comparison or essential genes in four libraries only including genes with gene names could be identified, which were 352 genes in DHB4, 361 genes in SHuffle K-12, 354 genes in T7 express and 360 genes in SHuffle B.

Table 3.2 Transposon library statistics for four libraries

TraDIS Library	Log likelihood			BioTraDIS pipeline			Manual*
	Essential	Ambiguous	Nonessential	Essential	Ambiguous	Nonessential	Essential
DHB4	329	109	3939	421	58	4036	389
SHuffle K-12	355	279	3995	472	56	4230	407
NEB T7 express	342	182	3784	467	180	3781	378
SHuffle B	382	474	3571	525	97	3931	378

*The number of manually checked essential genes contains nonannotated genomes that were not presented in Figure 3.5.

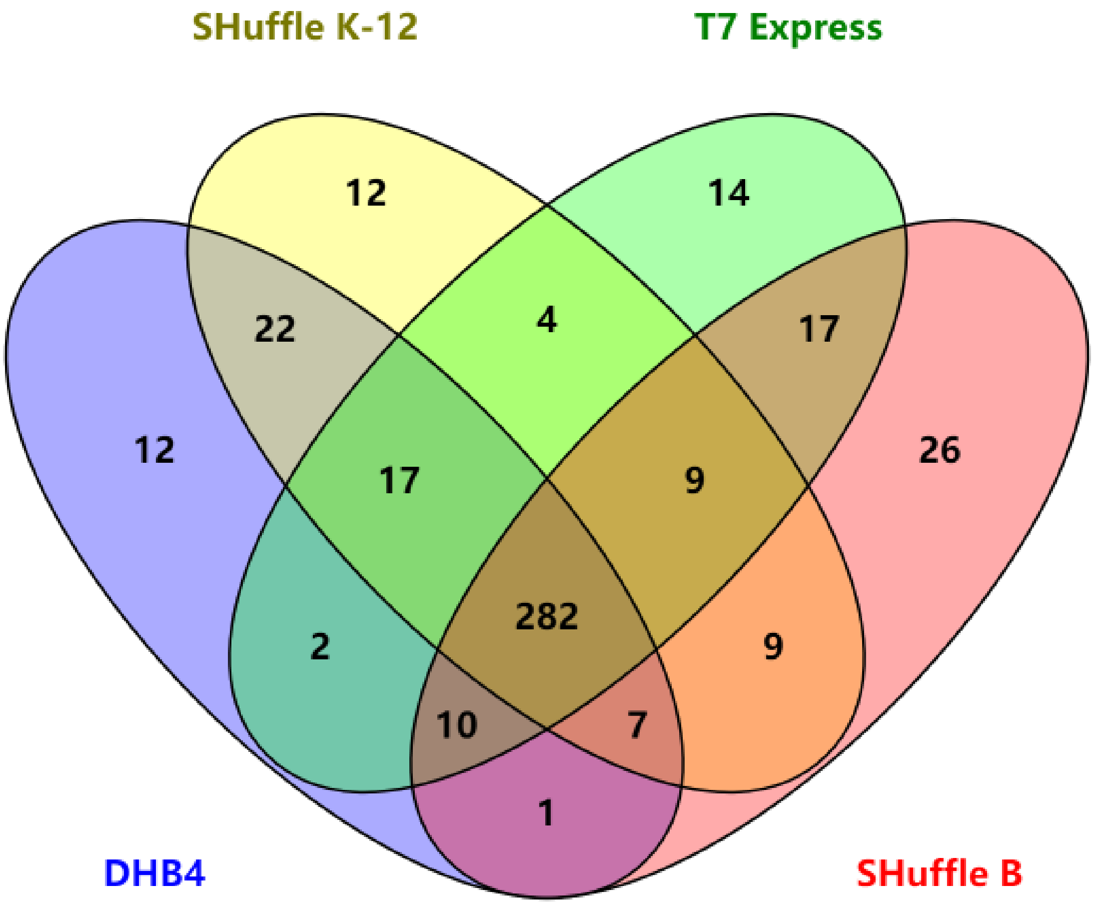


Figure 3.5 Venn diagram to compare the final determined essential genes in the libraries of DHB4, SHuffle K-12, T7 Express and SHuffle B. 282 genes were

common in four stains. No significant discrepancy was found in four strains compared to each other, 12 (2.7%) genes in DHB4, 12 (2.7%) genes in SHuffle K-12, 14 (3.2%) genes in T7 express and 26 (5.9%) genes in SHuffle B were found unique essential to each strain.

Table 3.3 Unique essential genes identified in four libraries by TraDIS

Library	Gene Name
DHB4	*ackA, cydX, glyA, metJ, nuoI, ratA, rluD, rppH, ruvA, xseB, ybeY, yffS*
SHuffle K-12	*atpC, gshA, mnmG, nfuA, rbfA, rsmH, secB, tusC, ubiF, ubiH, yacG, ynaE*
T7 Express	*cydC, folK, ldcA, lipB, rpmF, ssrS, sufA, tolB, tolQ, tolR, waaC, ybfB, yciM, yddM*
SHuffle B	*ahpF, arnE, arnF, cspA, cysQ, dksA, ihfB, kilR, marB, msrB, nohD, oxyR, pstB, sapD, tonB, ybcV, ybcW, ybeD, yceQ, ydcD, ydcX, ydiE, yeaC, yebV, ymjC, yneM*

3.2.3 Classification of the unique essential genes required for growth in four strains

The comparison of the essential genes in each transposon library provided unique genes required for growth in DHB4, SHuffle K-12, T7 Express and SHuffle B. It's noticeable that there were twice as many genes that were found to be essential exclusively in the library of SHuffle B (Figure 3.5; Table 3.3). To determine whether these unique essential genes in each library were involved in any functional enrichment, Gene Ontology (GO) enrichment test based on EcoCyc database was used to classify sets of genes associated with molecular function (MF) and biological process (BP). Cellular component is another orthogonal ontology in GO enrichment analysis but was not applied in this study. The relative abundance of each GO term for each library was shown in Figure 3.6. The GO terms were assigned to proteins with known function, genes encoding proteins with unannotated functions were counted and termed a category of uncharacterized protein.

According to results analyzed by GO enrichment test, there were no significant differences in the categories classified based on the molecular functions that genes were involved in. It is noticeable that there were 8/26 (31%) genes encoding for uncharacterized protein in SHuffle B, suggesting the importance of investigating the unknown function of the unique essential genes for deep understanding of SHuffle B. Integrating that SHuffle B had 2-fold unique essential genes than other strains, SHuffle B was more distinct, even compared to SHuffle K-12 strain containing same genetic modifications. This was further proved in the biological process GO enrichment test. Except for 8 uncharacterized proteins, 13/18 (72%) genes essential in SHuffle B encoding proteins response to cellular stress. SHuffle K also showed increased numbers of genes involved in cell response compared to the wild type, suggesting that the SHuffle modifications could lead to increased stress therefore need more proteins against it for viability.

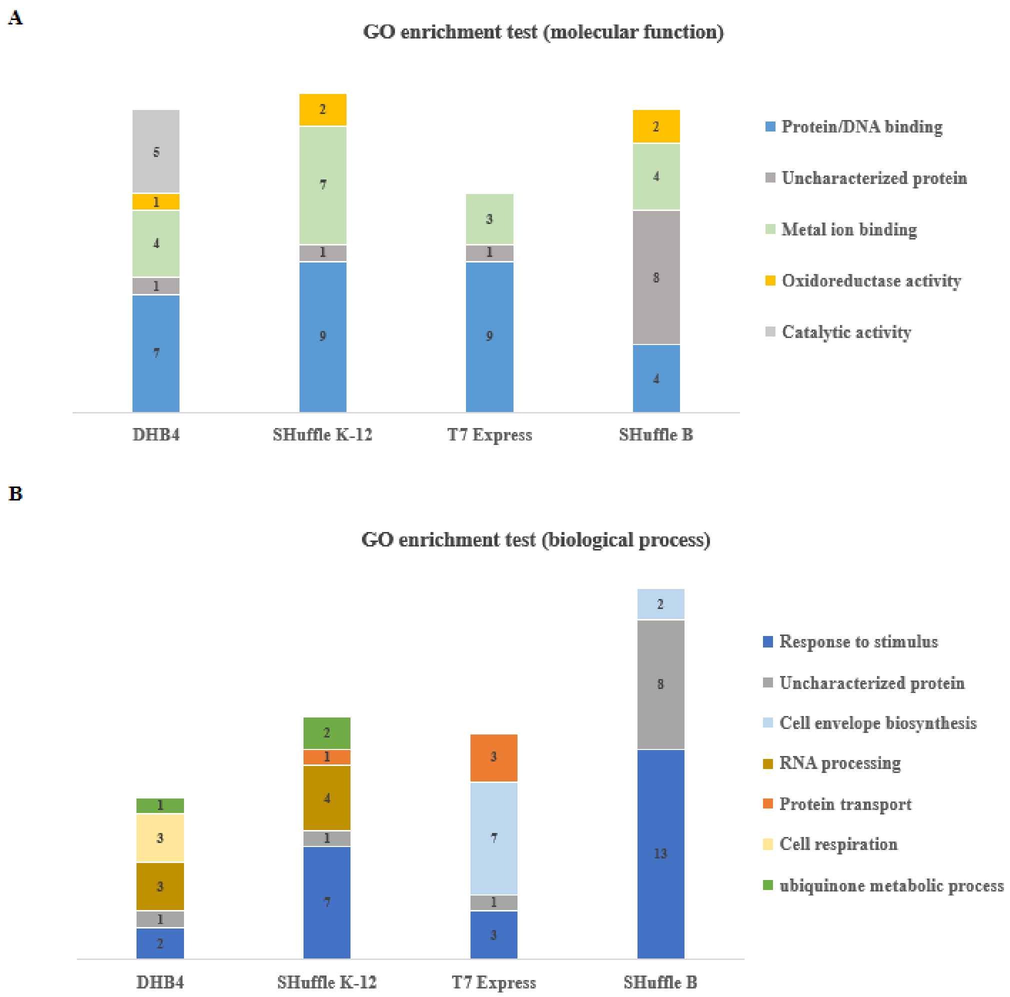


Figure 3.6 GO enrichment analysis to understand the discrepancy between datasets. Genes unique to transposon libraries of DHB4, SHuffle K-12, T7 express and SHuffle B were grouped into different categories according to molecular function (MO) (A) and biological process (BP) (B). Gene with unknown function in EcoCyc database were included in uncharacterized protein.

Upon inspection of genes involved in cellular response that had become essential in the SHuffle strains, 7 genes and 13 genes were identified in the SHuffle K-12 library and SHuffle B library, respectively. There were 4/7 genes, *gshA, nfuA*, *ubiF*, *ubiH* in SHuffle K-12 are required to response to oxidative stress. Gene g*shA*, as glutamate-cysteine ligase, catalyzes the synthesis of glutathione, which is a protein facilitating the formation of disulfide bonds, and prevents cells from oxidative damage. Gene *nufA* encodes iron-sulfur cluster carrier protein, the null mutant conferred server growth defect under oxidative stress (Angelini et al., 2008). Gene *ubiF* and *ubiH* function in ubiquinone (UQ) biosynethesis and are involved in cell redox. More importantly, mutations containing deficient UQ exhibited resistance to high temperatures (Collis and Grigg, 1989). The other 3 genes, *rbfA* and *secB* were implicated in cold shock response, mutations in each gene caused cold sensitive phenotype. *mnmG* was found response to radiation and assist the DNA repair.

By comparison, 13 unique essential genes in the SHuffle B library which were involved in cell response were identified in similar subcategories. 5/13 genes, *ahpE*, *kilR*, *marB*, *msrB*, *OxyR*, were shown to respond to oxidative stress, and 6/13 genes respond to abnormal temperatures. Among them, *cspA* and *infB* were cold-responsive (Jones et al., 1987; Goldstein et al., 1990; Brandi et al., 2019), *yneM*, *ybeD* were identified to increase thermotolerance and regulated by either RpoE (σE) or RpoH (σH). *PstB* and *ortT* (*ydcX*) were found regulated by σE or σH as well, they were respectively responsible for envelope-stress and heat shock response. The other two genes, *arnE* and *arnF*, respond to iron mediated cellular toxicity and are regulated by BasS-BasR two-component system. It was noticeable that the 5 genes not included in cell response to stimulus were all associated with metal ion response or transport but were not included in the GO terms. As in SHuffle K-12, there was a gene dedicated to DNA repair, *dksA*. It was shown bound to zinc ion and involved in translation regulation of RpoS against the environmental sources of stress.

To directly compare the differences between each SHuffle strain and its wild type, the comparisons between the essential genes in DHB4 and SHuffle K-12, T7 express and SHuffle B and SHuffle K-12 and SHuffle B were applied. The results were mostly covered in the 4-way comparison (Figure 3.5) and there were increased number of genes that were involved in the cellular response in the SHuffle strains compared to the wild type strains as expected. The only differences in SHuffle B library were genes that involved in regulation of cell shape and division turned out to be not essential compared to the wild type, this could be caused by increased oxidative stress in the cell.

There were 9 common genes found to be essential in both SHuffle strains but not in the wild-type strains (Table 3.4), which could provide insight into the physiology of SHuffle strains. Genes *aphC* and *grxA* were identified which confirmed the validity of the results. Since the lethal phenotype caused by double deletion of the *trxB* and *gor* to inactivate the cytoplasmic reducing pathways was suppressed by mutation in the *ahpC* gene (*ahpC**). This suppression was achieved through the reductive activity of AhpC* which could reduce the Grx1 encoded by *grxA*. The presence of *ssrA* and *insE1* could be a result of BioTraDIS pipeline, which leads to identification of some genome features that could be removed. Gene *sdhC* is involved in the electron transfer pathway in aerobic respiration by interacting with UQ, which would protect the SHuffle strains from oxidative stress caused by SHuffle mutations. There was no straightforward biological reason for other strains that were essential unique to SHuffle, but it could be useful for engineering the next-generation SHuffle strains.

Table 3.4 Essential genes unique to the SHuffle strains

Gene	**Function**
ahpC	Alkyl hydroperoxide reductase
grxA	Reduced glutaredoxin 1
lacI	DNA-binding transcriptional repressor LacI
ldrC	Small toxic polypeptide LdrC
rlmE	23S rRNA 2'-O-ribose U2552 methyltransferase
sdhC	Succinate:quinone oxidoreductase
ylcI	DLP12 prophage
ssrA	tmRNA
insE1	IS3 element protein InsE

3.2.4 Comparison of transposon libraries in BW25113, DHB4 and T7 Express

The *E. coli* BW25113 transposon library was of high density with 8279302 sequence reads and 901383 unique insertion sites, and 358 essential genes were identified (Goodall et al., 2018). It was decided to compare the essential genes between the library of BW25113 with DHB4 and T7 express to identify the genomic differences. Since the analysis of the BW25113 library used the log likelihood method, the same method was used to generate the essential genes for other two libraries. Similar numbers of essential genes were identified in three library datasets, since the T7 express strain was not annotated, there could be a number of overlooked essential genes. The results could also be affected by the overall density of the library as the densities of the libraries of DHB4 and T7 Express were lower than BW25113. The comparison was presented in Figure 3.7.

There were 288 common essential genes for three libraries, 21 unique essential genes were identified in DHB4 library and 27 genes were identified unique to T7 express (*E. coli* B). Some of the unique essential genes in the libraries of DHB4 and T7 express were included in the ambiguous genes identified for BW25113 (Table 3.5). The *cysE* gene encoding for serine acetyltransferase was identified as essential in the DHB4, which carries out the first step of cysteine biosynthesis. However, it was not essential in the SHuffle K-12, suggesting that the function of CysE could be complemented by other genes under oxidative stress. It is noticeable that genes encoding for proteins that involved in Tol-Pal system (*tolABQR, pal*) were identified to be unique essential to T7 express, indicating that *E. coli* B strain might face to more cell envelope stress compared to the K-12 strain.

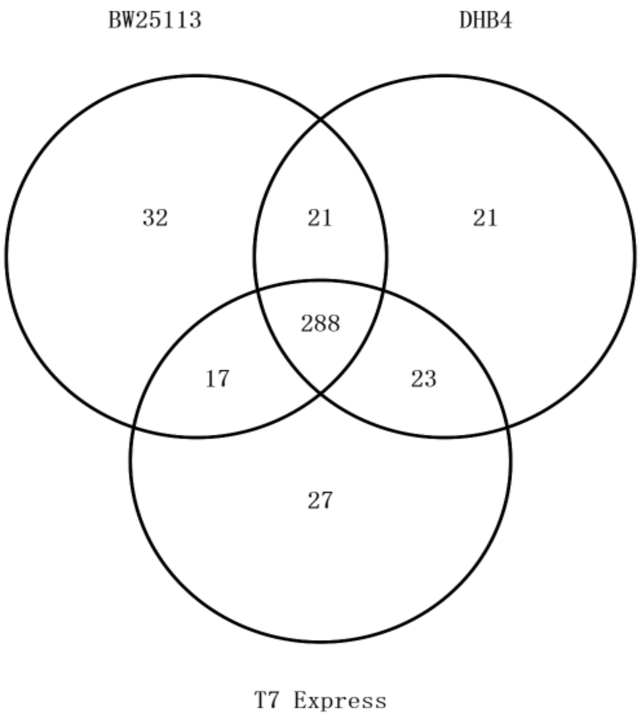


Figure 3.7 Comparison of the essential genes of the transposon libraries in BW25113, DHB4 and T7 Express. 288 common essential genes were identified for three libraries. There were 32 (7.5%) genes, 21 (4.9%) genes and 27 (6.3%) genes unique to the libraries of BW25113, DHB4 and T7 express respectively.

Table 3.5 Comparison of genes identified essential in DHB4 and T7 Express compared to genes which are ambiguous in BW25113

Condition	**Gene Name**
Essential in DHB4 ambiguous in BW25113	*hns, minE, nuoI, ratA, rseP, secM, xseB*
Essential in DHB4 nonessential in BW25113	*ackA, cysE, degS, insC1, metJ, rplI, rppH, ruvA, thiD, thiE, tpiA, tusB, yajC, yejL*
Essential in T7 Express ambiguous in BW25113	*crr, folK, hscB, icd, rnc, ybfB, yciM, yddM*
Essential in T7 Express nonessential in BW25113	*arcA, ffs, gmhB, ldcA, ldrB, lipB, pal, pinQ, pinR, ssrS, sufA, tdcR, tolA, tolB, tolQ, tolR, uof, waaC, ydfK*

3.3 Discussion

In this chapter, essential genes required for survival in the libraries of two SHuffle strains (SHuffle K-12 and SHuffle) and their wild-type strains (DHB4 and T7 Express) were determined by TraDIS, and the essential genes identified in each library dataset were compared by Venn diagram then further investigate the physiological differences between four strains. The essential genes unique to each library were involved in diverse biological processes, however, a large portion of genes that are involved in the cellular response to oxidative stress and abnormal temperatures have become essential in the SHuffle strains. Since most of the cytoplasmic reducing pathways were blocked by SHuffle mutations, the oxidative stress in the cytoplasm was significantly enhanced. Higher temperature normally increases the cell membrane fluidity and enzyme activities may aggravate the cell envelope stress and caused diminished cell viability. This could explain why SHuffle strains were recommended to produce most of the recombinant proteins at 30 °C and conferred growth defect at 42 °C. (Nasiri et al., 2017; Ren et al., 2019). Both SHuffle strains contained two unique essential genes that respond to cold shock, which could facilitate the cell growth and protein production at lower temperatures. There were additional 4 genes encoding for heat-shock proteins that become essential in SHuffle B, indicating the need of heat-shock proteins to defence the oxidative stress compounded by heat stress.

SHuffle K-12 in this study was shown to have normal cell growth at 42 °C except for cell being slightly larger in size but SHuffle B conferred severe growth defect at the same temperature (Figure 4.1), this did not match the temperature sensitive phenotype stated in the previous study (Ren et al., 2019). The differences could be caused by different strain background. There were several differences exhibited between SHuffle K-12 and SHuffle B. The redox differences were found in the *ahpC** mutations and the cytoplasmic *dsbC*. The *ahpC** mutation in *E. coli* K-12 was a triplet codon expansion, however in *E. coli* B is inversely a triplet codon contraction (Lobstein et al., 2012; Anton et al., 2016), indicating two mutations might be involved in distinct biological roles. The single copy insertion of *dsbC* in plasmid results in triplicate and duplicate copy of insertion in SHuffle K-12 and SHuffle B, respectively (Anton et al., 2016).

Triplicate copies of *dsbC* could enhanced the ability of SHuffle K-12 to resist cellular oxidative stress. In addition, the enzyme activity was found to be consistently higher in SHuffle B. All of these differences could be the reason why SHuffle K-12 was more resistant to higher temperatures than SHuffle B. However, this could not be further proved by the TraDIS result.

The UQ biosynthesis was predicted to play an important role in the SHuffle strains. Gene *sdhC* was shown to be essential in both SHuffle K-12 and SHuffle B, which encodes succinate:quinone oxireductase SdhC. It transports electrons from succinate to oxygen relying on the interaction with UQ. Gene *ubiF* and *ubiH* encoding the UQ biosynthesis proteins UbiF and UbiH were found to be essential unique to the SHuffle K-12. Since UQ was proposed to protect *E. coli* cells from oxidative stress (Soballe and Poole, 2000), and stabilize the cytoplasmic membrane (Sevin and Sauer, 2014). Moreover, deficient UQ leads to resistance to high temperatures (Collis and Grigg, 1989). The evidence suggests that the role of UQ in SHuffle strains, particularly in SHuffle K-12, is associated with both the oxidative stress and temperature sensitivity.In the *E. coli* protein oxidative folding pathways, UQ was shown to be an electron acceptor for the membrane protein DsbB to oxidize the DsbA and facilitate the disulfide bonds formation in the periplasm, it could be possible that DsbB is also involved in the production of disulfide bond in the cytoplasm.

There were a lot of genes encoding for uncharacterized proteins in the essential gene lists for both SHuffle strains, as well as genes that could not be identified in the TraDIS analysis, understanding the function of these genes could provide insight into the physiology of SHuffle strains. In addition, the library of DHB4 and T7 express were compared to the BW25113 library with regard to the essential genes, which could provide insight into different *E. coli* backgrounds. The unique essential genes involved in cell envelope biogenesis in T7 Express suggested B strains could be more susceptible than K-12 to certain stress conditions.

Chapter 4
Identification of the temperature sensitivity of SHuffle B

Declaration

The initial TraDIS library of SHuffele B was constructed by me, the construction of libraries grown at different temperatures was achieved by me and a master student, Sravanthi veerla. The sequencing data were mapped to the SHuffle B genome by Sravanthi veerla using the scripts provided by Mathew Milner. The subsequent analysis presented in this chapter was completed by me. Mutations in the Keio library was provided by Banzhaf's lab. The cross-linking for the mutations transduced from the transposon library of DHB4 to SHuffle B was performed by me and a master student, Michael Cagney and presented in his thesis.

4.1 Introduction

SHuffle strains are temperature sensitive, especially SHuffle B. The previous studies demonstrated that SHuffle strains confer a severe growth defect under 42 ℃ but grow at 37 ℃ and below. Since SHuffle strains can be used to produce a number of antibodies, such as antigen-binding fragment (Fab) and immuoglobin (IgG), as well as other recombinant proteins containing disulfide bonds (Krittanai et al., 2020, Eaglesham et al., 2021), the optimizing of protein expression conditions for SHuffle strains has never stopped. Before optimizing the protein expression conditions, the ideal SHuffle strains should be chosen. Previous studies proved that expression of proteins in SHuffle B strains was shown in greater yield than in SHuffle K-12 (Lobstein et al., 2012), and the main reason could be distinct cellular responses to oxidative stress, which could also cause the differences in temperature sensitivity. Therefore, understanding genes or biological processes caused the temperature sensitivity could provide insight into the oxidative stress in different strain background and be useful for constructing the next-generation of SHuffle strains to increase the protein production yield. Moreover, although the grow and expression of protein in SHuffle strains are recommended at 30 ℃, especially the protein containing multiple disulfide bonds to avoid misfolding, the temperature sensitivity causes a limitation for the expression of proteins which the higher temperatures are necessary.

The previous understanding of the temperature sensitivity of SHuffle strains were based on the expression level of certain proteins, the possible reason that low temperature that conferred improved yields could be explained by higher temperature resulting in higher metabolic activity, and leading to growth disadvantage in cells under oxidative stress. In the latest experiment performed by our lab, the transposon library of SHuffle B started to grow at 42 ℃ at approximately 10^{-3}-fold dilution of the original library (Peros, 2022). The *E. coli* B strain has of ~4500 genes, so that there could be 5-10 genes caused temperature resistance. Depending on the different size of the gene, there could be more.

To further investigate the genetic reasons causing the temperature sensitivity of SHuffle B and the mutations that could recover this phenotype, the transposon mutagenesis library of SHuffle B (Chapter 3) was used to screen for conditionally essential genes at different temperatures by TraDIS. Since only SHuffle K-12 was found to grow at 42 °C on the LB plates under laboratory conditions, the genomic DNA in the transposon insertion library of DHB4 (Chapter 3) was transduced into SHuffle B to identify the genomic differences or mutations that could recover the growth of SHuffle B at 42 °C.

4.2 Results

4.2.1 Construction of the transposon libraries of SHuffle B grown at different temperatures

To test the temperature sensitivity of SHuffle strains, the growth of SHuffle strains and their wild types were streaked on LB agar plates and grown at different temperatures overnight. All of the strains exhibited normal growth at 30 °C and 37 °C, however, SHuffle K-12 had a negligible growth defect, but SHuffle B conferred severe growth limitation at 42 °C (Figure 4.1).

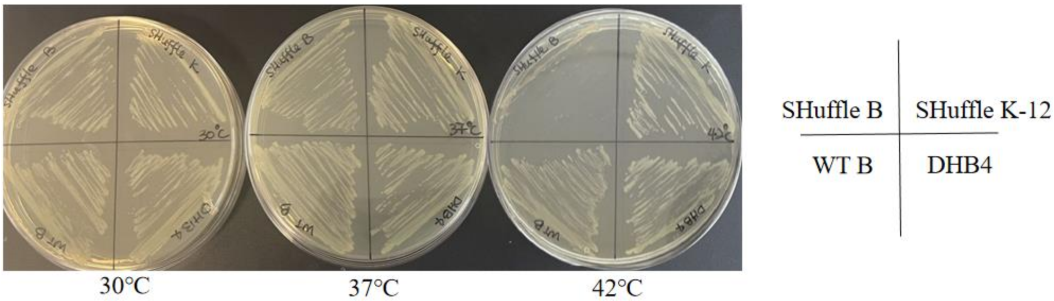


Figure 4.1 SHuffle B strains has limited growth at 42 °C. Growth of SHuffle K, SHuffle B and their isogenic wild-type strains on the LB plate over 8 hours at 30 °C, 37 °C and 42 °C.

As the transposon libraries were constructed by collecting the original library, it is necessary to determine the appropriate number of the sequence reads required for comprehensive coverage of the library. 10^{-4}-fold dilution of the SHuffle B library

(Chapter 3) was used in this experiment to ensure the single colony could be harvested. With the maximum number of unique insertion sites within the SHuffle B library of 5,585,567 mapped reads and 473,763 unique insertion sites (Chapter 3), according to the equation presented in Goodall et al., (2018) and described in Section 2.7.11, the number of insertions covered by total reads could be estimated (Figure 4.2).

Diluted SHuffle B transposon library was plated at 30 °C, 37 °C and 42 °C before being pooled, for plates grown at 30 °C and 37 °C, each plate contained approximately 8000 to 10000 colonies. Each colony was considered as a cell containing a single transposon insertion, to ensure full coverage of the original transposon library, according to the estimation of the percentage saturation and the sequence reads (Table 4.1), 300 plates were plated for each library to ensure the 100% coverage of sufficient mutant sampling, which was at least 2520000 reads per sample (Table 4.2). The libraries were sequenced and mapped to the genome using a TraDIS protocol described in Section 2.7.

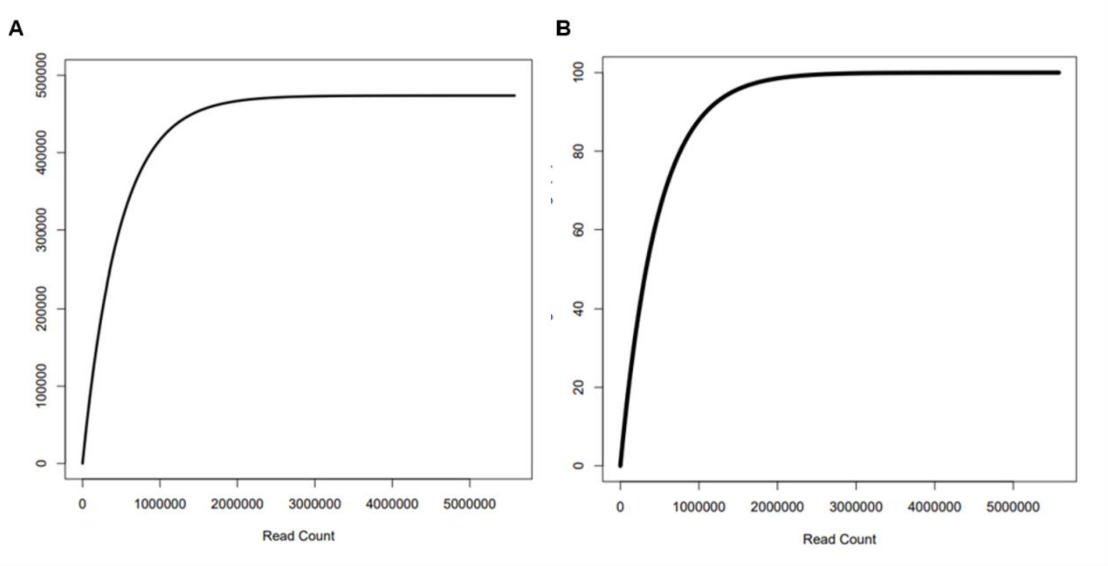


Figure 4.2 Estimated number of reads required to obtain a set number of unique insertions. A) Use of the equation described in 2.7.11 to estimate the number of reads needed to achieve certain number of insertions and the B) percentage of insertion coverage of the unique insertions in the initial library relies on the sequencing reads.

Table 4.1 Estimated reads and insertions corresponding to the percentage saturation

Reads	Percentage saturation*	Estimated insertions
770000	80%	380500
1100000	90%	427290
1420000	95%	4501111
2520000	100%	471443

* Percentage saturation refers to the coverage of the estimated insertions in the sub-library to the original library.

Table 4.2 Sequence reads aligned to the genome for each condition

Library	Total sequence reads
SHuffle B 30 °C	3265489
SHuffle B 37 °C	2865895
SHuffle B 42 °C	2158789

4.2.2 Sequencing and comparison of the transposon libraries grown at different temperatures

The sequencing data was passed through particular scripts to determine the position of the insertion sites and reads corresponding to each position. The BioTraDIS pipeline (Section 2.7.8) was used to generate the conditional essential genes combined with manual inspection on the Artemis browser. The transposon insertion sites were visibly depleted in the libraries grown at 37 °C and 42 °C, suggesting that insertion in certain genes resulted loss of fitness under higher temperatures. Since SHuffle strains are under constitutive oxidative stress, higher temperatures may stress the cells by altering the membrane stability or increasing the oxidative stress. Therefore, there would be more genes which are normally sublethal at 30 °C screened as unique essential or conditional essential under higher temperatures.

By comparing the essential genes in the SHuffle B libraries that grown at different temperatures, there was a large crossover in the genes predicted to be essential. The conditional essential genes identified in the library of 42 °C were 3-fold more than the other two libraries (Table 4.3). GO enrichment test based on the molecular function of genes were applied to the essential genes at 42 °C (Figure 4.3). Since SHuffle B was not well annotated in the previous studies, the essential genes in this analysis only included annotations which were present in the NCBI database. Most of the genes were functioned in oxidoreduction, which is consistent with the oxidative stress increased under 42 °C. There were also large portions of genes involved in DNA binding and protein synthesis. Considering SHuffle B strain has ~4500 genes, about half of the genes were determined to be essential or ambiguous, it seemed unlikely to understand the mechanisms when SHuffle B grown at 42 °C by analysis of the conditional essential genes.

Table 4.3 Summary of the outputs of SHuffle B libraries grown at different temperatures generated by BioTraDIS pipeline and manual inspection

Library	**Essential**	**Ambiguous**	**Non-essential**
SHuffle B 30 °C	739	152	3662
SHuffle B 37 °C	764	84	3705
SHuffle B 42 °C	2178	468	1907

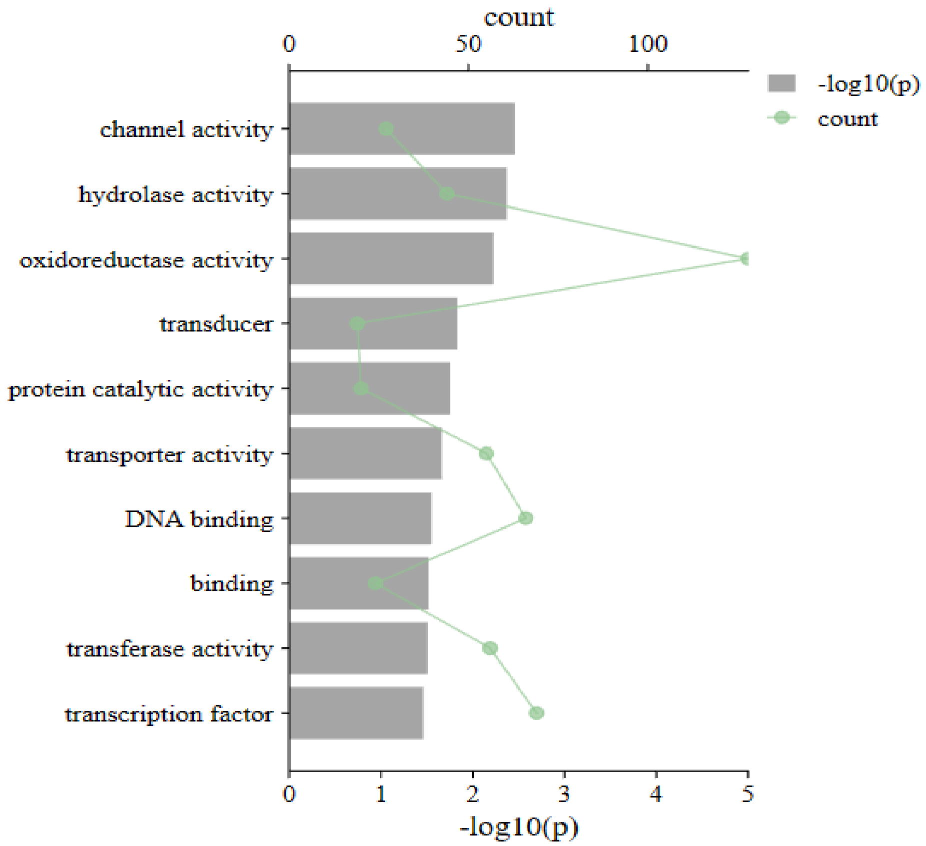


Figure 4.3 Go enrichment test of the essential genes identified in the SHuffle B transposon library grown at 42 °C. GO enrichment test on Ecocyc website was applied to the conditional essential genes for SHuffle B transposon library grown at 42 °C. Only genes present in the NCBI database that have been annotated were included. The plot is generated by online tool (https://www.bioinformatics.com.cn/).

Therefore, rather than understanding the conditional essential genes in the library at 42 °C, the insertion index scores which refer to the number of unique insertion sites normalized by gene length in the libraries of 37 °C and 42 °C against 30 °C were compared to identify the decreased or increased insertions. This is done by considering the insertions within the gene have a normalized read count in two different libraries. The fold-change metric in read counts for each gene can be used to determine the significantly declined or increased insertions that affect fitness (figure 4.4). In this analysis, 401 genes in 37 °C library and 852 genes in 42 °C library were determined to become essential compared to 30℃, suggesting more genes were required for the viability of SHuffle B strain grown under stress caused by higher temperature. Furthermore, the fold change in insertions frequency in genes that conferred fitness advantage and disadvantge were both identified. Since SHuffle B is temperature sensitive, genes with enriched insertions which attribute to resistance to high temperature, especially 42 °C, were considered to be worth to analyze.

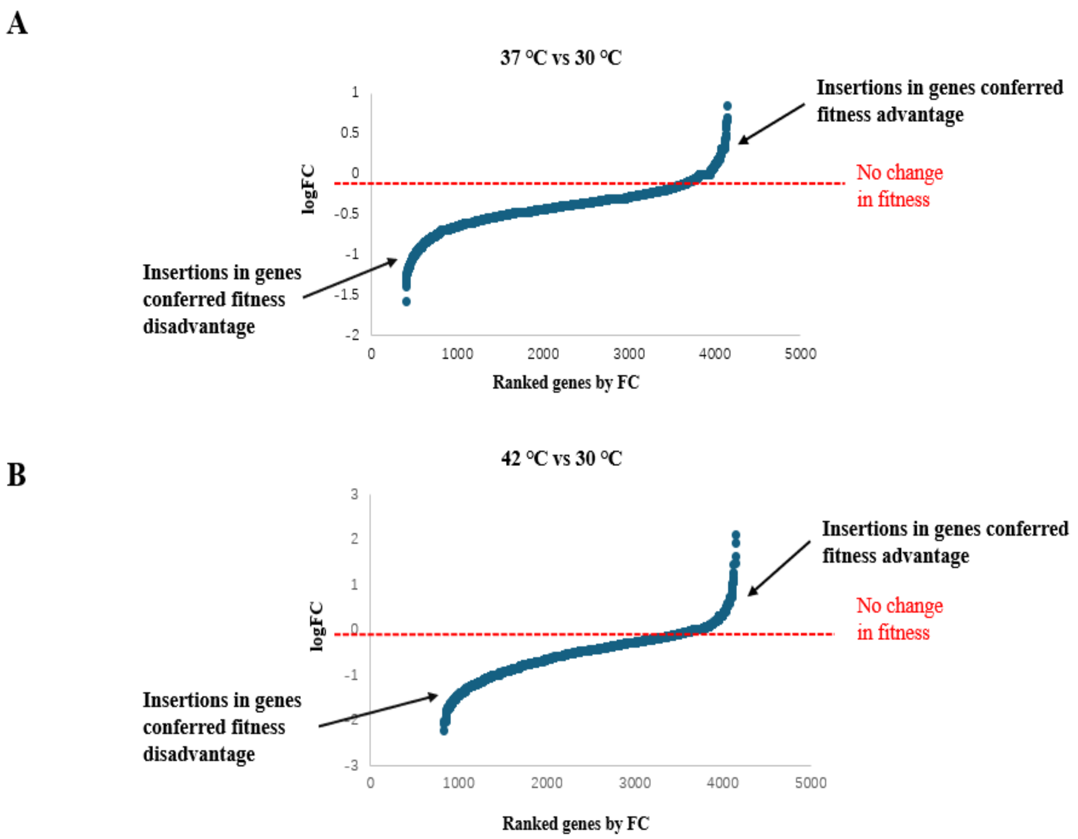


Figure 4.4 Fold change (FC) metric of the insertion frequencies within genes in SHuffle B library under 37 °C (A) and 42 °C (B) compared to 30 °C. Comparison of the frequency of the transposon insertion sites within genes to analyze the effect of different temperatures on fitness in each library dataset. Decreased insertions suggest mutations in that gene conferred detrimental impact on fitness, increased insertion conversely contributed to fitness advantage. The essential genes which the insertion index score were zero (no insertion) were not presented.

Genes with significant enriched insertions (logFC > 0.5) in 37℃ and 42℃ compared to 30℃ were identified. Genes containing enriched insertions in the library of 37 ℃ exhibited limited similarities compared to those genes within the library of 42 ℃. Insertions in 13 known genes (annotated genes in the NCBI database) were identified to have advantageous fitness in the 37℃ library, and they were mainly involved in the metal ion binding and protein activities. There were 97 genes were identified in the 42 ℃ library, compared to the enrichment test based on the molecular function of genes identified at 37 ℃, they were more associated with oxidoreduction activities acting on NADH dehydrogenase (quinone) activity and transmembrane transporter activity, genes involved in the metal ion binding also accounted for a large proportion. NADH combined with the quinone involved in the respiration chain, in the SHuffle strains are responsible for donating electrons for the reduction of the AhpC* and further involved in the reduction of Grx1 to suppress the lethal phenotype caused by double deletion of *gor* and *trxB*, suggesting that at 42 ℃ cells were suffered from extensive oxidative stress.

To further understanding genes containing enriched insertions conferred fitness advantage at 42 ℃, the frequency of insertions in the 97 genes identified as well as other genes were manually inspected on the Artemis browser. The analysis of insertion index scores to identify the increased or decreased insertions was possibly generated false-positives, visual inspection could avoid overestimation caused by discrepancies in gene length or insertion regions, and also be useful to discover genes dismissed in the statistical analysis. For example, genes with significant accumulated insertions in the *fis* gene was seen with a relatively lower logFC, but the logFC was two-fold higher in *fur* (Figure 4.5). In total, 50/97 genes were identified with significant enriched insertions in the 42 ℃ library compared to the library of 30 ℃ by manual inspection (Table 4.4).

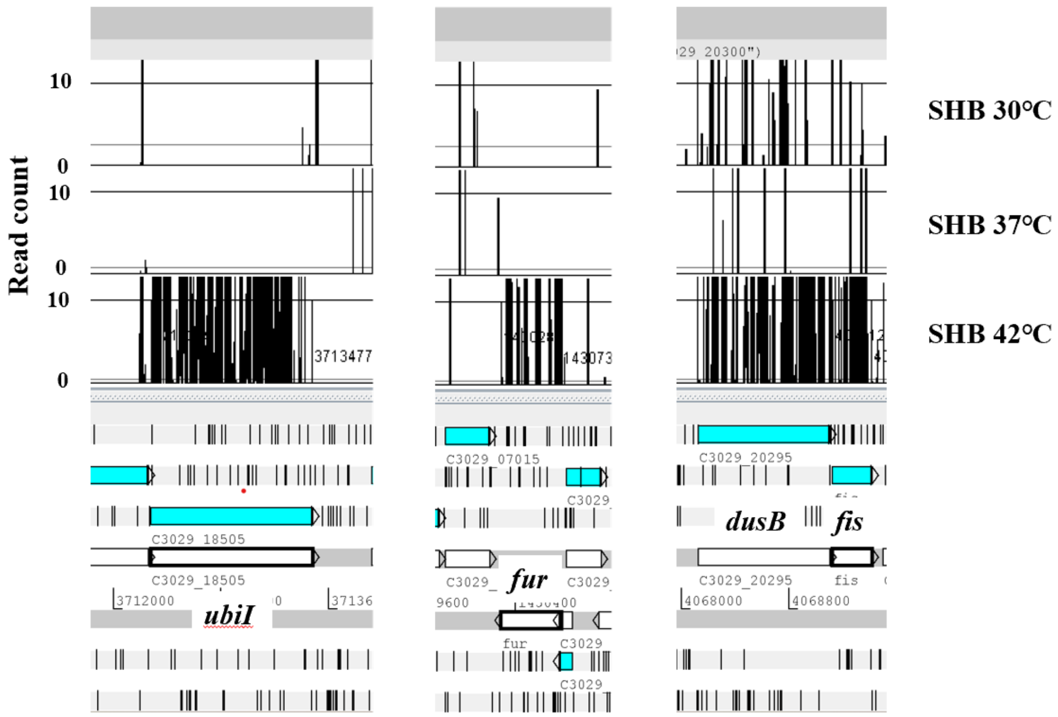


Figure 4.5 Visualization of selected genes with enriched transposon insertion sites in the SHuffle B library at 42 °C on Artemis. The frequency and reads of the transposon insertion sites were significantly increased in the *ubiI*, *fur, dusB* and *fis* gene in the SHuffle B tranpson library at 42 °C compared to the library at 30 °C and 37 °C. The insertion sites within each gene were shown in the same scale.

Table 4.4 Genes with enriched transposon insertion sites in the SHuffle B transposon library grown at 42 °C

Gene[1]	**Function**[2]
aroB	3-dehydroquinate synthase
aroK	Shikimate kinase
gpp	Exopolyphosphatase
ubiJ	Eypothetical protein
typA	GTP-binding protein TypA
dgkA	Diacylglycerol kinase
basS	Sensor protein BasS/PmrB
basR	Two-component system response regulator
deoB	Phosphopentomutase

cyoE	Protoheme IX farnesyltransferase
cyoD	Cytochrome o ubiquinol oxidase subunit IV
cyoC	Cytochrome o ubiquinol oxidase subunit III
cyoB	Cytochrome ubiquinol oxidase subunit I
cyoA	Cytochrome ubiquinol oxidase subunit II
ndh	NADH dehydrogenase
ydjN	L-cystine transporter tcyP
nuoN	NADH:ubiquinone oxidoreductase subunit N
nuoM	NADH:ubiquinone oxidoreductase subunit M
nuoL	NADH-quinone oxidoreductase subunit L
nuoK	NADH:ubiquinone oxidoreductase subunit K
nuoJ	NADH:ubiquinone oxidoreductase subunit J
nuoH	NADH:ubiquinone oxidoreductase
nuoG	NADH-quinone oxidoreductase subunit G
nuoF	NADH dehydrogenase
nuoE	NADH dehydrogenase
nuoC	NADH-quinone oxidoreductase subunit C/D
nuoB	NADH dehydrogenase
nuoA	NADH-quinone oxidoreductase subunit A
ratA	Hypothetical protein
sdaC	HAAAP family serine/threonine permease
pepP	Xaa-Pro aminopeptidase
ubiI	2-octaprenylphenol hydroxylase
yqiC	Hypothetical protein
pnp	Polyribonucleotide nucleotidyltransferase
mlaE	ABC transporter permease
mlaF	ABC transporter ATP-binding protein
arcB	Aerobic respiration two-component sensor histidine kinase ArcB
rng	Ribonuclease G
dusB	tRNA dihydrouridine synthase DusB
fis	Fis family transcriptional regulator
glpR	Transcriptional regulator
prlC	Oligopeptidase A
cpxP	Repressor CpxP
cyaA	Adenylate cyclase

cysM	Cysteine synthase
fre	NAD(P)H-flavin reductase
fur	Ferric uptake regulator
pstA	Phosphate transporter permease subunit PtsA
rluD	23S rRNA pseudouridylate synthase
truB	tRNA pseudouridine(55) synthase TruB

[1] The rank of gene was based on the location within the genome of SHuffle B

[2] Functions were taken from the Ecocyc website (https://ecocyc.org/)

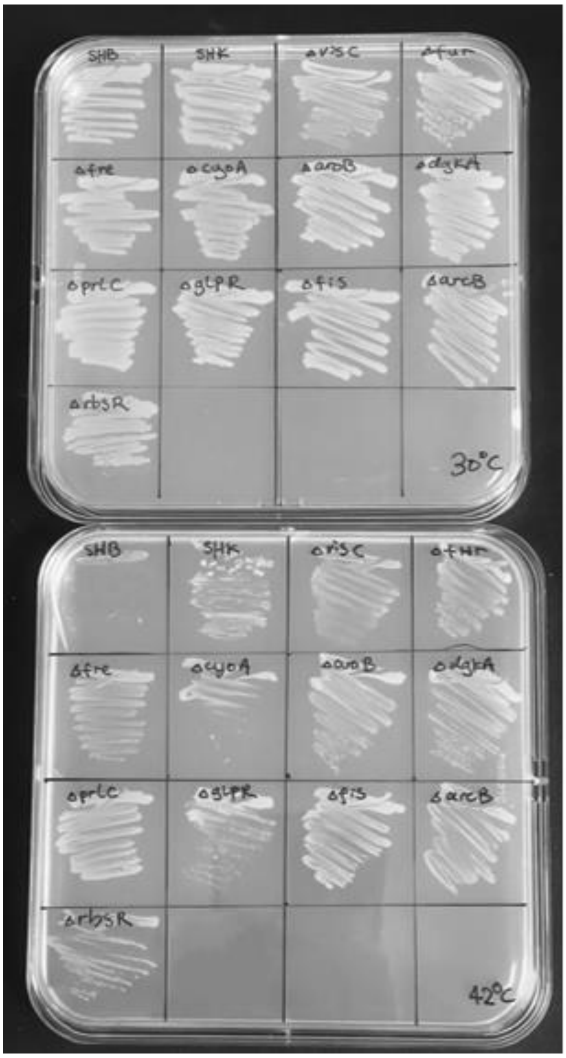
4.2.3 Confirmation of enriched insertions in relevant genes in SHuffle B leading to temperature resistance at 42 °C

To confirm the enriched insertions in genes identified conferred significant fitness advantages to SHuffle B grown at 42 °C, the top 10 genes based on their LogFC were selected, insertions in genes that encoded on the same operon were only picked the one on the 5' end. The single knockout mutations of the selected genes in *E. coli* K-12 strain BW25113 were obtained from the Keio library. The mutations were transduced into SHuffle B strains by P1 transduction to test whether these mutations could supress the temperature sensitivity. Since SHuffle K-12 strain shared the same genomic modifications with SHuffle B but showed relative resistance to higher temperatures, the mutations were also transduced into SHuffle K-12 strain as control.

Table 4.5 Top 10 genes selected with enriched insertions in the SHuffle B transposon library at 42 °C

Gene	**Function**	**logFC**
aroB	3-dehydroquinate synthase	0.927
dgkA	Diacylglycerol kinase	0.409
cyoA	Cytochrome ubiquinol oxidase subunit II	1.041
arcB	Aerobic respiration two-component sensor histidine kinase ArcB	1.031
fis	Fis family transcriptional regulator	0.611
glpR	Transcriptional regulator	0.927
prlC	Oligopeptidase A	0.855
fre	NAD(P)H-flavin reductase	1.62
fur	Ferric uptake regulator	1.929
ubiI	2-octaprenylphenol hydroxylase	1.217

SHuffle B and Shuffle K-12 carrying mutations in each gene were screen for viability at 30℃ and 42℃. Compared to the wild type, SHuffle B carrying selected mutations were found to be temperature resistant as expected (Figure 4.6), suggesting that the disruption of these genes in SHuffle B were able to increase the fitness advantage when faced to the oxidative stress under 42 °C. SHuffle K-12 strains carrying mutations in relevant genes showed no difference in growth compared to the wild type strain. The mutations in each strain were confirmed by colony PCR using corresponding primers (Section 2.4.3; Table 2.2).



SHuffle B	SHuffle K-12	ΔvisC	Δfur
Δfre	ΔcyoA	ΔaroB	ΔdgkA
ΔprlC	ΔgplR	Δfis	ΔarcB
ΔrbsR			

Figure 4.6 Mutations in selected genes in SHuffle B conferred temperature resistance at 42 °C. SHuffle B carrying mutations in relevant genes were screened for the growth at 30 °C and 42 °C. Growth of the wild type SHuffle B strain was set as a negative control, which had limited growth at 42 °C, SHuffle K-12 strain was used as positive control which had normal growth at 42 °C.

SHuffle B was demonstrated to produce certain proteins containing disulfide bond in greater yield than SHuffle K-12 in a normal growth condition (30 °C), and it was assumed to be resulted by higher enzyme activities in SHuffle B (Lobstein et al., 2012). Disruption of the reductive pathways make SHuffle strains cannot combat the highly oxidative stress therefore cannot survive at elevated temperatures higher than 42 °C. Therefore, it was considered whether the mutations recovered the temperatures sensitivity of SHuffle B at 42 °C could improve the production of the disulfide bond. To address this, alkaline phosphate (PhoA) assay was used to analyze the cytoplasmic disulfide bond formation in SHuffle K-12 strain carrying mutations in genes sated above. PhoA is activated by the oxidation of cysteine residue to form the disulfide bond. The plasmid pDH811, a derivative of pTrc99A, which expressed alkaline phosphate without signal sequence (Δ*ssphoA*) was transformed into the transductants carrying mutations in relevant genes in SHuffle K-12. SHuffle K-12 lacks chromosomal *phoA* therefore the detectable PhoA activities were all attributed by the disulfide bond formation.

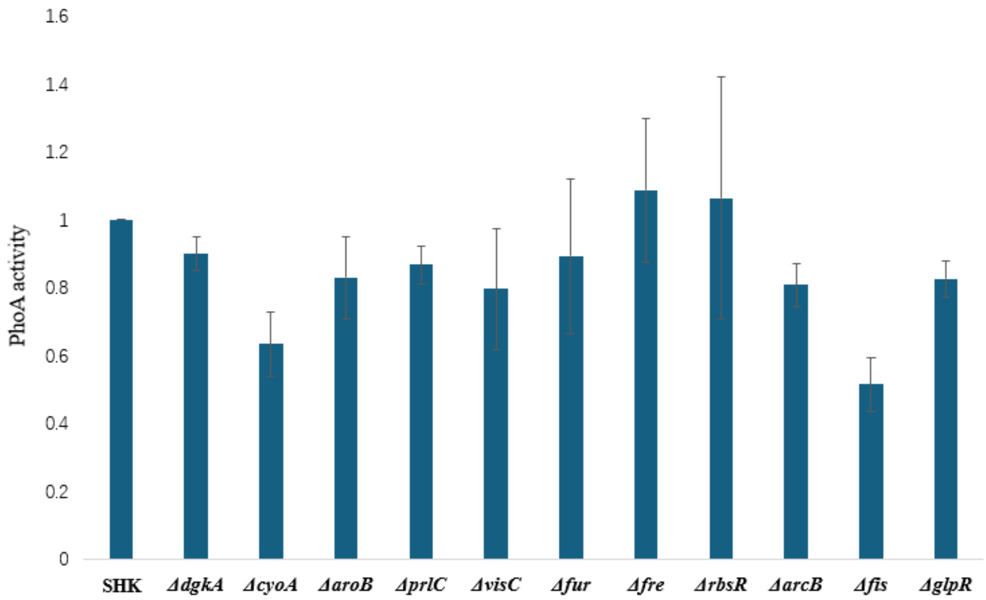


Figure 4.7 PhoA activities of the SHuffle K-12 (SHK) and SHK containing mutations in genes that could recover the temperature sensitivity of SHuffle B. The pDH811 which expressed alkaline phosphate without signal sequence (Δ*ssphoA*) was transformed into the SHuffle K-12 containing single-gene knock out of the genes which could recover the temperature sensitivity of SHuffle B. The expression was induced by 0.5 mM IPTG. The PhoA activity in SHK was set up as a control and normalized to 1, only the mutation in *fre* and *rbsR* showed increased PhoA activity compared to it was in SHK, and the p values of < 0.05. The data presented was the mean +/- SD of 3 independent replicates.

4.2.4 Genomic differences and insertion mutants in the DHB4 transposon library that were able to restore the temperature resistance in SHuffle B

To investigate why SHuffle K-12 was more tolerant to the oxidative stress caused by higher temperature even contained the same genetic modification as SHuffle B, the genetic differences was determined by transducing the transposon mutations from the transposon library of DHB4 into the SHuffle B. Since the bacteria background was different, which increased the difficulties in transduction, transduction was performed 10 timesfollowing the workflow shown in the Figure 4.7 and 6 individual mutants were isolated.

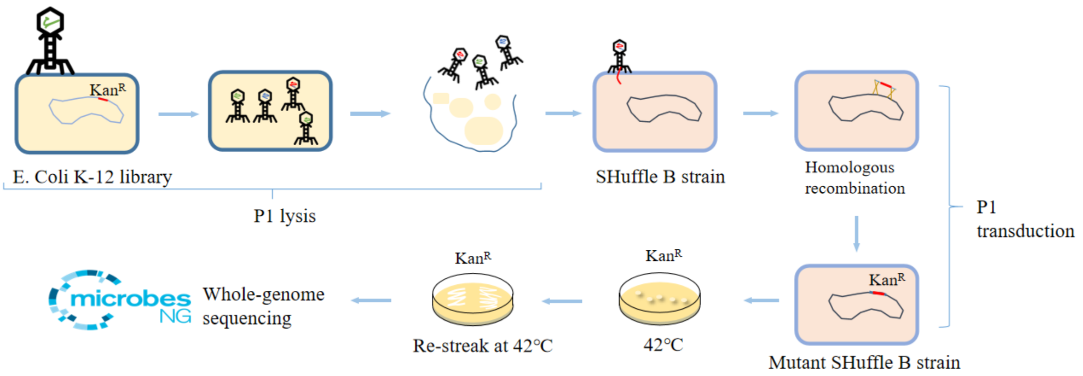


Figure 4.8 Workflow of the experiment to screen for mutations in DHB4 that can recover the temperature sensitivity of SHuffle B. The P1 lysate was grown in the diluted DHB4 transposon library and transduced the DNA fragments (~100 kb) including kanamycin cassettes from the transposon into SHuffle B by P1 transduction. The single colonies on the selective LB agar plates grown at 30 °C were restreaked to the LB plates with 50 mg/ml Kanamycin and grown at 42 °C to confirm the resistance to both kanamycin and higher temperature. The mutants were sent for whole-genome sequencing to identify the location in genes.

The transposon insertion sites of the 6 selected colonies (TR1-6) were determined by whole-genome sequencing and aligned to the genome of DHB4 and SHuffle by using NCBI BLAST (Table 4.6). The inserted genes were confirmed by colony PCR followed by Sanger sequencing. Since the P1 transduction normally accommodates ~100 kb of

DNA, linked genes that are close on the chromosome are inclined to be transferred. Whether the temperature resistance was caused by the disruption of a certain gene or genetic differences between DHB4 and SHuffle B was confirmed by linkage mapping. It estimated the distance between the identified insertions containing the kanamycin cassettes and the region causing temperature resistance by analyzing the recombination frequency. P1 lysate was prepared for each mutant strain and the region including the transposon insertion was transduced into SHuffle B. The tranductants were screened for kanamycin resistance and 24 individual colonies were selected to screen for the temperature sensitivity at 42 °C. The frequency of viable colonies was identified and the distance in two marker genes were calculated (Table 4.7). The temperature-resistance loci were determined by browsing genes that were apart from the transposon inserted gene around estimated distance on genome map on Ecocyc.

Table 4.6 Identified transpson insertion sites transduced from the DHB4 transposon library that recovered the temperature sensitivity of SHuffle B

Mutant	Gene insertion*	Description
TR1	*prlC*::Tn5..*gor*	Encode oligopeptidase A
TR2,5	*dmsA*::Tn5..*trxB* (different orientations)	Dimethyl sulfoxide reductase subunit A
TR3	*ycaK*::Tn5..*trxB*	Putative oxidoreductase
TR4	*hdeD*::Tn5..*gor*	Acid-resistance membrane protein
TR6	*rbsA*::Tn5	Ribose ABC transporter ATP binding subunit

* Gene was annotated if the location was close to *gor* or *trxB*, which are genes has been mutated in the SHuffle strains

Table 4.7 Linkage map to identify the distance between the location of the transposon insertion site and the temperature-resistance loci

Mutant	Viable colonies	Frequency (%)	Estimated distance (kb)
TR1	24/24	100	0
TR2	18/24	75	9.1
TR3	9/24	37.5	27.8
TR4	17/24	71	10.8
TR5	18/25	75	9.1
TR6	20/24	83	6

TR1 showed 100% transduction frequency so that the temperature resistance was considered to be resulted by disruption of the *prlC* gene, which is consistent with the results in Section 4.3. However, the *gor* gene encoding glutathione reductase was found tightly linked to *prlC* in the chromosome, deletion of *gor* and *trxB* in DHB4 and wild type B genome was the functional modification when engineered the SHuffle strains. Considering the accuracy of the experiment, the temperature resistance in mutant 1 could also be caused by recovering the mutation in *gor* and turned the SHuffle B back to the wile-type T7 Express, therefore conferring no growth defect at 42℃. The insertion sites in TR2 and TR5 were both found in *dmsA* gene but in different orientations, the estimated temperature resistance loci were considered to restore *trxB*, which encodes for the thioredoxin reductase. TR3 was predicted to restore *macB* or *makE*, however, neither of these two genes in DHB4 showed any genetic difference compared to B strain genome. As *trxB* was found located close to the estimated locus in genome, mutant 3 was assumed to restore the *trxB*. TR4 was determined to restore *gor* and TR6 restored *rbsR*.

Since all of the mutations were found to restore either *gor* or *trxB*, which partially recovered the genotype of the wild type to recover the temperature sensitivity of SHuffle B, except TR6, it is of interest to determine whether the temperature resistance was caused by genetic differences in this region between DHB4 and SHuffle B, or *rbsR* gene itself. To identify the genetic difference in the *rbs* region between DHB4, B strains and TR6, the genome sequence of DHB4 was aligned to wild-type B and TR6 then visualized on the Artemis browser. By comparison, there were several unique genomic features found in the *rbs* region in DHB4 and E. coli B (Figure 4.8). There was an insertion element IS3 located at the upstream of *rbsD* in *E. coli* B but an IS3-like IS2 element located at the downstream of *rbsR* in DHB4.

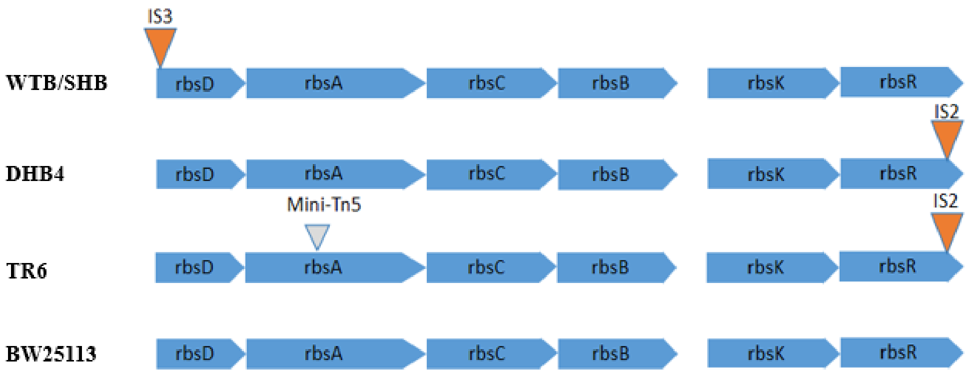


Figure 4.9 Genetic differences in the rbs region of the relevant strains in this study. The genome sequence of DHB4 was aligned to wild-type B strain (WTB; T7 Express), TR6, the SHuffle B mutant transduced from DHB4 transposon library which conferred temperature resistance, and BW25113. There was an insertion element IS3 located at the upstream of *rbsD* in *E. coli* B but an IS3-like IS2 element located at the downstream of *rbsR* in DHB4. The TR6 contained an IS2 insertion element the same as DHB4, whereas BW25113 had no insertion element in the *rbs* region.

The IS2 element in DHB4 that has been introduced into SHuffle B was predicted to suppress the temperature sensitivity. Considering the accuracy of estimation of distance may affect the prediction of the temperature-resistant loci, all of the rbs genes (*rbsD*, *rbsA*, *rbsC*, *rbsK* and *rbsR*) in DHB4 were amplified by colony PCR and have been inserted into the pDSW204 by plasmid cloning (Section 2.4.7, 8, 9). The *rbsR* gene in DHB4 was not able to be amplified after many attempts. The plasmid containing genetic copy of other *rbs* genes were transformed into SHuffle B and selected on LB plates with ampicillin. 4 random individual colonies were selected from the plates of SHuffle B containing plasmid with targeted genes to screen for temperature sensitivity, however, none of them were observed to be temperature resistant at 42℃. Therefore, the suppression of temperature sensitivity was caused by mutations in *rbsR*. To prove this, BW25113 containing Δ*rbsR* mutant was obtains from the Keio library. The mutations were transduced into SHuffle B and selected for kanamycin resistance. Random colonies were selected to screen for the growth at 42 ℃. The Δ*rbsR* was able to recover the temperature sensitivity of SHuffle B (Figure 4.5) and there was no unique insertion elements within the *rbs* region of the BW25113 genome (Figure 4.8).

4.3 Discussion

In summary, the TraDIS data revealed a 3-fold increase conditional essential genes in the SHuffle B grown at 42 ℃ compared to 30 ℃ and 37 ℃, this could be explained by that disruption of genes lower the tolerance of SHuffle B to the oxidative stress under higher temperatures. Since SHuffle B cannot grow at 42 ℃ in general, rather than comparing the essential genes, insertions that conferred temperature resistance was more reliable to investigate the temperature sensitivity of SHuffle B. There were 50 genes (Table 4.4), including *rbsR*, that contained significantly enriched transposon insertions in the 42 ℃ library compared to libraries grown at 30 ℃ and 37 ℃. Majority of the genes were involved in the respiratory electron transport chain, especially the NADH to quinone, suggesting that disrupting the aerobic respiration could relieve the detrimental effects of oxidative stress at 42 ℃.

The electron-transport chains in *E. coli* are composed of different dehydrogenase and terminal reductase or oxidase linked by quinones. The operons In aerobic respiration, oxygen is the preferred electron acceptor, the operons encoding for NADH-dehydrogenase (*nuo* genes) and enzymes required for the biosynthesis of cofactors (*cyo* genes) were all shown to have enriched insertion sites in this TraDIS analysis. Aerobic respiration in *E. coli* is based on NADH interacting with NADH:quinone oxidoreductase to reduce the UQ, the cytochromes ubiquinol oxidase oxidized UQ and reduce the oxygen (Borisov et al., 2011). There were 4 genes encoding for NADH dehydrogenase, 9 genes encoding for NADH:quinone oxidoreductase and 4 genes encoding for cytochromes ubiquinol oxidase subunit, with enriched transposon insertions in which increased the temperature resistance in SHuffle B at 42 °C. In addition, insertions in *arcB*, which encode for the sensor histidine kinase that mediates anaerobic repression of enzymes involved in the aerobic respiration also conferred temperature resistance (Brown et al., 2022). The best way to confirm this is to check the viability of SHuffle B strains in an anerobic condition at 42 °C. In addition, FNR is a cytoplasmic oxygen responsive regulator required for anaerobic respiration (Unden and Bongaerts, 1997), if disruption of the aerobic respiration could suppress the temperature sensitivity of SHuffle B, *fnr* gene should be essential at 42 °C.

Mutations in the top 10 genes with the highest logFC were selected to confirm the recovery of the cell growth of SHuffle B at 42 °C, all of the single-gene knockout of these genes conferred temperature resistance. These genes were involved in diverse biological processes but could be divided into genes encoding for enzymes (*cyoA, ubiI, fre, dgkA, aroB and prlC*) and genes involved in the regulatory networks (*fur, fis, glpR and arcB*). This observation indicates that diverse enzyme activities in SHuffle B could be a reason leading to the temperature sensitivity of SHuffle B. Global regulators such as Fis are supposed to regulate abundant metabolic pathways, so that deficient in metabolic activity could also increase the temperature resistance.

The mutation in *rbsR* gene was found to restore the growth of SHuffle B at 42 °C in the genetic screening for temperature resistance caused by transformation of the genome feature from the DHB4 transposon library into SHuffle B. Mutations in the *rbsR* gene resulted in temperature resistance was determined by linkage mapping (Table 4.6) and *rbsR* is also included in the 50 genes with enriched transposon insertions in the SHuffle

B 42 °C library. Gene *rbsR* encodes the transcriptional regulator RbsR, the latest RNAseq experiment for SHuffle B Δ*rbsR* in Huber's lab revealed downregulated *ilv* genes, one explanation is mutation in *rbsR* gene could affect the pentose phosphate pathway, which is the major source of NADPH in the cell or shutting down the branched-chain amino acid synthesis results in accumulation of homoserines and also result in inhibition of the NADPH. It could be a factor that increased the temperature resistance in SHuffle B Δ*rbsR* or it could be irrelevant. According to the genome alignment results (Figure 4.8), the temperature resistance is possibly a result of the genetic differences in insertion elements between *E. coli* K-12 and *E. coli* B, which was further identified by plasmid cloning. The cloning of genes in the *rbs* region except *rbsR* from DHB4 to SHuffle K could not suppress the temperature sensitivity and the amplification of the *rbsR* did not work.

The aim of this study was to investigate the physiology of SHuffle strains and further improve the production of disulfide bonds in the cytoplasm. The oxidative stress caused by SHuffle mutations could be the main result of the temperature sensitivity as the higher temperatures normally lead to high enzyme activities and enhance the metabolic processes, which accentuate the oxidative stress. However, enriched insertions in 50 genes including genes involved in the respiration chains suppress the temperature sensitivity of SHuffle B. It was useful to construct a strain could survive at higher temperatures and produce disulfide bonded protein in higher yield. The phoA assay was used to analyze the disulfide bond formation in the SHuffle strains containing selected mutations. However, only mutations in *rbsR* and *fre* slightly increased the formation of disulfide bond. Whether these two mutations or other undefined mutations could be used to produce proteins in SHuffle strains under higher temperatures need to be further investigated.

Chapter 5
Identification of transposon insertion mutants conferring azide resistance

5.1 Introduction

Sodium azide is a well-studied inhibitor of SecA, mutations conferred azide resistance map in the *secA* gene. Azide inhibits the ATP hydrolysis by interacting with the catalytic region of F-ATPase and preventing the release of ADP (Oliver et al., 1990). The mechanism of inhibition is thought to be similar to SecA, majority of the mutations in *secA* conferred increased azide resistance result in amino acid substitutions in one of the two NBDs of SecA and facilitate the nucleotide exchange in vitro (Huie and Silhavy, 1995; Schmidt et al., 2000, Figure 1.6). These results supported that azide inhibits protein-stimulated SecA-mediated translocation by inhibiting the ATP turnover. Azide inhibits deinsertion of SecA from the SecYEG translocon by stabilizing the SecA bound ADP (Vasilyeva et al., 1982; Minkov and Strotmann, 1989; Wolk et al., 1997).

However, recent studies using TraDIS to screen for mutations that affect the susceptibility of *E. coli* to azide suggested that azide inhibits SecA by disrupting the structure of the CTT (Chandler et al., 2018; Cranford-Smith et al., 2020). In most cases, insertions disrupting *sec* genes caused a defect in Sec-dependent protein translocation and generally increased the sensitivity to azide, but mutants containing insertions truncating the C-terminus of SecA become enriched during growth in the presence of azide. CTT of SecA is composed of a FLD and a MBD, and MBD is responsible for interacting with metal ions, SecB, ribosome and phospholipids. Disruption of the CTT in the presence of azide has been shown to interfere the binding of the MBD to iron and membrane phospholipid without addition of ATP (Chandler et al., 2018), and therefore the defect in translocation is not caused by inhibition of rate of ATP turnover.

Truncation of proteins produced by proteolytic cleavage of a small region of the annotated protein (typically the N-terminal domain or C-terminal domain) has been used in the analysis of regulating signal transduction or other cell activities (Papanikou and Economou, 2005). Recent work suggests that cells producing the CTT-truncated form of SecA conferred moderately higher resistance to azide (Chandler et al., 2018).

In this study, a transposon library of MG1655 Δ*tig*Δ*dsbC* was used to screen for the mutants that increase the resistance to sodium azide, and how these mutations affect the expression of SecA was identified. Previous studies generally focused on the effect attributed to mutations in SecA, investigating mutations unlinked to *secA* could support the previous ideas or generate a new approach for the mechanism of action of azide. The lab-made CTT-truncated SecA was used as a control to identify the SecA isoforms generated in the western blot.

5.2 Results

5.2.1 Genetic screen for azide-resistant mutants in the transposon library of MG1655 Δ*tig*Δ*dsbC*

All known mutations that cause azide resistance are identified in the *secA* gene. However, it was possible that mutations at multiple loci can cause azide resistance, but only mutations in the *secA* gene are sufficiently strong to cause high-level resistance. For example, screens for azide resistant mutants normally used very high concentrations of azide (typically ≥ 3mM), mutations for moderately increased resistance to azide may have been missed. However, the background growth at lower azide concentrations could interfere with the isolation of mutants with increased resistance to azide. One way to avoid interference from background growth is to select for mutants in a strain that has increased susceptibility to azide (Cranford-Smith et al., 2020). Previous work suggested that mutations in the *tig* and *dsbC* genes increased susceptibility to azide. To confirm that mutations in *tig* and *dsbC* increase susceptibility to azide, *E. coli* MG1655 and its derivative Δ*tig*, Δ*dsbC*, Δ*tig*Δ*dsbC* double mutant strains were grown in the presence of subinhibitory concentrations of azide (Figure 5.1). Both Δ*tig* and Δ*dsbC* mutations detectably increased the susceptibility of *E. coli* to sodium azide, and the Δ*tig*Δ*dsbC* double mutant showed strongest growth limitation, which is consistent with the plate test. The growth of MG1655 was inhibited by addition of 1.0 mM azide, cells containing Δ*tig*, Δ*dsbC* and Δ*tig*Δ*dsbC* mutations were completely inhibited by addition of 0.75 mM azide.

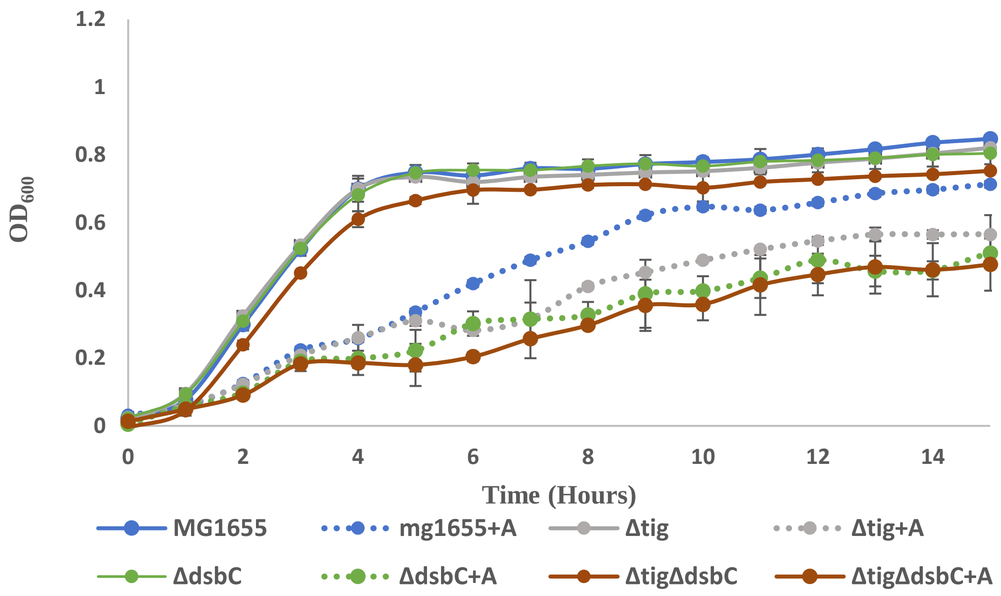


Figure 5.1 Mutations in *tig* and *dsbC* increased susceptibility of *E. coli* to sodium azide. Growth curve of MG1655 strain, Δ*tig*. Δ*dsbC*, and Δ*tig*Δ*dsbC* double mutant strain over 15 hours in LB broth and LB broth supplement with 0.5 mM sodium azide (A). OD_{600} was taken at each hour timepoint. The data was the mean +/- SD of 3 independent replicates.

To determine the concentration of sodium azide to screen for transposon mutant with increased resistance, the transposon library of MG1655 Δ*tig*Δ*dsbC* was inoculated on the LB plates supplemented with 0.5 mM, 0.75 mM, 1mM, and 1.5 mM sodium azide. Six transposon insertion mutants (AR1 - 6; Table 2.1) were isolated on the plate with 1.5 mM sodium azide and confirmed for resistance by restreaking on LB plates containing 1.5 mM sodium azide and by filter disc assay (Figure 5.2). All of the mutants showed complete resistance (no clear zone) when exposed to 20 µl of 0.5 M sodium azide.

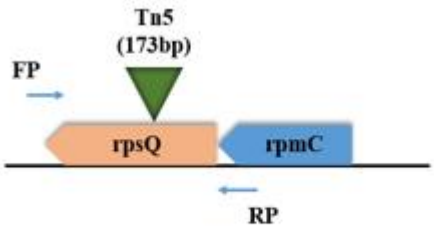
To confirm the increased azide resistance was caused by the transposon insertion mutations, the insertions were transduced into the MG1655 Δ*tig*Δ*dsbC* and another *E. coli* K-12 derivative, BW25113. In all cases, the freshly transduced strains displayed increased resistance to sodium azide compared to the recipient strain, indicating that the azide resistance caused by these mutations were not specific to MG1655 or Δ*tig*Δ*dsbC* double mutant.

The locations of the insertions in AR1-AR5 were determined by arbitrary PCR using a primer specific to the mini-Tn*5* transposon and Sanger sequencing. The location of insertion in AR6 was determined by whole genome sequencing. The identified insertions are listed in Table 5.1. The locations of the insertion genes were confirmed by PCR using the Tn*5*-speicific primer and a primer specific to each insertion site. For example, the location of the transposon insertion site in the *rpsQ* gene and PCR validation using primers specific to *rpsQ* were shown in Figure 6.2.

Table 5.1 Locations of the transposon insertion site in the azide-resistant mutants isolated from the mutagenesis library of MG1655 Δ*tig*Δ*dsbC*

Mutant	Gene insertion	Description
AR1, 2, 4	*rpsQ*	Encode 30S ribosomal protein S17; upstream *secY*
AR3	*rpsA*	Translation initiation; *ssF* mutant within *rpsA* suppress SecYts24
AR5	*srmB*	RNA helicase; facilitates an early step in the assembly of the 50S subunit of the ribosome
AR6	*nusB*	Encode transcription antitermination protein NusB; correct ribosome biogenesis; required for transcription of ribosomal RNA (rRNA) genes, mutant suppress SecYts24

A



B

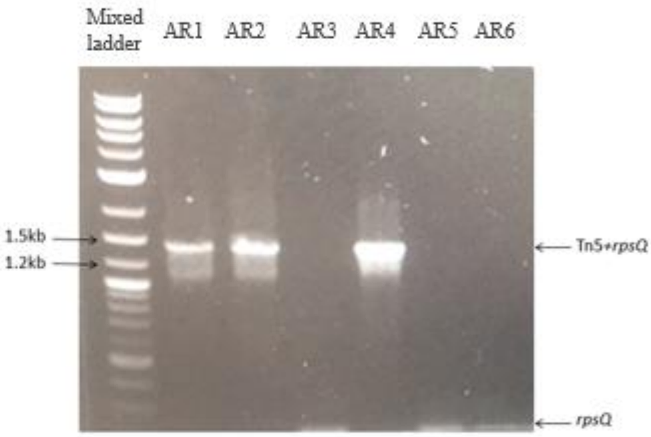


Figure 5.2 Identification of the transposon insertion site in the *rpsQ* gene. A) Cartoon representation of the Tn5 transposon mutants conferred azide resistance. The Tn5 insertion (green triangle) was shown at precise base pair relative to the start codon of the gene located in the chromosome. FP and RP correspond to the forward primer and reverse primer used in this study to confirm the insertions. B) Agarose gel of colony PCR to map the mutation using designed *rpsQ* forward and reverse primers (resolved on a 1% agarose gel). Location of the transposon insertion site in AR1, 2 and 4 were in the *rpsQ* gene which shows the size of the *rpsQ* gene and the Tn5 transposon (1221 bp). AR3, 5 and 6 contained no transposon insertion showed the normal size of the *rpsQ* gene (255bp).

5.2.2 Sensitivity of the azide-resistant mutants to other antibiotics

Mutation in *rpsQ* can cause resistance to the antibiotic neamine. To determine whether other azide-resistant mutations isolated in this study could be resistant to neamine, the susceptibility of the AR1-6 as well as MG1655 and MG1655 Δ*tig*Δ*dsbC* to neamine was determined by filter disc diffusion assay. All of the azide-resistant mutations also caused increased resistance to neamine. Neamine is a breakdown product of neomycin and inhibits bacterial translation, suggesting that the mutations cause azide resistance by affecting protein synthesis. To further investigate the molecular mechanism of azide-resistance, susceptibility of the AR1-6 to the antibiotics such as streptomycin, rifampicin and chloramphenicol were compared to the double mutant and the MG1655 (Table 5.2). All of these three antibiotics inhibit protein synthesis: streptomycin inhibit 30s ribosome subunit binding to mRNA (Waters and Tadi, 2023), rifampicin inhibits the bacteria RNA synthesis (Calvori et al., 1965) and chloramphenicol inhibits protein chain elongation. However, none of the AR1-6 caused detectable resistance to the selected antibiotics but tended to be slightly sensitive to chloramphenicol, suggesting that azide only shared a mechanism of action with neamine. To confirm this, neamine resistance was determined on the azide-resistant mutants linked to the *secA* gene by filter disc assay (Table 2.1, Table 5.2). However, none of the selected mutations conferred azide resistance in *secA* caused increased resistance to neamine.

Table 5.2 Susceptibility of the azide-resistant mutants, MG1655, and MG1655 ΔtigΔdsbC to different antibiotics

Strain	1.5 mM Sodium azide	Antibiotics			
		Neamine*	Streptomycin*	Chloramphenicol*	Rifampicin*
DRH-855	+	+	++	+++	+++
DRH-1292	-	++	+	++	+++
AR1	++++	++++	+	++	+++
AR2	++++	++++	+	++	+++
AR3	++++	++++	+	++	+++
AR4	++++	++++	++	++	+++
AR5	++++	++++	+	++	+++
AR6	++++	++++	++	++	+++
DRH-503	-	+	-	-	-
DRH-997	++++	+	-	-	-
DRH-998	++++	+	-	-	-

+ Corresponds to the quantification of inhibition (size of clear zone), where ++++ correspond to complete resistance or no clear zone and + corresponds to no resistance or clear zone > 30mm

– Corresponds to N/A

* 15 µl of each antibiotic was added on the 6 mm filter disc, the stock concentrations are listed in Table 2.4.

5.2.3 Expression of SecA in the azide-resistant mutants

Although the azide-resistant mutants isolated in this study were not linked to SecA, the insertion mutations isolated in the genetic screen caused azide resistance could also translationally control the expression of the SecA protein. The steady state levels of SecA in the mutants and MG1655 grown in the absence or presence of subinhibitory concentrations of azide were determined by western blotting. Consistent with the previous reports (Liebke, 1988), SecA migrated as two distinct isoforms in SDS-PAGE (Figure 5.3). By comparison, accumulation of large isoform was found in the wild-type strain and the azide-resistant mutants after treatment with 0.5 mM sodium azide, the small isoform was also relatively increased.

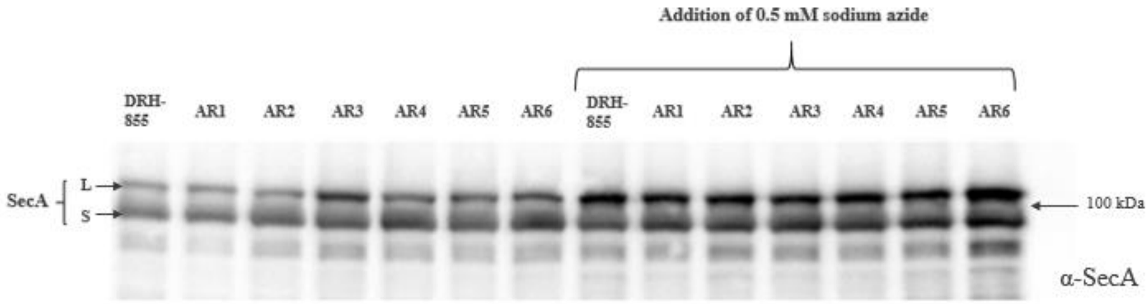


Figure 5.3 Expression of SecA isoforms was affected by the transposon mutations in the presence of sodium azide. The steady state levels of SecA in MG1655 (DRH-855) and azide-resistant mutants (AR1-6) were analyzed by western blotting against SecA in this work. Cells were grown in LB broth absence or presence of 0.5 mM sodium azide. SecA migrated as two distinct isoforms, large isoform (L) at ~110 kDa and small isoform (S) at ~90 kDa. The expression of bot SecA isoforms were detectable increased in the presence of sodium azide.

5.2.4 Identification of the two isoforms of SecA

One possibility was that the larger band was full-length SecA and the smaller band was a truncated protein. In order to investigate this possibility, the migration of SecA in a wild-type strain (DRH-503), and strain containing plasmid induced SecA (DRH-725), biotin-tagged SecA (DRH-839) and YFP-tagged SecA (DRH-991) was detected by western blotting (Table 2.1; Figure 5.4A). Biotin and YFP were tagged to the C-terminus SecA, but only the large isoform of SecA in DRH-839 and DRH-911 migrated slower compared to it was in the DRH-503 and DRH725, suggesting that the large isoform was full-length SecA and the small isoform could be C-terminus truncation of SecA. Since biotin is smaller in size, the migration was not significant. The expression of SecA in these strains with the treatment of azide for 30 min and at the beginning of the cell growth were measured. The large form was aggregated in varying degress in the presence of azide, but no consistent change was found in the small isoform.

To prove the small isoform corresponds to the C-terminus truncated SecA, the lab-made strain containing an insertion which truncating the CTT of SecA (DRH-1064; Table 2.1) was blotting against SecA comparing to DRH-503, DRH-839 and DRH-991 (Figure 5.4B). DRH-1064 only had the small isoform which was the same size as the small isoform in other strains as expected, suggesting that the large isoform corresponds to full-length SecA and small isoform represents CTT-truncated SecA. This was further confirmed by western blotting against biotin (Figure 5.4C), only DRH-839 had a single band in the blot and at the same size as the large isoform, suggesting that only the large band contained C-terminus of SecA. The aggregation of full-length SecA (large isoform) was significant in DRH-503 and DRH-839 when cell growth in the presence of sodium azide but not in DRH-991. Cells containing CTT-truncated SecA were suggested to be less susceptible to azide in the previous study, however, no difference in expression of the SecA was found in DRH-1064.

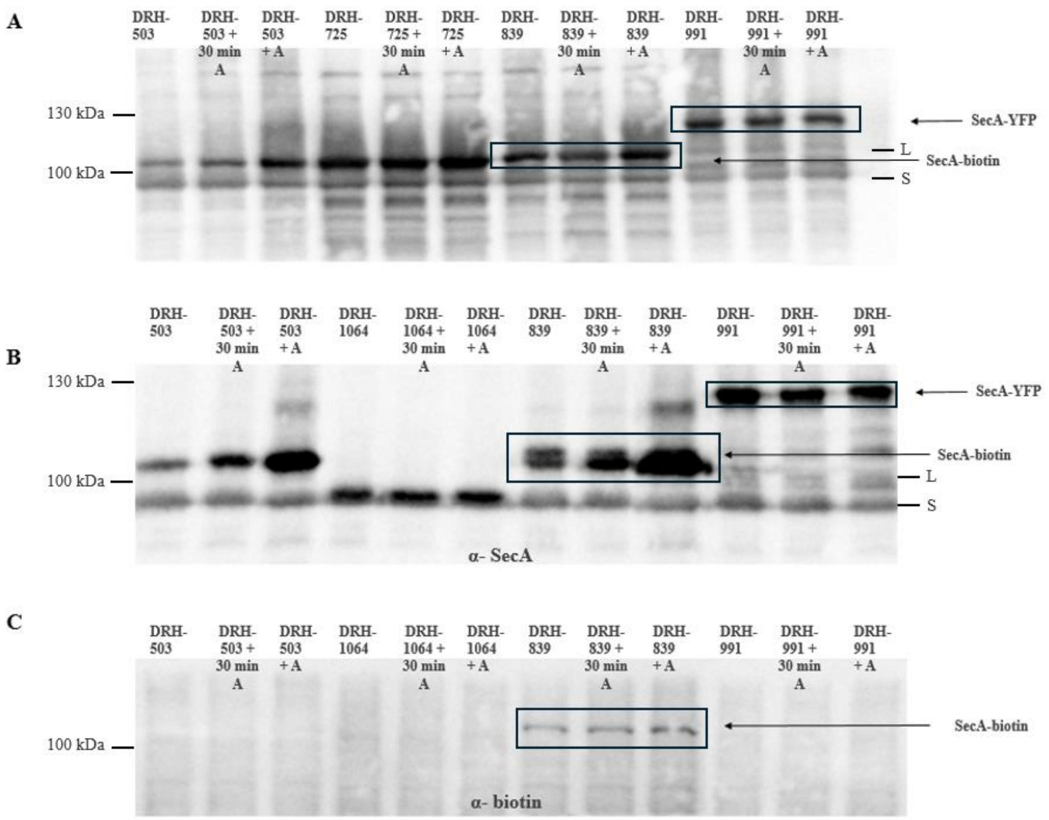


Figure 5.4 Identification of the small isoform of SecA. (A) *E. coli* strain DRH-503 (MC4100), DRH 725 (pTrc-*secA*), DRH 839 (*secA-biotin*), and DRH 991 (*secA-yfp*) were grown in the LB broth absence of sodium azide, presence of 30 min 0.5 mM azide (30 min A) and 0.5 mM azide (A). Cultures were grown to mid-log phase and the cell lysastes were analyzed by western blotting against SecA. The expression of SecA on the plasmid was induced by 1 mM IPTG. (B)DRH-503, DRH 1064 (*secA827*), DRH 839, and DRH 991 were grown in the LB broth treated with, no azide, 30 min 0.5 mM azide (30 min A) and 0.5 mM azide and the expression of SecA was analyzed by western blotting against SecA or (C) biotin.

5.2.5 Effect of mutations in *secA* gene on production of the small SecA isoform

The results above indicated that azide resistance can be caused by increased production of the small SecA isoform. To determine whether azide-resistant mutations in the *secA* gene can affect production of the small isoform, the expression of SecA was analyzed in several lab-stock *secA* mutations that conferred azide resistance treated with azide

(Figure 5.5). Compared to the wild-type strains, there was no consistent change in the levels of the large SecA isoform. However, all of the azide-resistant *secA* mutants exhibited increased expression in the C-terminus truncated SecA (smaller isoform), which is consistent with the disruption of the C-terminal tail in SecA making the cell less susceptible to azide. Indeed, production of the smaller SecA isoform was even more pronounced in the *secA* mutants than in the transposon mutants isolated in our screen, consistent with the level of azide resistance caused by the two different classes of mutation. The expression levels of the C-terminus truncated SecA in the mutations isolated from the transposon library were less than the point mutations in secA, which could explain why these point mutations in *secA* exhibited higher azide resistance level.

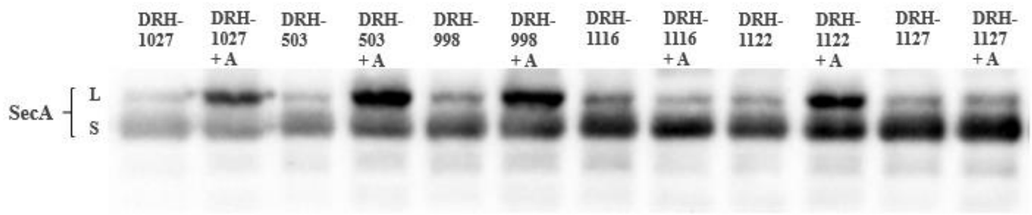


Figure 5.5 Effect of azide on the production of the small and large SecA isoforms in azide-resistant *secA* mutants. Strains containing *secA* mutations conferred azide resistance (DRH-998, DRH-1116, DRH-1122, and DRH-1127) as well as their wild-type strains (DRH-1027 and DRH-503) were grown in the LB broth in the absence or presence of 0.5 mM sodium azide (A). DRH-998 is the derivative of DRH-503, others are derivatives of DRH-1027. The protein samples were prepared from each strain and the expression of SecA was identified by western blotting against SecA. Aggregation of the full-length SecA was found in DRH-1027, DRH-503, DRH-998 and DRH-1122, which was not consistent in all of the *secA* mutations. The expression level of CTT-truncated SecA increased in the presence of azide.

5.2.6 Effect of SecM on azide resistance

Expression of SecA is controlled by translational stalling of *secM*, a gene upstream of *secA* which encodes for a SecA translation regulator. All of the azide-resistant mutants isolated in this study were found to interfere genes involved in ribosomal biosynthesis, although increased azide resistance did not appear to be result of a change in the expression of SecA, it was nonetheless possible that SecM arrested ribosomes were required for the increased production of the small isoform of SecA. To determine whether the azide-resistant insertion mutants affected translational stalling by SecM, the mutations in the azide-reisistant mutants were transduced into DHB4 (Δ*phoA*) with a transformed plasmid containing a *secM-secMarres*t-phoA fusion (DRH-736). The PhoA activities were determined (Section 2.8.3) but no consistent change was detected in the mutant strains (Figure 5.6), suggesting that the transposon mutants do not affect the SecM-induced translationally stalling.

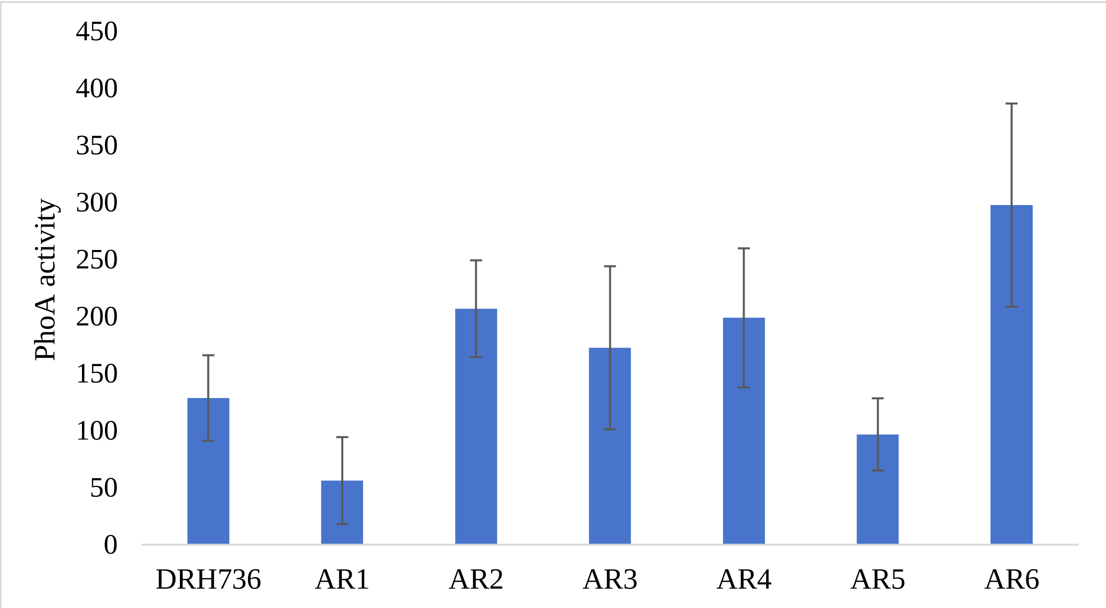


Figure 5.6 Alkaline phosphatase (PhoA) activity of the azide-resistant mutants. To determine the SecM-induced translationally stalling in the azide-resistant mutants (AR1-6), the mutations were transduced in DRH736, which is DHB4 with a plasmid containing a *secM-secMarres*t-phoA fusion and measured the PhoA activity in the presence of 1 mM IPTG and 0.5 mM sodium azide. The data presented was the mean +/- SD of 3 independent replicates.

To determine whether increased expression of SecM, which leads to more SecM-arrested ribosomes, could cause increased resistance to azide, a plasmid producing a SecM-PhoA fusion protein was transformed into the *E. coli* MC4100 (MC736). The growth of DRH-736 was compared to the negative controls (DRH-503 and MC717) in the absence or presence of sodium azide (Figure 5.7). Overexpression of SecM in cells conferred no growth advantage in the presence of azide, suggesting that the expression of SecM has no effect on azide resistance.

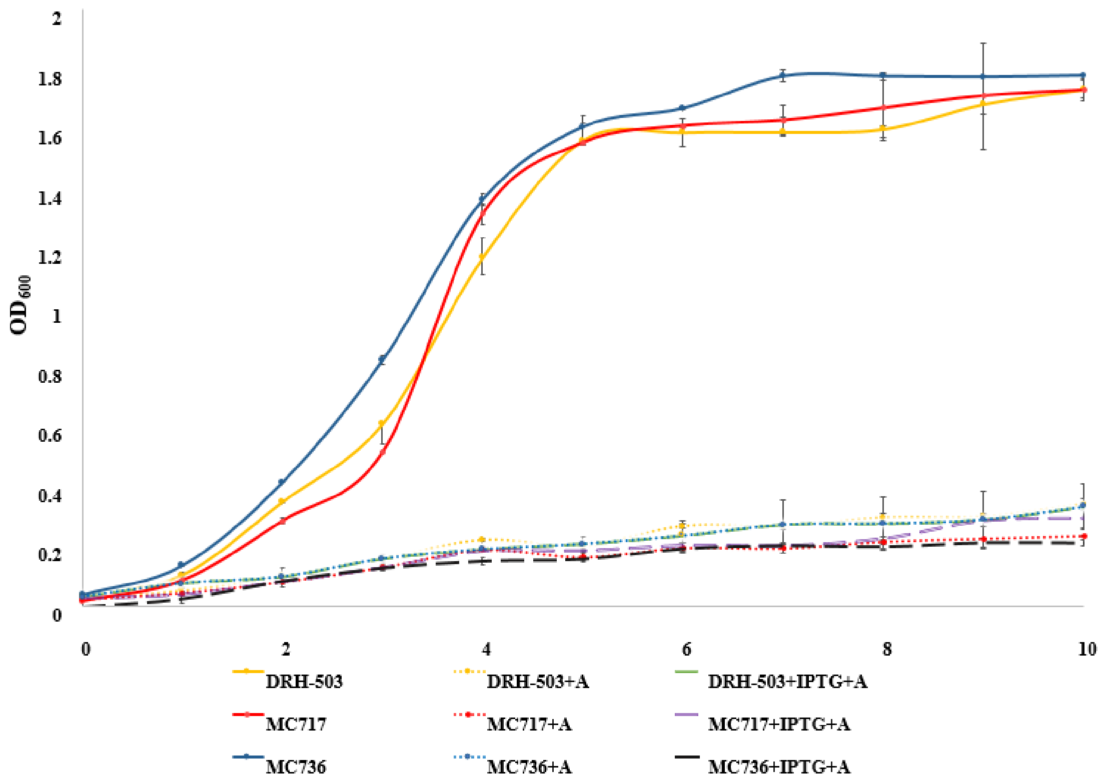


Figure 5.7 Expression of SecM had no effect on the azide resistance. Growth curves of DRH-503 (MC4100), MC717 (MC4100 with plasmid containing *phoA*) and MC736 (MC4100 with plasmid containing *secM-phoA*) over 10 hours in LB, LB addition of 0.5 mM sodium azide, LB addition of 0.5 mM sodium azide and of 0.5 mM IPTG. OD_{600} of culture was taken at every hour time point, the data presented was the mean +/- SD of 3 independent replicates.

5.3 Discussion

The results presented in this chapter indicated that *E. coli* produces two isoforms of SecA under normal growth conditions, which is consistent with results in the previous study (Liebke, 1986). It suggested that SecA protein may exist more than two isoforms. There were two significant charged SecA species identified by 2D gel electrophoresis and two minor species appeared after longer exposure time. In this study, two isoforms of SecA were separated using SDS-PAGE based on molecular weight. The large isoform of SecA was identified to be the full-length SecA, and the small isoform was identified to be the C-terminus truncation form of SecA lacking CTT.

A number of proteins in *E. coli* are known to have isoforms, such as glyceraldehy-3-phospate dehydrogenase (GAPDH) (Wu et la., 2012), IF2 (Plumbridge et al., 1985) and ribosomal protein S6 (Reeh and Pedersen., 1979), the isoforms normally vary in length. Most of the isoforms perform similar functions or have been involved in a common biological pathway, in some cases the function could be unique or completely opposite. In addition, many bacteria species such as Mycobacteria produce two paralogous SecA protein (SecA1 and SecA2) with non-redundant functions (Braunstein et al., 2001), both of which are highly conserved, SecA2 assists SecA1 in the Sec-dependent protein translocation, but could have specific role for certain proteins. Two isoforms of SecA were also found in the *Staphylococcus aureus* (*S. aureus*) (Mazmanian et al., 2001). SecA1 was found to be approximately 20 kda larger in size compared to SecA2, and SecA2 was proved to have a shorter C-terminus by sequencing alignment, which is consistent with the two SecA isoforms identified in this study (Braunstein et al., 2001). More importantly, cells lacking *secA2* tended to be supersensitive to sodium azide (Braunstein et al., 2001), suggesting that removing C-terminus makes cells less susceptible to azide. There were also studies proposed the difference in CTT between SecA1 proteins and SecA2 proteins, as well as the decreased susceptibility of SecA2 proteins to sodium azide (Chandler et al., 2018; Jamshad et al., 2019; Cranford-Smith et al., 2020). This discovery is complementary to the results in this study to confirm that the azide-resistance was caused by increased expression of the small isoform, which is proposed to be C-terminus truncated SecA.

Protein isoforms normally originated from a single gene by post-transcriptional modifications, ribosomal pausing or encoded by two or more closely related genes (Gunning and Hardeman, 2018). However, from the genome sequence, two SecA isoforms were originated from the same *secA* gene. In addition, the results indicated that two isoforms cannot functionally replaced by each other. Previous studies indicated that only the full-length SecA is essential for cell viability, whereas cell lacking other isoforms only caused a defect in protein export. They shared some of the protein substrates specialized interacting with particular proteins (Braunstein et al., 2001). Whether the presence or absence of SecA isoforms could affect the expression of full-length SecA was yet to be identified. Therefore, the increased production of the small isoform could be a cause or a secondary consequence of azide resistance. It could be possible that the azide-resistant mutants altered the ratio of the functions of two isoforms, alternatively, it may affect the activation of SecA. Since all of the transposon insertions were identified in genes involved in the ribosomal biothsynthesis, which may affect the ribosome interacting with SecA.

The mechanism of how CTT affects the activity of SecA has been studied in previous studies (Jamshad et al., 2019). CTT is responsible for altering the confirmation of SecA and regulating its activity. The FLD exhibit autoinhibition of SecA, which could be moderately relieved by MBD interaction with ribosome (Jamshad et al., 2019). The *secA* mutations conferred azide-resistant destabilized the autoinhibition, but the transposon mutations do it indirectly by producing more stalled ribosomes. Overexpression of SecM also resulted in arrested ribosomes, but did not cause any azide resistance. A possible explanation was that even interacting with ribosome can unlock the autoinhibition conformation, the truncation of CTT caused increased ATP activity compared to the wild type. The azide-resistance in the transposon insertion mutants could be more associated with the production of the small isoform, CTT-truncated SecA.

Two recent studies indicated that disruption of the CTT by transposon insertions conferred fitness advantages in the presence of sodium azide (Chandler et al., 2018; Cranford-Smith et al., 2020). Azide was supposed to disrupt the interaction between CTT with iron and interfere the interaction between SecA and the cytoplasmic membrane (Chandler et al., 2018). The metal binding domain (MBD) of the CTT was proved to be disputed in the second study (Cranford-Smith et al., 2020). In this study, the large isoform was aggregated in the presence of azide and the amount of small isoform was relatively increased, suggesting the expression level of both isoforms could affect the susceptibility to sodium azide. Therefore, the two isoforms may have different functions.

In addition, all of these mutants also conferred resistance to neamine but not the other aminoglycoside, streptomycin, suggesting the mechanism of action of sodium azide and neamine could have something in common. However, the neamine resistance was later proved to be affected by the kanamycin resistance of the mutant strains. Selected *E. coli* strains conferred kanamycin resistance were all resistant to neamine. Since the inhibitory mechanism of neamine is not fully understood, it could share a highly similar mechanism which interferes with protein synthesis leading to incorrect alignment with the mRNA and nonfunctional toxic peptides.

Chapter 6
Identification of the possible role for TF and DsbC in MG1655 using TraDIS

Declaration

The transposon library of MG1655 and the essential gene lists were provided by Dr. Mathew Milner in Lund's lab.

6.1 Introduction

In the previous chapter, the transposon library of MG1655 Δ*tig* Δ*dsbC* mutant was used to identify candidate mutations that have increased the resistance to sodium azide. Genes that were essential for viability in MG1655 Δ*tig*Δ*dsbC* were determined by TraDIS. Since the analysis of the essential genes in *E. coli* MG1655 was described (Milner, 2022), it was decided to compare the difference between these two datasets regarding essential genes and further investigate the possible function of *tig* and *dsbC*. To this end, the transposon mutagenesis libraries were constructed for MG1655 Δ*tig* mutant and MG1655 Δ*dsbC* mutant, and the essential genes for each library was determined using TraDIS.

TF, encoded by *tig*, is known as a molecular chaperone assisting the folding of nascent proteins by interacting with both the nascent polypeptide chains and ribosome during the translation (Kramer et al., 2002). It either facilitates spontaneous folding or cooperates with DnaK or GroEL for protein assembly. TF was shown to functionally overlap with DnaK, and double mutations in *tig* and *dnaK* resulted in lethal phenotype (Deurling et al., 1999), which could be used to confirm the gene essentiality determined by TraDIS. TF has been implicated in the mechanism of Sec-dependent protein translocation, but the exact role is disputed. It was identified based on the ability to assist protein translocation in vitro, but disruption of TF enhanced the translocation and can suppress the translocation defect caused by mutations in SecB. Therefore, analysis of the essential genes of Δ*tig* mutant is needed to identify the complementary genes or mechanism with prevailed roles when loss if TF under normal conditions, and it could be possible to promote the understanding of the interactions between TF and Sec machinery.

DsbC is encoded by the *dsbC* gene, which is a periplasmic disulfide oxireductase and catalyzed the protein disulfide bond isomerization in bacteria (Rietsch et al., 1996). Overexpression of DsbC induced the formation of correctly folded disulfide bonds, this discovery was in generation of the SHuffle strain that was described in Chapter 3 and 4. Deletion of *dsbC* was found to result in the accumulation of misoxidized periplasmic proteins and decreased efficiency of folding of envelope proteins containing disulfide bonds (Missiakas et al., 1994). Even the mechanism of action of DsbC cooperated with DsbA, DsbB and DsbD has been well understood in the previous studies, the experiments were restricted to a narrow range, system of the disulfide bond formation.

Both TF and DsbC have been implicated in cell envelope biogenesis, however, the mechanisms are not clear. TraDIS provides a novel method to investigate genes required for specific biological processes that have become essential for cell growth in the mutant strains compared to the wild type, which elucidates the mechanisms that compensate for the lack of the function of relative genes.

6.2 Results

6.2.1 Construction of the transposon libraries in MG1655 containing Δ*tig* and Δ*dsbC* mutations

The transposon libraries in MG1655 Δ*tig* mutant and MG1655 Δ*dsbC* mutant, and Δ*tig*Δ*dsbC* double mutant were constructed following the protocol described in Section 2.7.1. The libraries were sequenced and the transposon insertions were mapped to the MG1655 genome (NC_000913.3), and the parameters for each library dataset including the library of MG1655 and MG1655 Δ*tig*Δ*dsbC* were listed in Table 6.1. To ensure the compatibility between each dataset, the unique insertions need to be relatively similar. Therefore, the outputs for the library of MG1655 Δ*dsbC* was not used in the comparison to identify the possible roles of DsbC in this section, the relevant outputs and the comparison with other mutant strains and MG1655 was presented in the Appendices.

Table 6.1 Outputs for each library dataset mapped to the MG1655 genome

TraDIS library	Mapped reads	Unique insertions
MG1655*	6011223	574250
MG1655 Δ*tig*Δ*dsbC*	5411223	341181
MG1655 Δ*tig*	6236967	413069
MG1655 Δ*dsbC*	3710689	28916

*The MG1655 library was constructed and analyzed by Dr. Milner in Lund's group

6.2.2 Identification and comparison of the essential genes in each library datasets

Milner previously described a detailed analysis of the essential genome MG1655, four different methods including log likelihood and BioTraDIS were used. Log likelihood was used in his thesis to compare libraries in *E. coli* MG1655 and BW25113 (Milner. 2022), which was used to generate the data he provided in this Section. Therefore, log likelihood (Section 2.7.8) was used to generate the essential genes within libraries of Δ*tig* mutant and Δ*tig*Δ*dsbC* double mutant followed by manual inspection on the Artemis browser. The bimodal distribution of the insertion index scores (transposon insertions normalized to the gene length) was shown in Figure 5.1. Gene with the insertion index scores fitted to the left exponential distribution modal were determined to be essential. 412 genes were determined to be essential in the Δ*tig* mutant library, 866 genes for Δ*dsbC* mutant library and 444 genes for Δ*tig*Δ*dsbC* double mutant library. Statistical analysis alone normally leads to false positive identifications, and overlooked genes affected by domain essentiality and translation orientations. After visual confirmation on Artemis, there were 396 genes determined to be essential in the Δ*tig* mutant library, 452 genes for Δ*tig*Δ*dsbC* double mutant library. Since the Δ*dsbC* library was in a low complexity with inefficient coverage of transposon insertions, there was increased false-positive identification of essential genes. There were increased number of ambiguous genes in the library of double mutant, which could be a result of lower density compared to the Δ*tig* mutant library.

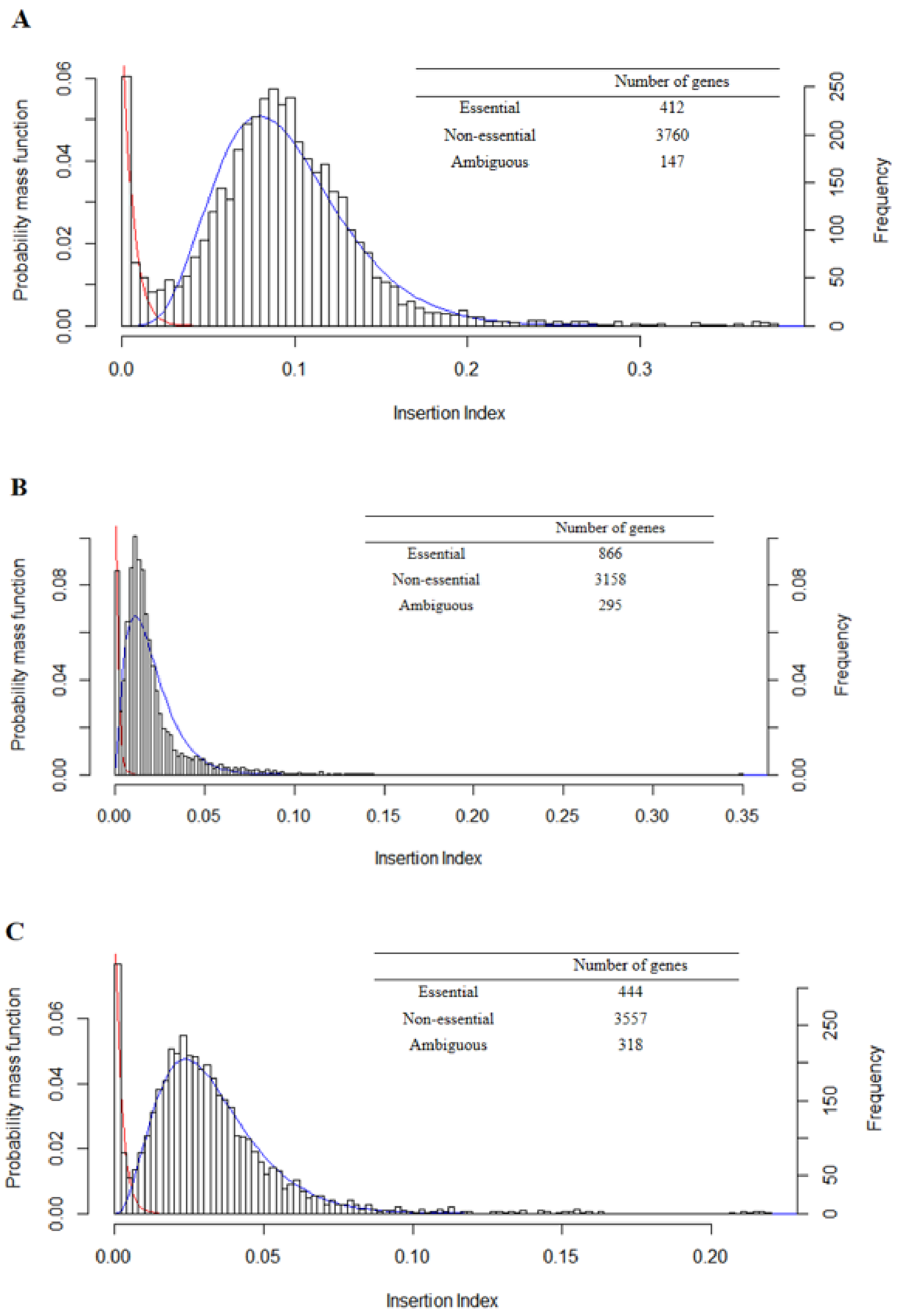


Figure 6.1 Bimodal distribution of insertion index scores to determine the essential gene. Bimodal distribution within the transposon libraries of (A) MG1655 Δ*tig,* (B) MG1655 Δ*dsbC* and (C) MG1655 Δ*tig*Δ*dsbC*. The exponential distribution (red line) fitted to essential mode and the gamma distribution (blue line) fitted to nonessential mode. The number of essential, nonessential and ambiguous genes are shown in the tables on the side. The library of MG655 Δ*dsbC* was in a low complexity with inefficient coverage of transposon insertions, therefore caused significant increased number of essential genes.

The overlapped essential genes in the Δ*tig* mutant library and double mutant library were compared to the library of their wild-type strain, MG1655 by a 3-way comparison using Venn diagram (Figure 6.2). There were 321 essential genes common in three libraries, 76 genes essential in the Δ*tig* mutant library were found included in the essential genes in the library of the double mutant, which kind of proved the data validity. If the functions of TF and DsbC were involved in completely different pathways, the 41 essential genes unique to the double mutant library could be used to identify the possible role of *dsbC*. However, TF and DsbC were both proposed to be involved in the membrane biosynthesis, therefore a high-density transposon library of Δ*dsbC* is useful for further studies. The two genes determined to be essential in MG1655 but not in the libraries of Δ*tig* and Δ*tig*Δ*dsbC* were identified to be *rpoE* and *rdlC*, which encode for a minor sigma factor in response to heat shock and an antisense regulatory RNA component.

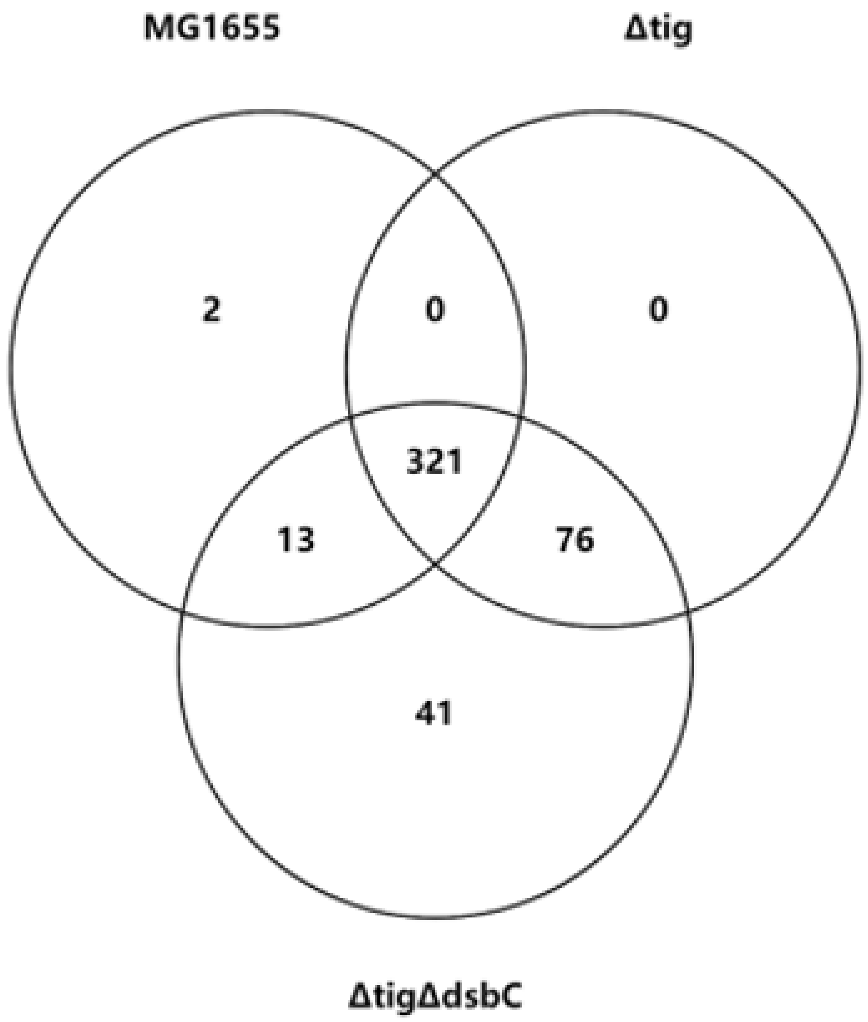


Figure 6.2 Venn diagram showing 3-way comparison of essential genes determined in the transposon libraries of MG1655, Δ*tig* mutant and Δ*tig*Δ*dsbC* mutant. There was an overlap of 321 essetnial genes in three libraries. Essential genes in the Δ*tig* mutant library were all covered by the other two libraries. There were 2 genes essential exclusively for the MG1655 library and 41 unique essential genes in the Δ*tig*Δ*dsbC* mutant library.

6.2.3 Analysis of the essential genes in MG1655 loss of function of TF

Since the 76 genes identified to be essential in the Δ*tig* mutant library but not in the MG1655 library were all covered in the double mutant library, analysis of the function and the biological pathways could provide insight into the function of TF. To analyze the functional enrichment and the interactions between proteins encoded by these genes, the GO enrichment test on EcoCyc was used in this analysis. 66/76 genes were classified according to the biological processes and the predominate GO terms was shown in Figure 6.3. There were overlapped regions when genes were classified by gene ontology, but each classification presented contained at least one unique gene. The overarching categories except the uncharacterized protein were regulation of gene expression, cellular response to temperature stimulus, protein-containing complex assembly, protein localization to membrane and chaperone cofactor-dependent protein refolding, suggesting these processes were functional important when TF is depleted. A large portion of genes were involved in the negative regulation pathways of gene expression or ribosomal protein biosynthesis, indicating that disruption of TF could result in the inhibition of ribosomal biosynthesis. The results were consistent with the predicted function of TF as a ribosome-associated chaperones which facilitating the co-translational folding and aggregation inhibiting.

Genes involved in protein assembly and localization supported the function of TF in protein folding and transportation. Genes encoding chaperones facilitating protein folding, DnaK and DnaJ, were essential in Δ*tig*, indicating the compensatory function of TF as expected. In addition, depleted TF caused enriched genes involved in cell response to cell stimulus. It was noticeable that the *cspI*, *ydfK* and *ynaE* genes were responsive to cold, suggesting depletion of TF conferred a growth defect at low temperatures. Genes *dnaK* and *dnaJ* were also included in this category, but as heat-shock proteins, this could be a reason why *rpoE* was dispensable in Δ*tig* but required for wild-type MG1655. The gene *yajC* in the Sec-dependent protein translocation was identified to be essential in the Δ*tig* mutant. YajC is not essential for protein secretion and cell viability, but SecDF-YajC facilitates the SecA function by stabilizing the SecG. Depletion of TF may facilitate the SecA-dependent protein export by inducing the expression of YajC.

Genes that had been essential for cell envelope biosynthesis, for example, the *rseX* and *micC* involved in the negative regulation of cell motility and *kilR* interferes cell division, which is consistent with that TF leading to defective filamentation and cell division. It was also noticeable that the majority of genes have become essential in Δ*tig* mutant library encoding for uncharacterized proteins, identification of the function of these gene could provide new perspective to analyze TF.

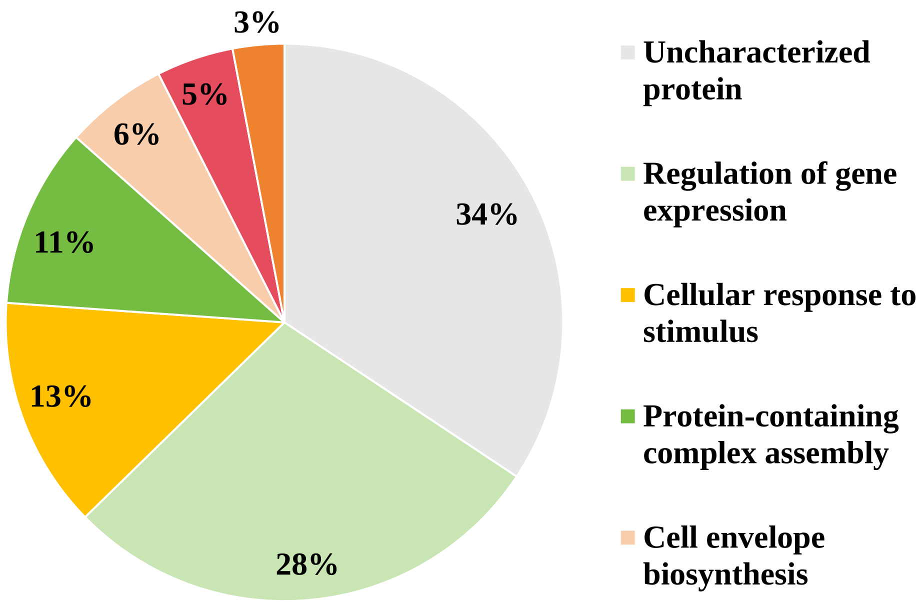


Figure 6.3 Classification of genes that have become essential when MG1655 loss of function of TF. Essential genes identified in the *Δtig* transpson library using TraDIS were compared to the essential gene datasets of MG1655 and Δ*tig*Δ*dsbC*. 76 genes were identified to be essential in libraries of *Δtig* and Δ*tig*Δ*dsbC* and were grouped into overarching categories based on biological process Gene Ontology. Genes encoding unannotated or uncharacterized proteins were grouped into "uncharacterized protein" and shown in gray. The p-value of genes annotated to each category was less than 0.05. There were 13 genes were shown to be essential in both libraries of MG1655 and Δ*tig*Δ*dsbC* nutant but not in the library of *Δtig* mutant (Figure 6.2; Table 6.2). The depletion of TF resulted in these genes to be not essential but turned out to be essential in the double mutant, suggesting it was caused by mutation in *tig* alone in unknown mechanisms.

Table 6.2 List of the common essential genes in the libraries of MG1655 and Δ*tig*Δ*dsbC* but not in the *Δtig* library

Gene	**Annotation**
guaA	GMP synthase
hda	DnaA regulatory inactivator Hda
ldrB	toxin Ldr, type I toxin-antitoxin system family protein
lipA	lipoyl synthase
lysS	lysine--tRNA ligase
mnmA	tRNA 2-thiouridine(34) synthase MnmA
thiL	thiamine monophosphate kinase
ybeD	hypothetical protein
ydiE	hypothetical protein
ygfZ	global regulator
yqfI	Unannotated protein
sdsN	small regulatory RNA SdsN
baxL	putative translational regulatory protein BaxL

6.2.4 Analysis of essential genes in the library of MG1655 Δ*tig*Δ*dsbC*

Since the essential genes in the *Δtig* library were all covered in the library of Δ*tig*Δ*dsbC*, the 41 unique essential genes in the double mutant library in the 3-way could be all related to the depletion of DsbC (Figure 6.2; Table 6.4). To determine whether these genes were involved in particular pathways, GO enrichment test in biology process was applied to identify the associated GO terms. However, 20/41genes were identified encoding uncharacterized proteins and the rest of genes were mostly involved in irrelevant biological processes. There were 2 genes classified into the GO term 'ATP biosynthetic process', 6 genes in 'phosphate-containing compound metabolic process' and 8 genes in 'nucleobase-containing compound metabolic process', respectively. The STRING database (https://string-db.org, version 12.0) was used to futher identify the interactions between these genes. There were 6 genes that have been identified to be physically or functionally interacted based on computational interaction predictions from database of genomic context, co-expression analysis and known pathways (Figure

6.4). Consistent with the result of GO enrichment test, *crr*, *tktA*, *zwf*, *tpiA*, *atpG*, *atpF* were all included in the GO term 'phosphate-containing compound metabolic process'. Moreover, *tktA*, *zwf*, *atpG*, *atpF* were genes overlapped in the GO term 'nucleobase-containing compound metabolic process'. Gene *crr* and *tpiA* encoding for enzymes which were involved in glycolysis, *zwf* encodes glucose-6-phosphate dehydrogenase (G6PDH), which is the first enazyme of the pentose phosphate pathway (PPP). Gene *tktA* encodes transketolase 1 which catalyzes the reversible reaction of producing xylulose-5-phosphate and ribose-5-phosphate, further provides a linkage between glycolysis and the PPP. Gene atpG and atpF were involved in ATP biosynthesis. Given that the glycolysis and PPP were functionally important for growth in *ΔdsbC*.

Table 6.3 Unique essential genes in the library of MG1655 Δ*tig*Δ*dsbC* in the comparison between datasets

Gene	Function*
atpF	ATP synthase Fo complex subunit b
atpG	ATP synthase F1 complex subunit gamma
crr	Enzyme IIAGlc
cydX	Cytochrome bd-I accessory subunit CydX
dapF	Diaminopimelate epimerase
dmsD	Redox enzyme maturation protein DmsD
flgO	Small RNA FlgO
lapA	Lipopolysaccharide assembly protein A
ldrC	Small toxic polypeptide LdrC
micA	Small regulatory RNA MicA
pssL	Protein PssL
racC	Protein RacC
rlmE	23S rRNA 2'-O-ribose U2552 methyltransferase
rnc	RNase III
rsgA	Ferritin iron-storage complex
seqA	Negative modulator of initiation of replication
sroE	Small RNA SroE
tktA	Transketolase 1
tpiA	Triose-phosphate isomerase
tusE	Sulfur transfer protein TusE
yadW	Protein YadW
ybfQ	Inactive transposase YbfQ

yccE	Uncharacterized protein YccE
yceQ	DUF2655 domain-containing protein YceQ
yciG	Stress-induced bacterial acidophilic repeat motifs-containing protein YciG
ydeO	DNA-binding transcriptional dual regulator YdeO
yejL	DUF1414 domain-containing protein YejL
ygeN	Putative type III secretion system protein YgeN
yjbL	Uncharacterized protein YjbL
yjbS	Uncharacterized protein YjbS
ykgP	Putative oxidoreductase
ymfD	Putative SAM-dependent methyltransferase YmfD
ymgI	Uncharacterized protein YmgI
ymjC	Uncharacterized protein YmjC
ynaM	Protein YnaM
yncL	Uncharacterized protein YncL
yncO	Uncharacterized protein YncO
ynfN	Protein YnfN
ynfQ	Protein YnfQ
ypdK	Uncharacterized membrane protein YpdK
zwf	NADP+-dependent glucose-6-phosphate dehydrogenase

*Functions taken from Ecocyc website: https://ecocyc.org/

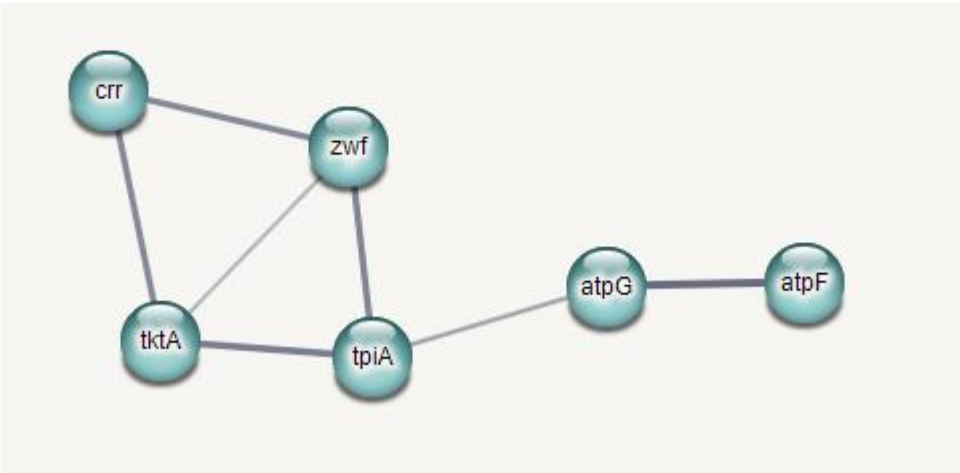


Figure 6.4 The interaction networks of essential genes unique to the library of MG1655 Δ*tig*Δ*dsbC*. The association of 41 genes were determined by STRING database, only genes that were grouped into a network were displayed. The edges represent associations and the thickness indicates the confidence of the associations.

6.3 Discussion

The results presented in this chapter indicated the possible function of TF and DsbC by comparing the essential genes in the transposon mutagenesis libraries of MG1655 *Δtig* mutant and Δ*tig*Δ*dsbC* double mutant to the MG1655 library. Inspection of genes required for growth in the mutant libraries compared to the wild type provided insight into the defective biological processes in the mutant strains and genes with overlapping functions. The essential genes identified in the *Δtig* mutant were all covered in the libraries of MG1655 and Δ*tig*Δ*dsbC* double mutant (Figure 6.2), which confirmed the validation of the results. Since the function of TF and DsbC are distinct from the current studies, genes unique to the Δ*tig*Δ*dsbC* double mutant library were used to identify the possible function of DsbC.

Previous studies indicated that disruption or overexpression of TF can interfere with the normal cell division in abnormal cell sizes and filamentation (Gutherie and Wickner, 1990). This was proved by genes, *rseX*, *micC*, *rdlB* and *kilR* that have become essential in the *Δtig* mutant library. Overexpression of RseX, MicC or RdlB decreases the flagella-based cell motility, and the single gene knock-out in *micC* and *rseX* caused reduction in biofilm formation (Bak et al., 2015). In addition, the inhibition in cell division by KilR was shown to be suppressed by overexpression of the essential cell division gene *ftsZ* (Conter and Dassain, 1996), the filamentation induced by TF, especially overexpression of TF was shown to be suppressed by FtsZ as well (Gutherie and Wickner, 1990), suggesting that the TF and KilR could be functional overlapping and complementary with FtsZ. Cell division needs a precise level of these genes, FtsZ can also cause filamentation in the same manner when being disrupted or overexpressed (Ward and Lutkenhaus, 1985). Cells lacking other chaperones, such as DnaK, GroEL also inhibit the cell division (Georgoupoulos and Eisen, 1972; Paek and Walker., 1987), which could be suppressed by overexpression of FtsZ, suggesting that the cell division requires the participation of molecular chaperones. Indeed, *secA* knockout can also led to filamentation which could be caused by inhibition of the translocation of desired proteins to the division site (Oliver and Beckwith, 1981). The gene bamD was also essential in the *Δtig* mutant, which is responsible for outer membrane lipoprotein, suggesting TF could be involved in the membrane assembly or insertion of the outer membrane proteins. By contrast, DsbC was shown to be involved in protecting the cell envelope from oxidative stress by reducing the AraF homodimer but not involved in the assemble of cell envelope.

Genes that compensated for the loss of function of TF could be predicted by comparing the essential genes for viability between Δ*tig* library and MG1655 library. Gene *rpoE* and *rdlC* were determined to be essential in the MG1655 but no longer required for a tig mutant. Gene *rpoE* was previously determined to be essential for viability in *E. coli* K-12 strain BW25113 (Baba et al., 2006; Goodall et al., 2018). It encodes for sigma E, which is a minor sigma factor in response to cellular stress such as heat shock and membrane stress. Sigma E is also induced in response to misfolded protein, which is functional overlapped with TF. However, whether the deletion of *tig* could suppress the Δ*rpoE* lethality needs to be further proved. In addition, Sigma E was found activated an unknown cell death pathway (Button et al., 2007), and TF was suggested facilitating the ClpXP-mediated degradation (Rizzolo et al., 2021), there could be a linkage between them. Expression of sigma E and RdlC was identified to decrease the cell motility (Smith et al., 2017; Bak 15), which further proved the role of TF in cell envelope biosynthesis. There were also 13 essential genes in the MG1655 library and the double mutant library but not in the *tig* single mutant library, however, there was no systematic biological pathways they were involved in.

The predicted functions of TF demonstrated by TraDIS could be involved in 4 biological pathways, cell division, ribosome-associated molecular chaperone, protein translocation and cellular response to cold. A large portion of genes that have become essential in the *Δtig* mutant library were involved in negative regulation of gene expression, especially in translation and protein folding, which could be caused by the conformational changes in the ribosome-nascent chain complex in absence of TF. This result proved the function of TF interacting with ribosome processes protein folding and targeting the subsequent translocation. Protein folding normally requires different chaperones, however, TF is the only chaperone directly interacting with ribosome (Deeng et al., 2016). Despite GroEL encoded by *groL,* which is responsible for proper protein folding, it is required for cell viability in the bacteria strains. DnaK, encoded by *dnaK*, and its cofactor DnaJ were identified to be essential when loss of TF, suggesting DnaK as a chaperone compensated the defect in protein folding process caused by deletion of TF. This result is consistent with their redundant function in protein folding and preventing the misfolding and aggregation (Deuerling et al., 1999). Lacking both DnaK and TF generally results in lethal phenotype except a specific range of temperature due to protein aggregation, inhibition of cell division and decreased the tolerance to heat. DnaK in *E. coli* also controls the heat-shock response by regulating sigma 32 (Chattopadhyay and Roy, 2002), which could explain why sigma E was not required in the *Δtig* mutant.

TF was previously deemed to mediate the SecA-dependent protein translocation with the help of the secretory chaperone SecB (Geyter et al., 2020). However, deletion of TF accelerated the protein translocation and resulted in more co-translational translocation (Oh et al., 2011) but the mechanisms remain unclear. Gene *yajC* involved in the Sec machinery was identified to be essential in the *Δtig* mutant library but not in the wild type, suggesting that YajC could be involved in maintaining the translocation competence of the nascent polypeptides since YajC also contains a conserved SecA-like MBD. Indeed, the SecDF-YajC complex promotes the SecA function by facilitating the ATP turnover and membrane insertion and deinsertion of Sec, which could be a reason caused the enhanced translocation efficiency.

Despite the function in proteostasis, TF is involved in the cellular response to a series of environmental stress, such as abnormal temperatures or antibiotics. Except for *dnaK* and *dnaJ*, other essential genes response to the temperature stimulus were all encoding for cold-shock proteins, which are compatible with TF as a cold-shock protein to enhance the tolerance of *E. coli* to low temperature (Kandror and Goldberg, 1997). Disruption of TF could probably cause heat stress in the cell. There were also other genes responses to other stimulus, such as magnesium ions or radiation, TF may have regulatory functions in a number of stress response genes, but no evidence support this yet. In addition, there was a large portion of genes that have become essential in the absence of TF that still need to be further investigated. Some of them, the function could be predicted by interaction with proteins with known functions. Investigating the function of other genes is possible to provide new insight into the function of TF.

Compared to extensive studies on the activities of TF, the role of DsbC was limited in periplasmic disulfide oxireductase as protein disulfide isomerase. Since the density of the transposon insertions in the Δ*dsbC* library was too low, and the function of TF and DsbC are involved in distinct biological pathways, the function of DsbC was investigated by analysis of genes have become essential in the Δ*tig*Δ*dsbC* double mutant library but not included in the Δ*tig* single mutant library. There were 6 genes in the comparison involved in glycolysis and PPP, suggesting the role of DsbC in the metabolic network. Glycolysis is responsible for regulating various metabolic functions and PP is parallel to glycolysis and responsible for generating NADPH and 5-carbon sugars. DsbC as the disulfide bond isomerase, absence of DsbC would lead to decreased reductive activity, NADPH could provide electrons for other reductive pathways to against oxidative stress. There were also three genes, *cydX*, *micA* and *rsgA* that involved in the cellular response to oxidative damage had become essential, suggesting the impact of DsbC in protecting cells against damage from oxidative stress. This result is consistent with that the DsbC associated reducing pathways could protect the cell envelope form the reactive oxygen species (Denoncin et al., 2014). The DSB proteins involved in the protein oxidative pathways were not identified to become essential in the absence of DsbC, even another isomerase DsbG, indicating that DsbC and DsbG catalyzes the rearrangement of misfolded disulfide bond in different ways or they may recognize different substrates.

Similar to the essential genes in the Δ*tig* library, there were about half of the genes encoding uncharacterized proteins. The function of TF and DsbC was well studied since all of the predicted functions identified by TraDIS analysis were proposed by previous research though some of the mechanisms are still unknown. Investigating the function of these less annotated functions of proteins could provide new insight into novel pathways or clues to mechanism of the existing pathways.

Chapter 7
Discussion

7.1 Summary of the project

In this thesis, the physiological differences between SHuffle strains and their isogenic parental strains were analyzed by comparison of the essential genes in the transposon library of each strain generated by TraDIS. The essential genes were predicted by two statistical methods, log likelihood and Biotradis pipeline combined with manual inspection on Artemis browser. The result revealed genes that had become essential in the SHuffle strains were mostly involved in the cellular response to oxidative stress and abnormal temperatures, which was consistent with the consequence of the SHuffle modifications and the temperature sensitivity conferred in SHuffle strains. UQ biosynthesis was assumed to be associated with both oxidative stress and temperature sensitivity, but the mechanism is unclear.

For further understanding the temperature sensitivity in SHuffle B, a transposon library of SHuffle B was grown at 30 ℃, 37 ℃ and 42 ℃ to construct new libraries. The essential genes for each library were identified by TraDIS, however, the severe oxidative stress diminished the cell viability at 42 ℃, therefore 3-fold more essential genes were identified in the library at 42 ℃ compared to other two libraries. Therefore, it was decided to analyze genes with enriched transposon insertions at 42 ℃. Upon inspection of these genes, insertions in a number of genes involved in aerobic respiration chain increased the resistance of SHuffle B to 42 ℃. Whether higher temperatures disrupting aerobic respiration and making cells survive in an anerobic condition could be confirmed by growing SHuffle B under anerobic condition at 42 ℃.

Genetic screens based on transposon mutagenesis library were used to identify the mechanism of sodium azide, which is an inhibitor of SecA. The results indicated that two isoforms of SecA were separated by SDS-PAGE: the large isoform was identified to be full-length SecA, and the small isoform was assumed to be C-terminus tail truncated SecA. Growth of *E. coli* in the presence of 0.5 mM sodium azide caused accumulation of the large isoform of SecA and relatively increased the small isoform.

Different from previous studies the azide-resistant mutants were all mapped in the *secA* gene (Figure 1.7), 6 isolated transposon insertion mutations that conferred increased resistance to azide in this study were involved in the ribosomal biosynthesis. However, the azide-resistant mutations in *secA* also contained an increased expression of the small isoform of SecA, suggesting that the expression of the small isoform could affect the susceptibility of *E. coli* to azide, and it could have a special physiological role distinct from that of the full length SecA.

Lastly, the function of TF and DsbC was investigated by analyzing the essential genes identified in the transposon library of a Δ*tig* mutant, Δ*dsbC* mutant, Δ*tig*Δ*dsbC* double mutant then compared to the wild-type MG1655 library. TF was found to be involved in cell division, ribosomal-associated protein folding, protein translocation and cellular response to cold. DsbC was found to be involved in cell redox and metabolic network. There were no substitution gene functions in disulfide bond isomerase, such as *dsbG*, turned to be essential to compensate the loss of DsbC, suggesting there could be more alternative pathways in *E. coli* to deal with the misfolded proteins.

7.2 Application and optimization of the SHuffle strains

SHuffle strains have been made to produce correctly folding disulfide bonded proteins in the cytoplasm, compared to other protein expression strains such as *BL21 (DE3)*, leading to greater yields in protein production irrespective of whether the protein contains disulfide bonds or not. For example, SHuffle strains can be used to produce the mammalian antibody IgG which contains multiple disulfide bonds (Eaglesham et al., 2021) and also showed superior expression of recombinant human fibroblast growth factor-1 with free cysteines but no disulfide bond (Nasiri et al., 2017). In addition, SHuffle strains as a platform were used to analyze the genetic modification in amino acids in the recombinant proteins (Gerding et al., 2019) and for cell free protein synthesis-based prototyping (Dopp and Reuel, 2020).

To optimize the production of proteins, multiple studies have tried to optimize the expression conditions such as temperature and cultivation media, or assistance of the co-expression helpers such as KatG (Lobstein et al., 2012, Ren et al., 2016; Khalilvand et al., 2022). However, the effects of these factors are all protein specific. In this study, the physiology of SHuffle strains and their wild type strains was investigate, which provide candidates for optimizing the strains to enhance the capability of producing recombinant proteins in higher yield. SHuffle B was previously proved to result in higher yield when producing certain proteins (Lobstein et al., 2012), this could be caused by higher enzyme activities in SHuffle B. Higher enzyme and metabolic activities are probably the reason aggravating the oxidative stress under higher temperatures, which caused increased temperature sensitivity at 42 °C compared to SHuffle K-12 and the wild-type strains (Figure 4.1). The TraDIS results of the SHuffle B transposon library grown at different temperatures indicated that a number of genes that involved in oxidative-reduction reactions had become essential at higher temperatures. In addition, enriched transposon insertion in certain genes could restore the growth of SHuffle B at 42 °C, which provide new directions to engineer a strain which could against the oxidative stress but produce more proteins, especially proteins expressed at high temperatures. However, in the tested mutations, only Δ*rbsR* and Δ*fre* showed slightly increased formation of disulfide bond in the SHuffle strain. There were still several "knock on" effects caused by SHuffle mutations could not be explained, for example, SHuffle strains cannot use glycerol as the sole carbon source. Further investigation of the physiology of SHuffle strains is useful to optimize the production of protein of interest.

The genome features of DHB4 were transduced into SHuffle B to screen for genomic differences that could recover temperature sensitivity. Due to time constraints, the plasmid cloning for the *rbsR* region has not been completed, there could be more observations in the continuous experiment. In addition, disruption of aerobic respiration was shown to increase the temperature resistance of SHuffle B, it is worth to identify if the anerobic condition could suppress the lethal phenotype at 42 °C and how it will affect the production of disulfide bonded proteins. TraDIS is limited in characterizing the essential genes that were mutated in SHuffle strains under different temperatures, therefore TraDIS-Xpress which combines inducible promoter into the transposon could provide insight into the expression and repression of all genes under different experimental conditions.

7.3 SecA isoforms and azide resistance

SecA isoforms were previously discovered in the Mycobacteria and *S. aureus* (Braunstein et al., 2001; Mazmanian et al., 2001), it proved the existence of two SecA isoforms in *E. coli* in this study. In this study, the large isoform was proposed to be the full-length SecA and the small isoform was a C-terminus truncation form of SecA, which was consistent with the previous studies. Since the genome sequence indicated the only one secA gene is present in *E. coli*, the two SecA isoforms are potentially a consequence of the alternative splicing or post-transcriptional modifications. The production of SecA is regulated by the translational stall of the SecM. In addition the azide resistance in *E. coli* was shown to be affected by increased expression of the small isoform in both the *secA* mutant conferred azide-resistance and the azide-resistance mutants (AR1-6) isolated in this study. However, the overexpression of SecM did not affect the azide resistance. Therefore, the isoforms could be a result of post-transcriptional modifications. Further sequence and structural analysis could provide more information about the formation and function of the small isoform.

There were two publications that indicated the effect of azide on the CTT of SecA (Chandler et al., 2018; Cranford-smith et al., 2020). In this study, the production of the CTT truncated SecA in *E. coli* was increased in the presence of azide. Since *E. coli* produces two isoforms of SecA under physiological growth conditions, CTT truncated SecA could have distinct functions. Although azide is known as An inhibitor of SecA, and previously identified azide-resistant mutations were all located in the *secA* gene, transposon insertions in a range of genes involved in different biological processes were found to affect the susceptibility to azide (Cranford-smith et al., 2020). Therefore the molecular mechanism of azide and the role of two isoforms of SecA need to be further investigated. As *secA* gene is required for cell viability, TraDIS-Xpress can be used to identify the expression of SecA as well as other essential genes in the presence of azide.

7.4 Applications and imitation of TraDIS

TraDIS is a useful tool in genetics to identify essential genes in the strain of interest by analyzing the location and frequency of the transposon insertions within a transposon mutagenesis library. Combined with the genome browser, it provides visualization of the insertion profile within the genome, even in the intragenic regions, which is useful for discovery of a putative but unannotated gene. TraDIS is also used to identify the conditional essential genes in a mutant strain to predict the function of a particular gene or to understand the mechanisms in the given condition, such as presence of antibiotics or grown at extreme temperature.

However, the main limitation of TraDIS in this study was owing to the characterization of the essential genes. The essential genes in this analysis were determined by absence of insertion sites in the population, it cannot be tracked how it was involved in the associated with a certain phenotype. TraDIS-Xpress could potential overcome this and affect the expression of essential genes by inducible promoter. Moreover, it might not be a limitation of TraDIS but gene annotation, but when compared the essential genes in libraries generated in different strain backgrounds, which makes it difficult to match genes in different genomes. In this study, GFF3 parse was used to convert the locus ID into the genome sequence obtained from NCBI database for each strain to the corresponding gene names in MG1655. However, a large number of genes could not be identified, even in the well-annotated DHB4 strain. In addition, TraDIS could be limited by construction of libraries in certain strains, for unknown reason the library construction in the MG1655 Δ*dsbC* mutant was difficult and the sequencing results was unfavorable. There are alternative methods which can also be used to assay the essential genes, such as CRISPRi and TraDIS-Xpress, which could provide a more comprehensive analysis of the gene essentiality.

References

Allen, W.J., Corey, R.A., Oatley, P., Sessions, R.B., Baldwin, S.A., Radford, S.E., Tuma, R. and Collinson, I. (2016). Two-way communication between SecY and SecA suggests a Brownian ratchet mechanism for protein translocation. *eLife*, 5. doi:https://doi.org/10.7554/elife.15598.

Anfinsen, C.B. and Haber, E. (1961). Studies on the reduction and re-formation of protein disulfide bonds. *The Journal of Biological Chemistry*, [online] 236, pp.1361–1363. Available at: https://pubmed.ncbi.nlm.nih.gov/13683523/ [Accessed 12 Jun. 2024].

Angelini, S., Deitermann, S. and Koch, H. (2005). FtsY, the bacterial signal-recognition particle receptor, interacts functionally and physically with the SecYEG translocon. *EMBO Reports*, 6(5), pp.476–481. doi:https://doi.org/10.1038/sj.embor.7400385.

Anton, B.P., Fomenkov, A., Raleigh, E.A. and Berkmen, M. (2016). Complete Genome Sequence of the Engineered Escherichia coli SHuffle Strains and Their Wild-Type Parents. *Genome Announcements*, 4(2). doi:https://doi.org/10.1128/genomea.00230-16.

Arne Rietsch, Bessette, P.H., Georgiou, G. and Beckwith, J. (1997). Reduction of the periplasmic disulfide bond isomerase, DsbC, occurs by passage of electrons from cytoplasmic thioredoxin. *Journal of Bacteriology*, 179(21), pp.6602–6608. doi:https://doi.org/10.1128/jb.179.21.6602-6608.1997.

Aslund, F., Ehn, B., Miranda-Vizuete, A., Pueyo, C. and Holmgren, A. (1994). Two additional glutaredoxins exist in Escherichia coli: glutaredoxin 3 is a hydrogen donor for ribonucleotide reductase in a thioredoxin/glutaredoxin 1 double mutant. *Proceedings of the National Academy of Sciences*, 91(21), pp.9813–9817. doi:https://doi.org/10.1073/pnas.91.21.9813.

Baars, L., Ytterberg, A.J., Drew, D., Wagner, S., Thilo, C., van Wijk, K.J. and de Gier, J.-W. (2005). Defining the Role of theEscherichia coli Chaperone SecB Using Comparative Proteomics. *Journal of Biological Chemistry*, 281(15), pp.10024–10034. doi:https://doi.org/10.1074/jbc.m509929200.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L. and Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, 2(1). doi:https://doi.org/10.1038/msb4100050.

Bader, M.A., Muse, W.B., Ballou, D.P., Gassner, C. and James C.A. Bardwell (1999). Oxidative Protein Folding Is Driven by the Electron Transport System. *Cell*, 98(2), pp.217–227. doi:https://doi.org/10.1016/s0092-8674(00)81016-8.

Bader, M.A., Muse, W.B., Zander, T. and James C.A. Bardwell (1998). Reconstitution of a Protein Disulfide Catalytic System. *Journal of Biological Chemistry*, 273(17), pp.10302–

10307. doi:https://doi.org/10.1074/jbc.273.17.10302.

Bader, M.W., Xie, T., Yu, C. and James C.A. Bardwell (2000). Disulfide Bonds Are Generated by Quinone Reduction. *Journal of Biological Chemistry*, 275(34), pp.26082–26088. doi:https://doi.org/10.1074/jbc.m003850200.

Bak, G., Lee, J., Suk, S., Kim, D., Young Lee, J., Kim, K., Choi, B.-S. and Lee, Y. (2015). Identification of novel sRNAs involved in biofilm formation, motility and fimbriae formation in Escherichia coli. *Scientific Reports*, 5(1). doi:https://doi.org/10.1038/srep15287.

Bardwell, J.C., Lee, J.O., Jander, G., Martin, N., Belin, D. and Beckwith, J. (1993). A pathway for disulfide bond formation in vivo. *Proceedings of the National Academy of Sciences*, [online] 90(3), pp.1038–1042. doi:https://doi.org/10.1073/pnas.90.3.1038.

Bardwell, J.C.A., McGovern, K. and Beckwith, J. (1991). Identification of a protein required for disulfide bond formation in vivo. *Cell*, 67(3), pp.581–589. doi:https://doi.org/10.1016/0092-8674(91)90532-4.

Batey, R.T., Rambo, R.P., Lucast, L., Rha, B. and Doudna, J.A. (2000). Crystal Structure of the Ribonucleoprotein Core of the Signal Recognition Particle. *Science*, 287(5456), pp.1232–1239. doi:https://doi.org/10.1126/science.287.5456.1232.

Beaudry, A. and Joyce, G. (1992). Directed evolution of an RNA enzyme. *Science*, 257(5070), pp.635–641. doi:https://doi.org/10.1126/science.1496376.

Bechtluft, P., Nouwen, N., Tans, S.J. and Driessen, A.J.M. (2009). SecB--a chaperone dedicated to protein translocation. *Molecular bioSystems*, [online] 6(4), pp.620–627. doi:https://doi.org/10.1039/b915435c.

Berg, D.E., Davies, J., Allet, B. and Jean-David Rochaix (1975). Transposition of R factor genes to bacteriophage lambda. *Proceedings of the National Academy of Sciences of the United States of America*, 72(9), pp.3628–3632. doi:https://doi.org/10.1073/pnas.72.9.3628.

Berkmen, M., Boyd, D. and Beckwith, J. (2005). The Nonconsecutive Disulfide Bond of Escherichia coli Phytase (AppA) Renders It Dependent on the Protein-disulfide Isomerase, DsbC. *Journal of Biological Chemistry*, 280(12), pp.11387–11394. doi:https://doi.org/10.1074/jbc.m411774200.

Berks, B.C. (1996). A common export pathway for proteins binding complex redox cofactors? *Molecular Microbiology*, 22(3), pp.393–404. doi:https://doi.org/10.1046/j.1365-2958.1996.00114.x.

Bessette, P.H., Aslund, F., Beckwith, J. and Georgiou, G. (1999). Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm. *Proceedings of the National Academy of Sciences*, 96(24), pp.13703–13708. doi:https://doi.org/10.1073/pnas.96.24.13703.

Borisov, V.B., Murali, R., Verkhovskaya, M.L., Bloch, D.A., Han, H., Gennis, R.B. and Verkhovsky, M.I. (2011). Aerobic respiratory chain of Escherichia coli is not allowed to work in fully uncoupled mode. *Proceedings of the National Academy of Sciences*, 108(42), pp.17320–17324. doi:https://doi.org/10.1073/pnas.1108217108.

Bowler, M.W., Montgomery, M.G., Leslie, A.G.W. and Walker, J.E. (2006). How azide inhibits ATP hydrolysis by the F-ATPases. *Proceedings of the National Academy of Sciences*, [online] 103(23), pp.8646–8649. doi:https://doi.org/10.1073/pnas.0602915103.

Brandi, A., Giangrossi, M., Paoloni, S., Spurio, R., Giuliodori, A.M., Pon, C.L. and Gualerzi, C.O. (2019). Transcriptional and post-transcriptional events trigger de novo infB expression in cold stressed Escherichia coli. *Nucleic Acids Research*, 47(9), pp.4638–4651. doi:https://doi.org/10.1093/nar/gkz187.

Braunstein, M., Brown, A.M., Kurtz, S. and Jacobs, W.R. (2001). Two Nonredundant SecA Homologues Function in Mycobacteria. *Journal of Bacteriology*, 183(24), pp.6979–6990. doi:https://doi.org/10.1128/jb.183.24.6979-6990.2001.

Breukink, E., Nouwen, N., van Raalte, A., Mizushima, S., Tommassen, J. and de Kruijff, B. (1995). The C Terminus of SecA Is Involved in Both Lipid Binding and SecB Binding. *Journal of Biological Chemistry*, 270(14), pp.7902–7907. doi:https://doi.org/10.1074/jbc.270.14.7902.

Brown, A.N., Anderson, M.T., Bachman, M.A. and Mobley, H.L.T. (2022). The ArcAB Two-Component System: Function in Metabolism, Redox Control, and Infection. *Microbiology and Molecular Biology Reviews*, v.86(2). doi:https://doi.org/10.1128/mmbr.00110-21.

Cabelli, R.J., Dolan, K.M., Qian, L.P. and Oliver, D.B. (1991). Characterization of membrane-associated and soluble states of SecA protein from wild-type and SecA51(TS) mutant strains of Escherichia coli. *Journal of biological chemistry/˜ The œ Journal of biological chemistry*, 266(36), pp.24420–24427. doi:https://doi.org/10.1016/s0021-9258(18)54245-9.

Calvori, C., Frontali, L., Leoni, L. and Tecce, G. (1965). Effect of Rifamycin on Protein Synthesis. *Nature*, 207(4995), pp.417–418. doi:https://doi.org/10.1038/207417a0.

Canard, B. and Sarfati, R.S. (1994). DNA polymerase fluorescent substrates with reversible 3′-tags. *Gene*, 148(1), pp.1–6. doi:https://doi.org/10.1016/0378-1119(94)90226-7.

Chandler, R., Jamshad, M., Yule, J., Robinson, A., Alam, F., Dunne, K.A., Nabi, N., Henderson, I. and Huber, D. (2017). Genetic screen suggests an alternative mechanism for azide-mediated inhibition of SecA. doi:https://doi.org/10.1101/173039.

Chattopadhyay, R. and Roy, S. (2002). DnaK-Sigma 32 Interaction Is Temperature-dependent. *Journal of Biological Chemistry*, 277(37), pp.33641–33647. doi:https://doi.org/10.1074/jbc.m203197200.

Chaudhuri, R.R. and Henderson, I.R. (2012). The evolution of the Escherichia coli phylogeny. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, [online] 12(2), pp.214–226. doi:https://doi.org/10.1016/j.meegid.2012.01.005.

Chen, J., Song, J., Zhang, S., Wang, Y., Cui, D. and Wang, C. (1999). Chaperone Activity of DsbC. *Journal of Biological Chemistry*, 274(28), pp.19601–19605. doi:https://doi.org/10.1074/jbc.274.28.19601.

Chen, M., Xie, K., Jiang, F., Yi, L. and Dalbey, R.E. (2002). YidC, a Newly Defined Evolutionarily Conserved Protein, Mediates Membrane Protein Assembly in Bacteria. *Biological Chemistry*, 383(10). doi:https://doi.org/10.1515/bc.2002.176.

Cho, S.-H., Porat, A., Ye, J. and Beckwith, J. (2007). Redox-active cysteines of a membrane electron transporter DsbD show dual compartment accessibility. *The EMBO Journal*, 26(15), pp.3509–3520. doi:https://doi.org/10.1038/sj.emboj.7601799.

Chung, C.T. and Miller, R.H. (1993). *[43] Preparation and storage of competent Escherichia coli cells*. [online] ScienceDirect. Available at: https://www.sciencedirect.com/science/article/pii/007668799318045E.

Chung, J., Chen, T. and Missiakas, D. (2000). Transfer of electrons across the cytoplasmic membrane by DsbD, a membrane protein involved in thiol-disulphide exchange and protein folding in the bacterial periplasm. *Molecular Microbiology*, 35(5), pp.1099–1109. doi:https://doi.org/10.1046/j.1365-2958.2000.01778.x.

Collis, C.M. and Grigg, G.W. (1989). An Escherichia coli mutant resistant to phleomycin, bleomycin, and heat inactivation is defective in ubiquinone synthesis. *Journal of Bacteriology*, [online] 171(9), pp.4792–4798. doi:https://doi.org/10.1128/jb.171.9.4792-4798.1989.

Conter, A., Bouché, J.P. and Dassain, M. (1996). Identification of a new inhibitor of essential division gene ftsZ as the kil gene of defective prophage Rac. *Journal of bacteriology*, 178(17), pp.5100–5104. doi:https://doi.org/10.1128/jb.178.17.5100-5104.1996.

Cranford-Smith, T., Jamshad, M., Jeeves, M., Chandler, R.A., Yule, J.V., Robinson, A., Alam, F., Dunne, K.A., Aponte, E.H., Mashael Alanazi, Carter, C., Henderson, I.R., Lovett, J.E., Winn, P.J., Knowles, T.J. and Huber, D. (2020). Iron is a ligand of SecA-like metal-binding domains in vivo. *Journal of Biological Chemistry*, 295(21), pp.7516–7528. doi:https://doi.org/10.1074/jbc.ra120.012611.

Crooke, E., Guthrie, B., Lecker, S., Lill, R. and Wickner, W. (1988). ProOmpA is stabilized for membrane translocation by either purified *E. coli* trigger factor or canine signal recognition particle. *Cell*, 54(7), pp.1003–1011. doi:https://doi.org/10.1016/0092-8674(88)90115-8.

Crooke, E. and Wickner, W. (1987). Trigger factor: a soluble protein that folds pro-OmpA into a membrane-assembly-competent form. *Proceedings of the National Academy of Sciences*, 84(15), pp.5216–5220. doi:https://doi.org/10.1073/pnas.84.15.5216.

Daegelen, P., Studier, F.W., Lenski, R.E., Cure, S. and Kim, J.F. (2009). Tracing Ancestors and Relatives of Escherichia coli B, and the Derivation of B Strains REL606 and BL21(DE3). *Journal of Molecular Biology*, 394(4), pp.634–643. doi:https://doi.org/10.1016/j.jmb.2009.09.022.

Darby, N.J. and Creighton, T.E. (1995). Disulfide Bonds in Protein Folding and Stability. *Methods Mol Biol*, pp.219–252. doi:https://doi.org/10.1385/0-89603-301-5:219.

Darby, N.J., Raina, S. and Creighton, T.E. (1998). Contributions of Substrate Binding to the Catalytic Activity of DsbC. *Biochemistry*, 37(3), pp.783–791. doi:https://doi.org/10.1021/bi971888f.

Dekker, C., Kruijff, B. de and Gros, P. (2003). Crystal structure of SecB from Escherichia coli. *Journal of Structural Biology*, 144(3), pp.313–319. doi:https://doi.org/10.1016/j.jsb.2003.09.012.

den Blaauwen, T. and Driessen, A.J. (1996). Sec-dependent preprotein translocation in bacteria. *Archives of Microbiology*, [online] 165(1), pp.1–8. doi:https://doi.org/10.1007/s002030050289.

Denks, K., Vogt, A., Sachelaru, I., Petriman, N.-A., Kudva, R. and Koch, H.-G. (2014). The Sec translocon mediated protein transport in prokaryotes and eukaryotes. *Molecular Membrane Biology*, 31(2-3), pp.58–84. doi:https://doi.org/10.3109/09687688.2014.907455.

Denoncin, K., Vertommen, D., Arts, I.S., Goemans, C., Rahuel-Clermont, S., Messens, J. and Collet, J.-F. (2014). A New Role for Escherichia coli DsbC Protein in Protection against Oxidative Stress. *Journal of Biological Chemistry*, 289(18), pp.12356–12364. doi:https://doi.org/10.1074/jbc.m114.554055.

Derman, A., Prinz, W., Belin, D. and Beckwith, J. (1993). Mutations that allow disulfide bond formation in the cytoplasm of Escherichia coli. *Science*, 262(5140), pp.1744–1747. doi:https://doi.org/10.1126/science.8259521.

Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A. and Bukau, B. (1999). Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature*, 400(6745), pp.693–696. doi:https://doi.org/10.1038/23301.

Dopp, J.L. and Reuel, N.F. (2020). Simple, functional, inexpensive cell extract for in vitro prototyping of proteins with disulfide bonds. *Biochemical Engineering Journal*, 164, p.107790. doi:https://doi.org/10.1016/j.bej.2020.107790.

Dutton, R.J., Boyd, D., Berkmen, M. and Beckwith, J. (2008). Bacterial species exhibit diversity in their mechanisms and capacity for protein disulfide bond formation. *Proceedings of the National Academy of Sciences*, 105(33), pp.11933–11938. doi:https://doi.org/10.1073/pnas.0804621105.

Dyotima, null, Abulaila, S., Mendoza, J. and Landeta, C. (2024). Development of a sensor for disulfide bond formation in diverse bacteria. *Journal of Bacteriology*, [online] 206(4), p.e0043323. doi:https://doi.org/10.1128/jb.00433-23.

Eaglesham, J.B., Garcia, A. and Berkmen, M. (2021). Production of antibodies in SHuffle Escherichia coli strains. *Methods in Enzymology*, Volume 659, pp.105–144.

Eckels, E.C., Chaudhuri, D., Chakraborty, S., Echelman, D.J. and Haldar, S. (2021). DsbA is a redox-switchable mechanical chaperone. *Chemical science*, 12(33), pp.11109–11120. doi:https://doi.org/10.1039/d1sc03048e.

Eisner, G., Koch, H.-G., Beck, K., Brunner, J. and Matthias Müller (2003). Ligand crowding at a nascent signal sequence. ˜ *The œjournal of cell biology/˜ The œJournal of cell biology*, 163(1), pp.35–44. doi:https://doi.org/10.1083/jcb.200306069.

Fekkes, P. and Driessen, A.J.M. (1999). Protein Targeting to the Bacterial Cytoplasmic Membrane. *Microbiology and Molecular Biology Reviews*, 63(1), pp.161–173. doi:https://doi.org/10.1128/mmbr.63.1.161-173.1999.

Fedoroff, N., Wessler, S. and Shure, M. (1983) 'Isolation of the transposable maize controlling elements AC and DS', *Cell*, 35(1), pp. 235–242. doi:10.1016/0092-8674(83)90226-x.

Frech, C., Wunderlich, M., R. Glockshuber and Schmid, F.X. (1996). Preferential binding of an unfolded protein to DsbA. *EMBO journal*, 15(2), pp.392–398. doi:https://doi.org/10.1002/j.1460-2075.1996.tb00369.x.

Gawronski, J.D., Wong, S.M., Giannoukos, G., Ward, D.V. and Akerley, B.J. (2009). Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for *Haemophilus* genes required in the lung. *Proceedings of the National Academy of Sciences of the United States of America*, 106(38), pp.16422–16427. doi:https://doi.org/10.1073/pnas.0906627106.

Gelis, I., Bonvin, A.M.J.J., Keramisanou, D., Koukaki, M., Gouridis, G., Karamanou, S., Economou, A. and Kalodimos, C.G. (2007). Structural Basis for Signal-Sequence Recognition by the Translocase Motor SecA as Determined by NMR. *Cell*, 131(4), pp.756–769. doi:https://doi.org/10.1016/j.cell.2007.09.039.

Genevaux, P., Keppel, F., Schwager, F., Langendijk-Genevaux, P.S., Hartl, F.U. and Georgopoulos, C. (2004). In vivo analysis of the overlapping functions of DnaK and trigger factor. *EMBO reports*, [online] 5(2), pp.195–200. doi:https://doi.org/10.1038/sj.embor.7400067.

Georgopoulos, C.P. and Eisen, H. (1974). Bacterial mutants which block phage assembly. *Journal of supramolecular structure*, 2(2-4), pp.349–359. doi:https://doi.org/10.1002/jss.400020224.

Gerding, H.R., Karreman, C., Daiber, A., Delp, J., Hammler, D., Mex, M., Schildknecht, S. and Leist, M. (2019). Reductive modification of genetically encoded 3-nitrotyrosine sites in alpha synuclein expressed in *E.coli*. *Redox Biology*, [online] 26, p.101251. doi:https://doi.org/10.1016/j.redox.2019.101251.

Gersteuer, F., Morici, M., Gabrielli, S., Fujiwara, K., Safdari, H.A., Helge Paternoga, Bock, L.V., Chiba, S. and Wilson, D.N. (2024). The SecM arrest peptide traps a pre-peptide bond formation state of the ribosome. *Nature communications*, 15(1). doi:https://doi.org/10.1038/s41467-024-46762-2.

Geyter, J.D., Portaliou, A.G., Srinivasu, B.Y., Krishnamurthy, S., Economou, A. and Spyridoula Karamanou (2020). Trigger factor is a *bona fide* secretory pathway chaperone that interacts with SecB and the translocase. *EMBO Reports*, 21(6). doi:https://doi.org/10.15252/embr.201949054.

Ghatak, S., King, Z.A., Sastry, A. and Palsson, B.O. (2019). The y-ome defines the 35% of *Escherichia coli* genes that lack experimental evidence of function. *Nucleic Acids Research*, 47(5), pp.2446–2454. doi:https://doi.org/10.1093/nar/gkz030.

Goldstein, J., Pollitt, N.S. and Inouye, M. (1990). Major cold shock protein of Escherichia coli. *Proceedings of the National Academy of Sciences*, 87(1), pp.283–287. doi:https://doi.org/10.1073/pnas.87.1.283.

Goodall, E.C.A., Robinson, A., Johnston, I.G., Jabbari, S., Turner, K.A., Cunningham, A.F., Lund, P.A., Cole, J.A. and Henderson, I.R. (2018). The Essential Genome of Escherichia coli K-12. *mBio*, 9(1). doi:https://doi.org/10.1128/mbio.02096-17.

Goodman, A.L., McNulty, N.P., Zhao, Y., Leip, D., Mitra, R.D., Lozupone, C.A., Knight, R. and Gordon, J.I. (2009). Identifying Genetic Determinants Needed to Establish a Human Gut Symbiont in Its Habitat. *Cell Host & Microbe*, 6(3), pp.279–289. doi:https://doi.org/10.1016/j.chom.2009.08.003.

Goulding, C.W., Sawaya, M.R., Parseghian, A., Lim, V., Eisenberg, D. and Missiakas, D. (2002). Thiol−Disulfide Exchange in an Immunoglobulin-like Fold: Structure of the N-Terminal Domain of DsbD,. *Biochemistry*, 41(22), pp.6920–6927. doi:https://doi.org/10.1021/bi016038l.

Green, B., Bouchier, C., Fairhead, C., Craig, N.L. and Cormack, B.P. (2012). Insertion site preference of Mu, Tn5, and Tn7 transposons. *Mobile DNA*, 3(1), p.3. doi:https://doi.org/10.1186/1759-8753-3-3.

Gunning, P.W. and Hardeman, E.C. (2018). Fundamental differences. *eLife*, 7.

doi:https://doi.org/10.7554/elife.34477.

Guthrie, B. and Wickner, W. (1990). Trigger factor depletion or overproduction causes defective cell division but does not block protein export. *Journal of Bacteriology*, 172(10), pp.5555–5562. doi:https://doi.org/10.1128/jb.172.10.5555-5562.1990.

Haebel, P.W., Goldstone, D., Beckwith, J. and Metcalf, P. (2002). The disulfide bond isomerase DsbC is activated by an immunoglobulin-fold thiol oxidoreductase: crystal structure of the DsbC-DsbDalpha complex. *The EMBO Journal*, 21(18), pp.4774–4784. doi:https://doi.org/10.1093/emboj/cdf489.

Halic, M., Gartmann, M., Schlenker, O., Mielke, T., Pool, M.R., Sinning, I. and Beckmann, R. (2006). Signal Recognition Particle Receptor Exposes the Ribosomal Translocon Binding Site. *Science*, 312(5774), pp.745–747. doi:https://doi.org/10.1126/science.1124864.

Han, M.-J. (2016). Exploring the proteomic characteristics of the Escherichia coli B and K-12 strains in different cellular compartments. *Journal of Bioscience and Bioengineering*, [online] 122(1), pp.1–9. doi:https://doi.org/10.1016/j.jbiosc.2015.12.005.

Han, M.-J., Jeong, H., Lee, C., Xia, X.-X., Lee, D.-H., Shim, J., Lee, S., Oh, T. and Kim, J.F. (2012). Comparative multi-omics systems analysis of Escherichia coli strains B and K-12. *Genome Biology*, 13(5), p.R37. doi:https://doi.org/10.1186/gb-2012-13-5-r37.

Hand, N.J., Klein, R., Laskewitz, A. and PohlschröderM. (2006). Archaeal and Bacterial SecD and SecF Homologs Exhibit Striking Structural and Functional Conservation. *Journal of Bacteriology*, 188(4), pp.1251–1259. doi:https://doi.org/10.1128/jb.188.4.1251-1259.2006.

Hatahet, F., Nguyen, V.D., Salo, K.E. and Ruddock, L.W. (2010). Disruption of reducing pathways is not essential for efficient disulfide bond formation in the cytoplasm of *E. coli*. *Microbial Cell Factories*, [online] 9(1). doi:https://doi.org/10.1186/1475-2859-9-67.

Heras, B., Edeling, M.A., Schirra, H.J., Raina, S. and Martin, J.L. (2004). Crystal structures of the DsbG disulfide isomerase reveal an unstable disulfide. *Proceedings of the National Academy of Sciences of the United States of America*, 101(24), pp.8876–8881. doi:https://doi.org/10.1073/pnas.0402769101.

Herrera, G., Martinez, A., Blanco, M. and O'Connor, J.-E. (2002). Assessment ofEscherichia coli B with enhanced permeability to fluorochromes for flow cytometric assays of bacterial cell function. *Cytometry*, 49(2), pp.62–69. doi:https://doi.org/10.1002/cyto.10148.

Hiniker, A. and James C.A. Bardwell (2004). In Vivo Substrate Specificity of Periplasmic Disulfide Oxidoreductases. *Journal of Biological Chemistry*, 279(13), pp.12967–12973. doi:https://doi.org/10.1074/jbc.m311391200.

Holmgren, A. (1976). Hydrogen donor system for Escherichia coli ribonucleoside-diphosphate reductase dependent upon glutathione. *Proceedings of the National Academy of*

Sciences, 73(7), pp.2275–2279. doi:https://doi.org/10.1073/pnas.73.7.2275.

Holmgren, A. (2002). Redox regulation by thioredoxin and glutaredoxin systems. *GBM Annual Fall meeting Halle 2002*, 2002(Fall). doi:https://doi.org/10.1240/sav_gbm_2002_h_000248.

Huber, D., Boyd, D., Xia, Y., Olma, M.H., Gerstein, M. and Beckwith, J. (2005). Use of Thioredoxin as a Reporter To Identify a Subset of *Escherichia coli* Signal Sequences That Promote Signal Recognition Particle-Dependent Translocation. *Journal of Bacteriology*, 187(9), pp.2983–2991. doi:https://doi.org/10.1128/jb.187.9.2983-2991.2005.

Huie, J.L. and Silhavy, T.J. (1995). Suppression of signal sequence defects and azide resistance in Escherichia coli commonly result from the same mutations in secA. *Journal of Bacteriology*, [online] 177(12), pp.3518–3526. doi:https://doi.org/10.1128/jb.177.12.3518-3526.1995.

Hutchings, M.I., Truman, A.W. and Wilkinson, B. (2019). Antibiotics: past, Present and Future. *Current Opinion in Microbiology*, 51(1), pp.72–80. doi:https://doi.org/10.1016/j.mib.2019.10.008.

Hyun Seo, J., Gyun Kang, D. and Joon Cha, H. (2003). Comparison of cellular stress levels and green-fluorescent-protein expression in several Escherichia coli strains. *Biotechnology and Applied Biochemistry*, 37(2), p.103. doi:https://doi.org/10.1042/ba20020041.

Imlay, J.A. (2013). The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. *Nature Reviews Microbiology*, 11(7), pp.443–454. doi:https://doi.org/10.1038/nrmicro3032.

Inaba, K., Murakami, S., Nakagawa, A., Iida, H., Kinjo, M., Ito, K. and Suzuki, M. (2009). Dynamic nature of disulphide bond formation catalysts revealed by crystal structures of DsbB. *The EMBO Journal*, 28(6), pp.779–791. doi:https://doi.org/10.1038/emboj.2009.21.

Jamshad, M., Knowles, T.J., White, S.R., Ward, D.G., Mohammed, F., Kazi Mizanur Rahman, Wynne, M., Hughes, G., Kramer, G., Bernd Bukau and Huber, D. (2019). The C-terminal tail of the bacterial translocation ATPase SecA modulates its activity. *elife*. doi:https://doi.org/10.7554/elife.48385.

Jander, G., Martin, N.L. and Beckwith, J. (1994). Two cysteines in each periplasmic domain of the membrane protein DsbB are required for its function in protein disulfide bond formation. *The EMBO Journal*, 13(21), pp.5121–5127. doi:https://doi.org/10.1002/j.1460-2075.1994.tb06841.x.

Jeong, H., Barbe, V., Lee, C.H., Vallenet, D., Yu, D.S., Choi, S.-H., Couloux, A., Lee, S.-W., Yoon, S.H., Cattolico, L., Hur, C.-G., Park, H.-S., Ségurens, B., Kim, S.C., Oh, T.K., Lenski, R.E., Studier, F.W., Daegelen, P. and Kim, J.F. (2009). Genome sequences of Escherichia coli B strains REL606 and BL21(DE3). *Journal of molecular biology*, [online] 394(4), pp.644–52. doi:https://doi.org/10.1016/j.jmb.2009.09.052.

Jiang, C., Wynne, M. and Huber, D. (2021). How Quality Control Systems AID Sec-Dependent Protein Translocation. *Frontiers in Molecular Biosciences*, 8. doi:https://doi.org/10.3389/fmolb.2021.669376.

Jones, P.G., VanBogelen, R.A. and Neidhardt, F.C. (1987). Induction of proteins in response to low temperature in Escherichia coli. *Journal of Bacteriology*, 169(5), pp.2092–2095. doi:https://doi.org/10.1128/jb.169.5.2092-2095.1987.

Kadokura, H. and Beckwith, J. (2009). Detecting Folding Intermediates of a Protein as It Passes through the Bacterial Translocation Channel. *Cell*, [online] 138(6), pp.1164–1173. doi:https://doi.org/10.1016/j.cell.2009.07.030.

Kadokura, H., Tian, H., Zander, T., James C.A. Bardwell and Beckwith, J. (2004). Snapshots of DsbA in Action: Detection of Proteins in the Process of Oxidative Folding. *Science*, 303(5657), pp.534–537. doi:https://doi.org/10.1126/science.1091724.

Kandror, O. and Goldberg, A.L. (1997). Trigger factor is induced upon cold shock and enhances viability of Escherichia coli at low temperatures. *Proceedings of the National Academy of Sciences*, 94(10), pp.4978–4981. doi:https://doi.org/10.1073/pnas.94.10.4978.

Kandror, O., Sherman, M., Rhode, M. and Goldberg, A.L. (1995). Trigger factor is involved in GroEL-dependent protein degradation in Escherichia coli and promotes binding of GroEL to unfolded proteins. *The EMBO Journal*, 14(23), pp.6021–6027. doi:https://doi.org/10.1002/j.1460-2075.1995.tb00290.x.

Kato, Y., Nishiyama, K. and Tokuda, H. (2003). Depletion of SecDF-YajC causes a decrease in the level of SecG: implication for their functional interaction. *FEBS Letters*, [online] 550(1), pp.114–118. doi:https://doi.org/10.1016/S0014-5793(03)00847-0.

Katzen, F. and Beckwith, J. (2000). Transmembrane Electron Transfer by the Membrane Protein DsbD Occurs via a Disulfide Bond Cascade. *Cell*, 103(5), pp.769–779. doi:https://doi.org/10.1016/s0092-8674(00)00180-x.

Khalilvand, A.B., Aminzadeh, S., Sanati, M.H. and Mahboudi, F. (2022). Media optimization for SHuffle T7 Escherichia coli expressing SUMO-Lispro proinsulin by response surface methodology. *BMC Biotechnology*, 22(1). doi:https://doi.org/10.1186/s12896-021-00732-4.

Kleckner, N., Chan, R.K., Tye, B.-K. and Botstein, D. (1975). Mutagenesis by insertion of a drug-resistance element carrying an inverted repetition. *Journal of Molecular Biology*, 97(4), pp.561–575. doi:https://doi.org/10.1016/s0022-2836(75)80059-3.

Kobayashi, T., Satoshi Kishigami, Sone, M., Hachiro Inokuchi, Tatsushi Mogi and Ito, K. (1997). Respiratory chain is required to maintain oxidized states of the DsbA-DsbB disulfide bond formation system in aerobically growing *Escherichia coli* cells. *Proceedings of the*

National Academy of Sciences of the United States of America, 94(22), pp.11857–11862. doi:https://doi.org/10.1073/pnas.94.22.11857.

Kramer, G., Boehringer, D., Ban, N. and Bukau, B. (2009). The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nature Structural & Molecular Biology*, 16(6), pp.589–597. doi:https://doi.org/10.1038/nsmb.1614.

Kramer, G., Rauch, T., Rist, W., Vorderwülbecke, S., Patzelt, H., Schulze-Specking, A., Ban, N., Deuerling, E. and Bukau, B. (2002). L23 protein functions as a chaperone docking site on the ribosome. *Nature*, [online] 419(6903), pp.171–174. doi:https://doi.org/10.1038/nature01047.

Kristensen, O. and Gajhede, M. (2003). Chaperone Binding at the Ribosomal Exit Tunnel. *Structure*, 11(12), pp.1547–1556. doi:https://doi.org/10.1016/j.str.2003.11.003.

Krupp, R., Chan, C. and Missiakas, D. (2001). DsbD-catalyzed Transport of Electrons across the Membrane ofEscherichia coli. *Journal of Biological Chemistry*, 276(5), pp.3696–3701. doi:https://doi.org/10.1074/jbc.m009500200.

Kumar, J.K., Tabor, S. and Richardson, C.C. (2004). Proteomic analysis of thioredoxin-targeted proteins in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 101(11), pp.3759–3764. doi:https://doi.org/10.1073/pnas.0308701101.

Kumazaki, K., Kishimoto, T., Furukawa, A., Mori, H., Tanaka, Y., Kishimoto, T., Dohmae, N., Ishitani , R., Tsukazaki, T. and Nureki, O. (2014). Crystal structure of Escherichia coli YidC, a membrane protein chaperone and insertase. *Scientific reports*, [online] 4. doi:https://doi.org/10.1038/srep07299.

Kusters, R., Lentzen, G., Eppens, E., van Geel, A., van, Wintermeyer, W. and Luirink, J. (1995). The functioning of the SRP receptor FtsY in protein-targeting in *E. coli* is correlated with its ability to bind and hydrolyse GTP. *FEBS Letters*, 372(2-3), pp.253–258. doi:https://doi.org/10.1016/0014-5793(95)00997-n.

Kyung Hee Paek and Walker, G.C. (1987). Escherichia coli dnaK null mutants are inviable at high temperature. *J. Bacteriol*, 169(1), pp.283–290. doi:https://doi.org/10.1128/jb.169.1.283-290.1987.

Langridge, G.C., Phan, M.-D. ., Turner, D.J., Perkins, T.T., Parts, L., Haase, J., Charles, I., Maskell, D.J., Peters, S.E., Dougan, G., Wain, J., Parkhill, J. and Turner, A.K. (2009). Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Research*, 19(12), pp.2308–2316. doi:https://doi.org/10.1101/gr.097097.109.

Laurent, T.C., Moore, E.C. and Reichard, P. (1964). ENZYMATIC SYNTHESIS OF DEOXYRIBONUCLEOTIDES. IV. ISOLATION AND CHARACTERIZATION OF THIOREDOXIN, THE HYDROGEN DONOR FROM ESCHERICHIA COLI B. *The Journal of Biological Chemistry*, [online] 239, pp.3436–3444. Available at:

https://pubmed.ncbi.nlm.nih.gov/14245400/.

Lederberg, J. and Tatum, E.L. (1946). DETECTION OF BIOCHEMICAL MUTANTS OF MICROORGANISMS. *Journal of Biological Chemistry*, 165(1), pp.381–382. doi:https://doi.org/10.1016/s0021-9258(17)41242-7.

Lee, H.-P. and Bernstein, H. (2001). The targeting pathway of *Escherichia coli* presecretory and integral membrane proteins is specified by the hydrophobicity of the targeting signal. *PNAS*, 98(6), pp.3471–3476. doi:https://doi.org/10.1073/pnas.051484198.

Lee, H.C. and Bernstein, H. (2002). Trigger Factor Retards Protein Export in Escherichia coli. *Journal of Biological Chemistry*, 277(45), pp.43527–43535. doi:https://doi.org/10.1074/jbc.m205950200.

Li, W., Schulman, S., Dutton, R.J., Boyd, D., Beckwith, J. and Rapoport, T.A. (2010). Structure of a bacterial homologue of vitamin K epoxide reductase. *Nature*, 463(7280), pp.507–512. doi:https://doi.org/10.1038/nature08720.

Liebke, H.H. (1987). Multiple SecA protein isoforms in Escherichia coli. *Journal of Bacteriology*, 169(3), pp.1174–1181. doi:https://doi.org/10.1128/jb.169.3.1174-1181.1987.

Lill, R., Crooke, E., Guthrie, B. and Wickner, W. (1988). The ‘trigger factor cycle’ includes ribosomes, presecretory proteins, and the plasma membrane. *Cell*, 54(7), pp.1013–1018. doi:https://doi.org/10.1016/0092-8674(88)90116-x.

Liu, J. and Rosen, B.P. (1997). Ligand Interactions of the ArsC Arsenate Reductase. *Journal of Biological Chemistry*, 272(34), pp.21084–21089. doi:https://doi.org/10.1074/jbc.272.34.21084.

Liu, X. and Wang, C.-C. (2001). Disulfide-dependent Folding and Export of Escherichia coli DsbC. *Journal of Biological Chemistry*, 276(2), pp.1146–1151. doi:https://doi.org/10.1074/jbc.m004929200.

Lobstein, J., Emrich, C.A., Jeans, C., Faulkner, M., Riggs, P. and Berkmen, M. (2012). SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microbial Cell Factories*, 11(1). doi:https://doi.org/10.1186/1475-2859-11-56.

Luria, S.E. and Delbrück, M. (1943). MUTATIONS OF BACTERIA FROM VIRUS SENSITIVITY TO VIRUS RESISTANCE. *Genetics*, 28(6), pp.491–511. doi:https://doi.org/10.1093/genetics/28.6.491.

Manta, B., Lundstedt, E., Garcia, A., Eaglesham, J.B. and Berkmen, M. (2022). *Disulfide bond formation in Escherichia coli*. [online] ScienceDirect. Available at: https://www.sciencedirect.com/science/article/pii/B9780323902199000212.

Martin, J.L., Bardwell, J.C.A. and Kuriyan, J. (1993). Crystal structure of the DsbA protein required for disulphide bond formation in vivo. *Nature*, [online] 365(6445), pp.464–468. doi:https://doi.org/10.1038/365464a0.

Mazmanian, S.K., Ton-That, H. and Schneewind, O. (2001). Sortase-catalysed anchoring of surface proteins to the cell wall of Staphylococcus aureus. *Molecular Microbiology*, 40(5), pp.1049–1057. doi:https://doi.org/10.1046/j.1365-2958.2001.02411.x.

McCarthy, A.A., Haebel, P.W., Törrönen, A., Rybin, V., Baker, E.N. and Metcalf, P. (2000). Crystal structure of the protein disulfide bond isomerase, DsbC, from Escherichia coli. *Nature Structural Biology*, 7(3), pp.196–199. doi:https://doi.org/10.1038/73295.

McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences*, 36(6), pp.344–355. doi:https://doi.org/10.1073/pnas.36.6.344.

Miller, J.H. (1972). *Experiments in molecular genetics.* Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. xvi, 466 p. pp.

Milner, M.T. (n.d.). *Can Transposon Directed Insertion-Site Sequencing Be Used To Predict Possible Outcomes Of Evolution?* PhD thesis. Available at: https://etheses.bham.ac.uk.

Minkov, I.B. and Heinrich Strotmann (1989). The effect of azide on regulation of the chloroplast H+-ATPase by ADP and phosphate. *Biochimica et biophysica acta. Bioenergetics*, 973(1), pp.7–12. doi:https://doi.org/10.1016/s0005-2728(89)80394-9.

Miranda-Vizuete, A., Damdimopoulos, A.E., Gustafsson, J.-Å. and Spyrou, G. (1997). Cloning, Expression, and Characterization of a NovelEscherichia coli Thioredoxin. *Journal of Biological Chemistry*, 272(49), pp.30841–30847. doi:https://doi.org/10.1074/jbc.272.49.30841.

Missiakas, D., Georgopoulos, C. and Raina, S. (1994). The Escherichia coli dsbC (xprA) gene encodes a periplasmic protein involved in disulfide bond formation. *The EMBO Journal*, 13(8), pp.2013–2020. doi:https://doi.org/10.1002/j.1460-2075.1994.tb06471.x.

Missiakas, D., Schwager, F. and Raina, S. (1995). Identification and characterization of a new disulfide isomerase-like protein (DsbD) in Escherichia coli. *The EMBO Journal*, 14(14), pp.3415–3424. doi:https://doi.org/10.1002/j.1460-2075.1995.tb07347.x.

Mitra, R. and Church, G.M. (1999). In situ localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Research*, 27(24), pp.34e34. doi:https://doi.org/10.1093/nar/27.24.e34.

Murén, E., Suciu, D., Topping, T.B., Kumamoto, C.A. and Randall, L.L. (1999). Mutational Alterations in the Homotetrameric Chaperone SecB That Implicate the Structure as Dimer of Dimers. *Journal of Biological Chemistry*, 274(27), pp.19397–19402.

doi:https://doi.org/10.1074/jbc.274.27.19397.

Narcis-Adrian Petriman, Jauß, B., Hufnagel, A., Franz, L., Ilie Sachelaru, Friedel Drepper, Warscheid, B. and Koch, H.-G. (2018). The interaction network of the YidC insertase with the SecYEG translocon, SRP and the SRP receptor FtsY. *Scientific Reports*, 8(1). doi:https://doi.org/10.1038/s41598-017-19019-w.

Nasiri, M., Babaie, J., Amiri, S., Azimi, E., Shamshiri, S., Khalaj, V., Golkar, M. and Fard-Esfahani, P. (2017). SHuffle^{TM} T7 strain is capable of producing high amount of recombinant human fibroblast growth factor-1 (rhFGF-1) with proper physicochemical and biological properties. *Journal of Biotechnology*, 259, pp.30–38. doi:https://doi.org/10.1016/j.jbiotec.2017.08.015.

Oh, E., Becker, Annemarie H., Sandikci, A., Huber, D., Chaba, R., Gloge, F., Nichols, Robert J., Typas, A., Gross, Carol A., Kramer, G., Weissman, Jonathan S. and Bukau, B. (2011). Selective Ribosome Profiling Reveals the Cotranslational Chaperone Action of Trigger Factor In Vivo. *Cell*, 147(6), pp.1295–1308. doi:https://doi.org/10.1016/j.cell.2011.10.044.

Oliver, D.B. and Beckwith, J. (1981). E. coli mutant pleiotropically defective in the export of secreted proteins. *Cell*, [online] 25(3), pp.765–772. doi:https://doi.org/10.1016/0092-8674(81)90184-7.

Oliver, D.B., Cabelli, R.J., Dolan, K.M. and Jarosik, G.P. (1990). Azide-resistant mutants of Escherichia coli alter the SecA protein, an azide-sensitive component of the protein export machinery. *Proceedings of the National Academy of Sciences*, [online] 87(21), pp.8227–8231. doi:https://doi.org/10.1073/pnas.87.21.8227.

Osborne, A.R., Rapoport, T.A. and van den Berg, B. (2005). PROTEIN TRANSLOCATION BY THE SEC61/SECY CHANNEL. *Annual Review of Cell and Developmental Biology*, 21(1), pp.529–550. doi:https://doi.org/10.1146/annurev.cellbio.21.012704.133214.

Owji, H., Nezafat, N., Negahdaripour, M., Hajiebrahimi, A. and Ghasemi, Y. (2018). A comprehensive review of signal peptides: Structure, roles, and applications. *European Journal of Cell Biology*, 97(6), pp.422–441. doi:https://doi.org/10.1016/j.ejcb.2018.06.003.

Paetzel, M., Dalbey, R.E. and Strynadka, N.C.J. (2002). Crystal Structure of a Bacterial Signal Peptidase Apoenzyme. *Journal of Biological Chemistry*, 277(11), pp.9512–9519. doi:https://doi.org/10.1074/jbc.m110983200.

Palmer, T. and Berks, B.C. (2012). The twin-arginine translocation (Tat) protein export pathway. *Nature Reviews Microbiology*, 10(7), pp.483–496. doi:https://doi.org/10.1038/nrmicro2814.

Peros, S. (2022). Genetic analysis of SHuffle B strain: Isolation and Characterisation of Temperature Resistant Mutants. MRes thesis.

Phue, J.-N., Lee, S.J., Kaufman, J.B., Negrete, A. and Shiloach, J. (2010). Acetate accumulation through alternative metabolic pathways in ackA − pta − poxB − triple mutant in *E. coli* B (BL21). *Biotechnology Letters*, 32(12), pp.1897–1903. doi:https://doi.org/10.1007/s10529-010-0369-7.

Plumbridge, J., Deville, F., Sacerdot, C., Hans Uffe Petersen, Cenatiempo, Y., Cozzone, A.J., M. Grunberg-Manago and John W.B. Hershey (1985). Two translational initiation sites in the infB gene are used to express initiation factor IF2 alpha and IF2 beta in Escherichia coli. *The EMBO Journal*, 4(1), pp.223–229. doi:https://doi.org/10.1002/j.1460-2075.1985.tb02339.x.

Pogliano, K.J. and Beckwith, J. (1994). Genetic and molecular characterization of the Escherichia coli secD operon and its products. *Journal of Bacteriology*, 176(3), pp.804–814. doi:https://doi.org/10.1128/jb.176.3.804-814.1994.

Poole, L.B. (2005). Bacterial defenses against oxidants: mechanistic features of cysteine-based peroxidases and their flavoprotein reductases. *Archives of Biochemistry and Biophysics*, [online] 433(1), pp.240–254. doi:https://doi.org/10.1016/j.abb.2004.09.006.

Porat, A., Cho, S.-H. and Beckwith, J. (2004). The unusual transmembrane electron transporter DsbD and its homologues: a bacterial family of disulfide reductases. *Research in Microbiology*, 155(8), pp.617–622. doi:https://doi.org/10.1016/j.resmic.2004.05.005.

Prieto-ÁlamoM.-J., Jurado, J., Gallardo-Madueño, R., Monje-Casas, F., Holmgren, A. and Pueyo, C. (2000). Transcriptional Regulation of Glutaredoxin and Thioredoxin Pathways and Related Enzymes in Response to Oxidative Stress. *Journal of Biological Chemistry*, 275(18), pp.13398–13405. doi:https://doi.org/10.1074/jbc.275.18.13398.

Prinz, W.A., Åslund, F., Holmgren, A. and Beckwith, J. (1997). The Role of the Thioredoxin and Glutaredoxin Pathways in Reducing Protein Disulfide Bonds in theEscherichia coliCytoplasm. *Journal of Biological Chemistry*, 272(25), pp.15661–15667. doi:https://doi.org/10.1074/jbc.272.25.15661.

Reeh, S. and Pedersen, S. (1979). Post-translational modification of Escherichia coli ribosomal protein S6. *MGG. Molecular and general genetics/MGG. Molecular & general genetics*, 173(2), pp.183–187. doi:https://doi.org/10.1007/bf00330309.

Ren, G., Ke, N. and Berkmen, M. (2016). Use of the SHuffle Strains in Production of Proteins. *Current Protocols in Protein Science*, 85(1). doi:https://doi.org/10.1002/cpps.11.

Ritz, D. (2001). Conversion of a Peroxiredoxin into a Disulfide Reductase by a Triplet Repeat Expansion. *Science*, 294(5540), pp.158–160. doi:https://doi.org/10.1126/science.1063143.

Ritz, D., Patel, H., Doan, B., Zheng, M., Åslund, F., Storz, G. and Beckwith, J. (2000). Thioredoxin 2 Is Involved in the Oxidative Stress Response inEscherichia coli. *Journal of Biological Chemistry*, 275(4), pp.2505–2512. doi:https://doi.org/10.1074/jbc.275.4.2505.

Rizzolo, K., Yu, A.Y.H., Ologbenla, A., Kim, S.R., Zhu, H., Ishimori, K., Thibault, G., Leung, E., Zhang, Y.W., Teng, M., Haniszewski, M., Miah, N., Phanse, S., Minic, Z., Lee, S., Caballero, J.D., Babu, M., Tsai, F.T.F., Saio, T. and Houry, W.A. (2021). Functional cooperativity between the trigger factor chaperone and the ClpXP proteolytic complex. *Nature Communications*, [online] 12(1), p.281. doi:https://doi.org/10.1038/s41467-020-20553-x.

Robinson, M.-P., Ke, N., Lobstein, J., Peterson, C., Szkodny, A., Mansell, T.J., Tuckey, C., Riggs, P.D., Colussi, P.A., Noren, C.J., Taron, C.H., DeLisa, M.P. and Berkmen, M. (2015). Efficient expression of full-length antibodies in the cytoplasm of engineered bacteria. *Nature Communications*, [online] 6(1), pp.1–9. doi:https://doi.org/10.1038/ncomms9072.

Saaranen, M.J. and Ruddock, L.W. (2013). Disulfide Bond Formation in the Cytoplasm. *Antioxidants & Redox Signaling*, 19(1), pp.46–53. doi:https://doi.org/10.1089/ars.2012.4868.

Sachelaru, I., Winter, L., Knyazev, D.G., Zimmermann, M., Vogt, A., Kuttner, R., Ollinger, N., Siligan, C., Pohl, P. and Koch, H.-G. (2017). YidC and SecYEG form a heterotetrameric protein translocation channel. *Scientific Reports*, 7(1). doi:https://doi.org/10.1038/s41598-017-00109-8.

Saio, T., Guan, X., Rossi, P., Economou, A. and Kalodimos, C.G. (2014). Structural Basis for Protein Antiaggregation Activity of the Trigger Factor Chaperone. *Science*, 344(6184). doi:https://doi.org/10.1126/science.1250494.

Sakr, S., Cirinesi, A.-M., Ullers, R.S., Schwager, F., Georgopoulos, C. and Genevaux, P. (2010). Lon Protease Quality Control of Presecretory Proteins in Escherichia coli and Its Dependence on the SecB and DnaJ (Hsp40) Chaperones. 285(30), pp.23506–23514. doi:https://doi.org/10.1074/jbc.m110.133058.

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchison, C.A., Slocombe, P.M. and Smith, M. (1977). Nucleotide sequence of bacteriophage φX174 DNA. *Nature*, [online] 265(5596), pp.687–695. doi:https://doi.org/10.1038/265687a0.

Sanger, F., Brownlee, G.G. and Barrell, B.G. (1965). A two-dimensional fractionation procedure for radioactive nucleotides. *Journal of Molecular Biology*, 14(1), p.303. doi:https://doi.org/10.1016/s0022-2836(65)80253-4.

Schibich, D., Gloge, F., Pöhner, I., Patrik Björkholm, Wade, R.C., Gunnar von Heijne, Bernd Bukau and Günter Krämer (2016). Global profiling of SRP interaction with nascent polypeptides. *Nature*, 536(7615), pp.219–223. doi:https://doi.org/10.1038/nature19070.

Schmidt, M.G. and Kiser, K.B. (1999). SecA: the ubiquitous component of preprotein translocase in prokaryotes. *Microbes and Infection*, 1(12), pp.993–1004. doi:https://doi.org/10.1016/s1286-4579(99)80517-6.

Schneider, D., Duperchy, E., Depeyrot, J., Coursange, E., Lenski, R.E. and Blot, M. (2002).

Genomic Comparisons among Escherichia Coli Strains B, K-12, and O157:H7 Using IS Elements as Molecular Markers. *BMC Microbiology*, 2(1), p.18. doi:https://doi.org/10.1186/1471-2180-2-18.

Scholz, C., Stoller, G., Zarnt, T., Fischer, G. and Schmid, F.X. (1997). Cooperation of enzymatic and chaperone functions of trigger factor in the catalysis of protein folding. *The EMBO Journal*, 16(1), pp.54–58. doi:https://doi.org/10.1093/emboj/16.1.54.

Seifert, H.S., Chen, E.Y., So, M. and Heffron, F. (1986). Shuttle mutagenesis: a method of transposon mutagenesis for Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences*, 83(3), pp.735–739. doi:https://doi.org/10.1073/pnas.83.3.735.

Sévin, D.C. and Sauer, U. (2014). Ubiquinone accumulation improves osmotic-stress tolerance in Escherichia coli. *Nature Chemical Biology*, 10(4), pp.266–272. doi:https://doi.org/10.1038/nchembio.1437.

Shanmugam, S.K. and Dalbey, R.E. (2019). The Conserved Role of YidC in Membrane Protein Biogenesis. *Microbiology Spectrum*, 7(1). doi:https://doi.org/10.1128/microbiolspec.psib-0014-2018.

Shibai, A., Takahashi, Y., Ishizawa, Y., Motooka, D., Nakamura, S., Ying, B.-W. and Tsuru, S. (2017). Mutation accumulation under UV radiation in Escherichia coli. *Scientific Reports*, [online] 7(14531). doi:https://doi.org/10.1038/s41598-017-15008-1.

Sims, G.E. and Kim, S.-H. (2011). Whole-genome phylogeny of Escherichia coli/Shigella group by feature frequency profiles (FFPs). *Proceedings of the National Academy of Sciences of the United States of America*, [online] 108(20), pp.8329–8334. doi:https://doi.org/10.1073/pnas.1105168108.

Slatko, B.E., Gardner, A.F. and Ausubel, F.M. (2018). Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology*, [online] 122(1), p.e59. doi:https://doi.org/10.1002/cpmb.59.

Smith, D., Price, J., Burby, P., Blanco, L., Chamberlain, J. and Chapman, M. (2017). The Production of Curli Amyloid Fibers Is Deeply Integrated into the Biology of Escherichia coli. *Biomolecules*, [online] 7(4), p.75. doi:https://doi.org/10.3390/biom7040075.

Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B.H. and Hood, L.E. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature*, [online] 321(6071), pp.674–679. doi:https://doi.org/10.1038/321674a0.

Søballe, B. and Poole, R.K. (2000). Ubiquinone limits oxidative stress in Escherichia coli. *Microbiology*, 146(4), pp.787–796. doi:https://doi.org/10.1099/00221287-146-4-787.

Sri Karthika Shanmugam, Backes, N., Chen, Y., Belardo, A., Phillips, G.J. and Dalbey, R.E. (2019). New Insights into Amino-Terminal Translocation as Revealed by the Use of YidC and Sec Depletion Strains. *Journal of Molecular Biology/Journal of molecular biology*, 431(5), pp.1025–1037. doi:https://doi.org/10.1016/j.jmb.2019.01.006.

St Johnston, D. (2002). The art and design of genetic screens: Drosophila melanogaster. *Nature Reviews Genetics*, 3(3), pp.176–188. doi:https://doi.org/10.1038/nrg751.

Staden, R. (1979). A strategy of DNA sequencing employing computer programs. *Nucleic Acids Research*, [online] 6(7), pp.2601–2610. doi:https://doi.org/10.1093/nar/6.7.2601.

Stewart, E.J. (1998). Disulfide bond formation in the Escherichia coli cytoplasm: an invivo role reversal for the thioredoxins. *The EMBO Journal*, [online] 17(19), pp.5543–5550. doi:https://doi.org/10.1093/emboj/17.19.5543.

Storz, G., Tartaglia, L.A., Farr, S.B. and Ames, B.N. (1990). Bacterial defenses against oxidative stress. *Trends in Genetics*, 6, pp.363–368. doi:https://doi.org/10.1016/0168-9525(90)90278-e.

Studier, F.W., Daegelen, P., Lenski, R.E., Maslov, S. and Kim, J.F. (2009). Understanding the Differences between Genome Sequences of Escherichia coli B Strains REL606 and BL21(DE3) and Comparison of the E. coli B and K-12 Genomes. *Journal of Molecular Biology*, 394(4), pp.653–680. doi:https://doi.org/10.1016/j.jmb.2009.09.021.

Sun, X. and Wang, C. (2000). The N-terminal Sequence (Residues 1–65) Is Essential for Dimerization, Activities, and Peptide Binding of Escherichia coli DsbC. *Journal of Biological Chemistry*, 275(30), pp.22743–22749. doi:https://doi.org/10.1074/jbc.m002406200.

Takahashi, Y., Inaba, K. and Ito, K. (2004). Characterization of the Menaquinone-dependent Disulfide Bond Formation Pathway ofEscherichia coli. *Journal of Biological Chemistry*, 279(45), pp.47057–47065. doi:https://doi.org/10.1074/jbc.m407153200.

Tang, M., Nesbitt, A.E., Sperling, L.J., Berthold, D.A., Schwieters, C.D., Gennis, R.B. and Rienstra, C.M. (2013). Structure of the Disulfide Bond Generating Membrane Protein DsbB in the Lipid Bilayer. *Journal of Molecular Biology/Journal of molecular biology*, 425(10), pp.1670–1682. doi:https://doi.org/10.1016/j.jmb.2013.02.009.

Teter, S.A., Houry, W.A., Ang, D., T. Tradler, D. Rockabrand, Fischer, G., Blum, P., Georgopoulos, C. and Hartl, F.U. (1999). Polypeptide Flux through Bacterial Hsp70. *Cell*, 97(6), pp.755–765. doi:https://doi.org/10.1016/s0092-8674(00)80787-4.

Topping, T.B. and Randall, L.L. (1994). Determination of the binding frame within a physiological ligand for the chaperone SecB. *Protein Science*, 3(5), pp.730–736. doi:https://doi.org/10.1002/pro.5560030502.

Tsirigotaki, A., De Geyter, J., Šoštaric´, N., Economou, A. and Karamanou, S. (2016). Protein export through the bacterial Sec pathway. *Nature Reviews Microbiology*, 15(1), pp.21–36. doi:https://doi.org/10.1038/nrmicro.2016.161.

Ullers, R.S., Ang, D., Schwager, F., Georgopoulos, C. and Genevaux, P. (2007). Trigger Factor can antagonize both SecB and DnaK/DnaJ chaperone functions in Escherichia coli. *Proceedings of the National Academy of Sciences*, 104(9), pp.3101–3106. doi:https://doi.org/10.1073/pnas.0608232104.

Ullers, R.S., Luirink, J., Harms, N., Schwager, F., Georgopoulos, C. and Genevaux, P. (2004). SecB is a bona fide generalized chaperone in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, [online] 101(20), pp.7583–7588. doi:https://doi.org/10.1073/pnas.0402398101.

Unden, G. and Bongaerts, J. (1997). Alternative respiratory pathways of Escherichia coli: energetics and transcriptional regulation in response to electron acceptors. Biochimica et Biophysica Acta (BBA) - Bioenergetics, 1320(3), pp.217–234. doi:https://doi.org/10.1016/s0005-2728(97)00034-0.

van der Sluis, E.O., Nouwen, N., Koch, J., de Keyzer, J., van der Does, C., Tampé, R. and Driessen, A.J.M. (2006). Identification of Two Interaction Sites in SecY that Are Important for the Functional Interaction with SecA. *Journal of Molecular Biology*, 361(5), pp.839–849. doi:https://doi.org/10.1016/j.jmb.2006.07.017.

van der Wolk, J.P.W., Wit, J.G. de and Driessen, A.J. (1997). The catalytic cycle of the Escherichia coli SecA ATPase comprises two distinct preprotein translocation events. *The EMBO Journal*, 16(24), pp.7297–7304. doi:https://doi.org/10.1093/emboj/16.24.7297.

van Opijnen, T., Bodi, K.L. and Camilli, A. (2009). Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature Methods*, 6(10), pp.767–772. doi:https://doi.org/10.1038/nmeth.1377.

van Opijnen, T. and Camilli, A. (2013). Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nature Reviews Microbiology*, 11(7), pp.435–442. doi:https://doi.org/10.1038/nrmicro3033.

Vasilyeva, E.A., Minkov, I.B., Fitin, A.F. and Vinogradov, A.D. (1982). Kinetic mechanism of mitochondrial adenosine triphosphatase. Inhibition by azide and activation by sulphite. *Biochemical Journal*, 202(1), pp.15–23. doi:https://doi.org/10.1042/bj2020015.

Vlasuk, G.P., Inouye, S., Ito, H., Itakura, K. and Inouye, M. (1983). Effects of the complete removal of basic amino acid residues from the signal peptide on secretion of lipoprotein in Escherichia coli. *Journal of Biological Chemistry*, 258(11), pp.7141–7148. doi:https://doi.org/10.1016/s0021-9258(18)32343-3.

von Heijne, G. (1986). A new method for predicting signal sequence cleavage sites. *Nucleic Acids Research*, 14(11), pp.4683–4690. doi:https://doi.org/10.1093/nar/14.11.4683.

VOORST, F.V. and DE KRUIJFF, B. (2000). Role of lipids in the translocation of proteins across membranes. *Biochemical Journal*, 347(3), p.601. doi:https://doi.org/10.1042/0264-

6021:3470601.

Walker, K.W. and Gilbert, H.F. (1997). Scanning and Escape during Protein-disulfide Isomerase-assisted Protein Folding. *Journal of Biological Chemistry*, 272(14), pp.8845–8848. doi:https://doi.org/10.1074/jbc.272.14.8845.

Wang, P. and Dalbey, R.E. (2011). Inserting membrane proteins: The YidC/Oxa1/Alb3 machinery in bacteria, mitochondria, and chloroplasts. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1808(3), pp.866–875. doi:https://doi.org/10.1016/j.bbamem.2010.08.014.

Ward, J.E. and Lutkenhaus, J. (1985). Overproduction of FtsZ induces minicell formation in *E. coli*. *Cell*, 42(3), pp.941–949. doi:https://doi.org/10.1016/0092-8674(85)90290-9.

Waters, M. and Tadi, P. (2023). *Streptomycin*. [online] PubMed. Available at: https://www.ncbi.nlm.nih.gov/books/NBK555886/.

Weiss, D.S., Chen, J.C., Jean-Marc Ghigo, Boyd, D. and Beckwith, J. (1999). Localization of FtsI (PBP3) to the Septal Ring Requires Its Membrane Anchor, the Z Ring, FtsA, FtsQ, and FtsL. *J Bacteriol* , 181(2), pp.508–520. doi:https://doi.org/10.1128/jb.181.2.508-520.1999.

Wu, K., Minshull, T.C., Radford, S.E., Calabrese, A.N. and Bardwell, J.C.A. (2022). Trigger factor both holds and folds its client proteins. *Nature Communications*, [online] 13(1), p.4126. doi:https://doi.org/10.1038/s41467-022-31767-6.

Wu, Y., Wu, M., He, G., Zhang, X., Li, W., Gao, Y., Li, Z., Wang, Z. and Zhang, C. (2012). Glyceraldehyde-3-phosphate dehydrogenase: A universal internal control for Western blots in prokaryotic and eukaryotic cells. *Analytical Biochemistry*, [online] 423(1), pp.15–22. doi:https://doi.org/10.1016/j.ab.2012.01.012.

Wynne, M.A. (2022). *Investigation of the structure and function of SecH,* a novel component of the Sec machinery in Escherichia Coli. PhD thesis. Available at: https://etheses.bham.ac.uk..

Xu, W., Jiang, X. and Huang, L. (2019). RNA Interference Technology. *Comprehensive Biotechnology*, [online] pp.560–575. doi:https://doi.org/10.1016/B978-0-444-64046-8.00282-2.

Yamamoto, Y., Ritz, D., Anne-Gaëlle Planson, Jönsson, T.J., Faulkner, M.J., Boyd, D., Beckwith, J. and Poole, L.B. (2008). Mutant AhpC Peroxiredoxins Suppress Thiol-Disulfide Redox Deficiencies and Acquire Deglutathionylating Activity. *Molecular Cell*, 29(1), pp.36–45. doi:https://doi.org/10.1016/j.molcel.2007.11.029.

Yoon, S.H., Jeong, H., Kwon, S.-K. and Kim, J.F. (2009). Genomics, Biological Features, and Biotechnological Applications of Escherichia coli B: ‘Is B for better?!’ *Systems Biology and Biotechnology of Escherichia coli*, pp.1–17. doi:https://doi.org/10.1007/978-1-4020-9394-4_1.

Zarnt, T., Tradler, T., Stoller, G., Scholz, C., Schmid, F.X. and Fischer, G. (1997). Modular structure of the trigger factor required for high activity in protein folding. *Journal of Molecular Biology/Journal of molecular biology*, 271(5), pp.827–837. doi:https://doi.org/10.1006/jmbi.1997.1206.

Zhang, Z., Li, Z.-H., Wang, F., Fang, M., Yin, C.-C., Zhou, Z.-Y., Lin, Q. and Huang, H.-L. (2002). Overexpression of DsbC and DsbG markedly improves soluble and functional expression of single-chain Fv antibodies in Escherichia coli. *Protein Expression and Purification*, 26(2), pp.218–228. doi:https://doi.org/10.1016/s1046-5928(02)00502-8.

Zheng, M. (1998). Activation of the OxyR Transcription Factor by Reversible Disulfide Bond Formation. *Science*, [online] 279(5357), pp.1718–1722. doi:https://doi.org/10.1126/science.279.5357.1718.

Zimmer, J., Nam, Y. and Rapoport, T.A. (2008). Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature*, 455(7215), pp.936–943. doi:https://doi.org/10.1038/nature07335.

Zimmer, J. and Rapoport, T.A. (2009). Conformational Flexibility and Peptide Interaction of the Translocation ATPase SecA. *Journal of Molecular Biology/Journal of molecular biology*, 394(4), pp.606–612. doi:https://doi.org/10.1016/j.jmb.2009.10.024.

Appendices

8.1 TraDIS experiments for SHuffle strains and the wild-type strains

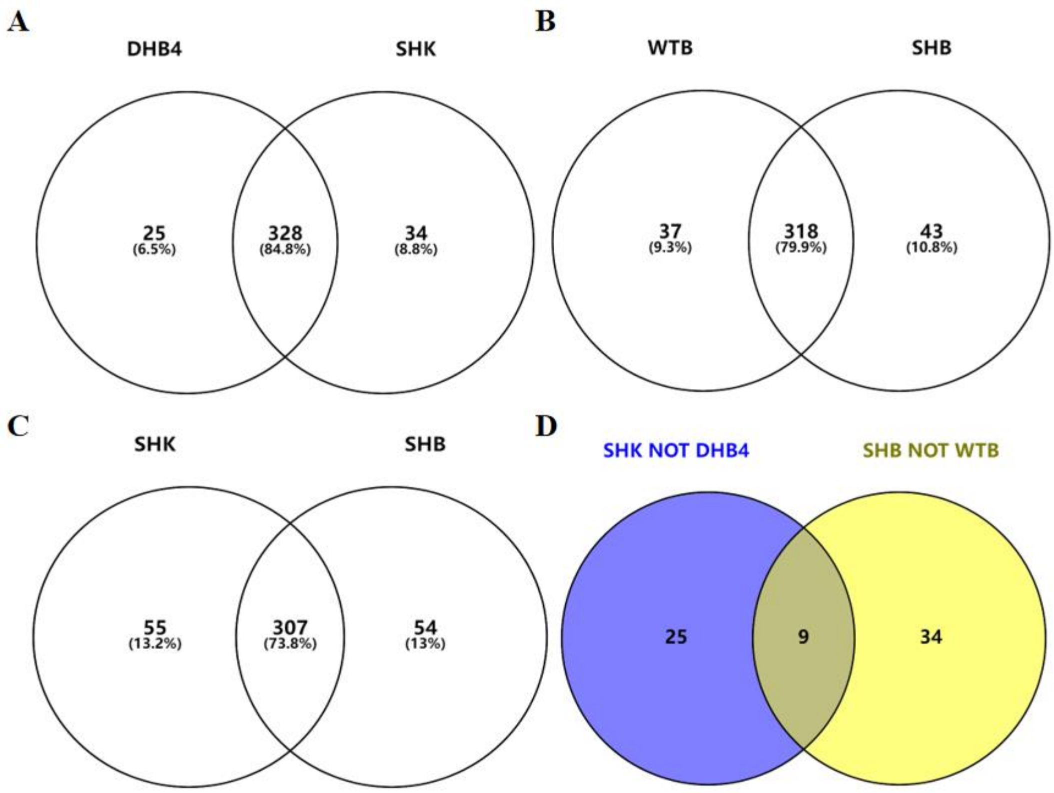
8.1.1 TraDIS data including the Illumina sequencing data and analysis output.

Available at:

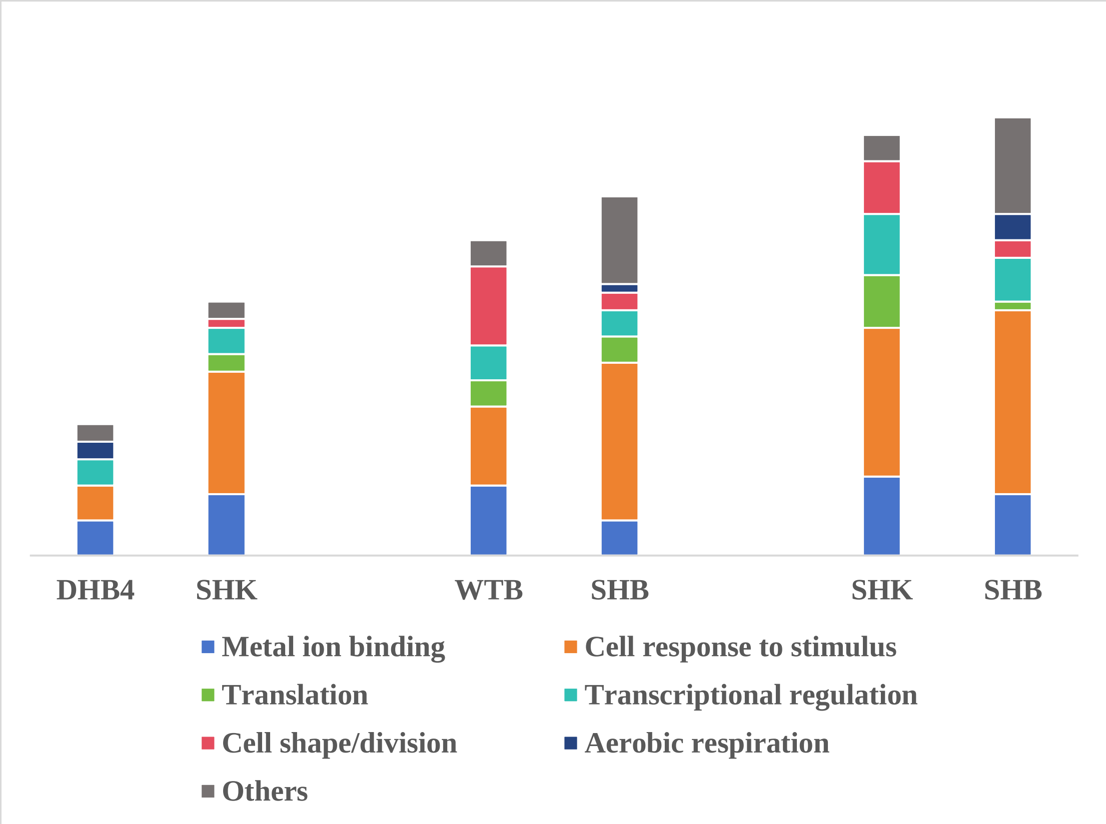
https://www.dropbox.com/scl/fo/eysqoctqjc6yvfi5gnlr7/APMb1EdLypm2SmbjQT9pSEI?rlkey=fti805tu1fwwd0lp3z69ybph9&st=6lumk9ts&dl=0

8.1.2 Manually checked final essential genes determined for four libaries (Table 8.1 below the text)

8.1.3 Comparison of the essential genes between the transposon libraries of A) DHB4 and SHuffle K-12 (SHK), B) T7 express (WTB) and SHuffle B (SHB), C)SHK and SHB, D) SHK but not in DHB4 and SHB but not in WTB.



8.1.4 Go enrichment analysis to reveal discrepancies in genes unique to each transposon libaries by two-way comparison in 8.1.3.



8.2 TraDIS experiments for SHuffle B transposon library grown at different temperatures.

8.2.1 TraDIS data including outputs and essential genes (no sequencing data).

Available at:
https://www.dropbox.com/scl/fo/ip92bo292yk5pp12nkrdm/ADBhh0LgmEgyNGzJDtHjgbk?rlkey=bsqroty9to11cbk20xczcq5ip&st=lyvs1r0e&dl=0

8.3 TraDIS experiments for MG1655 Δ*tig* , MG1655 Δ*dsbC* and MG1655 Δ*tig*Δ*dsbC*

8.3.1 TraDIS data including the Illumina sequencing data and analysis output.

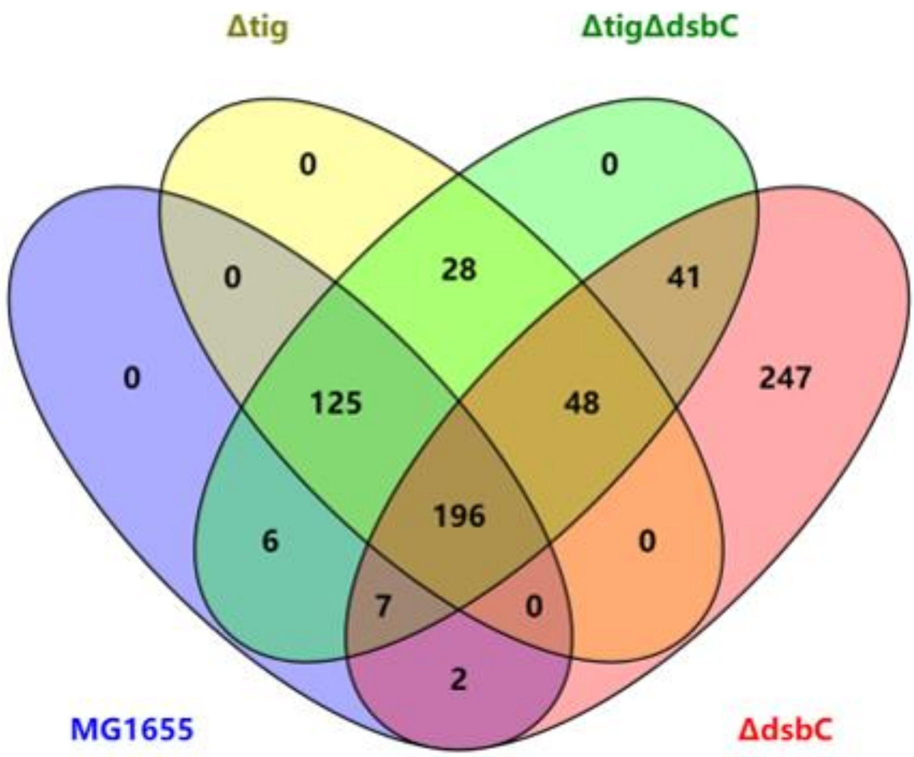
Available at:

https://www.dropbox.com/scl/fo/xswfs6of9e0plwl2dg3g5/AC_-ke10Jq9xnTJlUTm6Zyo?rlkey=trov64wzrhf97ho697t2efa3s&st=2h3c64lg&dl=0

8.3.2 Manually checked final essential genes determined for three libraries generated in this study and the MG1655 obtained from Dr. Milner (Table 8.2 below text).

8.3.3 Venn diagram showing 4-way comparison of essential genes determined in the transposon libraries of MG1655, Δ*tig*, Δ*dsbC* and Δ*tig*Δ*dsbC*.

The library of Δ*dsbC* was not used to analyze the potential function of DsbC due to the low density of transposon insertions in the population, therefore not comparable. However, all of the essential genes identified in the double mutant library were included either in the Δtig library or in the Δ*dsbC* library.



8.3.4 The functional and regulatory interactions among proteins encoding by unique essential genes in the library of Δ*tig* mutant.

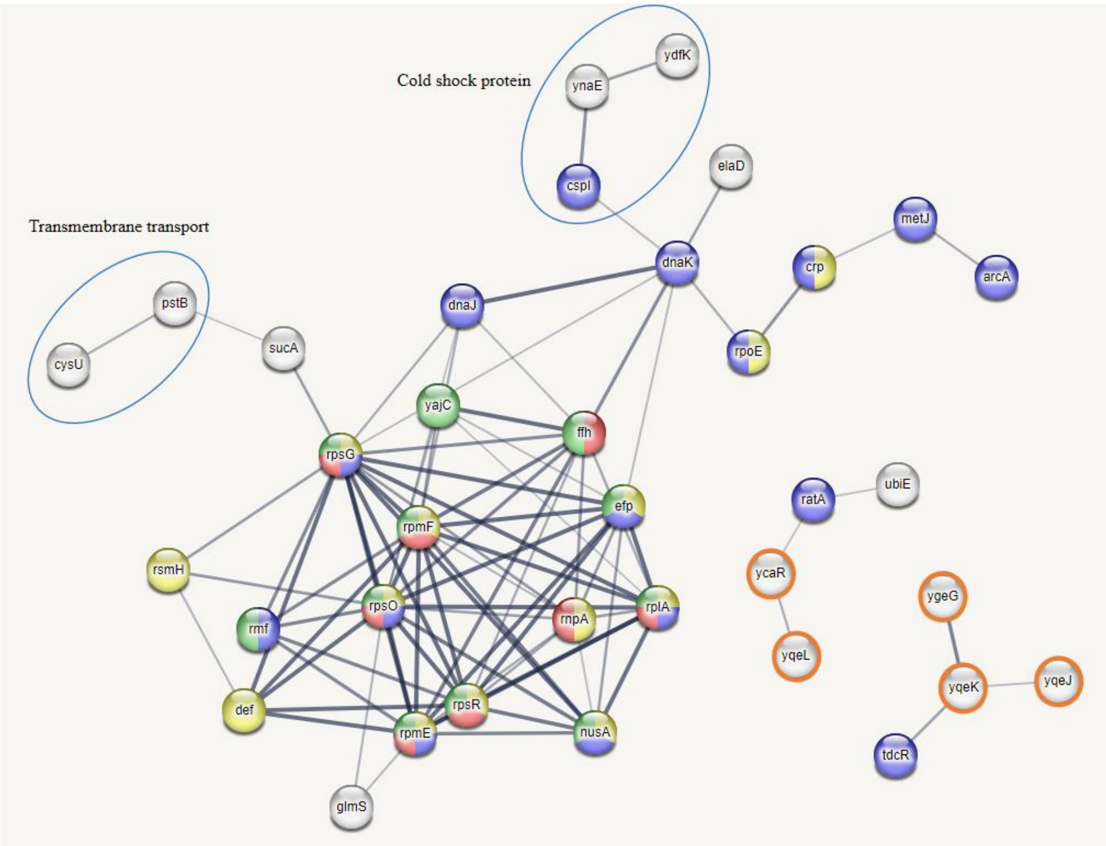


Figure 8.3.4 Physical or functional interactions among proteins encoding by conditional essential genes in library of *Δtig* compared to the wild type (MG1655) generated by STRING database. 3 independent clusters of genes were determined to be associated, the thickness of lines between genes indicates the confidence of the associations. No enriched terms were found for genes which nodes in gray color, cysU and pstB were identified to be involved in transmembrane transport inclusively. cspI, ynaE and ydfK were identified to encode cold shock proteins. In addition, genes involved in ribonucleoprotein complex (red), gene expression (yellow), regulation of gene expression (navy), translation and protein transport (green) were highlighted. Genes encoding for uncharacterized proteins were bordered in orange circle.

Table 8.1 Essential genes generated for libraries of SHuffle strains and their wild type strains

DHB4	SHuffle K-12	T7 Express	SHuffle B
accA	*accA*	*accA*	*accA*
accB	*accB*	*accB*	*accB*
accC	*accC*	*accC*	*accC*
accD	*accD*	*accD*	*accD*
ackA	*acpP*	*aceF*	*aceF*
acpP	*acpS*	*acpP*	*acpP*
acpS	*adk*	*acpS*	*acpS*
adk	*ahpC*	*adk*	*adk*
alaS	*alaS*	*alaS*	*ahpC*
argS	*argS*	*arcA*	*ahpF*
asd	*asd*	*argS*	*alaS*
asnS	*asnS*	*asd*	*arcA*
aspS	*aspS*	*asnS*	*argS*
bamA	*atpC*	*aspS*	*arnE*
bamD	*bamA*	*bamA*	*arnF*
birA	*bamD*	*bamD*	*asd*
can	*birA*	*can*	*asnS*
cdsA	*can*	*cca*	*aspS*
cmk	*cca*	*cdsA*	*bamA*
coaA	*cdsA*	*cmk*	*bamD*
coaD	*cmk*	*coaD*	*can*
coaE	*coaA*	*coaE*	*cca*
crp	*coaD*	*crp*	*cdsA*
csrA	*coaE*	*crr*	*cmk*
cydA	*crr*	*cydA*	*coaA*
cydB	*csrA*	*cydB*	*coaD*
cydX	*cysB*	*cydC*	*coaE*
cysB	*cysS*	*cydD*	*crp*
cysE	*dapA*	*cysB*	*crr*
cysS	*dapB*	*cysS*	*cspA*
dapA	*dapD*	*dapA*	*csrA*
dapB	*dapE*	*dapB*	*cydA*
dapD	*dcd*	*dapD*	*cydB*
dapE	*def*	*dapE*	*cydD*
dcd	*degS*	*def*	*cysB*
def	*der*	*der*	*cysE*
degS	*dfp*	*dfp*	*cysQ*
der	*dicA*	*dnaA*	*cysS*
dfp	*dnaA*	*dnaB*	*dapA*
dicA	*dnaC*	*dnaC*	*dapB*
dnaA	*dnaE*	*dnaE*	*dapD*
dnaB	*dnaG*	*dnaG*	*dapE*
dnaC	*dnaK*	*dnaK*	*def*

dnaE	*dnaN*	*dnaN*	*der*
dnaG	*dnaQ*	*dnaQ*	*dfp*
dnaK	*dnaT*	*dnaT*	*dksA*
dnaN	*dnaX*	*dnaX*	*dnaA*
dnaQ	*dut*	*dut*	*dnaB*
dnaT	*dxr*	*dxr*	*dnaC*
dnaX	*dxs*	*dxs*	*dnaE*
dut	*efp*	*efp*	*dnaG*
dxr	*eno*	*eno*	*dnaN*
dxs	*era*	*era*	*dnaQ*
efp	*erpA*	*erpA*	*dnaT*
eno	*fabA*	*fabA*	*dnaX*
era	*fabB*	*fabB*	*dut*
erpA	*fabD*	*fabD*	*dxr*
fabA	*fabG*	*fabG*	*dxs*
fabB	*fabH*	*fabH*	*eno*
fabD	*fabI*	*fabI*	*era*
fabG	*fabZ*	*fabZ*	*erpA*
fabH	*fbaA*	*fbaA*	*fabA*
fabI	*ffh*	*fdx*	*fabB*
fabZ	*ffs*	*ffh*	*fabD*
fbaA	*fldA*	*ffs*	*fabG*
fdx	*fmt*	*fldA*	*fabH*
ffh	*folA*	*fmt*	*fabI*
fldA	*folB*	*folA*	*fabZ*
fmt	*folC*	*folB*	*fbaA*
folA	*folD*	*folC*	*fdx*
folB	*folE*	*folD*	*ffh*
folC	*folP*	*folE*	*ffs*
folD	*frr*	*folK*	*fldA*
folE	*ftsA*	*folP*	*fmt*
folP	*ftsB*	*frr*	*folA*
frr	*ftsH*	*ftsA*	*folB*
ftsA	*ftsI*	*ftsB*	*folC*
ftsB	*ftsL*	*ftsH*	*folD*
ftsH	*ftsQ*	*ftsI*	*folE*
ftsI	*ftsW*	*ftsL*	*folP*
ftsL	*ftsY*	*ftsQ*	*frr*
ftsQ	*ftsZ*	*ftsW*	*ftsA*
ftsW	*fusA*	*ftsY*	*ftsB*
ftsY	*gapA*	*ftsZ*	*ftsH*
ftsZ	*glmM*	*fusA*	*ftsI*
fusA	*glmS*	*gapA*	*ftsQ*
gapA	*glmU*	*glmM*	*ftsW*
glmM	*glnS*	*glmS*	*ftsY*
glmS	*gltX*	*glmU*	*ftsZ*
glmU	*glyQ*	*glnS*	*fusA*
glnS	*glyS*	*gltX*	*gapA*
gltX	*gmhA*	*glyQ*	*glmM*

glyA	*gmk*	*glyS*	*glmS*
glyQ	*groL*	*gmhA*	*glmU*
glyS	*groS*	*gmhB*	*glnS*
gmhA	*grpE*	*gmk*	*gltX*
gmk	*grxA*	*groL*	*glyQ*
groL	*gshA*	*groS*	*glyS*
groS	*gyrA*	*grpE*	*gmhA*
grpE	*gyrB*	*gyrA*	*gmhB*
gyrA	*hda*	*gyrB*	*gmk*
gyrB	*hemA*	*hemA*	*groL*
hda	*hemB*	*hemB*	*groS*
hemA	*hemC*	*hemC*	*grpE*
hemB	*hemD*	*hemD*	*grxA*
hemC	*hemE*	*hemE*	*gyrA*
hemD	*hemG*	*hemG*	*gyrB*
hemE	*hemH*	*hemH*	*hemA*
hemG	*hemL*	*hemL*	*hemB*
hemH	*hipB*	*higA*	*hemC*
hemL	*hisS*	*hipB*	*hemE*
higA	*hldD*	*hisS*	*hemG*
hipB	*hldE*	*hldD*	*hemH*
hisS	*hns*	*hldE*	*hemL*
hldD	*holA*	*holA*	*hipB*
hldE	*holB*	*holB*	*hisS*
hns	*ileS*	*hscA*	*hldE*
holA	*infA*	*hscB*	*hns*
holB	*infB*	*icd*	*holA*
ileS	*infC*	*ileS*	*holB*
infA	*insC1*	*infA*	*hscA*
infB	*insE1*	*infB*	*hscB*
infC	*iscS*	*infC*	*icd*
insC1	*iscU*	*iscA*	*ihfB*
iscA	*ispA*	*iscS*	*ileS*
iscS	*ispB*	*iscU*	*infA*
iscU	*ispD*	*ispA*	*infB*
ispA	*ispE*	*ispB*	*infC*
ispB	*ispF*	*ispD*	*insE1*
ispD	*ispG*	*ispE*	*iscA*
ispE	*ispH*	*ispF*	*iscS*
ispF	*ispU*	*ispG*	*iscU*
ispG	*kdsA*	*ispH*	*ispA*
ispH	*kdsB*	*ispU*	*ispB*
ispU	*lacI*	*kdsA*	*ispD*
kdsA	*ldrC*	*kdsB*	*ispE*
kdsB	*lepB*	*ldcA*	*ispF*
lepB	*leuS*	*ldrB*	*ispG*
leuS	*lexA*	*lepB*	*ispH*
lexA	*lgt*	*leuS*	*ispU*
lgt	*ligA*	*lgt*	*kdsA*

ligA	*lipA*	*ligA*	*kdsB*
lnt	*lnt*	*lipA*	*kilR*
lolA	*lolA*	*lipB*	*lacI*
lolB	*lolB*	*lnt*	*ldrB*
lolC	*lolC*	*lolA*	*ldrC*
lolD	*lolD*	*lolB*	*lepB*
lolE	*lolE*	*lolC*	*leuS*
lpd	*lpd*	*lolD*	*lgt*
lptA	*lptA*	*lolE*	*ligA*
lptB	*lptB*	*lpd*	*lnt*
lptC	*lptD*	*lptA*	*lolA*
lptD	*lptE*	*lptB*	*lolB*
lptE	*lptF*	*lptC*	*lolC*
lptF	*lptG*	*lptD*	*lolD*
lptG	*lpxA*	*lptE*	*lolE*
lpxA	*lpxB*	*lptF*	*lpd*
lpxB	*lpxC*	*lptG*	*lptA*
lpxC	*lpxD*	*lpxA*	*lptB*
lpxD	*lpxH*	*lpxB*	*lptC*
lpxH	*lpxK*	*lpxC*	*lptD*
lpxK	*lpxL*	*lpxD*	*lptE*
lpxL	*lspA*	*lpxH*	*lptF*
lspA	*map*	*lpxK*	*lptG*
map	*metG*	*lpxL*	*lpxA*
metG	*metK*	*lspA*	*lpxB*
metJ	*minE*	*lysS*	*lpxC*
metK	*mnmA*	*map*	*lpxD*
minE	*mnmG*	*metG*	*lpxH*
mraY	*mraY*	*metK*	*lpxK*
mrdA	*mrdA*	*mnmA*	*lspA*
mrdB	*mrdB*	*mraY*	*lysS*
mreB	*mreB*	*mrdA*	*map*
mreC	*mreC*	*mrdB*	*marB*
mreD	*mreD*	*mreB*	*metG*
msbA	*msbA*	*mreC*	*metK*
mtn	*mtn*	*mreD*	*minE*
mukB	*mukB*	*msbA*	*mnmA*
mukE	*mukE*	*mtn*	*mraY*
mukF	*mukF*	*mukB*	*mrdB*
murA	*murA*	*mukE*	*msbA*
murB	*murB*	*mukF*	*msrB*
murC	*murC*	*murA*	*mtn*
murD	*murD*	*murB*	*mukB*
murE	*murE*	*murC*	*mukE*
murF	*murF*	*murD*	*mukF*
murG	*murG*	*murE*	*murA*
murI	*murI*	*murF*	*murB*
murJ	*murJ*	*murG*	*murC*
nadD	*nadD*	*murI*	*murD*

nadE	*nadE*	*murJ*	*murE*
nadK	*nadK*	*nadD*	*murF*
nrdA	*nfuA*	*nadE*	*murG*
nrdB	*nrdA*	*nadK*	*murI*
nuoI	*nrdB*	*nrdA*	*murJ*
nusA	*nusA*	*nrdB*	*nadD*
nusG	*nusB*	*nusA*	*nadE*
obgE	*nusG*	*nusB*	*nadK*
orn	*obgE*	*nusG*	*nohD*
parC	*orn*	*obgE*	*nrdA*
parE	*parC*	*orn*	*nrdB*
pdxH	*parE*	*pal*	*nusB*
pgk	*pdxH*	*parC*	*nusG*
pgsA	*pgk*	*parE*	*orn*
pheS	*pgsA*	*pdxH*	*oxyR*
pheT	*pheS*	*pgk*	*pal*
plsB	*pheT*	*pgsA*	*parC*
plsC	*plsC*	*pheS*	*parE*
ppa	*ppa*	*pheT*	*pdxH*
ppiB	*ppiB*	*pinQ*	*pgk*
prfA	*prfA*	*pinR*	*pgsA*
prfB	*prfB*	*plsB*	*pheS*
priA	*priB*	*plsC*	*pheT*
priB	*prmC*	*ppa*	*pinQ*
prmC	*proS*	*ppiB*	*pinR*
proS	*prs*	*prfA*	*plsB*
prs	*psd*	*prfB*	*plsC*
psd	*pssA*	*priA*	*ppa*
pssA	*pth*	*priB*	*ppiB*
pth	*ptsH*	*proS*	*prfA*
ptsH	*ptsI*	*prs*	*prfB*
ptsI	*purB*	*psd*	*priA*
pyrG	*pyrG*	*pssA*	*priB*
pyrH	*pyrH*	*pth*	*proS*
racR	*racR*	*ptsH*	*prs*
ratA	*rbfA*	*ptsI*	*psd*
relB	*relB*	*purB*	*pssA*
rho	*rho*	*pyrG*	*pstB*
ribA	*ribA*	*pyrH*	*pth*
ribC	*ribC*	*rho*	*ptsH*
ribD	*ribD*	*ribA*	*ptsI*
ribE	*ribE*	*ribC*	*pyrG*
ribF	*ribF*	*ribD*	*pyrH*
rimM	*rimM*	*ribE*	*rho*
rluD	*rlmE*	*ribF*	*ribA*
rne	*rnc*	*rimM*	*ribC*
rnpA	*rnpA*	*rnc*	*ribD*
rnpB	*rnpB*	*rne*	*ribE*
rplA	*rplA*	*rnpA*	*ribF*

rplB	*rplB*	*rnpB*	*rimM*
rplC	*rplC*	*rplA*	*rlmE*
rplD	*rplD*	*rplB*	*rnpA*
rplE	*rplE*	*rplC*	*rnpB*
rplF	*rplF*	*rplD*	*rplA*
rplI	*rplI*	*rplE*	*rplB*
rplJ	*rplJ*	*rplF*	*rplC*
rplK	*rplK*	*rplJ*	*rplD*
rplL	*rplL*	*rplK*	*rplE*
rplM	*rplM*	*rplL*	*rplF*
rplN	*rplN*	*rplM*	*rplJ*
rplO	*rplO*	*rplN*	*rplK*
rplP	*rplP*	*rplO*	*rplL*
rplQ	*rplQ*	*rplP*	*rplM*
rplR	*rplR*	*rplQ*	*rplN*
rplS	*rplS*	*rplR*	*rplO*
rplT	*rplT*	*rplS*	*rplP*
rplU	*rplU*	*rplT*	*rplQ*
rplV	*rplV*	*rplU*	*rplR*
rplW	*rplW*	*rplV*	*rplS*
rplX	*rplX*	*rplW*	*rplT*
rplY	*rplY*	*rplX*	*rplU*
rpmA	*rpmA*	*rplY*	*rplV*
rpmB	*rpmB*	*rpmA*	*rplW*
rpmC	*rpmC*	*rpmB*	*rplX*
rpmD	*rpmD*	*rpmC*	*rplY*
rpmE	*rpmE*	*rpmD*	*rpmA*
rpmG	*rpmG*	*rpmE*	*rpmB*
rpmH	*rpmH*	*rpmF*	*rpmC*
rpmI	*rpmI*	*rpmG*	*rpmD*
rpmJ	*rpmJ*	*rpmH*	*rpmE*
rpoA	*rpoA*	*rpmI*	*rpmH*
rpoB	*rpoB*	*rpmJ*	*rpmI*
rpoC	*rpoC*	*rpoA*	*rpmJ*
rpoD	*rpoD*	*rpoB*	*rpoA*
rpoE	*rpoE*	*rpoC*	*rpoB*
rpoH	*rpoH*	*rpoD*	*rpoC*
rppH	*rpsA*	*rpoH*	*rpoD*
rpsA	*rpsB*	*rpsA*	*rpoH*
rpsB	*rpsC*	*rpsB*	*rpsA*
rpsC	*rpsD*	*rpsC*	*rpsB*
rpsD	*rpsE*	*rpsD*	*rpsC*
rpsE	*rpsF*	*rpsE*	*rpsD*
rpsF	*rpsG*	*rpsF*	*rpsE*
rpsG	*rpsH*	*rpsG*	*rpsF*
rpsH	*rpsI*	*rpsH*	*rpsG*
rpsI	*rpsJ*	*rpsI*	*rpsH*
rpsJ	*rpsK*	*rpsJ*	*rpsI*
rpsK	*rpsL*	*rpsK*	*rpsJ*

rpsL	*rpsM*	*rpsL*	*rpsK*
rpsM	*rpsN*	*rpsM*	*rpsL*
rpsN	*rpsO*	*rpsN*	*rpsM*
rpsO	*rpsP*	*rpsP*	*rpsN*
rpsP	*rpsQ*	*rpsR*	*rpsP*
rpsQ	*rpsR*	*rpsS*	*rpsQ*
rpsR	*rpsS*	*rpsT*	*rpsS*
rpsS	*rpsT*	*rpsU*	*rpsT*
rpsT	*rpsU*	*rrf*	*rpsU*
rpsU	*rrf*	*safA*	*rrf*
rrf	*rseP*	*secA*	*safA*
rseP	*rsmH*	*secD*	*sapD*
ruvA	*sdhC*	*secE*	*sdhC*
secA	*secA*	*secF*	*secA*
secD	*secB*	*secY*	*secD*
secE	*secD*	*serS*	*secE*
secF	*secE*	*ssb*	*secF*
secM	*secF*	*ssrS*	*secY*
secY	*secM*	*sucA*	*serS*
serS	*secY*	*sucB*	*ssb*
ssb	*serS*	*sufA*	*ssrA*
sucA	*ssb*	*suhB*	*sucB*
sucB	*ssrA*	*tadA*	*suhB*
suhB	*sucA*	*tdcR*	*tdcR*
tadA	*sucB*	*thrS*	*thrS*
thiD	*suhB*	*thyA*	*thyA*
thiE	*tadA*	*tilS*	*tilS*
thiL	*thiD*	*tmk*	*tmk*
thrS	*thiE*	*tolA*	*tolA*
thyA	*thiL*	*tolB*	*tonB*
tilS	*thrS*	*tolQ*	*topA*
tmk	*thyA*	*tolR*	*trmD*
topA	*tilS*	*topA*	*trpS*
tpiA	*tmk*	*trmD*	*tsaB*
trmD	*topA*	*trpS*	*tsaC*
trpS	*tpiA*	*tsaB*	*tsaD*
tsaB	*trmD*	*tsaC*	*tsaE*
tsaC	*trpS*	*tsaD*	*tsf*
tsaD	*tsaB*	*tsaE*	*tusE*
tsaE	*tsaC*	*tsf*	*tyrS*
tsf	*tsaD*	*tusE*	*ubiA*
tusB	*tsaE*	*tyrS*	*ubiB*
tusE	*tsf*	*ubiA*	*ubiE*
tyrS	*tusB*	*ubiB*	*ubiG*
ubiA	*tusC*	*ubiD*	*ubiX*
ubiB	*tusE*	*ubiE*	*uof*
ubiD	*tyrS*	*ubiG*	*valS*
ubiE	*ubiA*	*ubiX*	*waaA*
ubiG	*ubiB*	*uof*	*wzyE*

ubiX	*ubiD*	*valS*	*yajC*
valS	*ubiE*	*waaA*	*ybcV*
waaA	*ubiF*	*waaC*	*ybcW*
wzyE	*ubiG*	*wzyE*	*ybeD*
xseB	*ubiH*	*ybfB*	*ycaR*
yajC	*ubiX*	*ycaR*	*yceQ*
ybeY	*uof*	*yciM*	*ydaE*
ycaR	*valS*	*ydaE*	*ydcD*
yejL	*waaA*	*yddM*	*ydcX*
yffS	*wzyE*	*ydfK*	*ydfK*
ygfZ	*yacG*	*ydhL*	*ydiE*
yidC	*yajC*	*ygfZ*	*yeaC*
ymiB	*ycaR*	*yidC*	*yebV*
yqcG	*ydfK*	*yihA*	*ygfZ*
yqgF	*ydhL*	*ymiB*	*yidC*
yrfF	*yejL*	*yncH*	*yihA*
	ygfZ	*yqgF*	*ylcI*
	yidC	*zipA*	*ymiB*
	yihA		*ymjC*
	ylcI		*yncH*
	ymiB		*yneM*
	ynaE		*yqcG*
	yqcG		*yqgF*
	yrfF		*zipA*
	zipA		

Table 8.2 Essential genes in the libraries of MG1655, Δ*tig* mutant, Δ*dsbC* mutant and Δ*tig*Δ*dsbC* double mutant

MG1655	*Δtig*	*ΔtigΔdsbC*	*ΔdsbC**
accA	*accA*	*accA*	*fabD*
accB	*accB*	*accB*	*hscB*
accC	*accC*	*accC*	*menI*
accD	*accD*	*accD*	*plsC*
acpP	*acpP*	*acpP*	*dxr*
acpS	*acpS*	*adk*	*dxs*
adk	*adk*	*alaS*	*rlmH*
alaS	*alaS*	*appY*	*ispF*
argS	*appY*	*arcA*	*era*
asd	*arcA*	*argS*	*rpsJ*
asnS	*argS*	*asd*	*rpsK*
aspS	*asd*	*asnS*	*rpsM*
bamA	*asnS*	*aspS*	*rpsN*
baxL	*aspS*	*atpF*	*rpsP*

birA	*azuC*	*atpG*	*rpsS*
can	*bamA*	*azuC*	*rpsT*
cca	*bamD*	*bamA*	*rpsU*
cdsA	*birA*	*bamD*	*rpsC*
coaA	*can*	*baxL*	*rpsD*
coaD	*cca*	*birA*	*rpsE*
coaE	*cdsA*	*can*	*rpsF*
csrA	*coaA*	*cca*	*rpsG*
cydA	*coaD*	*cdsA*	*rpsH*
cydB	*coaE*	*coaA*	*rpsI*
cydC	*crp*	*coaD*	*rutC*
cydD	*cspI*	*coaE*	*ubiF*
cysB	*csrA*	*crp*	*fabZ*
cysS	*cydA*	*crr*	*fabB*
dapA	*cydB*	*cspI*	*fabH*
dapB	*cydC*	*csrA*	*fabG*
dapD	*cydD*	*cydA*	*ispE*
dapE	*cysB*	*cydB*	*ispH*
dcd	*cysE*	*cydC*	*dapB*
der	*cysS*	*cydD*	*dapA*
dfp	*cysU*	*cydX*	*rplJ*
dicA	*dapA*	*cysB*	*rplK*
dnaA	*dapB*	*cysE*	*rplL*
dnaB	*dapD*	*cysS*	*rplM*
dnaC	*dapE*	*cysU*	*rplN*
dnaE	*dcd*	*dapB*	*rplP*
dnaG	*def*	*dapD*	*rplQ*
dnaN	*der*	*dapE*	*rplR*
dnaT	*dfp*	*dapF*	*rplS*
dnaX	*dicA*	*dcd*	*rplB*
dut	*dnaA*	*def*	*rplT*
dxr	*dnaB*	*der*	*rplU*
dxs	*dnaC*	*dfp*	*rplV*
eno	*dnaE*	*dicA*	*rplW*
era	*dnaG*	*dnaA*	*rplX*
erpA	*dnaJ*	*dnaB*	*rplY*
fabA	*dnaK*	*dnaC*	*rpmA*
fabB	*dnaN*	*dnaE*	*rpmB*
fabD	*dnaT*	*dnaG*	*rpmC*
fabG	*dnaX*	*dnaJ*	*rpmD*
fabH	*dut*	*dnaK*	*rpmF*
fabI	*dxr*	*dnaN*	*rpmG*
fabZ	*dxs*	*dnaT*	*rpmH*
fbaA	*efp*	*dnaX*	*rpmI*

fdx	*elaD*	*dut*	*rpmJ*
ffh	*eno*	*dxr*	*ykgO*
fldA	*era*	*dxs*	*rplD*
fmt	*erpA*	*efp*	*rplE*
folA	*fabA*	*elaD*	*rplF*
folB	*fabB*	*eno*	*der*
folC	*fabD*	*era*	*nudG*
folD	*fabG*	*erpA*	*rrfH*
folE	*fabH*	*fabA*	*rrfG*
folP	*fabI*	*fabB*	*rrfF*
frr	*fabZ*	*fabD*	*rrfD*
ftsA	*fbaA*	*fabH*	*rrfC*
ftsB	*fdx*	*fabI*	*rrfA*
ftsH	*ffh*	*fbaA*	*rrfB*
ftsI	*ffs*	*ffh*	*ribE*
ftsL	*fldA*	*ffs*	*ssrS*
ftsQ	*fliX*	*fldA*	*accA*
ftsW	*fmt*	*flgO*	*yqgB*
ftsY	*folA*	*fliX*	*msyB*
ftsZ	*folB*	*folA*	*acpP*
fusA	*folC*	*folB*	*adk*
gapA	*folD*	*folC*	*alaS*
glmM	*folE*	*folD*	*iraM*
glmU	*folK*	*folE*	*flgM*
glnS	*folP*	*folK*	*higA*
gltX	*frr*	*folP*	*hipB*
glyA	*ftsA*	*frr*	*lnt*
glyQ	*ftsB*	*ftsA*	*argS*
glyS	*ftsH*	*ftsB*	*asnS*
gmk	*ftsI*	*ftsH*	*asd*
gnsB	*ftsL*	*ftsI*	*aspS*
groL	*ftsQ*	*ftsL*	*atpD*
groS	*ftsW*	*ftsQ*	*atpG*
grpE	*ftsY*	*ftsW*	*atpB*
guaA	*ftsZ*	*ftsY*	*atpF*
gyrA	*fusA*	*ftsZ*	*atpE*
gyrB	*gapA*	*fusA*	*dnaN*
hda	*glmM*	*gapA*	*fabA*
hemA	*glmS*	*glmM*	*spoT*
hemB	*glmU*	*glmS*	*ubiE*
hemC	*glnS*	*glmU*	*folC*
hemD	*gltX*	*glnS*	*ribF*
hemE	*glyA*	*gltX*	*accC*
hemG	*glyQ*	*glyA*	*pgsA*

hemH	*glyS*	*glyQ*	*cdsA*
hemL	*gmk*	*glyS*	*ftsB*
higA	*gnsB*	*gmk*	*ftsZ*
hipB	*groL*	*gnsB*	*yqcG*
hisS	*groS*	*groL*	*mreC*
holA	*grpE*	*groS*	*mreD*
holB	*gyrA*	*grpE*	*cbpM*
igaA	*gyrB*	*guaA*	*groL*
ileS	*hemA*	*gyrA*	*mukE*
infA	*hemB*	*gyrB*	*mukF*
infB	*hemC*	*hda*	*cobS*
infC	*hemD*	*hemA*	*btuR*
iraM	*hemE*	*hemB*	*ydfK*
iroK	*hemG*	*hemC*	*cspI*
iscS	*hemH*	*hemD*	*pyrG*
iscU	*hemL*	*hemG*	*cydH*
ispA	*higA*	*hemH*	*cydX*
ispB	*hipB*	*hemL*	*appX*
ispD	*hisS*	*higA*	*murF*
ispE	*holA*	*hipB*	*alsE*
ispF	*holB*	*hisS*	*dcd*
ispG	*igaA*	*holA*	*coaE*
ispH	*ileS*	*holB*	*dapF*
ispU	*infA*	*igaA*	*folA*
kdsA	*infB*	*ileS*	*sucB*
kdsB	*infC*	*infA*	*ispU*
lapB	*iraM*	*infB*	*dinI*
ldrB	*iroK*	*infC*	*yebG*
lepB	*iscS*	*iraM*	*ligA*
leuS	*iscU*	*iroK*	*dnaE*
lexA	*ispA*	*iscS*	*dnaG*
lgt	*ispB*	*iscU*	*dnaC*
ligA	*ispD*	*ispA*	*appY*
lipA	*ispE*	*ispB*	*tdcR*
lnt	*ispF*	*ispE*	*arcA*
lolA	*ispG*	*ispF*	*dicA*
lolB	*ispH*	*ispH*	*hns*
lolC	*ispU*	*ispU*	*sutR*
lolD	*kdsA*	*kdsA*	*lexA*
lolE	*kdsB*	*kdsB*	*rcbA*
lpd	*kilR*	*kilR*	*yqjC*
lptA	*lapB*	*lapA*	*ydaF*
lptB	*lepB*	*lapB*	*ybcV*
lptC	*leuS*	*ldrB*	*ycgX*

lptD	*lexA*	*ldrC*	*ydfO*
lptE	*lgt*	*lepB*	*yjjZ*
lptF	*ligA*	*leuS*	*bhsA*
lptG	*lnt*	*lexA*	*mcbA*
lpxA	*lolA*	*lgt*	*yebV*
lpxB	*lolB*	*ligA*	*ynjH*
lpxC	*lolC*	*lipA*	*yggU*
lpxD	*lolD*	*lnt*	*YqgE*
lpxH	*lolE*	*lolA*	*yfiM*
lpxK	*lpd*	*lolB*	*yecH*
lpxL	*lptA*	*lolC*	*ybaM*
lspA	*lptB*	*lolD*	*ygdB*
lysS	*lptC*	*lolE*	*ybfA*
map	*lptD*	*lpd*	*yhfG*
metG	*lptE*	*lptA*	*yaaY*
metK	*lptF*	*lptB*	*yceQ*
mnmA	*lptG*	*lptC*	*yqgD*
mraY	*lpxA*	*lptD*	*yaiZ*
mrdA	*lpxB*	*lptE*	*yecJ*
mrdB	*lpxC*	*lptF*	*yfcZ*
mreB	*lpxD*	*lptG*	*yeaL*
mreC	*lpxH*	*lpxA*	*yqcC*
msbA	*lpxK*	*lpxB*	*ybeD*
mukB	*lpxL*	*lpxC*	*yncH*
mukE	*lspA*	*lpxD*	*ypjC*
mukF	*map*	*lpxH*	*dut*
murA	*metG*	*lpxK*	*fusA*
murB	*metJ*	*lpxL*	*ridA*
murC	*metK*	*lspA*	*ralR*
murD	*mgrB*	*lysS*	*chpB*
murE	*micC*	*map*	*ybeY*
murF	*mliC*	*metG*	*eno*
murG	*mraY*	*metJ*	*fabI*
murI	*mrdA*	*metK*	*wzyE*
murJ	*mrdB*	*mgrB*	*xseB*
nadD	*mreB*	*micA*	*crcB*
nadE	*mreC*	*micC*	*ydiT*
nadK	*mreD*	*mliC*	*hemH*
nrdA	*msbA*	*mnmA*	*flgN*
nrdB	*mukB*	*mraY*	*fliQ*
nusB	*mukE*	*mrdA*	*flhE*
nusG	*mukF*	*mrdB*	*fldA*
orn	*murA*	*mreB*	*fdnI*
parC	*murB*	*mreC*	*cyaY*

parE	*murC*	*mreD*	*ribD*
pdxH	*murD*	*msbA*	*ispA*
pgk	*murE*	*mukB*	*ydgC*
pgsA	*murF*	*mukE*	*glnS*
pheM	*murG*	*mukF*	*nrdH*
pheS	*murI*	*murA*	*gapA*
pheT	*murJ*	*murB*	*glyQ*
plsB	*nadD*	*murC*	*glyS*
plsC	*nadE*	*murD*	*ribA*
polA	*nadK*	*murE*	*yihA*
ppa	*nrdA*	*murF*	*gmk*
ppiB	*nrdB*	*murG*	*hlyE*
prfA	*nusA*	*murI*	*hisL*
prfB	*nusB*	*murJ*	*acpS*
priB	*nusG*	*nadD*	*gloB*
prmC	*orn*	*nadE*	*ilvL*
proS	*parC*	*nadK*	*ybfQ*
prs	*parE*	*nrdA*	*mliC*
pssA	*pbl*	*nrdB*	*ppa*
pth	*pdxH*	*nusA*	*ihfA*
purB	*pgk*	*nusB*	*pqiC*
pyrG	*pgsA*	*nusG*	*rseP*
pyrH	*pheL*	*orn*	*ymfR*
racR	*pheM*	*parC*	*erpA*
rcbA	*pheS*	*parE*	*insA2*
rdlC	*pheT*	*pbl*	*insA3*
relB	*plsB*	*pdxH*	*ileS*
rho	*plsC*	*pgk*	*gloA*
ribA	*polA*	*pgsA*	*mnaT*
ribC	*ppa*	*pheL*	*ulaB*
ribD	*ppiB*	*pheS*	*lpxL*
ribE	*prfA*	*pheT*	*yddK*
ribF	*prfB*	*plsB*	*fucU*
rimM	*priB*	*plsC*	*lpd*
rluD	*prmC*	*polA*	*lapB*
rnt	*proS*	*ppa*	*lptD*
rplB	*prs*	*ppiB*	*lptE*
rplC	*psaA*	*prfA*	*lptF*
rplD	*psd*	*prfB*	*lptG*
rplE	*pssA*	*priB*	*lolE*
rplF	*pstB*	*prmC*	*lspA*
rplJ	*pth*	*proS*	*tsaC*
rplK	*purB*	*prs*	*manA*
rplL	*pyrG*	*psaA*	*yidC*

rplM	*pyrH*	*psd*	*yidD*
rplN	*racR*	*pssA*	*tisB*
rplO	*ratA*	*pssL*	*map*
rplP	*rcbA*	*pstB*	*yeaK*
rplQ	*rdlB*	*pth*	*moaD*
rplR	*relB*	*purB*	*hicA*
rplS	*rho*	*pyrG*	*relE*
rplT	*ribA*	*pyrH*	*mdtJ*
rplU	*ribC*	*racC*	*marB*
rplV	*ribD*	*racR*	*tsaD*
rplW	*ribE*	*ratA*	*tsaE*
rplX	*ribF*	*rcbA*	*nadK*
rplY	*rimM*	*rdlB*	*nuoK*
rpmA	*rluD*	*rho*	*napD*
rpmB	*rmf*	*ribA*	*ybiA*
rpmC	*rnpA*	*ribC*	*nadD*
rpmD	*rnpB*	*ribD*	*glnK*
rpmH	*rnt*	*ribF*	*lolA*
rpmI	*rplA*	*rimM*	*lolB*
rpmJ	*rplB*	*rlmE*	*slyB*
rpoA	*rplC*	*rluD*	*bamA*
rpoB	*rplD*	*rmf*	*bamD*
rpoC	*rplE*	*rnc*	*prfA*
rpoD	*rplF*	*rnpA*	*prfB*
rpoE	*rplJ*	*rnpB*	*def*
rpoH	*rplK*	*rnt*	*ftsW*
rpsA	*rplL*	*rplA*	*pth*
rpsB	*rplM*	*rplB*	*cutC*
rpsC	*rplN*	*rplC*	*ycaR*
rpsD	*rplO*	*rplD*	*yfbV*
rpsE	*rplP*	*rplE*	*ybdD*
rpsF	*rplQ*	*rplF*	*ycgL*
rpsH	*rplR*	*rplJ*	*ycfP*
rpsI	*rplS*	*rplM*	*ybaY*
rpsJ	*rplT*	*rplQ*	*ydiE*
rpsK	*rplU*	*rplR*	*ypfN*
rpsL	*rplV*	*rplS*	*pspB*
rpsM	*rplW*	*rplT*	*pspD*
rpsN	*rplX*	*rplU*	*pheL*
rpsP	*rplY*	*rplV*	*paaD*
rpsQ	*rpmA*	*rplW*	*pheS*
rpsS	*rpmB*	*rplX*	*pheT*
rpsT	*rpmC*	*rplY*	*pheM*
rpsU	*rpmD*	*rpmA*	*mgrB*

rseP	*rpmE*	*rpmB*	*pstB*
safA	*rpmF*	*rpmC*	*lgt*
sdsN	*rpmH*	*rpmD*	*pssA*
secA	*rpmI*	*rpmE*	*glmM*
secD	*rpmJ*	*rpmF*	*yqaE*
secE	*rpoA*	*rpmH*	*proS*
secF	*rpoB*	*rpmI*	*torI*
secY	*rpoC*	*rpmJ*	*efp*
serS	*rpoD*	*rpoA*	*tsf*
ssb	*rpoH*	*rpoB*	*evgL*
sucB	*rpsA*	*rpoC*	*gnsB*
suhB	*rpsB*	*rpoD*	*iroK*
tadA	*rpsC*	*rpoH*	*pssL*
thiL	*rpsD*	*rpsA*	*racC*
thrS	*rpsE*	*rpsC*	*rem*
thyA	*rpsF*	*rpsD*	*secA*
tilS	*rpsG*	*rpsF*	*yabR*
tmk	*rpsH*	*rpsG*	*yacM*
topA	*rpsI*	*rpsJ*	*yadW*
trmD	*rpsJ*	*rpsK*	*yadX*
trpL	*rpsK*	*rpsL*	*yahV*
trpS	*rpsL*	*rpsO*	*yaiA*
tsaB	*rpsM*	*rpsQ*	*ybgV*
tsaC	*rpsN*	*rpsR*	*ydcA*
tsaD	*rpsO*	*rpsS*	*ydfR*
tsaE	*rpsP*	*rpsT*	*yfiS*
tsf	*rpsQ*	*rpsU*	*ygeI*
ttcC	*rpsR*	*rseP*	*yicU*
tyrS	*rpsS*	*rseX*	*ykfO*
ubiA	*rpsT*	*rsgA*	*ykiE*
ubiB	*rpsU*	*rsmH*	*ylcJ*
ubiD	*rseP*	*safA*	*yldA*
ubiG	*rseX*	*sdsN*	*yliM*
ubiX	*rsmH*	*secA*	*yljB*
valS	*safA*	*secD*	*ymdG*
waaA	*secA*	*secE*	*ymgC*
wzyE	*secD*	*secF*	*ymgK*
ybeD	*secE*	*secY*	*ymgL*
ybfB	*secF*	*seqA*	*ymgN*
ydaE	*secY*	*serS*	*ymiC*
yddK	*serS*	*sroE*	*ymiD*
ydfO	*ssb*	*ssb*	*ynaM*
ydiE	*sucA*	*sucA*	*ynaN*
yffS	*sucB*	*sucB*	*yneP*

ygeI	*suhB*	*suhB*	*ynfN*
ygfZ	*tadA*	*tadA*	*ynfP*
yidC	*tdcR*	*tdcR*	*ynfQ*
yihA	*tff*	*tff*	*ynfR*
ykfM	*thrS*	*thiL*	*yoaI*
ymfk	*thyA*	*thrS*	*yoaL*
ymiB	*tilS*	*thyA*	*yoaM*
ynbG	*tmk*	*tilS*	*ypaB*
yncH	*topA*	*tktA*	*yqeJ*
yobI	*trmD*	*tmk*	*yqeK*
yqcG	*trpL*	*topA*	*yqfI*
yqfI	*trpS*	*tpiA*	*yqgC*
yqgF	*tsaB*	*trmD*	*yqgG*
zipA	*tsaC*	*trpS*	*yqhI*
	tsaD	*tsaB*	*yqiM*
	tsaE	*tsaC*	*yrfJ*
	tsf	*tsaD*	*ysaE*
	ttcC	*tsaE*	*ytcB*
	tyrS	*tsf*	*ytgB*
	ubiA	*ttcC*	*ythB*
	ubiB	*tusE*	*ytiE*
	ubiD	*tyrS*	*ytjE*
	ubiE	*ubiA*	*prmC*
	ubiG	*ubiB*	*hemG*
	ubiX	*ubiD*	*ttcC*
	valS	*ubiE*	*psaA*
	waaA	*ubiG*	*ygeF*
	wzyE	*ubiX*	*pbl*
	yabR	*valS*	*ygeL*
	yacM	*waaA*	*sgrT*
	yahV	*wzyE*	*ynjD*
	yajC	*yabR*	*rdlA*
	ybfB	*yacM*	*rdlB*
	ycaR	*yadW*	*rdlC*
	ychT	*yahV*	*yhcO*
	yciE	*yajC*	*bcsF*
	ydaE	*ybeD*	*dicC*
	ydcD	*ybfB*	*ymfT*
	yddK	*ybfQ*	*nohD*
	ydfK	*ycaR*	*eutN*
	ydfO	*yccE*	*fixX*
	yfdF	*yceQ*	*yehC*
	yffS	*ychT*	*ycbV*

yfjW	*yciE*	*insL3*
ygeF	*yciG*	*yifL*
ygeG	*ydaE*	*yagP*
ygeI	*ydcD*	*ykgR*
ygeL	*yddK*	*ybcO*
yidC	*ydeO*	*ynfG*
yihA	*ydfK*	*nohA*
ykfM	*ydfO*	*rzoR*
ylcJ	*ydiE*	*ymiB*
ymfE	*yejL*	*ymfK*
ymfK	*yfdF*	*ymfD*
ymgC	*yffS*	*sokB*
ymgK	*yfjW*	*narS*
ymgL	*ygeF*	*tff*
ymiB	*ygeG*	*yjbJ*
ynaE	*ygeI*	*ibsB*
ynbG	*ygeL*	*argL*
yncH	*ygeN*	*baxL*
ynfR	*ygfZ*	*ariR*
ynfU	*yidC*	*ygeN*
yoaM	*yihA*	*yddL*
yobI	*yjbL*	*yeeW*
ypjC	*yjbS*	*yzgL*
yqcG	*ykfM*	*pdxH*
yqeJ	*ylcJ*	*dicF*
yqeK	*ymfD*	*kilS*
yqeL	*ymfE*	*grxA*
yqgF	*ymfK*	*dnaB*
ytiC	*ymfk*	*rnhA*
ytjE	*ymgC*	*yqgF*
zipA	*ymgI*	*prs*
	ymgK	*rimM*
	ymgL	*rpe*
	ymjC	*rpoD*
	ynaE	*rpoA*
	ynaM	*rpoB*
	ynbG	*rnpB*
	yncH	*uof*
	yncL	*secD*
	yncO	*secF*
	ynfN	*yajC*
	ynfQ	*secE*
	ynfR	*secG*

ynfU	*secY*
yoaM	*secM*
yobI	*glyA*
ypdK	*aroK*
ypjC	*lepB*
yqcG	*ffs*
yqeJ	*ffh*
yqeK	*pmrD*
yqeL	*ryeA*
yqfI	*mgtS*
yqgF	*3ETSleuZ*
ytiC	*sibC*
ytjE	*chiX*
zipA	*chiZ*
zwf	*dsrA*
dmsD	*glmY*
ykgP	*ispZ*
acpS	*istR*
dapA	*micA*
fabG	*micC*
fabZ	*micF*
fdx	*micL*
fmt	*omrA*
hemE	*rseX*
ispD	*rydB*
ispG	*rydC*
pheM	*sdhX*
relB	*sdsN*
ribE	*sdsR*
rplK	*spf*
rplL	*zbiJ*
rplN	*agrA*
rplO	*bhsB*
rplP	*eyeA*
rpsB	*flgO*
rpsE	*fliX*
rpsH	*ryjB*
rpsI	*sibE*
rpsM	*sokE*
rpsN	*sraB*
rpsP	*sroA*
trpL	*sroE*
ymiB	*hokA*

ldrA
ldrB
ldrC
ldrD
rsxB
ssb
yciG
ychH
sdhC
thiS
tusA
tusB
lpxK
yibG
thrL
thyA
tmk
ssrA
tnaC
ibsC
ibsE
ykfI
ygeG
nusB
cspE
rho
infA
tadA
trmD
mnmA
trpL
safA
ubiB
lpxC
murB
murD
yoaH
pyrH
azuC
ilvX
ybfB
ybfC
ybjH

yccE
ydfB
ydfC
yehK
yhaL
yjbL
yjbS
yjeV
ykfM
ymfE
ymjD
ynaE
ynbG
yncL
yobI
yoeI
ypfM
yqeL
yshB
ythA
ybjG
yciZ
valS
yefM
ygdT
ydaE
minC
minE
crr
lapA
rlmE
rnc
rsgA
seqA
tktA
tpiA
tusE
ydeO
yejL
ymgI
ymjC
yncO
ypdK

zwf
dmsD
ykgP

*The low insertion density of the Δ*dsbC* library resulted in overestimation of the essential genes.