Exploring the evolvability of resistance determinants in bacteria

A dissertation submitted in partial fulfilment for the degree of

Doctor of Philosophy

In Biochemistry

At Massey University, Albany New Zealand

Paulina Hanson-Manful

-2013-

ABSTRACT

Antibiotic resistance poses a major risk to human and animal health. It has impacted on morbidity and mortality rates and has increased hospitalisation costs worldwide. To stand a fighting chance in this losing battle, it is important to understand the biochemical and evolutionary pathways that give rise to resistance. The goals of this thesis were to gain mechanistic insights into new resistance pathways and to explore whether protein functions associated with resistance could be evolved artificially.

In the first part of this thesis, I report that many pre-existing gene products in a non-pathogenic bacterium (*E. coli*) are able to impart resistance towards different classes of antibiotics. Investigating the relationship between fitness and resistance revealed that it was complex and as a result, it was difficult to predict how resistance would impact on bacterial fitness. It is apparent that the reservoir of resistance elements in non-pathogenic bacteria is much larger than previously appreciated, and may provide a rich source of resistance genes that could be co-opted by pathogenic bacteria.

In the second part of the thesis, specific examples were selected to explore the evolvability of resistance functions. Weak resistance activities of carbonic anhydrase and three proteins of unknown native function (YeaD, YdfW and YejG) could not be improved by directed evolution experiments in this study. Nonetheless, a novel multi-step route to high level tobramycin resistance involving YejG was discovered. Over-expression of YejG in the presence of a chromosomal mutation in the *fusA* gene (G1478T) endowed the bacterium with maximal resistance to tobramycin. In addition, a function of YejG in mediating an early entry into the log phase of growth has been uncovered.

Overall, the results in this thesis have revealed that there are even more pathways to resistance than previously appreciated. Weak resistance functions may not necessarily be easily evolvable, however, they may facilitate the emergence of mutations that confer higher levels of resistance. Ultimately, it is hoped that gaining a deeper insight into resistance pathways and how they evolve will help in the development of the new drugs that we so desperately need.

[Blank page]

Dedicated to my Father, Mr Alex Hanson-Manful, thank you for being my source of inspiration and for giving me the opportunity and investing in me all these years.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my Lord and Saviour Jesus Christ, without whom none of this would have been possible. Secondly I would like to thank my lovely, kind and patient husband, Danny, for his tremendous help and for always putting a smile on my face and making this journey as smooth and stress free as it was. I also extend my gratitude to my lovely parents, Alex and Helena, for their support (both emotional and financial) and to my siblings, Richard, Erasmus and Albert for always reminding me to laugh. I also thank my extended family (the Mwiinga's, Josephine, Emmanuel and El Shaddai) for their love and support.

I thank my awesome supervisor, Dr. Wayne Patrick, for helping me pursue my passion in science and for making this experience enjoyable. His kindness (always buying us meals), constructive feedback and willingness to help his students excel have taught me more than just science. His incredible skill of recognising and selecting interesting, talented but insane students to join our lab massively contributed to this fun and exciting journey. I would also like to thank his wife Dr. Monica Gerth for always being willing to help me whenever I needed her. I thank all of the Patrick lab members (both current and old: Ilana, Ankita, Hayden, Susan, Danni, Mack, Matteo and Matilda) for allowing me to forget that I was doing a PhD because we had so much fun in the lab. I can honestly say that there was not one dull moment during my studies. My experience in New Zealand would have simply not been the same without their friendship, advice and help. I would also like to thank Laura Nigon and Natasha Le Roux for their kindness, support and friendship and Dr. Valerie Soo for her time and help during my PhD research.

I thank my secondary supervisor, Prof. Paul Rainey, and my adopted supervisor, Dr. Austen Ganley, for making time when I needed their help and for their contribution. I also thank Prof. Greg Cook (University of Otago), Dr. Andrew Cridge (University of Otago) and Dr. Heather Hendrickson (Massey University) for helping me with the membrane potential, ribosome interaction and phage transduction experiments, respectively. I am grateful to Laura Nigon, Ilana Gerber and Susan Morton for providing technical assistance for the fitness experiments.

I thank everyone in Building 9 and 11 for their friendship and especially to all those who sacrificed their time to read my thesis (you know who you are), I really appreciate it. A big thank you to Dr. Peter Lind for the time he spent with me discussing my thesis. I also thank the Institute Administrators Muharram, Vesna and Colleen for helping me out with various paper works.

Finincial support for my PhD is from the Royal Society of New Zealand's Marsden Fund/Maurice & Phyllis Paykel Trust/Lottery Health Research Committee. Travel grants were from the NZ Society for Biochemistry and Molecular Biology, the Ministry of Science and Innovation and JSPS (Japan Society for the Promotion of Science). I am grateful to these institutions/societies for funding me.

Finally, I would like to thank all of my friends (all over the world) for their support and encouragement. I know that this journey would not have been the same without any one of you.

Contents

List of illustrations

List of Tables

Abbreviations

Abbreviation	Description
ASKA	A complete Set of *Escherichia coli* K-12 Open Reading Frame Achive
CA	Carbonic anhydrase
Cfu	Colony-forming unit
DMSO	Dimethyl sulfoxide
EDTA	Ethylenediaminetetraacetic acid
epPCR	Error-prone PCR
GFP	Green fluorescent protein
HGT	Horizontal gene transfer
His_6	Hexa-histidine
IPTG	β-D-1-thiogalactopyranoside
IS	Insertion sequences
lacZ	β-galactosidase
LB	Lysogeny broth
metC	Cystathionine β-lyase
MIC	Minimum Inhibitory Concentration
NoIns	pCA24N plasmid without any insert (negative control plasmid)
OD_{600}	Optical density measured at 600 nm
SE	Standard Error
SOC	Super Optimal Broth with Catabolite repression
Tris	Tris(hydroxymethyl)aminomethane
W	Relative fitness
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

[Blank page]

CHAPTER 1

Introduction

Background

Since the introduction of antibiotics into clinical settings in the 20th century, we have been able to fight against many diseases and infections caused by bacteria (e.g., wound infections, septicaemia and tuberculosis). The accidental discovery of penicillin by Sir Alexander Fleming in 1929 and its use in the treatment of infections during the Second World War revolutionised modern medicine and paved the way for the development of many more antibiotics. However, over the years the success of treatment against infections has been jeopardised by the emergence of antibiotic resistance.

Currently, many bacteria that were once effectively managed with antibiotics are now major threats in human diseases; about 70 % of the bacteria that cause infections in hospitals are resistant to at least one of the antibiotics commonly used for treatment (Bisht et al., 2009). Further impacting on the problem is the existence of multidrug resistance (MDR) micro-organisms, which include *Klebsiella pneumonia*, vancomycin resistant enteroccoci (VRE) and methicillin-resistant *Staphylococcus aureus* (MRSA) (Alekshun & Levy, 2007). We are now faced with a long list of microbes that have more ways to evade antibiotics than the therapeutic interventions themselves.

1.1 Antibiotics

1.1.1 The discovery and usage of antibiotics worldwide

From 1890 to 1930, the main human pathogens (bacteria, fungi and viruses) had been isolated and identified and the search for agents that would target and kill these microbes began. Success came with the discovery of the first anti-microbial agent, arsphenamine (Salvarsan; anti-syphilitic agent). This drug was discovered in Paul Ehrlich's lab in the early 1900s after they screened many chemical compounds with the hope of discovering one that possessed anti-microbial activity. This was followed by the discovery of penicillin (1928) and sulphonamides (1932) (Davies, 2007), and from there, the rapid discovery and development of anti-microbial compounds ensued (Figure 1).

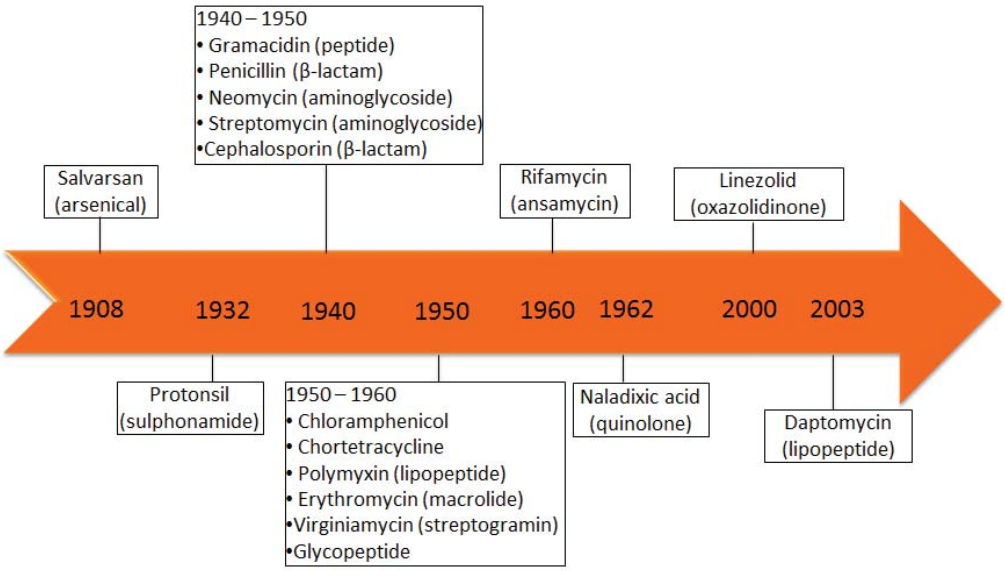


Figure 1. The history of anti-microbial agents (redrawn from Wright, 2007)

To date, there are many antibiotics available to us that inhibit bacterial DNA replication and repair, cell wall synthesis and protein synthesis (Figure 2). Anti-bacterial drugs target cell structures and biological processes that are different to those in human cells. They include the cell wall, 70 S ribosome and DNA biosynthesis (Walsh, 2000). Just to mention a few, there are more than 50 penicillins, 70 cephalosporins, 12 tetracyclines and 8 aminoglycosides (Neu, 1992). However, the pace for the development and release of new drugs has slowed dramatically. Daptomycin (belonging to the lipopeptide class of

antibiotics) was the last new antibiotic to be released into the market (in 2003), however, this drug was discovered in the 1980s (Wright, 2007).

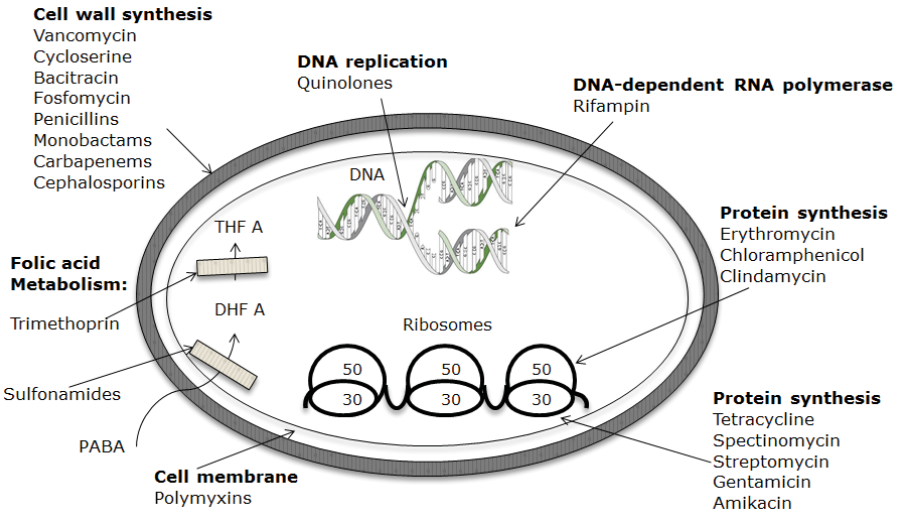


Figure 2. Sites of action of anti-microbial agents (modified from Neu, 1992)

Antibiotic targets of various drugs; mRNA, messenger RNA; tRNA, transfer RNA; PABA, *p*-aminobenzoic acid; DHFA, dihydrofolic acid; THFA, tetrahydrofolic acid.

Today, major pharmaceutical companies have reduced their efforts for finding new antibiotics. This is due to a number of reasons including the failure of screening-based approaches to find new compounds and the high cost of discovery research (~ US $1 billion including clinical trials for a new drug). Commercial considerations also play a major role; antibiotics are among the cheapest drugs compared to drugs that treat chronic diseases (e.g., hypertension, arthritis and dementia), which are more profitable (Davies, 2007; Marinelli & Tomasz, 2010; Nathan & Goldberg, 2005; Spellberg et al., 2004; Wright, 2007). In addition, newly developed anti-microbials are not used as first-line drugs but instead are considered 'drugs of last resort', thus negatively impacting on sales (Spellberg et al., 2004). Furthermore, it is recognized that drug resistance is inevitable (Wright, 2010), therefore, there is a risk that companies may not recover the costs incurred from pharmaceutical research and development before the drug becomes ineffective and loses its value in the market.

In 2002, the estimated use of antibiotics worldwide was 100,000-200,000 tonnes per annum (Wise, 2002). Less than half of these antibiotics are provided for use in humans while the majority are used in agriculture (Davies & Davies, 2010). In humans, 20 % of antibiotics are used in hospitals and 80 % are used in the community (outpatients). In agriculture, only 20 % are employed for therapeutic purposes, the remainder are used for growth promotion (Wise et al., 1998). It has been estimated that 20-50 % of human and 40-80 % of agricultural antibiotic use is unnecessary or questionable (Wise et al., 1998). This frequent and sometimes unnecessary use of antibiotics has been shown in many studies. For example, Nyquist and colleagues (Nyquist, 1998) found that more than 40 % of adult children with common colds, upper respiratory infections (URIs) and bronchitis were prescribed antibiotics, conditions that do not generally benefit from antibiotics. As a consequence of seventy years of use and misuse of antibiotics, bacteria with resistance to multiple drugs have now emerged.

1.2 Antibiotic resistance

1.2.1 Emergence and dissemination of antibiotic resistance

There is no doubt that the discovery of antibiotics revolutionised the field of medicine. However, the use of antibiotics inevitably selects for resistant microbes (Smith et al., 2005). The most notable example is the use of penicillin and methicillin to treat infections caused by *Staphylococcus aureus* that led to the emergence of penicillin and eventually methicillin-resistant *Staphylococcus aureus* (MRSA). Vancomycin became the drug treatment of last resort, however, because of an increase in the use of the drug, vancomycin resistant strains also appeared. When reports of an MRSA strain with reduced susceptibility to vancomycin first emerged in Japan, a minimum inhibitory concentration (MIC; the lowest concentration of an anti-microbial that will inhibit the visible growth of a microorganism after overnight growth) of 8 μg/mL was recorded (Hiramatsu et al., 1997). Since then, strains with MIC values of 64 μg/mL have been observed (Thati et al., 2011). To further add to the problem, over the years, selective pressure exerted by the use of the drug has resulted in other microbes also acquiring resistance to vancomycin (e.g., vancomycin resistant enteroccoci (VRE) and *Enterococcus faecium*) (Levy & Marshall, 2004).

Common infections that used to be easily treatable with antibiotics are proving more and more difficult to manage, resulting in an increase in morbidity, higher rates of mortality and greater hospitalization costs (Davies, 2007). At present, we are faced with many multiple resistant organisms such as, *Acinetobacter baumannii*, *Klebsiella pneumonia*, and *Mycobacterium tuberculosis*, which are extremely difficult to treat (Alekshun & Levy, 2007). In Europe, the European Centre for Disease Prevention and Control (ECDC) reported that 25,000 people die each year from antibiotic resistant bacteria (www.ema.europa.eu). In the US alone, hospital costs associated with antibiotic resistance are estimated at more than $20 billion per year (Cooper & Shlaes, 2011). As it stands, the world health organization (WHO) has warned of a 'post-antibiotic era' (who.int). In a recent report (2013), the chief medical officer of England (Prof. Dame Sally Davies) stated that the rise in drug resistant infections was as serious as the threat of global warming (http://www.bbc.co.uk/news/health-21178718).

There are a number of studies that have investigated the impact of antibiotic use on the emergence of antibiotic resistance. Van de Sande-Bruinsma and colleagues (2008) looked at the consumption of anti-microbial drugs (penicillin, macrolide and fluoroquinolone) and acquired resistance in *Streptococcus pneumonia* and *E. coli* in Europe from the year 2000 through 2005 and found a good correlation between antibiotic consumption and reduction in drug susceptibility. Furthermore, others have shown that countries with the highest per capita antibiotic usage have the highest rates of antibiotic resistance (Cizman, 2003).

The agricultural use of antibiotics (for example in animal feeds) has been identified as having a major impact on the emergence of resistant bacteria in the gut of chickens, swine and other animals as well as humans (Levy, 2002; Lipsitch et al., 2002; Smith et al., 2005). Although zoonotic infections of humans are rare it is one of the ways resistant strains can be transmitted between humans, for example from the ingestion of contaminated meat (Lipsitch et al., 2002). Engberg and colleagues performed a molecular epidemiological study to investigate the relationship between the use of fluoroquinolones in food production and the increase in antibiotic resistant bacterial infections in humans. They found identical strains of chicken *Campylobacter* isolates to those responsible for human infections (Engberg et al., 2004). Antibiotic resistant bacteria have also been found in wild boars (Poeta et al., 2009) as well as wild rodents (Gilliver et al., 1999). However, it is uncertain whether this is a result of scavenging, exposure to materials associated with humans (e.g.,

eating human food) or whether the microbes form part of the animal's normal flora (Wright, 2010). Other causes attributed to the emergence and spread of bacterial resistance are the partial treatment of infections (suboptimal drug doses, especially in third world countries) and the increase in frequent travelling around the world (Okeke & Edelman, 2001).

1.2.2 The origin of antibiotic resistance

While it is essential to investigate factors that contribute to the rise and spread of antibiotic resistance, it is also important to understand the origins of resistance. Studies have shown that 'resistance proteins' may serve other functions in the bacterium. For example, one study showed that 2′-*N*-acetyltransferase in *Providencia stuartii*, which functions to modify bacterial peptidoglycan, could inactivate aminoglycosides (Macinga & Rather, 1999). The ability of bacteria to use regular gene products for resistance suggests that resistance predates the clinical use of antibiotics. D'Costa and colleagues (2011) have shown this in a recent study. They sampled ancient DNA from Beringian permafrost sediments (30, 000 years old) and found genes that coded for resistance to β-lactam, tetracycline and glycopeptides antibiotics. Meanwhile, these drugs were first introduced into the clinics less than 100 years ago (Figure 1).

Indeed, bacterial exposure to antibiotics is not a modern phenomenon. Cyanobacteria are a group of photosynthetic microorganisms that produce a number of bioactive chemicals including antibiotics (Singh et al., 2008). Fossil records have shown that stromatolite mats, that are remnants of cyanobacterial communities, are > 3.45 billion years old (Allwood et al., 2006; Wright, 2010). In reality, many of the natural compounds produced in the environment by micro-organisms themselves (e.g., kanamycin) are the same drugs used in clinical settings; two-thirds of prescribed antibiotics are natural bacterial products or semisynthetic derivatives of these molecules (Clardy et al., 2006). Therefore, it should not be surprising that the microbes have evolved many ways of protecting themselves against drugs, since they have been exposed to antibiotics for a long time. Some of these mechanisms include the production of enzymes that inactivate the antibiotic, active removal of the antibiotic from the bacterial cell (pumps) or alteration of the cellular target (Hawkey, 1998) (Figure 3). It is possible that bacteria are able to use proteins naturally found in them to perform these functions and attain resistance. In this

thesis, the ability of regular gene products in bacteria to act as resistance proteins is explored.

Studies involving the screening of soil bacteria have revealed an underappreciated reservoir of antibiotic resistance carried by these microbes. For example, D'Costa and colleagues (2006) surveyed spore forming bacteria isolated from various soil samples. For their study, 480 strains were screened against 21 antibiotics (natural, semi-synthetic and synthetic products) including ones that had only recently been clinically approved. They found that all of the strains were multidrug resistant to seven or eight antibiotics on average. Dantas and colleagues (2008) expanded on this work and showed that a number of soil bacteria were able to subsist on antibiotics as their sole carbon source. Studies like these have reinforced the idea of a bacterial 'resistome', which consist of a collection of genes that contribute directly or indirectly to resistance (D'Costa et al., 2006; Wright, 2007). Housekeeping genes also form part of the bacterial resistome and have the potential to give rise to resistance either by mutation or over-expression (Wright, 2012).

Antibiotic resistance is not restricted to pathogenic bacteria. There are an estimated 5×10^{30} bacteria on the planet (Whitman, 1998) and the majority of these microbes are non-pathogenic (Wright, 2010). This means that resistance genes that are present or that arise in non-pathogenic microbes can be transferred to clinically import bacteria. A number of studies have shown that non-pathogenic bacteria such as human microflora harbour a large reservoir of antibiotic resistant genes (Sommer et al., 2009; Marshall et al., 2009). Therefore, the reservoir of resistance elements in non-pathogenic bacteria warrants further investigation. This is addressed in this PhD research.

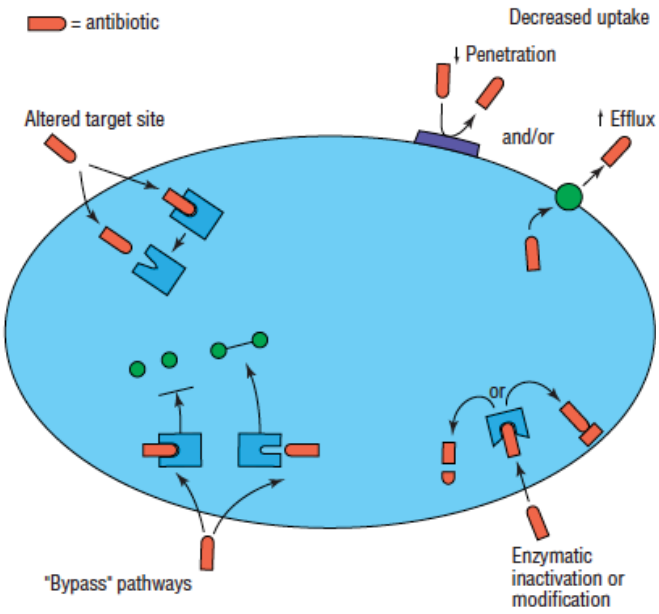


Figure 3. Molecular mechanisms of antibiotic resistance (taken from Hawkey, 1998)

Some of the mechanisms used to escape the effects of antibiotics: prevention of the drug from entering the cell or pumping it out; inactivation of the drug; production of an alternative target that is resistant to inhibition, whilst continuing to produce the original sensitive target (bypass); or alteration of the primary site of action.

1.2.3 Antibiotic resistance via acquisition and mutation

Aside from the intrinsic bacterial resistome, antibiotic resistance can be achieved by horizontal acquisition of a gene carrying a resistant trait (by transduction, transformation or conjugation) or by mutations in a pre-existing gene (Ochman et al., 2000). Major contributors to horizontal gene transfer (HGT) are conjugative plasmids, which generally carry genes that are useful to the host bacterium (e.g., resistance genes) (Norberg et al., 2011). Plasmid-borne resistance genes are thought to have originated from environmental microbes; genes conferring resistant traits derived from the chromosomes of other species may have sequestered into plasmids during events of horizontal gene transfer as a result of selective pressure (presence of a toxin) (Davies, 2007; Kado, 1998). Today, HGT is recognised as being largely responsible for the rise in antibiotic resistant infections worldwide (Warnes et al., 2012).

Resistance can also be acquired spontaneously (point mutations). Since DNA polymerase has an error rate of ~ 1 base in 10^9 during DNA replication, this can lead to spontaneous mutations every 200 cell divisions in an *E. coli* cell (genome size of ~ 5 x 10^6

base pairs) (Wright, 2010). Certainly, many will be deleterious mutations, however, the occasional beneficial one may emerge, especially because during a bacterial infection, a large number of cells are involved. A beneficial mutation in the bacterial chromosome may alter the drug target and reduce the binding of the antibiotic or if in a promoter sequence, might result in the upregulation of genes that favour the survival of the host bacterium.

Gene duplication and amplification (GDA) is an adaptive mechanism in response to the presence of antibiotics. This process can confer resistance by causing an overproduction of efflux pumps, target molecules or antibiotic modifying enzymes (Sandegren & Andersson, 2009). Duplication events are even more common than point mutations in bacterial populations, and are estimated to occur at a rate of 10^{-2} to 10^{-5} per cell generation (Sandegren & Andersson, 2009). The mechanisms suggested to be responsible for duplication include non-equal homologous recombination (e.g., between long repetitive palindromic sequences) and RecA-independent mechanisms (e.g., between short- or non-homologous regions). After duplication has occurred, higher-level amplification takes place by RecA-dependent recombination or by rolling circle duplication if there is no duplication intermediate. The copy number of amplified units can range between 2 and 40 (Sandegren & Andersson, 2009).

The first characterized example of antibiotic resistance resulting from an increased gene copy number was observed in R plasmids in the early 1970s. Here, the R-plasmid (NR1) within *Proteus mirabilis* increased in size when grown in the presence of the antibiotics to which the plasmid conferred resistance. The increase in plasmid size was a result of amplified resistance determinants (Rownd & Mickel, 1971). Gene amplification in chromosomal genes was first described for the AmpC β-lactamase, which was amplified in ampicillin-resistant strains of *E. coli* (Normark et al., 1977). It is believed that GDA can play a significant role in facilitating the development of stable genetic changes that lead to clinical resistance (Sandegren & Andersson, 2009).

1.3 Fitness cost of antibiotic resistance

The connection between fitness and resistance is important but remains poorly understood. Antibiotic resistance can affect the fitness of the cell in the absence of the drug, which may be relevant in a clinical setting. To gain a better understanding of this, it will be investigated in this thesis.

Resistance is clearly beneficial to the bacterium when an antibiotic is present. However, in the absence of the antibiotic, there is often a fitness cost (i.e., reduction in the efficiency at which the bacteria multiply) associated with the resistance phenotype (Andersson & Levin, 1999). One of the reasons for this is that the mutations are sometimes in genes that perform essential functions. For example, resistance to the antibiotic kanamycin can be acquired by mutating the *fusA* gene (Hou et al., 1994). *fusA* codes for elongation factor-G, which promotes the translocation step of protein synthesis (Agrawal et al., 1998). *fusA* resistant mutants show a reduction in growth because of a decreased rate of protein synthesis (Hou et al., 1994). Another reason for reduced fitness is that for example in the case of plasmid-encoded resistance functions, the bacteria is required to synthesise additional proteins, thus imposing an energy burden on the cell (Da Silva & Bailey, 1986). However, compensatory mutations can ameliorate the fitness cost whilst allowing the bacterium to maintain resistance (Björkman et al., 2000).

On the other hand, some studies have discovered mutations that increase resistance without any measurable fitness cost (Sander et al., 2002). Enne and colleagues (2004) discovered a *sul2*-encoding plasmid that provided resistance to the sulphonamide class of antibiotics and also imparted a fitness benefit to *E. coli*.

In laboratory studies, bacterial fitness is usually determined by performing competition experiments between clonally related susceptible and resistant strains in the absence of the antibiotic (Andersson, 2003). Although the correlation of measured fitness in a laboratory setting to fitness in real life is still uncertain, it is assumed that if a mutation imparts a fitness cost under favourable *in vitro* conditions, there will probably also be a cost under a more stressful environment *in vivo* (Andersson, 2003). While care must be taken when interpreting fitness results obtained *in vitro*, there are nevertheless advantages to measuring fitness under laboratory conditions. This includes assessing whether resistance determinants are likely to be relevant in clinical settings.

1.4 Promiscuity and origins of new functions

1.4.1 Promiscuity

Not only is antibiotic resistance incredibly important to understand public health, but it is also a paradigm for the evolution of new function. Secondary activities in proteins can be uncovered through resistance events. Thus, antibiotic resistance can provide an understanding of how these secondary activities can serve as starting points for new protein functions. This is the focus of my thesis.

The prevailing model for the evolution of new functions was proposed by Jensen (1976). In his model, he assumed that primordial cells possessed a small number of genes that encoded enzymes with numerous functions. This enabled a wide range of reactions to be performed using a limited number of enzymes. Through duplication, mutation and selection, enzymes with specialised functions diverged from their ancestors. The existence of enzyme families and superfamilies that have evolved new functions are an example of this (Figure 4). In these enzyme families, the primary function of one member is frequently identified as a secondary (or 'promiscuous') activity in other family members. For example, alkaline phosphatase and arylsulfatase are members of the alkaline phosphatase superfamily (O'Brien & Herschlag, 2001). The alkaline phosphatase superfamily comprises a large group of hydrolytic metalloenzymes (Galperin et al., 1998). Alkaline phosphatase has a primary phosphomonoesterase activity, but the enzyme also exhibits a secondary sulfatase activity (O'Brien & Herschlag, 2001). On the other hand, arylsulfatase has a primary sulfatase activity and can also catalyse the hydrolysis of phosphate monoesters (Olguin et al., 2008).

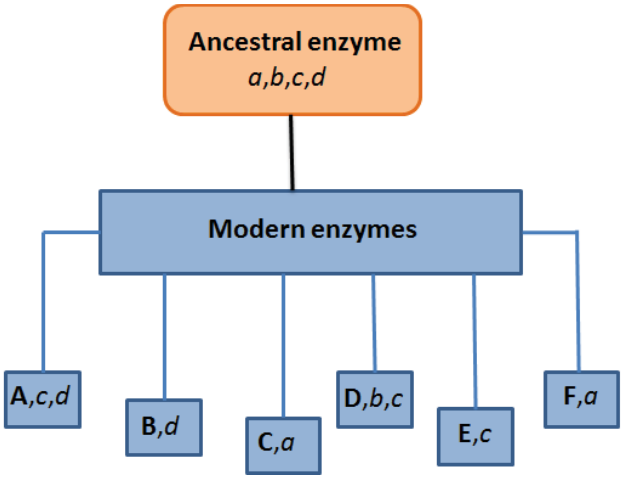


Figure 4. Divergence of modern enzymes from a multi-functional ancestral enzyme (modified from from Khersonsky et al., 2006)

Jensen's model proposes that ancestral enzymes were multi-functional and displayed a range of activities (lower case letters; *a*, *b*, *c* and *d*). Selection pressures for those activities has led to duplication and divergence of modern enzymes with specialized functions (upper case; A, B, C and D) that have retained some functions of their ancestor (lower case letters). New activities that are absent in the ancestor can emerge in the course of divergence (E and F).

Jensen's model is based on the idea that primordial enzymes were multi-functional. It is becoming increasingly apparent that modern enzymes are not completely specific, but instead retain vestiges of their multi-functional ancestry. Promiscuity is defined as the ability of an enzyme to catalyse other reactions in addition to the one it has evolved to carry out (Khersonsky & Tawfik, 2010). Many proteins are known that are able to catalyse reactions other than the ones they are specialised to perform. For example, carbonic anhydrase catalyses the reaction of carbon dioxide to bicarbonate, but it is also known to function as an esterase (Pocker & Stone, 1965). Promiscuity can be described in different terms to classify enzymes. For example, 'catalytic promiscuity' refers to enzymes that are able to catalyse chemically distinct reactions, while 'substrate ambiguity' is used to describe an enzyme that is able to catalyse the same chemical reaction for a range of different substrates (Babtie et al., 2010). For the purpose of this study, I broadly define promiscuity as the ability of a protein (including enzymes) to catalyse other reactions (secondary functions) that to date, have not been selected for throughout evolution. This may include 'catalytic promiscuity' and 'substrate-ambiguity'.

The mechanistic aspect of promiscuity is an interesting one, especially with regards to how specificity and promiscuity coincide within a single active site. Some enzymes

achieve the native and promiscuous activities by using different active-site configurations depending on the substrate present. For example, isopropylmalate isomerase is an enzyme that catalyses the stereo-specific isomerization of α-isopropylmalate to β-isopropylmalate in the leucine biosynthesis pathway (Cole et al., 1973). The enzyme also functions as homoaconitase in the lysine biosynthesis pathway to convert homocitrate to homoisocitrate (Yasutake et al., 2004; Drevland et al., 2008). To perform these two activities, a loop structure in isopropylmalate isomerase is important for recognizing the two substrates (isopropylmalate and homocitrate).

When the same configuration is used for the native and promiscuous functions, the amino acids that bind the substrate can be different, for example for the promiscuous hydrolysis of phosphodiesters by alkaline phosphatase (Catrina et al., 2007). However, when the same catalytic residues are used, they can act in different protonation states, as seen in the case of 4-oxalocrotonate tautomerase's weak promiscuous catalysis of hydratase activity (Wang et al., 2003). In other cases, different subsites within the same active site are used to catalyse the promiscuous reaction [e.g., serum paraoxonase PON1 with promiscuous esterase activity (Khersonsky et al., 2006)]. Additionally, enzymes could change their specificity by introducing metal substitutions, for example, the substitution of native zinc by manganese to catalyse styrene epoxidation by carbonic anhydrase (Fernández-Gacio et al., 2006).

Promiscuity is thought to be an inherent feature of enzymes and proteins in general. This view is strengthened by the observation that even specific enzymes such as those involved in DNA or protein synthesis exhibit measurable substrate infidelities (Khersonsky & Tawfik, 2010; Brustad et al., 2008). However, there is no single detection method available for a high-throughput screen of all promiscuous enzymatic activity (because of the range of different substrates and reactions). Regulation at the level of expression is believed to mask the (sometimes undesirable) effects of promiscuity (Jensen, 1976; Khersonsky & Tawfik, 2010). However, in the presence of an antibiotic where there is a high selection pressure an enzyme's 'hidden' promiscuous activity (the ability to deactivate the antibiotic) that allows the cells survival becomes unveiled. The Innovation, Amplification and Divergence (IAD) model (Bergthorsson et al., 2007) has been proposed to connect genetic processes with enzyme promiscuity, in the evolution of new functions (discussed below).

1.4.2 Evolution of new genes

While the biochemical basis of promiscuity is plausible, it remains to be explained how promiscuity can lead to the evolution of new functions at the genetic level. Ultimately, new enzyme functions are the result of new genes. Early ideas on the origin of new genes were developed by Ohno (1970). He suggested that duplication of a given gene would generate a redundant copy that would be retained long enough, while relieved from the burden of selection to accumulate beneficial mutations, eventually leading to a new function. The view that duplication would precede the emergence of a gene with a new function was shared by many (Kimura & Ota, 1974) and has been called the mutation during non-functionality (MDN) model. However, a newly duplicated gene that is selectively neutral would have a small probability of being fixed in a population (Zhang, 2003) and is more likely to be lost by mutation, drift or segregation (Bergthorsson et al., 2007). Furthermore, if the redundant copy was to last long enough in a population, it is more likely to acquire common deleterious mutations than rarer beneficial mutations (Roth et al., 2006). Bergthorsson and colleagues (2007) addressed this dilemma in their IAD model (Figure 5).

The IAD model comprises three steps, Innovation, Amplification and Divergence, with selection acting at all stages. The parent gene encodes for a protein with a weak promiscuous activity (the Innovation). In an environment where this minor activity can be useful, such as in the presence of an antibiotic, there is selection for an increase in the expression level of the promiscuous protein.

As mentioned previously, duplication and amplification are common events that occur at higher rates than point mutations. Therefore, the most probable way to increase the level of the minor promiscuous activity is by amplification of the parent gene. Its frequency is raised in the population and maintained by selection for its weak activity. Under these circumstances, gene amplification relieves the new copies from the original function, so long as some copies are still performing that function. This is the second step, known as Amplification.

The final step in this model is Divergence. The amplified gene copies provide many opportunities for improvement by selection; more targets become available for the acquisition of beneficial mutations to generate a protein with improvements in the promiscuous function. Genes that allow an increase in cell fitness will be selected for, whilst the others will be lost, most likely through a loss of the amplification and inactivating

mutations. Eventually, one of the copies will produce a protein with a specialised activity, thus creating a new function. The original gene function will still be maintained by selection in at least one of the copies.

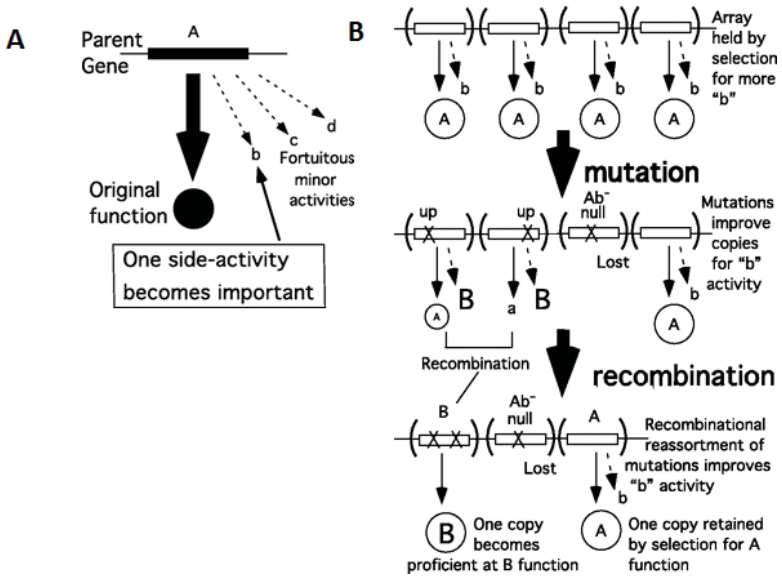


Figure 5. Diagram shows innovation as an initial step before amplification (taken from Bergthorsson et al., 2007)

(**A**) The IAD model suggests that the parent gene has weak side activities that become important in a given situation, which then results in its duplication. (**B**) New functions are improved as explained in Section 1.4.2.

1.5 Investigating promiscuity

Multicopy suppression studies have proved to be a good way to uncover promiscuity (Patrick et al., 2007; Miller & Raines, 2004). In this approach genes are cloned onto a high copy number expression plasmid and transformed into a strain that carries a certain mutation. The genes that circumvent the defective phenotype can then be identified. Results from such studies have shown that promiscuity is common within proteins and can be readily accessed for adaptation. This is in agreement with Jensen's model (multi-functionality of proteins) and the IAD model (Innovation and Amplification). Furthermore, these studies have shown that suppression can occur with unrelated genes. A study by Patrick and colleagues (2007) showed that out of 104 single-gene knockout strains screened, 20 % of auxotrophs were rescued by the over-expression of at least one

noncognate *E. coli* gene. Their results demonstrated that first, amplification can be used as a mechanism for adaptation and second, many bacterial genes are multi-functional and are capable of performing functions different to their native ones.

A number of studies have investigated the role of promiscuity in protein evolution and demonstrated that promiscuous functions are readily evolvable, at least *in vitro* (Divergence part of IAD model). In some cases, a single point mutation was enough to substantially improve the enzyme's new function (O'Brien & Herschlag, 1999). Laboratory directed-evolution has proved a useful approach for this type of research and has successfully demonstrated how promiscuous activities can serve as starting points towards a specialized function (Bershtein & Tawfik, 2008). In this approach, the promiscuous activity of an enzyme is increased by introducing random mutations into the target gene then selecting for the desired trait. For example, Aharoni and colleagues (2005) targeted the promiscuous activities of serum paraoxonase (PON1), bacterial phosphotriesterase (PTE) and carbonic anhydrase II (CAII) using laboratory-directed evolution. They were able to improve the promiscuous activities in all three cases.

Although studies such as those performed by Aharoni et al. (2005) are useful in investigating promiscuity in general, not many studies have (i) screened the bacterial proteome to uncover the role of promiscuity in antibiotic resistance and (ii) explored whether these promiscuous resistance activities are 'evolvable'; i.e., whether they can provide the starting points for the evolution of new antibiotic resistance functions by gradually acquiring mutations that convert the weak promiscuous function into a primary function. My colleague, Valerie Soo, began to address the first question in her PhD research (2012). My goal was to investigate the second question using a laboratory-directed evolution approach.

There have been a number of suggestions to help tackle the issue of antibiotic resistance, including searching for new drugs in order to preserve the pool of effective antibiotics. However, focusing on research that could predict genes capable of endowing bacteria with resistance prior to their emergence seems a worthwhile investment and an important initial step in addressing the problem of clinical resistance. By identifying these genes, we can begin to understand the bacterial resistance reservoir.

Some researchers have confronted this by creating transposon insertion libraries to detect genes responsible for increased sensitivity to one or more antibiotics, in both

Acinetobacter baylyi (Gomez & Neyfakh, 2006) and in *Pseudomonas aeruginosa* (Breidenstein et al., 2008). However, this type of experimental approach does not unveil the full potential of the bacterial intrinsic resistome; many promiscuous activities are latent and are not detectable unless protein expression is increased to elevate the level of the promiscuous activity. Thus, generating a random transposon gene-knockout library would not uncover many promiscuous functions in the bacterium. Therefore, to detect promiscuous resistance functions in the cell, there needs to be a direct test for resistance rather than susceptibility. Our lab has directly tested for resistance in the presence of several antibiotics. A number of novel proteins exhibiting promiscuous activity have been identified. I hope to investigate whether these promiscuous functions can provide the raw materials for the evolution of new functions.

1.6 Aims of this thesis

1.6.1 The origin of this project

Previous work in the laboratory showed that over-expression of many native *E. coli* proteins increased resistance to an array of antibacterial agents and toxins (including antifungal and antiparasitic agents; (Soo et al., 2011). PhD student Valerie Soo made use of the ASKA (A complete Set of *Escherichia coli* K-12 Open Reading Frame Achive) library, which comprises every *E. coli* open reading frame cloned into the expression vector pCA24N (Kitagawa et al., 2005). She transformed the pooled plasmids of the ASKA library into *E. coli* DH5α-E cells and used the pool to inoculate 96-well Phenotype Microarray plates (Bochner et al., 2001), that contained a total of 237 toxins (including antibacterial agents). The growth rate of the ASKA pool in each well was compared to a negative control (*E. coli* harbouring an empty pCA24N vector). If an ASKA gene endowed a bacterium with resistance, then that strain would grow faster than the control in the presence of the toxin.

Soo was able to isolate and identify individual ASKA clones that reproducibly grew better than the control, in 86 of the 237 toxin-containing environments that were tested. There were a total of 78 antibacterial agents present in the Phenotype Microarray plates and the ASKA pool grew better than the control in 50 of these (64 %). From these results, a list of 115 putative resistance genes (with diverse mechanisms of resistance) was compiled (Soo et al., 2011).

1.6.2 Specific aims

The goals of this study were to gain mechanistic insights into some of the previously-identified resistance pathways and to test the 'evolvability' of protein functions associated with antibiotic resistance. The specific aims were:

1) To quantify the level of resistance conferred by a variety of proteins discovered in Soo's experiments and to investigate whether over-expression of the genes coding for these proteins impart a fitness cost in the absence of the relevant antibiotic (Chapter 2).

2) To perform directed evolution experiments to improve the drug resistance functions of four selected proteins: carbonic anhydrase, YeaD, YdfW (Chapter 3) and YejG (Chapter 4).

3) To probe the resistance mechanism of YejG, and in particular, to study its connection with a new mutation in the *fusA* gene (Chapter 5 and 6).

CHAPTER 2

Drug resistance and its impact on bacterial fitness

The Minimum Inhibitory Concentrations and the fitness assay experiments in this chapter were performed with Valerie Soo. Technical assistance in the pairwise fitness assays was provided by Ilana Gerber, Laura Nigon and Susan Morton.

2.1 Minimum Inhibitory Concentration and Fitness assays

To assess the susceptibility of a bacterial strain to an anti-microbial agent, the Minimum Inhibitory Concentration (MIC) is determined. An MIC is defined as the lowest concentration of an anti-microbial that will inhibit the visible growth of a microorganism after incubation for a defined period of time (typically 16-20 h) (Wiegand et al., 2008). This test allows you to determine whether a strain of interest is resistant to a particular anti-microbial agent by comparing its MIC value to that of a control strain. The higher the MIC value, the more resistant the clone is to the toxin being tested. This method is considered the 'gold standard' for anti-microbial susceptibility testing and is used by many diagnostic laboratories (Andrews, 2001).

While resistance is beneficial in an environment that contains an antibiotic, it often imparts a fitness cost in the absence of the toxin, although secondary mutations that ameliorate this fitness burden have been reported (Levin et al., 2000). The fitness of a bacterium is defined as the cell's ability to survive and reproduce in a given environment and can be measured by its growth rate (Sandegren & Andersson, 2009). Fitness assays are used to assess the impact of a given trait on the microorganism's competitive ability. The assays measure the relative fitness of two competitors that are identical, except that one of the strains possesses the trait of interest (and also, one of the strains carries a neutral marker to distinguish between them) (Leroi et al., 1994). The competitors are grown together in a pool of limited resources for several hours. The change in ratio of the two strains from the start and the end of the experiment is used to calculate their relative fitness. Thus, the impact of the trait under consideration, in the environmental condition used in the experiment, can be evaluated. The connection between fitness and resistance remains poorly understood therefore it was investigated in this study.

2.2 Focus of this chapter

My first task was to extend the results from the library-on-library screen, by assessing the contribution to resistance of individual ASKA genes selected from the list compiled by Valerie Soo. To do this, the MIC values of strains carrying the plasmid-encoded gene of interest were measured. The MIC value of the *E. coli* strain over-expressing this target gene was then compared to the MIC value of a control strain (strain carrying

plasmid without any insert). The aim of this work was to confirm, in individual cases, that the artificial up-regulation of specific *E. coli* genes could confer resistance. To accomplish this, the MIC values for 25 different strains were determined, in a total of 29 toxin/strain combinations (some strains were measured for resistance to more than one antibiotic), covering a variety of toxins (with different mechanisms of action). The 25 strains were selected with the aim of investigating diverse modes of resistance as well as incorporating drugs with different mechanisms of action. Next, the relative fitness values (W) of the same set of *E. coli* strains were determined, by performing fitness assays between the ASKA strains and the control strain.

2.3 Results

2.3.1 Determining the MIC values of the ASKA strains

The strains used in this experiment were i) *E. coli* DH5α-E cells over-expressing selected ASKA genes [fused to a C-terminal green fluorescent protein (GFP) tag; similar to the strains used in the original library screening study (Soo et al., 2011)] and ii) a negative control strain harbouring an empty pCA24N vector (pCA24N-GFP; over-expressing GFP with no fusion partner). The GFP protein was used for the negative control in order to ensure that this strain was also over-expressing a high level of protein. Protein expression was induced by the addition of 50 μM of isopropyl β-D-1-thiogalactopyranoside (IPTG). The MICs were determined by either plate-based (E-test) or by broth-based methods (Section 2.5.1).

The results showed that in most cases, the ASKA strain was more resistant than the control strain, with only two out of the 25 strains tested not showing an increase in MIC (Table 1). The majority of the strains showed a moderate increase in resistance (< 6-fold). The three strains with the highest fold-increases in resistance were the dihydrofolate reductase (FolA) and multi-drug transporter-expressing strains (Cmr and MdtM) (Tables 1 and 2). The MIC value of the FolA-expressing strain for trimethoprim displayed by far the highest increase, with at least a 500-fold increase in resistance compared to the control strain. Trimethoprim is a folate analog that binds tightly to the product of *folA* and inhibits its action (Huovinen et al., 1995). In this case, resistance appears to be associated with the over-expression of trimethoprim's intracellular target. Overproduction of this enzyme is indeed one of the common mechanisms associated with trimethoprim-resistant phenotypes (Smith & Calvo, 1982). In one study, a clinically isolated *E. coli* strain that overproduced dihydrofolate reductase showed resistance to more than 1000 $\mu g.mL^{-1}$ of trimethoprim (Flensburg & Skold, 1987).

In a few cases, the ASKA strains attained clinically relevant levels of resistance. A clinical breakpoint is defined as 'the particular MIC that differentiates susceptible, and assumingly treatable, from resistant and assumingly untreatable organisms' (Wiegand et al., 2008). Over-expression of a small uncharacterized protein, YcgZ (Table 2), increased the MIC value for cefuroxime from 6 to 16 $\mu g.mL^{-1}$ (Table 1). The clinical breakpoint for resistance to this drug is 8 $\mu g.mL^{-1}$ (The European Committee on Anti-microbial Susceptibility Testing [EUCAST], 2012). This means that the strain showed double the

MIC value that would be associated with a high likelihood of therapeutic failure, simply through over-expression of this uncharacterized protein. The clinical breakpoint for resistance to trimethroprim is 2 $\mu g.mL^{-1}$, therefore the FolA-expressing strain would be considered highly resistant to the drug in a clinical environment (MIC value of > 32 $\mu g.mL^{-1}$). The YejG-expressing strain showed an MIC value of 1.5 $\mu g.mL^{-1}$ for tobramycin, approaching a clinically relevant level of resistance for this aminoglycoside (EUCAST breakpoint = 2 $\mu g.mL^{-1}$).

Out of the two strains that did not show increases in drug resistance compared to the control, only the TfaX-expressing strain (predicted prophage protein; Table 2) was more susceptible to the antibiotic (nitrofurantoin) than the control (Table 1). Its MIC fell by a factor of 2, suggesting that TfaX expression led to an increase in fitness in the context of the library-on-library screen, without increasing resistance to nitrofurantoin when a clonal population of the cells was tested in isolation.

Table 1. Drug resistance data for *E. coli* DH5α-E cells over-expressing ASKA genes

MIC values were measured using either plate-based [E-test; shown with an asterisk (*)] or broth-based methods.

ASKA gene	Antibiotic	MIC (μg.mL^{-1}) for ASKA gene	MIC (μg.mL^{-1}) for control	Fold-increase in MIC
β-Lactams (Bacterial cell wall synthesis inhibitors)				
bdm	*Cephalothin	16	8	2.0
bfd	*Penicillin G	48	32	1.5
rbsR	*Aztreonam	0.19	0.064	3.0
yeaD	*Aztreonam	0.094	0.064	1.5
ycgZ	*Aztreonam	0.38	0.064	5.9
ycgZ	*Cefuroxime	16	6	2.7
Aminoglycosides (Protein synthesis inhibitors)				
yejG	Sisomicin	0.19	0.125	1.5
yejG	*Tobramycin	1.5	0.5	3.0
Anti-folates (Folic acid synthesis inhibitors)				
folA	*Trimethoprim	>32	0.064	>500
luxS	Sulfadiazine	24	16	1.5
nudB	Sulfadiazine	32	16	2.0
Quinolones (DNA replication and transcription inhibitors)				
cynT	Enoxacin	0.25	0.19	1.3
marA	*Ciprofloxacin	0.125	0.032	3.9
galE	*Nalidixic acid	96	32	3.0
yfdO	*Nalidixic acid	48	32	1.5
Tetracyclines (Protein synthesis inhibitors)				
bcr	Oxytetracycline	6	2	3.0
Macrolide-Lincosamide -Streptogramin (Protein synthesis inhibitors)				
cmr	Lincomycin	1,556	192	8.0
rbsR	Lincomycin	384	192	2.0
ydfW	Spiramycin	36	24	1.5
ydfW	*Erythromycin	96	32	3.0
ydaC	*Erythromycin	128	32	4.0
Glycopeptides (Bacterial cell wall synthesis inhibitors)				
mzrA	Vancomycin	256	128	2.0
Nitrofurans (Damage bacterial DNA)				
tfaX	*Nitrofurantoin	0.094	0.19	0.5
Rifamycins (Bacterial RNA synthesis inhibitors)				
feoC	*Rifampicin	6	4	1.5
tusE	Rifampicin	6	4	1.5
Other toxins				
mdtM	Puromycin	128	8	16.0
puuD	Benzethonium chloride	6	6	1.0
ybcD	Benzethonium chloride	8	6	1.3
yhiK	Benzethonium chloride	12	6	2.0

Table 2. Products of the ASKA genes

Gene	Gene product	Reference(s)
bcr	Bcr multidrug transporter	(Keseler et al., 2011; The Uniprot Consortium, 2012)
bdm	Biofilm-dependent modulation protein	(Francez-Charlot et al., 2005)
bfd	Bacterioferritin-associated ferredoxin	(Quail et al., 1996)
cmr	MdfA/Cmr multidrug transporter	(Edgar & Bibi, 1997)
cynT	Carbonic anhydrase	(Guilloton et al., 1992)
feoC	Transcriptional regulator of ferrous iron transport	(Cartron et al., 2006)
folA	Dihydrofolate reductase (DHFR)	(Keseler et al., 2011; The Uniprot Consortium, 2012)
galE	Uridine diphosphate (UDP)-glucose 4-epimerase	(Keseler et al., 2011; The Uniprot Consortium, 2012)
luxS	S-ribosylhomocysteine lyase	(Li et al., 2007)
marA	MarA transcriptional activator	(Cohen et al., 1993)
mdtM	MdtM multidrug transporter	(Paulsen et al., 2000)
mzrA	Modulator of EnvZ and OmpR (osmoregulatory two-component signal transduction system)	(Gerken et al., 2009)
nudB	Dihydroneopterin triphosphate pyrophosphatase (DHNTPase)	(Gabelli et al., 2007)
puuD	γ -glutamyl- γ -aminobutyrate hydrolase	(Kurihara et al., 2005)
rbsR	Ribose transcriptional repressor	(Mauzy & Hermodson, 1992)
tfaX	Uncharacterized protein from DLP12 prophage	(Keseler et al., 2011; The Uniprot consortium, 2012)
tusE	Sulfurtransferase	(Keseler et al., 2011; The Uniprot consortium, 2012)
ybcD	DLP12 prophage; predicted replication protein fragment	(Keseler et al., 2011; The Uniprot Consortium, 2012)
ycgZ	Uncharacterized small protein	(Tschowri et al., 2009)
ydaC	Uncharacterized protein from Rac prophage	(Keseler et al., 2011; The Uniprot Consortium, 2012)
ydfW	Uncharacterized protein from Qin prophage	(Keseler et al., 2011; The Uniprot Consortium, 2012)
yeaD	Putative mutarotase	(You et al., 2010)
yejG	Uncharacterised protein	Unknown
yfdO	Uncharacterized protein from cryptic prophage CPS-53/KpLE1	(Keseler et al., 2011; The Uniprot Consortium, 2012)
yhiK	Uncharacterized protein	Unknown

2.3.2 Relative fitness of the ASKA strains in the absence of antibiotic

The fitness of the ASKA strain in the absence of toxin (but in the presence of IPTG) was measured relative to a neutrally marked negative control. To do this, equal amounts of ASKA and the control strains were grown together in a pool of limited nutrients and the change in the rate of their relative abundance was determined (Section 2.5.2.2). In each fitness experiment, eight replicates (n = 8) of ASKA and control strains were used. The relative fitness (W) was calculated as the change in ratio of the ASKA clone relative to the control (W = 1) from the initial (t = 0) to the final (t = 24 h) time points. The levels of significance (P value) of the fitness measurements were then determined using Student's t-tests. In this experiment, $W < 1$ corresponds to a fitness cost to the host cell from the protein being over-expressed while $W > 1$ represents a fitness improvement. A value of $W = 1$ indicates that the two competitors have equal fitness.

The results revealed that 14 out of 25 ASKA strains had significant fitness improvements ($W > 1$; $P < 0.01$) and two showed a significant reduction ($W < 1$; $P < 0.01$) compared to the control (Figure 6). Nine clones were not statistically different from the control ($P > 0.01$).

The strains expressing either TfaX or PuuD were the only ones that did not show increases in antibiotic resistance compared to the control (Table 1). However, both strains had statistically significant fitness advantages in the absence of the antibiotic (TfaX: W = 1.18 ± 0.02; PuuD: W = 1.12 ± 0.03). The FolA-expressing strain on the other hand displayed the highest fold-increase in resistance but it was one of the least fit clones in the absence of antibiotic (W = 0.84 ± 0.03; P = < 0.01). The other clone that showed a significant reduction in fitness was the NudB-expressing strain (encodes dihydroneopterin triphosphate pyrophosphatase; Table 2) (W = 0.68 ± 0.04; P = < 0.01). Interestingly, both the *folA* and *nudB* gene products are involved in folate biosynthesis. This suggests that the over-expression of enzymes in the folate synthesis pathway is costly to the cell.

There were some cases where the fitter clones were also more resistant. For example, three genes (*ycgZ*, *rbsR* and *yeaD*) conferred increased resistance to aztreonam (Table 1), with the YcgZ-expressing strain showing both the highest MIC and the highest fitness (W = 1.72 ± 0.05; Figure 6). However, the opposite was also true in other cases. For example the NudB-expressing strain showed a higher resistance to sulfadiazine than the

clone expressing LuxS (32 versus 24 µg.mL^{-1}; Table 1). However, the NudB-expressing strain revealed the lowest fitness (0.68 ± 0.04; Figure 6).

Overall, the results show that in the majority of cases, there was no fitness cost associated with over-expressing an ASKA gene, and in fact, the ASKA strains outcompeted the control.

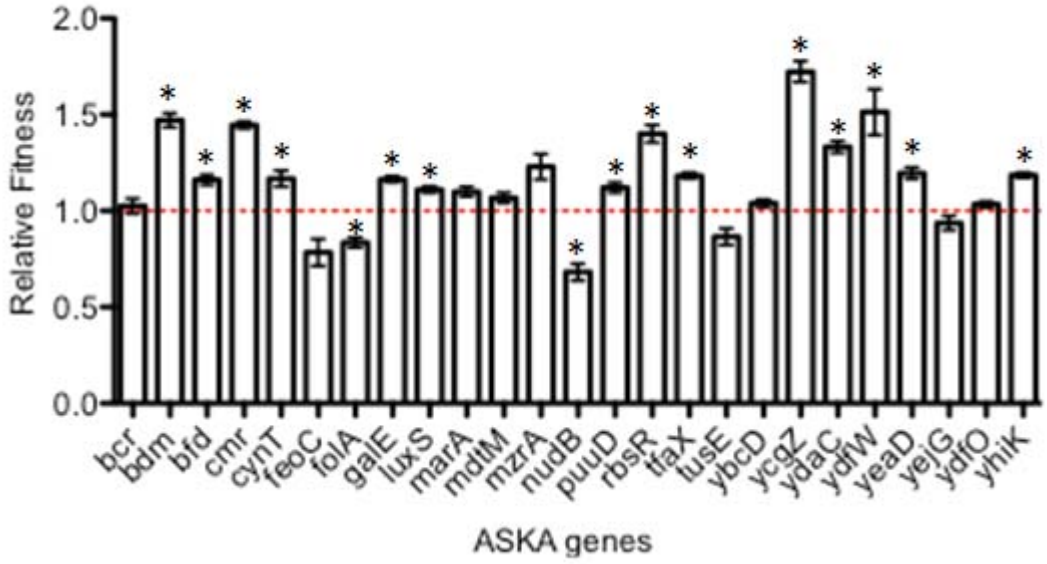


Figure 6. Relative fitness, W, of *E. coli* DH5α-E cells over-expressing ASKA genes in competition with a neutrally marked control (*lacZ*-tagged).

Fitness of ASKA strains in the absence of the antibiotic, relative to a negative control. Fitness values below 1 indicate a fitness cost while those above 1 confer a fitness benefit. Data are from eight replicates in each case (mean ± SE). Strains with statistically significant levels of $P < 0.01$ for a two-tailed test with the null hypothesis that $W = 1$, calculated using a t distribution, are indicated with an asterisk (*).

2.4 DISCUSSION

The first aim of this chapter was to assess the contribution of individual ASKA genes to antibiotic resistance. This was achieved by comparing the MIC values of strains over-expressing the target gene and a control strain. The MIC values of 25 different strains were determined in a total of 29 toxin/strain combinations. The results revealed that in the majority of cases, the ASKA strain was more resistant than the control and showed moderate increases in resistance (< 6-fold). The presence of low-level resistance in bacteria is not insignificant. It is thought to play an important role in the emergence of high-level antibiotic resistance, by increasing the possibility of a clinically relevant resistance phenotype appearing (Baquero et al., 1997; Martinez & Baquero, 2000). Higher levels of resistance (≥ 8-fold) were acquired either by over-expressing efflux proteins (*cmr* and *mdtM*) or the drug target (*folA*) (Table 1).

These results have demonstrated that over-expressing pre-existing genes in non-pathogenic bacteria (represented here by a laboratory *E. coli* strain) is able to impart resistance. This is in agreement with studies that have suggested that antibiotic resistance is not restricted to pathogenic bacteria (Sommer et al., 2009; Marshall et al., 2009). Some of the ASKA strains (i.e., those over-expressing *ycgZ*, *folA* and *yejG*) attained or came close to attaining clinically relevant levels of resistance (Table 1), that is, a high likelihood of therapeutic failure. The clinical breakpoint for resistance to trimethoprim is 2 $\mu g.mL^{-1}$. FolA (dihydrofolate reductase) is an enzyme involved in the biosynthesis of tetrahydrofolate, which is required for RNA-, DNA- and protein-synthesis (Myllykallio et al., 2003). Over-expression of this protein revealed an MIC value of > 500 $\mu g.mL^{-1}$ for trimethoprim (Table 1). This is at least 250 times the clinical breakpoint. More interestingly, over-expression of an uncharacterized protein, YcgZ, increased the MIC value for cefuroxime from 6 to 16 $\mu g.mL^{-1}$ (Table 1), which is double the clinical breakpoint for this drug (EUCAST breakpoint = 8 $\mu g.mL^{-1}$). The YejG-expressing clone on the other hand showed an MIC value of 1.5 $\mu g.mL^{-1}$ for tobramycin and came very close to being considered a clinically resistance strain (EUCAST breakpoint = 2 $\mu g.mL^{-1}$). Over-expression of the latter two proteins (YcgZ and YejG) was able to endow bacteria with levels of resistance that would be of concern in a clinical setting, and yet their functions are unknown. In this study, eight genes of unknown or uncharacterized function (*ybcD*, *ycgZ*, *ydaC*, *ydfW*, *yeaD*, *yejG*, *yfdO* and *yhiK*; Table 2) were found to confer moderate increases in resistance in the presence of a total of eleven antibiotics (Table 1). These findings

highlight the potential importance of such genes in contributing to clinically relevant levels of resistance.

Different types of proteins were able to contribute to resistance. In some cases, resistance was achieved by simply over-expressing the drug target or by up-regulating enzymes involved in the same biosynthetic pathway as the drug target. For example, sulfadiazine is a structural analog of *p*-aminobenzoic acid (PABA). This compound is required for the synthesis of folate and a substrate of the bacterial enzyme dihydropteroate synthase (DHPS) (Huovinen et al., 1995). Sulfadiazine competes with PABA for the active site of DHPS (Houvinen et al., 1995). NudB [dihydroneopterin triphosphate pyrophosphohydrolase (DHNTPase)] is an enzyme that catalyses the committed-step of folate synthesis (Gabelli et al., 2007). Over-expression of this protein leads to a 2-fold increase in resistance to sulfadiazine, compared to the control (Table 1).

In other cases, over-expression of the ASKA gene was likely to mediate pleiotropic effects. That is, it resulted in a downstream effect of altering the expression of one or more genes associated with drug resistance. For example, RbsR is a repressor protein that controls the transcription of the *rbs* operon for ribose catabolism and transport (Mauzy & Hermodson, 1992). Over-expressing this protein endowed the cell with a 2-fold increase in resistance to lincomycin (Table 1), a drug that binds to the 50 S ribosomal subunit to inhibit protein synthesis (Fitzhugh, 1998). There were also cases where some of the genes that conferred resistance coded for enzymes that were possibly displaying some form of promiscuity. For example, carbonic anhydrase (CA; encoded by *cynT*) is an enzyme that catalyses a reversible hydration reaction and has been previously shown to possess a promiscuous esterase activity (Tripp et al., 2001; Gould & Tawfik, 2005). The CA-expressing strain was more resistant (1.3-fold) to enoxacin (inhibits the bacterial enzyme DNA gyrase; Ruiz, 2003) than the control strain (Table 1). Altogether, this shows that bacteria possess even more mechanisms/pathways than previously realised.

One of the factors that affects the emergence and survival of resistant strains is the biological cost of resistance. A resistance mutation that decreases fitness will result in the resistant strain being out-competed by susceptible strains in the absence of the drug (Björkman & Andersson, 2000; Andersson 2003). Antibiotic resistance is frequently associated with a biological fitness cost (Andersson & Levin, 1999; Nagaev et al., 2001; Andersson 2003). This is because altered targets may not function as efficiently as their

progenitors and the synthesis of additional proteins can lead to metabolic costs (Da Silva & Bailey, 1986; Schulz zur Wiesch et al., 2010). However, there have been reports that this is not always the case. Gagneux and colleagues (2006) showed that rifampin-resistance in certain clinical isolates of *Mycobacterium tuberculosis* engendered little or no fitness cost. Sander and colleagues (2002) on the other hand, have observed a strong selection pressure for resistance mutations that incur little or no fitness cost in clinical isolates of *Mycobacterium tuberculosis*. To assess whether drug resistance is associated with a fitness cost, the relative fitness of each of the ASKA clones and the control strain was determined.

The results revealed that for the majority of the strains, over-expression of the ASKA gene did not incur any fitness cost. Instead, most of the ASKA clones out-competed the negative control (Figure 6). In light of the studies that have reported that resistance does not always confer a cost on the bacteria, it was not surprising that some of the ASKA clones were relatively fitter than the control. However, what was surprising was that most of the ASKA clones showed this. Other studies have also found no fitness cost associated with resistance to some of the drugs tested in this work. For example, Marcusson and colleagues (2009) reported a ciprofloxacin-resistant *E. coli* strain (24-fold increase in MIC) that did not display a reduction in fitness (W = 1.01). In agreement with this study, the MarA-expressing strain showed an increase in resistance to the same drug (3.9-fold) and exhibited a small fitness advantage (W = 1.08) (Table 1 and Figure 6).

On the other hand, other groups have shown a fitness cost associated with resistance. For example, Rozen and colleagues (2007) reported that mutations that conferred ciprofloxacin resistance engendered a fitness cost. Their work involved *Streptococcus pneumoniae* strains and drug resistance was conferred by point mutations in the topoisomerase IV genes (*parC* or *parE*) and the DNA gyrase genes (*gyrA* or *gyrB*). Therefore, possible explanations for the differences in the fitness results might be the type of mutation conferring resistance and/or the bacterial strain used (i.e., fitness cost might depend on the genetic background of the strain) (2006).

It is possible that the 'increased fitness' of the ASKA clones was in fact a 'decreased fitness' of the negative control. Each ASKA clone harboured a gene insert fused to a C-terminal GFP tag cloned into the pCA24N vector. The control strain carried the same vector but without an insert; that is, it expressed GFP (with no fusion partner). Protein over-expression was induced by the addition of 50 μM IPTG. It has been reported that the

expression of proteins unneeded for growth burden cells for their capacity to grow (known as cost). For example, Shachrai et al., (2010) found that there was a cost associated with GFP production (an unneeded protein). The cost increased with the level of GFP production; ~ 35 % relative growth reduction when GFP reached ~ 15 % of the cells' total protein.

Malakar and Venkatesh (2012) also showed that the unneeded expression of the Lac proteins in the presence of glycerol imparted a reduction in growth. In their study, IPTG was used for protein over-expression. They revealed that this cost increased with increasing concentrations of IPTG; although β-galactosidase activity saturated at 25 μM IPTG, the reduction in growth reached 5 % at 50 μM IPTG. This elevated to 15 % with the addition of 100 μM IPTG and beyond. Thus, IPTG also appeared to contribute to the cost. It is thought that the uptake of IPTG involves the use of proton pumps, which may impose an energy burden on the cell (Krzewinski et al., 1996; Malakar & Venkatesh, 2012).

Kitagawa and colleagues (2005) reported that induced expression of the ASKA gene products with a high concentration of IPTG (1 mM) lead to growth defects. 51 % of the ASKA clones showed severe growth defects, and a further 28 % revealed moderate growth defects. Therefore, in the work presented here, a lower IPTG concentration (50 μM) that would minimize these effects and be sufficient to activate the pCA24N T5-*lacO* promoter was used. Taken together, it is likely that the control strain used in the fitness experiments had a reduction in fitness associated with expressing GFP, but that this was likely to be small. Nevertheless, the results show that most of the ASKA clones exhibited improved growth compared to the control. For future work, it will useful to carry out experiments that test explicitly whether GFP expression carries a fitness cost in this system (*E. coli* DH5α-E harbouring pCA24N-GFP tag in the presence of 50 μM IPTG).

The relationship between resistance and fitness appeared to be complex and it was hard to establish an obvious correlation between the two. Sometimes the fittest strain was also the most resistant and other times, a higher fitness advantage did not guarantee the highest level of resistance. The results suggest that it is not necessarily straightforward to determine how resistance will impact on bacterial fitness. Furthermore, it appears that sometimes, mutations that offer relatively lower fitness may in fact afford higher levels of resistance.

Overall, the results have shown that mutations enabling the over-expression of naturally occurring proteins from bacteria are gateways to a diversity of adaptive mechanisms. This work has shown that small increases in resistance can be associated with significant changes in fitness (both up and down). As a next step, a few examples from this Chapter were studied in more detail, to try and understand the underlying resistance/fitness mechanisms.

2.5 Materials and Methods

All reagents were from Sigma, unless otherwise specified. The materials and methods described here are specific to this chapter. Common materials and molecular biology techniques are described in Appendix I.

Enoxacin and sulfadiazine were dissolved in 1 N NaOH. Oxytetracycline (Duchefa Biochemie) was dissolved in water: ethanol (2:1). Spiramycin was dissolved in ethanol. Benzethonium chloride, lincomycin (Duchefa Biochemie), puromycin (Melford), sisomicin and vancomycin (Duchefa Biochemie) were dissolved in water.

The concentrations of the antibiotic stocks were: benzethonium chloride, 20 mg/mL; enoxacin, 8 mg/mL; lincomycin, 100 mg/mL; oxytetracycline, 20 mg/mL; puromycin, 20 mg/mL; sisomicin, 20 mg/mL; spiramycin, 20 mg/mL; sulfadiazine, 100 mg/mL; vancomycin, 50 mg/mL.

Powdered X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside; ForMedium) was prepared as 20 mg/mL stocks. It was dissolved in dimethyl sulfoxide (DMSO).

2.5.1 Antibiotic susceptibility testing

The ASKA clone and a negative control (*E. coli* DH5α-E with pCA24N-GFP) were used to inoculate 2 mL Lysogeny broth [LB; 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl (ForMedium)] supplemented with chloramphenicol (34 $\mu g.mL^{-1}$ final concentration) and incubated at 37 ºC overnight. Two hundred µL of each overnight culture was used to inoculate 10 mL LB supplemented with chloramphenicol (34 $\mu g.mL^{-1}$ final concentration) and incubated at 37 ºC with shaking. At early-log phase (OD_{600} ~0.2), IPTG (50 µM final concentration) was added to both cultures to induce protein over-expression for 2 h at 37 ºC. The cells were harvested by centrifugation (2000 xg, for 10 min at 4 ºC) and the pellets were washed with 6 mL sterile water and centrifuged again. The cells were resuspended to OD_{600} ~ 0.2 in sterile water. To determine the MIC values, one of two methods (Andrews, 2001; Wiegand et al., 2008) described below were used. The average of at least two independent experiments was determined as the MIC value.

2.5.1.1 MIC E-Test method

An E-test (AB bioMérieux) consists of a predefined antibiotic concentration gradient on a plastic strip. They were available for the following antibiotics; cephalothin, penicillin G, aztreonam, cefuroxime, tobramycin, trimethoprim, ciprofloxacin, nalidixic acid, erythromycin, nitrofurantoin and rifampicin.

One hundred µL of resuspended cells (see above) of the ASKA and control clones were transferred into 1.5 mL microcentrifuge tubes containing IPTG (50 µM final concentration) and the tubes were incubated at room temperature for 30 min. All of the cells were spread on Mueller-Hinton agar plates (BD Difco™; 2 g/L beef heart infusion, 17.5 g/L acid hydrolysate of casein, 1.5 g/L starch, 17 g/L agar) supplemented with chloramphenicol (34 $\mu g.mL^{-1}$ final concentration) and IPTG (50 µM final concentration). An E-test was placed in the middle of the plate on top of the bacterial lawn and incubated upside down at 37 ºC overnight. The MIC value was reported as the point the ellipse meets the strip from the zone of inhibition produced.

2.5.1.2 MIC microdilution method

The microdilution method was used to measure the MIC values for the following antibiotics; sisomicin, sulfadiazine, enoxacin, oxytetracycline, lincomycin, spiramycin, vancomycin, puromycin and benzethonium chloride. A 96-well microtitre plate containing Mueller-Hinton broth (200 µL final volume) supplemented with chloramphenicol (34 $\mu g.mL^{-1}$ final concentration), IPTG (50 µM final concentration) and serial dilutions of the antibiotic being tested in triplicates was prepared. Each well was inoculated with ~10^5 cells and monitored at OD_{600} using a Synergy 2 plate reader (Biotek Instruments; growth recorded every 10 min for 16 h). The number of cells was calculated based on the approximation of $OD_{600} \sim 1 = 2.5 \times 10^8$ cfu/mL (Sambrook & Russell, 2001). The MIC was defined as the lowest concentration of an antibiotic that prevented cell growth (which was established at $OD_{600} < 0.1$).

2.5.2 Fitness experiment

2.5.2.1 Creating lacZ-marked control strain

This experiment was performed in order to distinguish between two competing strains in the fitness experiments. Two plasmids were required; the first was the delivery plasmid (pUC18R6K mini-TN7T-*Gm-lacZ*) that contained the Tn*7* ends (TN*7*L and TN*7*R) flanking *lacZ* and was inserted in the *att*Tn7 site downstream of the chromosomal *glmS* gene. The second plasmid was pUXBF13, a helper plasmid that contained five genes *tns*ABCDE that encoded the transposase complex (Choi et al., 2005).

An aliquot of electrocompetent *E. coli* DH5α-E cells was transformed (Appendix I.13 and I.14) with the two plasmids at equal amounts (250 ng). The cells were electroporated and allowed to recover in 500 µL LB medium at 37 °C for 4 h. The long incubation period allowed for expression of the transposase complex, and the transposition event. Aliquots of cells (2 x 100 µL) were spread on LB agar plates supplemented with gentamicin (15 $\mu g.mL^{-1}$ final concentration) and X-gal (40 $\mu g.mL^{-1}$ final concentration). The plates were incubated at 37 °C overnight. The presence of blue colonies on the plates indicated successful integration of the ~5.2-kb *lacZ*/*GmR* fragment. This was also confirmed by PCR using a primer binding to the *att*Tn7 site (ecTn7R.for; Table I.3, Appendix I.5) and an *E. coli glmS* gene-specific primer (ecGlmS.rev; Table I.3, Appendix I.5), yielding a 700 bp product.

The gentamicin marker was flanked by FRT sites (FLP recognition sites) and was removed by Flp-mediated excision to create an unmarked mini-Tn*7* insertion. To do this, the pCP20 plasmid was employed (Table I.2, Appendix I.4). This plasmid contained ampicillin and chloramphenicol resistance markers, a thermo-inducible FLP recombinase expression system that acted on the FRT sites and a heat-sensitive replication origin (R6K) (Cherepanov & Wackernagel, 1995). The plasmid maintained low copy at 30 °C but was lost when incubated at 37 °C. The *lacZ*/Gm^R-tagged strain was made electrocompetent and transformed with the pCP20 plasmid. The cells were recovered with 500 µL SOC [(super optimal broth medium: 20 g/L tryptone, 5 g/L yeast extract, 10 mM NaCl, 2.5 mM KCl) and 20 mM glucose] and incubated at 30 °C for 1 h. Twenty µL of the culture was spread on an LB agar plate supplemented with ampicillin (100 $\mu g.mL^{-1}$ final concentration). The next day, ten colonies were inoculated in LB-ampicillin medium (100 $\mu g.mL^{-1}$ final concentration) and incubated overnight at 30 °C. One hundred µL of each overnight culture

was added to 10 mL LB broth and incubated at 30 °C until OD_{600} ~ 0.1. The cultures were shifted to 42 °C and incubated until OD_{600} ~ 1 (FLP synthesis and loss of the pCP20 plasmid are induced at this temperature). Four µL of each culture was spread on LB agar plates and incubated at 37 °C overnight. Individual colonies were picked and tested for loss of all antibiotic resistance.

Finally the *lacZ*-tagged strain was made electrocompetent and transformed with the pCA24N-GFP plasmid to construct the neutrally marked control strain.

2.5.2.2 Fitness assays

The ASKA strain and the negative control (marked with a *lacZ* gene) strain were streaked on LB agar plates supplemented with chloramphenicol (34 µg.mL^{-1} final concentration) and incubated overnight at 37 °C. The next day, 8 individual colonies of each strain were inoculated in 5 mL LB broth supplemented with chloramphenicol (34 µg.mL^{-1} final concentration) and IPTG (50 µM final concentration) and incubated at 37 °C for 24 h to pre-condition the clones to the competition medium.

Each overnight culture was diluted in LB to OD_{600} ~0.5-0.6 and then equal amounts of each strain [~ 10^6 cfu per strain; calculated based on the approximation of OD_{600} ~ 1 = 2.5 x 10^8 cfu/mL (Sambrook & Russell, 2001)] were mixed together (to obtain a 1:1 ratio) in a fresh 5 ml aliquot of LB medium supplemented with chloramphenicol (34 µg.mL^{-1} final concentration); this corresponded to the T_0 (initial) sample. This step was performed for the 8 replicates. The T_0 samples were then incubated at 37 °C for 24 h, this corresponded to the T_{24} (final) sample. The cell densities for the initial and final cultures were determined by diluting the cultures and spreading aliquots on LB agar plates supplemented with chloramphenicol (34 µg.mL^{-1} final concentration), IPTG (50 µM final concentration) and X-gal (5-bromo-4-chloro-indolyl-β-D-galactopyranoside) (40 µg.mL^{-1} final concentration). The rate of change in their relative abundance is used to calculate the relative fitness, *W* [estimate of number of doublings from initial (t = 0) to final (t = 24 h) time points, divided by number of doublings of the *lacZ*-marked negative control]. Thus, the impact of the trait being investigated can be assessed (Lenski et al., 1991).

CHAPTER 3

Mimicking evolutionary mechanisms using directed evolution

3.1 Introduction

3.1.1 Directed evolution on *cynT*, *yeaD* and *ydfW* genes

From the list of proteins in Table 1, a total of four was selected for further study: carbonic anhydrase (CA, product of the *cynT* gene), YeaD, YdfW and YejG (Chapter 4). The goal of this chapter was to determine whether the drug resistance functions of CA, YeaD and YdfW could be improved using directed evolution techniques. These experiments allow us to gain insight into how resistance functions evolve. When resistance is the result of a promiscuous function (e.g., CA enzyme), directed evolution experiments can be used to investigate the role of promiscuity in the evolution of new functions and to better understand protein evolution. In cases where proteins of unknown function (e.g., YeaD, YdfW and YejG) impart resistance, directed evolution can be used to investigate their function (Mastrobattista et al., 2005).

In the experiments described in this chapter, random mutations were introduced into copies of the *cynT*, *yeaD* and *ydfW* genes by error-prone PCR. The mutagenised inserts were ligated into the expression vector, pCA24N (carries a chloramphenicol resistance marker), and the pooled plasmids for the individual genes were used to transform *E. coli* DH5α-E cells. Mutants with the desired properties (in this case, improved antibiotic resistance) were selected on LB agar plates supplemented with IPTG, chloramphenicol (to maintain the pCA24N plasmid) and a second antibiotic (i.e., the drug used to test for the improved protein function) at concentrations that would identify mutants with enhanced resistance. To eliminate the possibility that improved resistance function was due to the proteins being fused to GFP (encoded by pCA24N-*cynT*, -*yeaD* and –*ydfW*), the GFP tags were removed (Section 3.4.1).

3.1.2 Directed evolution

Directed evolution mimics the process of natural evolution on a single gene, on a laboratory timescale. The first step in this technique is to create a large and diverse library of variants by introducing random mutations into copies of the target sequence. Variants within the library that possess the desired traits can then be identified by genetic selection or high-throughput screening.

Mutations can be introduced into the target gene by altering the reaction conditions of standard PCR or through the use of error-prone DNA polymerases. In the first method, the fidelity of the polymerase is reduced by adding Mn^{2+} or unbalanced ratios of dNTPs (Cadwell & Joyce, 1992), thus enhancing the mutation rate during the amplification step. In most cases, *Taq* is the polymerase of choice, in part because of its cheap cost. However, this polymerase is heavily biased towards introducing mutations at AT base pairs and not GC base pairs (Shafikhani, 1997). For this reason, the second method was used.

The error-prone PCR was performed using the GeneMorph II Random Mutagenesis Kit (Agilent technologies) that employs Mutazyme II DNA polymerase (a blend of Mutazyme I DNA polymerase and a novel *Taq* DNA polymerase mutant), which exhibits increased misinsertion and misextension frequencies compared to wild-type *Taq*. The combination of these two polymerases is thought to reduce bias and produce equivalent mutation rates at AT and GC sites, therefore promoting variation in the mutation type (http://www.genomics.agilent.com).

3.1.3 Carbonic Anhydrase

The *cynT* gene is part of the *cynTSX* operon, which includes three genes in the following order: *cynT* (carbonic anhydrase), *cynS* (cyanase) and *cynX* (putative cyanate transporter). The expression of the *cynTSX* operon is induced by cyanate and enables *E. coli* to use cyanate as a nitrogen source (Lamblin & Fuchs, 1994; Guilloton et al., 1992). Cyanase catalyses bicarbonate-dependent decomposition of cyanate into ammonia and carbon dioxide. The substrate of cyanase (bicarbonate) is formed by the hydration of carbon dioxide, which is catalysed by carbonic anhydrase (Tripp et al., 2001). This reaction ensures adequate levels of endogeneous bicarbonate inside the cell since carbon dioxide diffuses out of the cell faster than the non-catalysed hydration back to bicarbonate (Guilloton et al., 1992) (Figure 7). Thus, the biological role for carbonic anhydrase is to prevent the depletion of endogenous bicarbonate. Carbonic anhydrase is a zinc metalloenzyme of 219 residues (Tripp et al., 2001).

Carbonic anhydrase has been found in eukaryotes, bacteria and archaea, where it plays an important role in processes such as photosynthetic carbon fixation and respiration (Kimber & Pai, 2000; Smith & Ferry, 2000).

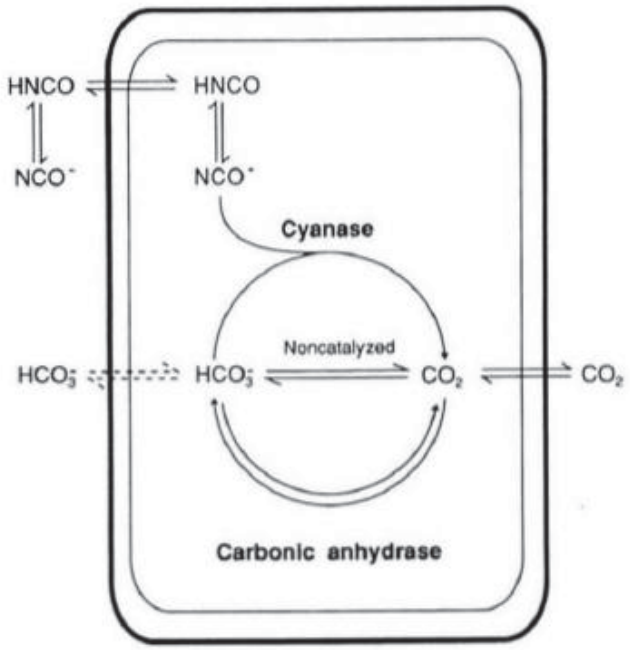


Figure 7. The function of carbonic anhydrase

Cyanate (NCO^-) diffuses freely into the cell (as cyanic acid, HNCO). Cyanate is then broken down by cyanase (encoded by *cynS*), resulting in the conversion of endogeneous bicarbonate (HCO^-_3) to carbon dioxide (CO_2). Carbonic anhydrase (encoded by *cynT*) catalyses the hydration of CO_2 to bicarbonate to prevent diffusion of CO_2 out of the cell. This diagram is taken from Guilloton et al. (1992).

In Chapter 2, it was determined that carbonic anhydrase-expressing cells show a 1.3-fold increase in resistance towards enoxacin compared to the control (Table 1). Enoxacin is a fluoroquinolone antibiotic that is used extensively in human medicine, mainly to treat urinary tract infections. It is a synthetic broad-spectrum bactericidal drug that acts by inhibiting bacterial gyrase, which is the DNA topoisomerase that negatively supercoils DNA (Ruiz, 2003).

Carbonic anhydrase was selected for further study because it is an enzyme that has been reported to have a promiscuous esterase activity (Pocker & Stone, 1965). This demonstrates that carbonic anhydrase can catalyse reactions other than its evolved function, suggesting that it may have a promiscuous activity that allows it to enzymatically modify enoxacin. In addition, unlike the other three proteins, carbonic anhydrase has a known function and has been well studied.

3.1.4 YeaD

The *E. coli yeaD* gene codes for a 294 amino acid protein that has been suggested to be a member of the galactose mutarotase-like superfamily (You et al., 2010). This family shares some structural similarities with known galactose mutarotases but there are differences between them in the substrate-binding pocket (You et al., 2010). Galactose mutarotase is an enzyme that plays an important role in the Leloir pathway. This pathway is involved in the metabolism of galactose to glucose-1-phosphate, which is then further converted into glucose 6-phosphate, an intermediate compound in glycolysis (Frey, 1996; You et al., 2010). Galactose mutarotase converts β-D-galactose to α-D-galactose, which is the substrate required for the next step in the Leloir pathway. This reaction is achieved by stereochemical inversion at carbon 1 of galactose (Holden et al., 2003).

Sequence analysis show that the *E. coli* YeaD protein shares 82 % sequence identity with the *Salmonella typhimurium* homologue, which is also considered a galactose mutarotase-like protein (Figure 8) (Chittori et al., 2007). I hypothesised that identifying mutants with enhanced activities might shed light on the function of YeaD.

YeaD over-expression endowed cells with a modest increase in resistance to aztreonam (1.5-fold) compared to the control (Table 1). Aztreonam is a bactericidal, monocyclic β-lactam (monobactam) originally isolated from *Chromobacterium violaceum*. The drug binds to the penicillin binding protein 3 to inhibit bacterial cell wall synthesis and shows broad activity against Gram-negative aerobes (Hamed et al., 2013). Aztreonam is used to treat respiratory and urinary tract infections, and is resistant to hydrolysis by the β-lactamases produced by many Gram-negative (and Gram-positive) pathogens (Sykes & Bonner, 1985).

YeaD was selected because although it has an unknown function, it is long enough to encode an enzyme-sized protein and it has been suggested to be a galactose mutarotase-like protein (You et al., 2010).

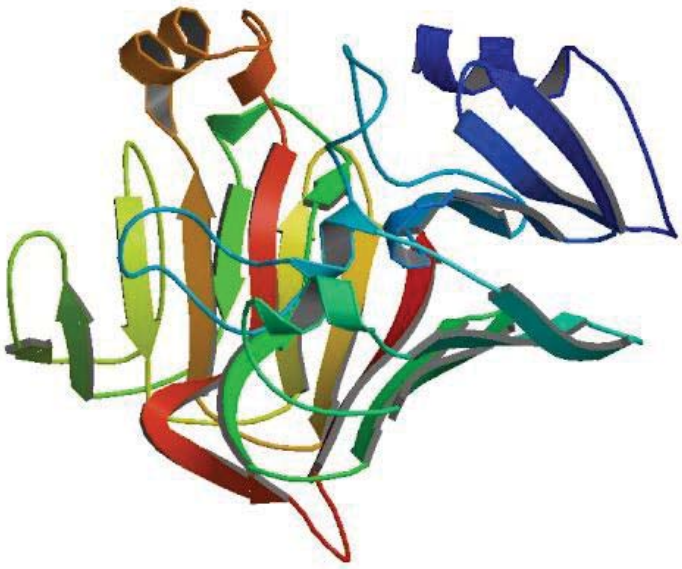


Figure 8. Structure of YeaD from *Salmonella typhimurium* (PDB code: 2HTB)

Salmonella typhimurium YeaD (294 amino acids) consists mainly of β-sheets (46 %) with a low α-helical content (7 %). The protein displays a β-sandwich fold similar to galactose mutarotases. This image is taken from Chittori et al. (2007).

3.1.5 YdfW

YdfW is an uncharacterized protein of 49 residues that is found in the Qin prophage (Keseler et al., 2011). It was the smallest out of the proteins chosen. YdfW-expressing cells were selected from environments containing three macrolide antibiotics (erythromycin, spiramycin, oleandomycin), and also chelerythrine (Soo et al., 2011). This suggests that the protein may mediate a broad-specificity response towards macrolides.

Macrolides are bacteriostatic drugs that reversibly bind to domain V of the 23S ribosomal RNA in the 50S subunit of the bacterial ribosome. They interfere with protein synthesis by preventing processes such as transpeptidation and translocation reactions (Hansen et al., 1999). Macrolides are most effective against Gram-positive bacteria and they are used to treat upper and lower respiratory tract infections, as well as sexually transmitted diseases such as gonorrhea and chlamydia (Zuckerman, 2004). Chelerythrine is an alkaloid extracted from the plant *Chelidonium majus L.* and is used in Chinese herbal medicine and in Europe to treat several diseases, including bronchitis, asthma and jaundice (Colombo & Bosisio, 1996).

YdfW was chosen because it was selected from environments containing three different macrolides (erythromycin, spiramycin and oleandomycin) and this is the first hint of any function for the protein.

3.2 Results

3.2.1 *cynT* error-prone PCR library

Directed evolution experiments were performed on *cynT* to determine whether variants that conferred increased resistance to enoxacin could be identified (Section 3.4.2 and 3.4.5). I hypothesised that this approach would demonstrate that this promiscuous activity is evolvable.

After introducing random mutations in the *cynT* gene by error-prone PCR and transforming *E. coli* DH5α-E cells with the pooled plasmids (Section 3.4.2), a library of *cynT* mutants, consisting of 1.4 x 10^7 clones (excluding vector-only background) was generated. To gain an overview of the mutational spectrum in the variants of the library, the *cynT* inserts from sixteen random clones were amplified by PCR using plasmid backbone-specific primers (pCA24N.for and pCA24N.rev2; Table I.3, Appendix I.5) and then sequenced (Section 3.4.3). Among the sixteen inserts sequenced, there was a total of 121 point mutations, as well as one insertion and two deletions; therefore, the dataset revealed a total of 124 mutations (Table 3). The number of mutations found in each gene ranged between 3 and 12. There were no wild-type genes present in the sequenced samples and as a consequence the overall mutation rate was ~ 8 mutations per *cynT* gene. However, there were more A→T and T→A mutations (27.4 %) than G→C and C→G mutations (3.2 %), indicating that, like *Taq*, the Mutazyme II polymerase was still biased towards intoducing mutations at AT base pairs.

To estimate the amino acid diversity from the mutational spectrum, the PEDEL-AA program was used (Firth & Patrick, 2008). PEDEL-AA calculated the mean number of amino acid substitutions to be ~ 5 per *cynT* variant in the library. This amino acid substitution rate is moderately high and generally not preferred because lower error rates, 1-2 amino acid substitutions per gene, minimizes the accumulation of deleterious mutations in each variant (Tracewell & Arnold, 2009). However, this moderately high mutation rate means that only ~3 % of the library is likely to consist of wild-type *cynT* (Table 4). Overall, the PEDEL-AA analysis estimated that the library was likely to encode ~7.4 × 10^6 different protein sequences. The mutation rate was as predicted from the manufacturer's protocol (Section 3.4.2.1). As a result, the number of wild-type sequences in this library was reduced.

Table 3. The mutational spectrum of the *cynT* mutant library

Type(s) of mutations	Frequency	Proportion of total
Transitions		
A→G, T→C	28	22.6%
G→A, C→T	28	22.6%
Transversions		
A→T, T→A	34	27.4%
A→C, T→G	9	7.3%
G→C, C→G	4	3.2%
G→T, C→A	18	14.5%
Insertions and deletions		
Insertions	1	0.8%
Deletions	2	1.6%
Summary of bias		
Transitions/transversions	0.9	NA[a]
AT→GC/GC→AT	1	NA[a]
A→N, T→N	71	57.3%
G→N, C→N	50	40.3%
Mutation rate		
Mutations per kb	11.8	NA[a]
Mutations per *cynT* gene[b]	7.8	NA[a]

[a]NA: not applicable.

[b]The cloned *cynT* insert was 657 bp.

Table 4. Composition of the *cynT* mutant library as estimated by PEDEL-AA

Property	Estimate
Total library size	1.4×10^7
Number of variants with no insertions, deletions, or stop codons	9.0×10^6
Mean number of amino acid substitutions per variant	5.5
Unmutated (wild-type) sequences (% of total library)	3.0%
Number of distinct, full-length proteins in the library	7.4×10^6

3.2.1.1 Selecting *cynT* 'winners'

To determine the number of mutant clones to sample under selective conditions, GLUE, a program that estimates the sampling statistics for randomised libraries in which each variant is equally likely, was used (Patrick et al., 2003). For my library of 1.4 x 10^7 *cynT* clones, GLUE estimated that spreading ~ 9.7 x 10^7 cells on enoxacin-containing plates would ensure that 99.9 % of these clones would be represented in the selection experiment.

The lowest concentration of enoxacin that the strain expressing wild-type carbonic anhydrase could not grow on was measured using the agar dilution method (Section 3.4.4). This was determined by spreading carbonic anhydrase-expressing cells directly onto LB agar plates that have incorporated different antibiotic concentrations of enoxacin. The MIC value of the strain expressing native carbonic anhydrase for enoxacin is 1 $\mu g.mL^{-1}$. Thus, mutants that are able to grow on plates containing $\geq$ 1 $\mu g.mL^{-1}$ of enoxacin indicates improved resistance.

The *cynT* mutant library was spread on LB agar plates containing enoxacin at concentrations of 1 and 4 $\mu g.mL^{-1}$. After 24 h of incubation, there were ~ 13,000 colonies present on the plate containing 1 $\mu g.mL^{-1}$ enoxacin (0.013 % survival rate) and no colonies on the 4 $\mu g.mL^{-1}$ plate. Individual colonies could not be easily picked off the plate with 1 $\mu g.mL^{-1}$ enoxacin, so the cells were scraped off the plate and used for further selection at enoxacin concentrations of 1.5, 2, 3 and 3.5 $\mu g.mL^{-1}$. This time, ~ 10^5 recovered cells were spread on each plate. After 24 h, there were 181, 150, 43 and 35 colonies on the 1.5, 2, 3 and 3.5 $\mu g.mL^{-1}$ plates, respectively.

3.2.1.2 DNA sequencing of *cynT* mutants

To confirm that the clones on the plates harboured the plasmid-encoded *cynT* gene (and were not contaminants), a total of ten colonies were picked at random off the 2, 3 and 3.5 $\mu g.mL^{-1}$ plates and screened by PCR using plasmid backbone-specific primers (pCA24N.for and pCA24N.rev2; Table I.3, Appendix I.5) (Section 3.4.6). The PCR yielded products of the expected size (~ 850 bp) and so they were sequenced.

Out of the ten colonies sequenced all of them were mutants; a total of five distinct mutants were present (Table 5). One of the mutants (mutant 1) was able to grow at enoxacin concentrations of up to 3.5 $\mu g.mL^{-1}$, exhibiting a 3.5 fold increase in resistance, after only

one round of error-prone PCR. Surprisingly, the sequence of mutant 1 revealed a truncated protein of only 95 amino acids (~ 60 % of protein truncated; full length CA is 219 amino acids), with three amino acid substitutions (Table 5). The other four mutants contained between 2 and 5 amino acid substitutions. There were no mutations present in the active site at the zinc ligands, Cys 160, His 220 and Cys 223 (Kimber & Pai, 2000). Mutant 3 contained a premature nonsense mutation, which terminated the protein after 73 amino acids (~ 70 % of protein truncated). With two out of the five proteins being truncated, it raised suspicions that the mutants might be false positives. To confirm whether the plasmid-encoded *cynT* gene was responsible for the improved resistance to enoxacin, a retransformation test was performed.

Table 5. Mutations in five selected carbonic anhydrase variants

Mutant	Amino Acid change	Type of mutation
1	P46P	*Silent*
	N65Y	Substitution
	Y70C	Substitution
	V85F	Substitution
	I93I	*Silent*
	C96*	Nonsense
2	E20G	Substitution
	L26P	Substitution
	I66N	Substitution
	E73K	Substitution
	S91S	*Silent*
	E151G	Substitution
	E172E	*Silent*
3	E73*	Nonsense
4	L22M	Substitution
	L57Q	Substitution
	S91Y	Substitution
	I93I	*Silent*
	E151G	Substitution
	R207R	*Silent*
5	V49V	*Silent*
	G71G	*Silent*
	L141L	*Silent*
	I154I	*Silent*
	E172K	Substitution
	A210D	Substitution
	/	*Silent*

3.2.1.3 Retransformation tests and potential mutations in chromosomal genes

A retransformation test serves to confirm the origin of an improved cell phenotype, such as increased resistance to enoxacin. By confirming that the improved phenotype is conferred by a plasmid-encoded gene, the test rules out a spontaneously-acquired chromosomal mutation as the source of resistance.

Plasmids were isolated from mutants 1-5 and introduced into fresh *E. coli* DH5α-E cells. The transformed cells were grown in the same manner as the mutant library and streaked on plates containing the same enoxacin concentrations as before, i.e., 1.5, 2, 3 and 3.5 $\mu g.mL^{-1}$. After overnight incubation, no colonies were present on any of the plates. There was still no growth after 72 h of incubation. The retransformation test was repeated several times with fresh plasmid isolates and fresh cells but without success. Interestingly, when mutants 1 and 2 (picked off the original library plates) were streaked on plates containing the same enoxacin concentrations, there was growth for each mutant on all of the plates. The results suggested that the mutated, plasmid-encoded *cynT* genes were not responsible for endowing the cells with resistance to enoxacin. Instead, it suggested that mutants 1-5 might have acquired beneficial chromosomal mutations that were the sources of resistance. To confirm this hypothesis, mutants 1 and 2 were passaged in the absence of chloramphenicol (which is the antibiotic used to maintain pCA24N-*cynT*), in an attempt to cure them of their plasmids. The cells could then be tested for resistance to enoxacin at the same concentrations as before. However, after passaging the cells for ~130 generations, the plasmid continued to persist inside the cells and/or the cultures were becoming contaminated. For this reason, this approach was abandoned and instead, it was investigated whether genes known to confer resistance to fluoroquinolones may have acquired beneficial mutations over the course of the experiments.

The *gyrA, gyrB, parC* and *parE* genes were selected as candidates because fluoroquinolone resistance is most commonly attributable to them (Blondeau, 2004; Hooper, 1999). The *gyrA* and *gyrB* genes code for DNA gyrase, a tetrameric type II topoisomerase composed of two GyrA and two GyrB subunits. The *parC* and *parE* genes code for a tetrameric topoisomerase (type IV) that is homologous to GyrAB (Ruiz, 2003). The *gyrA* and *gyrB* genes are 2,628 bp and 2,415 bp in length, respectively, and have been mapped to widely separated regions of the *E. coli* chromosome (Lampe & Bott, 1985). Three sets of primers (gyrA_, gyrB_, parC, and parE_setA to C; Table I.3, Appendix I.5) were designed to amplify each gene (Figure 9). To cover the entire sequence, one of the primers was designed to bind 30 bp upstream and another 24 bp downstream of *gyrA*, and a total region of 2,682 bp was amplified. For *gyrB*, one of the primers bound 36 bp upstream and the other bound 25 bp downstream and a total region of 2,476 bp was amplified. The *parC* and *parE* genes are 2,259 bp and 1,893 bp in length, respectively, and are also located on separate regions of the chromosome (ecocyc.org). To amplify each gene, again three

sets of primers were designed to amplify a total region of 2,308 bp and 1,960 bp for *parC* and *parE*, respectively. Twenty-three bp upstream and 26 bp downstream of *parC* and 35 bp upstream and 32 bp downstream of *parE* were targeted, to ensure that both genes were sufficiently covered.

The *gyrA*, *gyrB*, *parC* and *parE* genes of mutant 1 and a wild-type *E. coli* DH5α-E clone (reference sequence) were amplified by PCR using the appropriate primers, and then sequenced. The results showed that the mutant and the *E. coli* DH5α-E clone had the same DNA sequences, indicating that the topoisomerase enzymes were unlikely to be responsible for the increase in resistance to enoxacin observed in mutant 1. A large number of other genes could also be responsible for the resistance phenotypes observed, including those involved in membrane permeability and/or efflux pump systems (Ruiz, 2003). Therefore, a decision was made not to sequence the other mutants. Since two rounds of selection had already been performed on the *cynT* mutant library and yet variants with increased resistance to enoxacin could not be identified, the focus was shifted on the other two selected proteins (YeaD and YdfW).

Figure 9. Approximate locations of primer pairs designed to amplify the *gyrA* gene

Cartoon showing the approximate locations of the three primer pairs designed to amplify the chromosomal *gyrA* gene for sequencing. Primers for the other three genes, *gyrB*, *parC* and *parE* were designed to bind in similar locations on each gene. The primers amplified gene fragments of ~ 600 to 1000 bp and are indicated by small arrows. Primer pairs are represented by the same colour.

3.2.2 *yeaD* error-prone PCR library

Using the same approach as described previously for *cynT*, random mutations were introduced into the *yeaD* gene by error-prone PCR (Section 3.4.2). A library of mutants consisting of 4 x 10^6 clones (excluding vector-only background) was generated. Ten random clones from the library were amplified by PCR using plasmid backbone-specific primers (pCA24N.for and pCA24N.rev2; Table I.3, Appendix I.5). The PCR products were sequenced to gain an overview of the diversity in mutation types (Section 3.4.3). A total of 36 point mutations were present. The number of mutations found in each clone ranged from 1 to 8. There were no insertion or deletion mutations present and the overall mutation rate was ~ 4 per *yeaD* gene. Again, G→C and C→G mutations (5.5 %) occurred much less frequently than A→T and T→A mutations (33.3 %) (Table 6). There was also a heavy bias towards A→N, T→N mutations (22 of 36 mutations) compared to G→N, C→N mutations (14 of 36 mutations).

The mean number of amino acid substitutions, as estimated by PEDEL-AA was ~ 3 per variant. This low mutation rate meant that ~16 % of the library was likely to consist of wild-type *yeaD* (Table 7) and that the library was likely to include multiple copies of some variants. This mutation rate was lower than predicted from the manufacturer's protocol (Section 3.4.2.1). Overall, PEDEL-AA analysis estimated that the library is likely to encode ~1.9 × 10^6 different protein sequences.

Table 6. The mutational spectrum of the *yeaD* mutant library

Type(s) of mutations	Frequency	Proportion of total
Transitions		
A→G, T→C	7	19.4%
G→A, C→T	6	16.6%
Transversions		
A→T, T→A	12	33.3%
A→C, T→G	3	8.3%
G→C, C→G	2	5.5%
G→T, C→A	6	16.6%
Insertions and deletions		
Insertions	0	0%
Deletions	0	0%
Summary of bias		
Transitions/transversions	0.6	NA[a]
AT→GC/GC→AT	0.8	NA[a]
A→N, T→N	22	61.1%
G→N, C→N	14	38.9%
Mutation rate		
Mutations per kb	4.1	NA[a]
Mutations per *yeaD* gene[b]	3.6	NA[a]

[a]NA: not applicable.
[b]The cloned *yeaD* insert was 882 bp.

Table 7. Composition of the *yeaD* mutant library as estimated by PEDEL-AA

Property	Estimate
Total library size	4.0×10^6
Number of variants with no insertions, deletions, or stop codons	3.4×10^6
Mean number of amino acid substitutions per variant	2.6
Unmutated (wild-type) sequences (% of total library)	15.8%
Number of distinct, full-length proteins in the library	1.9×10^6

3.2.2.1 Selecting and testing *yeaD* 'winners'

The same method used in the *cynT* experiment was employed to determine the lowest aztreonam concentration that the strain expressing wild-type YeaD could not grow on (Section 3.4.4**)**. The MIC value of this strain for aztreonam is 0.1 $\mu g.mL^{-1}$. Mutants from the *yeaD* library that are able to grow on plates containing aztreonam concentrations of $\geq$ 0.1 $\mu g.mL^{-1}$ show enhanced resistance.

GLUE estimated that spreading ~ 2.7 x 10^7 cells would ensure 99.9 % sampling of the *yeaD* library (4 x 10^6 clones). The *yeaD* mutant library was spread on LB agar library plates containing aztreonam at concentrations of 0.14, 0.19, 0.38 and 0.80 $\mu g.mL^{-1}$. After 48 h of incubation, only 4 colonies and 6 colonies grew on the 0.14 and 0.19 $\mu g.mL^{-1}$ plates, respectively. There were no colonies on the 0.38 and 0.80 $\mu g.mL^{-1}$ plates. In spite of the extremely low survival rate, retransformation tests were performed.

Plasmid DNA was isolated from the 10 clones that grew on the plates containing 0.14 and 0.19 $\mu g.mL^{-1}$ aztreonam, and then it was used to transform fresh *E. coli* DH5α-E cells. To confirm the presence of the pCA24N plasmid, the transformed cells were spread on LB agar plates supplemented with chloramphenicol at a concentration (34 $\mu g.mL^{-1}$) that inhibits the growth of cells free of plasmid. The transformed cells were grown in the same manner as the mutant library had been, and streaked on plates containing aztreonam at the same concentrations. After 48 h of incubation, one out of the 10 tranformants gave rise to a total of 2 and 3 colonies on the 0.14 $\mu g.mL^{-1}$ and 0.19 $\mu g.mL^{-1}$ plates, respectively. Given that a large number of transformed cells had been spread on each plate (Section 3.4.7), this was an extremely low survival rate and suggested that the plasmid-encoded *yeaD* variant was not responsible for growth at these aztreonam concentrations. To gain a better understanding of the results, two of the original mutants were grown in the same manner as the retransformed cells, and streaked on plates containing aztreonam at the same concentrations as before. After 24 h, both mutants grew as a lawn up to 0.19 $\mu g.mL^{-1}$ aztreonam. These results suggested that the cells were acquiring beneficial chromosomal mutations during the course of my experiments (similar to the *cynT* results). Therefore, YdfW was investigated to determine whether this protein would yield different results.

3.2.3 *ydfW* error-prone PCR library

The same approach used to create the *cynT* and *yeaD* libraries was employed. Random mutations were introduced into the *ydfW* gene by error-prone PCR (Section 3.4.2) and a *ydfW* mutant library of 3.8 x 10^6 clones (excluding vector-only background) was generated. Among ten inserts sequenced, there was a total of 14 mutations and no insertions or deletions present. There were fewer mutations here than for the *yeaD* clones (a total of 36 mutations), which is not surprising since *ydfW* is a much smaller gene (150 bp). Indeed, the mutation rate per kb was over twice as high in the *ydfW* library as it was in the *yeaD* library (Table 6). The number of mutations found in each clone ranged between 0 and 4. Three out of the ten genes were wild-type *ydfW*, and the overall mutation rate was 1.4 mutations per gene. There were far more A→T and T→A substitutions than any other transversion mutation (purine to pyrimidine substitution or vice versa) (6 out of 14 mutations). As a result, A→N and T→N mutations were overrepresented (10 out 14 mutations) (Table 8).

The mean number of amino acid substitutions calculated by PEDEL-AA was low, at 1 per *ydfW* gene, and meant that a large percentage of the library, ~ 40 %, was likely to consist of wild-type *ydfW* (Table 9). Overall, the library was estimated to contain ~ 3.5 x 10^5 distinct protein sequences.

Table 8. The mutational spectrum of the *ydfW* mutant library

Type(s) of mutations	Frequency	Proportion of total
Transitions		
A→G, T→C	3	21.4%
G→A, C→T	2	14.3%
Transversions		
A→T, T→A	6	42.9%
A→C, T→G	1	7.14%
G→C, C→G	1	7.14%
G→T, C→A	1	7.14%
Insertions and deletions		
Insertions	0	0%
Deletions	0	0%
Summary of bias		
Transitions/transversions	0.5	NA[a]
AT→GC/GC→AT	1.3	NA[a]
A→N, T→N	10	71.4%
G→N, C→N	4	28.6%
Mutation rate		
Mutations per kb	9.5	NA[a]
Mutations per *ydfW* gene[b]	1.4	NA[a]

[a]NA: not applicable.

[b]The cloned *ydfW* insert was 147 bp.

Table 9. Composition of the *ydfW* mutant library as estimated by PEDEL-AA

Property	Estimate
Total library size	3.8×10^6
Number of variants with no insertions, deletions, or stop codons	3.3×10^6
Mean number of amino acid substitutions per variant	1.0
Unmutated (wild-type) sequences (% of total library)	39.6%
Number of distinct, full-length proteins in the library	3.5×10^5

3.2.3.1 Selecting and testing *ydfW* 'winners'

The selection experiment was performed in the presence of a macrolide (spiramycin) because macrolides are clinically relevant drugs. Furthermore, the YdfW-expressing strain was selected more times in the presence of macrolides and the MIC values had already been measured for this class of antibiotic (Table 1).

Again, the same method used earlier (agar dilution method; Section 3.4.4**)** was employed to determine the MIC value of the strain expressing wild-type YdfW for spiramycin (MIC value = 345 $\mu g.mL^{-1}$). *ydfW* mutants that are able to grow on plates containing spiramycin concentrations of $\geq$ 345 $\mu g.mL^{-1}$ indicates improved resistance.

For the *ydfW* library of 3.8 x 10^6 clones, GLUE estimated that spreading ~ 2.6 x 10^7 cells would give 99.9 % coverage. The library was spread on plates that contained spiramycin at concentrations of 345, 350, 360 and 370 $\mu g.mL^{-1}$. After 24 h of incubation, there were ~3,700, ~3,600, ~2,800 and ~2,100 colonies on the 345, 350, 360 and 370 $\mu g.mL^{-1}$ plates, respectively.

Retransformation tests were performed using plasmid DNA isolated from clones picked off each of the library plates. Plasmid DNA was isolated from a total of 10 random clones (two off the 345 and 360 $\mu g.mL^{-1}$ plates, and three off the 350 and 370 $\mu g.mL^{-1}$ library plates), and introduced into fresh *E. coli* DH5α-E cells. To confirm that the cells were carrying pCA24N-*ydfW*, three random plasmids out of the ten were amplified by PCR using plasmid backbone-specific primers (pCA24N.for and pCA24N.rev2; Table I.3, Appendix I.5) and yielded PCR products of the expected size (~ 350 bp). The transformed cells were grown in the same manner as the mutant library and streaked on spiramycin plates with the same concentrations as before (345, 350, 360 and 370 $\mu g.mL^{-1}$). After 24 h of incubation, there were lawns of colonies on all of the plates, for all of the clones.

To confirm that the results were not false positives, fresh stocks of spiramycin and chloramphenicol were prepared. In addition, fresh transformations with the ten isolated plasmids, using new aliquots of electrocompetent *E. coli* DH5α-E cells, were performed. As before, the freshly transformed cells, along with the (wild-type) YdfW-expressing strain and the negative control strain (carrying an empty pCA24N plasmid) were streaked on plates containing spiramycin at concentrations of 345, 350, 360, 370, 400, 500 and 690 $\mu g.mL^{-1}$. After 24 h of incubation, there were lawns on all of the plates, for all of the clones.

These results were very unusual since the MIC values for the wild-type and the negative control clones had been measured previously and determined to be 345 $\mu g.mL^{-1}$.

The experiment was repeated a final time, this time with only one transformed clone and the wild-type and the negative control (carrying an empty pCA24N plasmid) strains. The cells were streaked on 370, 500 and 690 $\mu g.mL^{-1}$ spiramycin plates. After 24 h of incubation, there were lawns for all of the three strains on the 370 $\mu g.mL^{-1}$ plate. There were 43, 53, and 90 colonies on the 500 $\mu g.mL^{-1}$ plate and 17, 26 and 53 colonies on the 690 $\mu g.mL^{-1}$ plate, for the transformed, wild-type and control strains, respectively. The clones appeared to have varying spiramycin resistances, even though they were grown and treated in the same manner, in each experiment. Due to these highly inconsistent results, no more time was spent working on YdfW.

3.3 Discussion

In this chapter, the aim was to evolve the antibiotic resistance functions of CA, YeaD and YdfW by directed evolution. The process involved introducing random mutations into copies of the genes coding for each of the three proteins to create libraries of mutants. This was followed by selection of variants with improved function in the presence of increasing antibiotic concentrations. Despite constructing libraries that contained a large pool of distinct variants for each protein, no improved variants were identified.

In earlier experiments, over-expression of CA, YeaD and YdfW endowed the bacterium with resistance to a fluoroquinolone (enoxacin), a β-lactam (aztreonam) and macrolides (erythromycin, spiramycin and oleandomycin), respectively (Table 1, Chapter 2). Since CA had previously been found to possess a promiscuous esterase activity (Pocker & Stone, 1965), it was believed that the enzyme might be exhibiting another form of promiscuous function. On the other hand, YeaD and YdfW have unknown functions, although YeaD is predicted to be a galactose mutarotase-like protein (You et al., 2010). Because of its small size (49 amino acids), YdfW was unlikely to be acting as an enzyme, instead the protein may have been interacting with some other protein that has an activity towards macrolides or be involved in a signaling pathway. It is possible that over-expressing the proteins mediates pleiotropic effects by altering the expression of downstream resistance determinants. I hypothesized that performing directed evolution experiments on CA, YeaD and YdfW would help gain insight into the catalytic mechanisms of CA and help shed light on the functions of both YeaD and YdfW. Indeed, one of the significant advantages of directed evolution is that neither structural nor catalytic information is required to guide the experiment. Directed evolution has been previously used to improve the activity of a protein of unknown function (Mastrobattista et al., 2005) and can provide useful insights into catalytic mechanisms of enzymes (Künzler et al., 2005).

Some key steps in directed evolution have been proposed (Romero & Arnold, 2009). First, is the identification of a good starting sequence; second, the introduction of mutations in this target gene to construct a library of variants; and third, the selection of variants with improved function. It is possible that any one of these steps might have affected the outcome of this study. Before considering the first step, I wish to examine the other two steps (library construction and selection of improved variants).

Libraries consisting of 7.4 x 10^6, 1.9 x 10^6 and 3.5 x 10^5 distinct protein sequences for CA, YeaD and YdfW, respectively, were created. The size and mutation spectra of these libraries are similar to those obtained in other studies. For example, in a work by Patrick and Matsumura (2008), directed evolution was performed on the glutamine phosphoribosylpyrophosphate amidotransferase (PurF) protein to improve its promiscuous phosphoribosylanthranilate isomerase (PRAI) activity. An evolved PurF protein with ~ 25-fold higher activity (k_{cat}/K_M) was identified from a library of 6.4 x 10^5 variants. The size of the *purF* library is smaller than the CA and YeaD libraries. The *purF* error-prone library also showed similar mutational bias; mutations at AT base pairs occurred more frequently than mutations at GC base pairs and certain types of mutations like A→T or T→A were more abundant than others, like A→C or T→G. In another study, the promiscuous esterase activity of the human CA II (hCAII) was successfully evolved by directed evolution Gould and Tawfik (2005). This group discovered a variant with 40-fold higher rates toward 2-naphthyl acetate relative to the wild-type hCAII. The average mutation rate per gene in this study (between 1 and 7) was comparable to the mutation rate observed in the *cynT* error-prone library in this study (~ 8 per gene). Therefore, the size or the mutational diversity of the *cynT*, *yeaD* and *ydfW* libraries is unlikely to be the reason why no improved variant was identified in this study.

It is possible that the selection system used in this study might not have been elegant enough. When the MIC values were measured by the plate-based (E-test) or broth-based methods (Section 2.5.1), the CA-, YeaD- and YdfW-expressing strains all showed an increase in resistance (between 1.3- to 3-fold) compared to the negative control (Table 1). However, when the agar dilution method was used to measure the MIC values, none of the clones displayed any resistance when compared to the negative control (expressing GFP); MIC value of CA-expressing strain and the negative control for enoxacin is 1 $\mu g.mL^{-1}$; MIC value of YeaD-expressing strain and the negative control for aztreonam is 0.1 $\mu g.mL^{-1}$ and; MIC value of YdfW-expressing strain and the negative control for spiramycin is 345 $\mu g.mL^{-1}$. This may have suggested that the system used for the selection process may not have been sensitive enough to distinguish small differences in resistance (up to 3-fold) for the antibiotics tested (enoxacin, aztreonam and spiramycin) and thus not sensitive enough to detect variants with improved antibiotic function.

An earlier study by Counts and colleagues (1977) also highlighted differences in MIC results between the broth and agar methods. They attributed this variance to a number of

factors, including the testing method, inoculum size and the medium used. It is also possible that *cynT*, *yeaD* and *ydfW* are all false positives. However, MICs are considered the 'gold standard' for determining the susceptibility of organisms to antibiotics (Andrews, 2001). The methods used to measure the MIC values in the first part of this study (Section 2.5.1) followed recommended procedures (Andrews, 2001; Wiegand et al., 2008). These experiments were performed in triplicate and the average of the experimental repeats, the strains expressing CA, YeaD and YdfW were more resistant than the negative control. This suggests that for future experiments, performing the library selection process in broth might be a better approach. This method was used by Barlow and Hall (2002) in a work involving the TEM-1 β-lactamase gene and yielded successful results.

A study has suggested that some proteins are more evolvable than others (O'Loughlin et al., 2006). Although the molecular mechanisms that contribute to evolvability have been discussed (Khersonsky et al., 2006; Khersonsky & Tawfik, 2010), indicators of evolvability have also been proposed. One such indicator is the natural functional diversity of a protein family. For example, cytochrome P450s exhibit functional diversity in nature and are easier to evolve) (Tracewell & Arnold, 2009).

Others have suggested that evolvability is a function of stability. This view has been confirmed experimentally in a study where marginally stable and thermostable variants of cytochrome P450 BM3 were mutated (Bloom et al., 2006). They showed that the mutated versions of the thermostabilized parent were more likely to evolve a novel function towards naproxen, as the thermostabilized parent was more likely to withstand the destabilising mutations that were required to gain this new function. Miller and colleagues (2006) on the other hand have shown that the adaptive landscape can be highly constrained, whereby certain evolutionary paths are not always accessible. Taken together, these studies suggest that not being able to evolve CA, YeaD and YdfW might not necessarily have been due to inadequate sampling of the protein sequence space i.e., space of all possible protein sequences. One explanation for the results obtained might be evolutionary constraints; that is, evolving the proteins towards the desired functions might not be easily ascertainable.

The outcome of directed evolution (or, indeed, evolution) is difficult to predict *a priori* (O'Loughlin et al., 2006). However, it is generally accepted that when a promiscuous activity is present, it provides the starting point for optimization towards a new function (O'Brien & Herschlag, 1999; Khersonsky et al., 2006). Directed evolution experiments

have supported this view and shown that the promiscuous activities of enzymes can easily be improved (Aharoni et al., 2005). However, in addition to creating new enzymes, directed evolution can also be used to explore the limits of protein function. For a better understanding of this, it is important to consider the source of promiscuity. Some proteins are able to catalyse reactions for which the presence of one amino acid residue in the right location is the only requirement. For example, the human serum albumin protein is able to perform a promiscuous activity [the Kemp elimination; conversion of benzisoxazoles to *o*-cyano phenol derivatives (D'Anna et al., 2008)] owing to the presence of a lysine residue (James & Tawfik, 2001). Although these activities may provide the starting point for the evolution of new functions, the catalytic machinery may be too rudimentary and may require major modifications (sometimes to the protein architecture), changes that may not be accessible by directed evolution.

Furthermore, directed evolution tends to be limited to small steps (i.e., single base changes), thereby restricting the amino acids that are sampled at each position (Tracewell & Arnold, 2009). Thus, with reference to the first key step in directed evolution, perhaps the genes selected for this study may not have been 'good starting' sequences. Thus, there were no improved variants in the mutant libraries created and the only way for the bacterium to survive in the presence of the antibiotic was to acquire rare beneficial mutations in the chromosome.

Failure to evolve improved variants CA, YeaD and YdfW does not take away from their potential importance as contributors to clinically relevant resistance. Gene amplification has been observed to play a role in drug resistance. Over-expression of latent resistance determinants may serve as an initial or immediate solution to adaptation (Sandegren & Andersson, 2009). This study provides evidence for the key role of these, low-level resistance determinants in nature. I hypothesise that CA, YeaD and YdfW can 'buy time' for other, larger-effect mutations to accumulate.

Finally, the experiments described in this Chapter allowed me to optimise protocols for constructing and analysing error-prone PCR libraries. These protocols have recently been published (Hanson-Manful & Patrick, 2013), and the reprint is provided in Appendix III.

3.4 Materials and Methods

All reagents were from Sigma, unless otherwise specified. The materials and methods described here are specific to this chapter. Common materials and molecular biology techniques are described in Appendix I.

3.4.1 Removing GFP tag

The pCA24N-*cynT*, -*yeaD* and -*ydfW* plasmids were amplified using the forward primer pCA24N_NoGFP.for (for all three plasmids) and either cynT_NoGFP.rev, yeaD_NoGFP.rev or ydfW_NoGFP.rev reverse primers (Table I.3, Appendix I.5). The reverse primers have a stop codon and an *Sfi*I recognition site incorporated at the 5′ end of the sequence. The forward and reverse primers were both phosphorylated at the 5′ ends. The forward primer bound at the end of the GFP region and the reverse primer bound to the 3′ ends of the *cynT*, *yeaD* and *ydfW* genes (before the start of the GFP region). A 50 µL PCR reaction was prepared and contained 1x Phusion HF buffer, 2 U Phusion DNA polymerase (Finnzymes) 0.2 mM dNTPs, 0.5 µM of each primer and ~ 20 ng of the purified plasmids. The thermocycling conditions were 98 °C for 30 s, 35 cycles of 98 °C (10 s), 72 °C (150 s) and finally 75 °C for 5 min. The amplified product was purified using a commercial spin column (Appendix I.10) and eluted in 40 µL EB.

The template DNA was digested with *Dpn*I for 1 h at 37 °C in a 20 µL reaction. The reaction contained 1x appropriate buffer (Appendix I.12), 10 U restriction enzyme and ~ 1000 ng of purified PCR products. The DNA was purified using a commercial spin column (Appendix I.10) and eluted in 30 µL EB. The DNA concentration was estimated by running 1 µL on an ethidium bromide-stained agarose gel, alongside a suitable DNA ladder and comparing the intensity of the samples with the intensity of the bands in the ladder. The bands on the gel corresponded to the expected DNA sizes: 5180 bp, 5404 bp and 4670 bp for the *cynT*, *yeaD* and *ydfW* products, respectively. The phosphorylated ends of the PCR products of each gene were ligated to form a circular plasmid. The ligation reaction (50 µL final volume) consisted of 1x T4 DNA ligase buffer, 5 U T4 DNA ligase (Fermentas) and 50 ng of linear DNA. The samples were incubated at 22 °C for 1 h and then at 70 °C for 5 min to inactivate the enzyme.

The ligated DNA was used purified using a commercial spin column (Appendix I.10) and eluted in 30 µL EB. The DNA was then used to transform (Appendix I.14) a 50 µL aliquot of electrocompetent *E. coli* DH5α-E cells (Appendix I.14). The cells were recovered with 500 µL SOC medium (20 g/L tryptone, 5 g/L yeast extract, 10 mM NaCl, 2.5 mM KCl and 20 mM glucose) and incubated at 37 °C for 1 h. Aliquots (20 µL and 100 µL) of the recovered cells were spread on LB-chloramphenicol (34 $\mu g.mL^{-1}$ final concentration) plates and incubated at 37 °C overnight.

The pCA24N plasmid (minus GFP) was re-purified from a single transformed colony. To confirm that GFP was absent from the plasmid, two sets of PCR reactions were performed using pCA24N.for and pCA24N.rev/pCA24N.rev2 primers (Table I.3, Appendix I.5). The forward primer binds upstream of the 6xHis region (before the start of the *cynT*, *yeaD* and *ydfW* genes). The pCA24N.rev primer binds to the GFP region and served as negative control whereas pCA24N.rev2 binds downstrean of the GFP region. The PCR products generated by the pCA24N.for and pCA24N.rev2 primers were sequenced (Appendix I.6); 857 bp, 1082 bp and 347 bp for *cynT*, *yeaD* and *ydfW* genes, respectively.

3.4.2 Creating a mutant library

The error-prone PCR was performed using the GeneMorph II Random Mutagenesis Kit from Agilent technologies that employs Mutazyme II DNA polymerase (http://www.genomics.agilent.com). Mutazyme II is a blend of two error-prone DNA polymerases, Mutazyme I DNA polymerase and a novel *Taq* DNA polymerase mutant.

3.4.2.1 Error-prone PCR

Random mutations were introduced in each target gene during PCR using the GeneMorph II Random Mutagenesis Kit (Agilent Technologies). The error-prone PCR (epPCR) reaction (50 µL final volume) consisted of 1x Mutazyme reaction buffer, 2.5 U Mutazyme II DNA polymerase, 0.2 mM dNTP mix, 0.4 µM forward and reverse primers (pCA24N.for and pCA24N.rev2; Table I.3, Appendix I.5) and purified plasmid DNA containing the gene that was to be mutagenized. The DNA amounts were 90 ng for *cynT* and 92 ng for the *yeaD* and *ydfW* genes. This amount of DNA (0.1-100 ng) was recommended by the manufacturer to achieve a high mutation rate (4.5-9 mutations/kb). The amount of plasmid template

added was calculated to correspond to the initial amount of target DNA required for the mutation frequency desired. For example, in total, the plasmid was 5,180 bp in size and the amplified *cynT* product (plus flanking sequences) was 857 bp. Therefore, the amount of template DNA in the reaction was 15 ng [(857 bp/5,180 bp) x 90 ng]. For *yeaD* and *ydfW* the amount was 18 ng [(1082 bp/5,405 bp) x 92 ng] and 7 ng [(347 bp/4,670 bp) x 92 ng], respectively. The epPCR was run at conditions 1 cycle at 95 °C for 1 min, 30 cycles at 94 °C (20 s), 56 °C (20 s), 72 °C (60 s) and a final cycle of 72 °C for 2 min for the *cynT* gene. For *yeaD* and *ydfW*, the times for the extension step were 70 s and 30 s, respectively.

The epPCR product (2 µL) was run on an ethidium bromide-stained agarose gel alongside a DNA ladder. The total yield of the products was estimated by comparing the intensity of the samples with the intensity of the bands in the ladder (Appendix I.11). The PCR efficiency (the probability that any particular sequence is duplicated in any one cycle of the PCR) was calculated from the total product yield and the amount of starting template. This was required for statistical analysis of the library composition (Section 3.4.3). To calculate this, first, the number of doubling in the PCR (*d*) was determined: d = [log(Product/Template)]/log2. The PCR efficiency (*eff*) was then calculated using the following equation; $eff = 2^{(d/n)} - 1$, where n is the number of cycles (30 in this experiment). The PCR efficiencies for the *cynT*, *yeaD* and *ydfW* reactions are as follows:

cynT = 90 ng → 1,437.5 ng; d = 3.99; PCR efficiency = 0.097

yeaD = 92 ng → 2,050 ng; d = 4.48; PCR efficiency = 0.109

ydfW = 92 ng → 2,050 ng; d = 4.48; PCR efficiency = 0.109

The remainder of epPCR products were purified using the QiaQuick PCR Purification Kit (Qiagen) and eluted in 30 µL elution buffer (EB). The concentration of the purified samples were estimated by running 1 µL on an ethidium bromide-stained agarose gel, alongside a suitable DNA ladder and comparing the intensity of the samples with the intensity of the bands in the ladder.

3.4.2.2 Vector and Insert preparation

The pCA24N vector backbone was used for the cloning and expression of the epPCR library. To obtain the vector backbone, pCA24N-*trpD* (6113 bp) (Tabe I.2, Appendix I.4) was digested with 20 U of *Sfi*I at 50 °C for 5 h (30 µL volume with ~3 µg DNA) (Appendix I.12). To minimize the number of 'vector only' clones in the final library, 10 U of *Bgl*II was added to the reaction (that cuts within the stuffer fragment) and incubated for a further 2 h at 37 °C. The epPCR products (~1 µg) were also treated with 20 U of *Sfi*I (30 µL volume) in order to introduce sticky ends and facilitate directional cloning. The reaction was also incubated at 50 °C for 5 h. In order to eliminate the methylated, un-mutated DNA template that may have carried over from the epPCR, 10 U of *Dpn*I was added to the reaction and incubated for a further 2 h at 37 °C.

The digested mutagenised insert and the linearised vector backbone were run on separate agarose gels (1.2 % agarose for the insert and 0.8 % agarose for the vector) and the bands corresponding to the digested inserts (670 bp, 895 bp and 160 bp for the *cynT*, *yeaD* and *ydfW* genes, respectively) and vector (4510 bp) were excised using a blue-light transilluminator (Invitrogen) and 1x SYBR Safe stain (Invitrogen) (Appendix I.9). The excised gel bands were then purified and eluted in 12 µL EB using the MinElute Gel Extraction Kit (Qiagen). The concentrations of the purified vector and inserts were determined by running 2 µL of each on an ethidium bromide-stained agarose gel, alongside a suitable DNA ladder as described earlier.

3.4.2.3 Preparation of a test library

Before constructing a large-scale mutant library, a test library was first prepared. The digested inserts were ligated to the pCA24N vector backbone at a 3:1 molar excess using 1 U of T4 DNA ligase (10 µL final volume) (Appendix I.12); for each of the reactions 22 ng, 30 ng, and 5 ng of *cynT*, *yeaD* and *ydfW* insert DNA, respectively, were used. In each case, 50 ng of vector DNA was added to the ligation reaction. A 'vector only' ligation reaction was also prepared in order to determine the frequency of vector recircularisation. This reaction contained the same ligation components but without insert DNA. The ligation reactions were incubated for a total of 16 h at 16 °C, heat inactivated (65 °C for 10 min).

One µL aliquot of each ligation reaction (the remaining 9 µL were stored at -20 °C) was used to transform 50 µL aliquots of electrocompetent *E. coli* DH5α-E (Appendix I.13) by electroporation using sterile Gene Pulser cuvettes (0.2 cm electrode gap, Bio-Rad) at 2.5 kV, 200 Ω and 25 µF in a Gene Pulser unit with Pulse Controller (Bio-Rad) (Appendix I.14). After electroporation, 500 µL SOC medium (20 g/L tryptone, 5 g/L yeast extract, 10 mM NaCl, 2.5 mM KCl and 20 mM glucose) was added to the cuvette before transferring the cultures into sterile 15 mL Falcon tubes. The cells were allowed to recover by incubating at 37 °C, with shaking, for 1 h. Aliquots of the recovery cultures (10 µL and 50 µL) were spread on LB-chloramphenicol (34 $\mu g.mL^{-1}$ final concentration) agar plates and incubated at 37 °C for 16 h. The results from the 'vector only' plates were used to calculate the fraction of the library that contains recircularised vector.

3.4.2.4 Construction of a full-size library

After calculating and determining that < 1 % of the library corresponded to the 'vector only' background (from the colony counts on the 'vector only' and 'vector + insert' agar plates), the ligation was scaled up to construct a full-sized library . The digested *cynT*, *yeaD* and *ydfW* insert DNA and pCA24N vector backbone from Section 3.4.2.2 were used. 'Vector only' and 'vector + insert' ligation reactions that were 10-fold larger than the test library (Section 3.4.2.2) were prepared and incubated for 16 h at 16 °C as before. The remainder (9 µL) of each test ligation was added to the scaled up 'vector only' and 'vector + insert' ligation reactions. The products from each ligation reaction were purified using the QiaQuick PCR Purification Kit (Qiagen) and the DNA was eluted in 42 µL EB.

Meanwhile, LB agar library plates (245 mm x 245 mm square bioassay dishes) supplemented with chloramphenicol (20 $\mu g.mL^{-1}$ final concentration), IPTG (50 µM final concentration) and varying concentrations of either enoxacin, aztreonam or spiramycin on which cells transformed with each library were going to be spread, were prepared. The plates were incubated at 37 °C for 4 h before plating since they are prone to 'sweating' when incubated.

Three µL aliquots of the 'vector + insert' ligation was used to transform multiple 50 µL aliquots of electrocompetent *E. coli* DH5α-E by electroporation. A total of 13, 9 and 13 aliquots of cells were transformed for the *cynT*, *yeaD* and *ydfW* libraries, respectively. A single 50 µL aliquot of electrocompetent *E. coli* DH5α-E was transformed with 3 µL of

each 'vector only' ligation. To determine transformation efficiency, pUC19 (10 pg) (Table I.2, Appendix I.4) was used to transform one 50 µL aliquot of electrocompetent *E. coli* DH5α-E. To recover the cells, 500 µL SOC medium was added to the cuvettes after electroporation, the cultures were transferred into sterile 15 mL Falcon tubes and incubated at 37 °C, with shaking, for 1 h.

The 'vector + insert' transformants of each library were pooled together in a sterile 15 mL Falcon tube and mixed briefly by inverting 2-3 times. Small volumes of this culture (5 µL and 20 µL) were spread on regular LB-chloramphenicol plates (34 $\mu g.mL^{-1}$ final concentration) and the remainder (7,125 µL, 4,925 µL and 7,125 µL for the *cynT*, *yeaD* and *ydfW* libraries, respectively) was spread on library plates. Two library plates were used for each of the libraries, with half of the culture volume spread on each plate. The library plates were incubated at 30 °C for 16 h to avoid a confluent lawn. Aliquots of the 'vector only' (5 µL and 20 µL) and pUC19 transformed cells (5 µL, 20 µL and 100 µL) were spread on regular LB-chloramphenicol (34 $\mu g.mL^{-1}$ final concentration) and LB-ampicillin (100 $\mu g.mL^{-1}$ final concentration) plates, respectively, and incubated at 37 °C for 16 h. Based on the number of colonies on the 'vector + insert' plates and the volume of cells spread on the library plates, the size of the *cynT*, *yeaD* and *ydfW* libraries was estimated.

3.4.3 Analysing the library composition

To analyse the composition of each library, 10-20 colonies were picked at random from the 'vector + insert' plates using 2 µL tips and transferred into individual thin-walled, 0.2 mL tubes containing 5 µL of sterile water. The cells were lysed by incubating the tubes at 95 °C for 5 min in a thermocycler (MJ Mini Bio-Rad). The randomly mutagenised gene inserts from each colony were amplified by PCR (Appendix I.8) using pCA24N backbone-specific primers (pCA24N.for and pCA24N.rev2; Table I.3, Appendix I.5) (25 µL final volume). Two µL aliquot of each PCR product was run on an ethidium bromide-stained agarose gel along with a DNA ladder. Products with sizes of ~ 900 bp (857 bp), ~ 1100 bp (1082 bp) and ~ 350 bp (347 bp) for the *cynT*, *yeaD* and *ydfW* PCR reactions, respectively, confirmed successful amplification. The remainder of the PCR product was purified using the QiaQuick PCR Purification Kit (Qiagen) and eluted in 30 µL EB. The DNA was sequenced (Appendix I.6) using the pCA24N.for and pCA24N.rev2 primers to determine the spectrum of mutations that arose in the epPCR. The sequenced PCR products were aligned with

known sequences of the unmutated parental genes (*cynT*, *yeaD* and *ydfW*) using MacVector (Appendix I.7) for analysis.

The mutations (point mutations, deletions and insertions) in the sequenced samples were tabulated for the *cynT*, *yeaD* and *ydfW* genes. The point mutations were grouped together by type: this data was used to examine the transition (purine to purine and pyrimidine to pyrimidine changes) and transversion (purine to pyrimidine and pyrimidine to purine chages) mutation frequencies to help calculate the overall mutation rate. To predict the amino acid diversity in the *cynT*, *yeaD* and *ydfW* libraries, PEDEL-AA (Firth and Patrick, 2008) was used; the sequence of each gene, the estimated size of the scaled up libraries (Section 3.4.2.4), the types of mutation present, the number of PCR cycles, the PCR efficiency and the mean number of insertions and deletions per gene were entered in the library analysis program. The library size of the scaled up libraries was estimated based on the number of colonies that were present on the 5 µL and 20 µL 'vector + insert' plates and the number of aliquots of cells used for each of the libraries (a total of 550 µL for each aliquot).

To store the full-sized library, the cells were recovered from each of the plates by pre-wetting with 4 mL LB-chloramphenicol (34 $\mu g.mL^{-1}$ final concentration). The cells were scraped and pooled together in one corner of the plate and then the cell suspension was pipetted into a 50 mL Falcon tube. This process was repeated once (i.e., 2 x 4 mL LB-chloramphenicol per plate). The cells were pelleted by centrifugation (3,000 xg, 4 °C for 15 min). The supernatant was removed and the pellet was resuspended in 1 mL LB-chloramphenicol (34 $\mu g.mL^{-1}$ final concentration). The cell density was measured at OD_{600} by mixing 1 µL of cell suspension with 999 µL of sterile water. The library was split into 100 µL aliquots and transferred into cryogenic vials containing 50 µL glycerol (50 % v/v), mixed and stored at -80 °C.

3.4.4 Determining MIC values using agar dilution method

The MIC values of the strains expressing wild-type carbonic anhydrase, YeaD and YdfW for enoxacin, aztreonam and spiramycin, respectively, were measured in order to determine the lowest concentrations these strains could not grow on. This way, mutants with improved resistance compared to their parent could be identified. This was achieved using the agar dilution method.

A 200 µL of overnight culture of cells harbouring either pCA24N-*cynT*(-GFP), -*yeaD*(-GFP) or –*ydfW*(-GFP) was used to inoculate 10 mL LB-chloramphenicol (34 µg.mL^{-1} final concentration). The culture was incubated at 37 °C with shaking until OD_{600} ~ 0.3. IPTG was added to the culture (50 µM final concentration) and incubated for a further 2 h at 37 °C. The OD_{600} of the culture was measured and the concentration of cells in the culture was calculated based on the approximation of OD_{600} ~ 1 = 2.5 x 10^8 cfu/mL (Sambrook & Russell, 2001). One thousand cells were spread on regular LB-agar plates supplemented with IPTG (50 µM final concentration), chloramphenicol (20 µg.mL^{-1} final concentration) and varying concentrations of enoxacin, aztreonam or spiramycin. The same volume of the culture was spread on two LB-chloramphenicol plus IPTG plates; based on the colony counts on the plate, the total number of cells plated for each experiment was determined. All of the plates were incubated at 37 °C overnight. The MIC value was determined as the plate with the lowest antibiotic concentration with less than 10 % growth. At least three experiments were performed. The MIC value was determined from the average of the experimental repeats.

3.4.5 Selection of 'winners' from library

A frozen aliquot of the mutant library was thawed on ice. Fifteen µL was used to inoculate 40 mL LB-chloramphenicol (20 µg.mL^{-1} final concentration). The culture was incubated at 37 °C with shaking until OD_{600} ~ 0.3 was reached. IPTG was added to the culture (50 µM final concentration) and incubated for further 1 h at 37 °C. The OD_{600} of the culture was measured. The concentration of cells in the culture was calculated based on the approximation of OD_{600} ~ 1 = 2.5 x 10^8 cfu/mL (Sambrook & Russell, 2001). To estimate the number of cells from each library to spread for 99.9 % coverage, GLUE (the library diversity and completeness programme) was used (Patrick et al., 2003).

LB agar library plates supplemented with IPTG (50 µM final concentration), chloramphenicol (20 µg.mL^{-1} final concentration) and varying concentrations of a second antibiotic (i.e., the drug used to test for the improved protein function) were prepared. In total, 3 or 4 library plates were prepared for each selection experiment. Based on the calculation of cells present in the culture and the estimated number of cells to spread for 99.9 % library coverage, the appropriate volume of culture was spread on the library plates. Another aliquot of the 40 mL culture was used to make a suitable dilution (~ 200 cells/µL).

Aliquots of this dilution, 2 µL and 4 µL (for 400 and 800 cells, respectively), were plated on two regular LB- chloramphenicol (20 µg.mL^{-1} final concentration) plates. These aliquots were also plated on regular LB-agar plates containing the same components as the library plates. These additional plates allow the total number of cells plated on the library plates to be calculated and also the survival percentage at each antibiotic concentration (the drug used to test for the improved protein function). The plates were incubated at 30 °C overnight.

3.4.6 Analysing sequences of 'winners'

A total of ten colonies were picked at random off the library selection plates (Section 3.4.5) and transferred into individual thin-walled, 0.2 mL tubes containing 6 µL of sterile water. Two µL of the cell suspensions were transferred into individual 1.5 mL tubes containing 300 µL LB-chloramphenicol (34 µg.mL^{-1} final concentration) and incubated at 37 °C for 3 h. The remainder (4 µL) was lysed by incubating the tubes at 95 °C for 5 min in a thermocycler (MJ Mini Bio-Rad). The colonies were PCR screened (Appendix I.8) using pCA24N backbone-specific primers (pCA24N.for and pCA24N.rev2; Table I.3, Appendix I.5) (25 µL final volume). If the PCR screen yielded the expected product size (857 bp, 1082 bp and 347 bp for the *cynT*, *yeaD* and *ydfW* genes, respectively), the PCR products were purified using the QiaQuick PCR Purification Kit (Qiagen) and eluted in 30 µL EB. The DNA was then sequenced (Appendix I.6) using the pCA24N.for and pCA24N.rev2 primers.

The growing cultures (300 µL) were used to inoculate 10 mL of LB-chloramphenicol (34 µg.mL^{-1} final concentration) and incubated at 37 °C overnight. The pCA24N plasmid was extracted from the cells in order to perform the retransformation test (see below).

3.4.7 Retransformation tests

This test was to confirm that the improved resistance phenotype is conferred by the plasmid-encoded *cynT*, *yeaD* and *ydfW* genes.

Aliquots (3x 1.5 mL and 1x 500 µL) of the overnight cultures (Selection 3.4.6) were centrifuged in a benchtop microcentrifuge (Heraeus Pico 17 centrifuge) at 17, 000 xg for 1 min (5 mL final volume) and the supernatant was discarded. The plasmid was extracted from each pellet [using QIAprep Spin Miniprep kit (Qiagen)] and eluted in 30 µL EB. The plasmid concentration was measured determined spectrophotometrically in an Eppendorf Biophotometer, using EB as a blank.

The plasmid DNA (~ 20 ng) was used to transform fresh electrocompetent *E. coli* DH5α-E cells (Appendix I.13). For each variant, one transformed cell was picked and inoculated in 3 mL LB-chloramphenicol (34 µg.mL^{-1} final concentration) and incubated overnight. Ten mL LB-chloramphenicol (34 µg.mL^{-1} final concentration) was inoculated with 200 µL of the overnight culture and incubated at 37 °C with shaking. When OD_{600} ~ 0.3 was reached, IPTG was added to the culture (50 µM final concentration) and incubated for further 2 h at 37 °C. The OD_{600} of the culture was measured again. Based on the approximation of OD_{600} ~ 1 = 2.5 x 10^8 cfu/mL (Sambrook & Russell, 2001) 50,000 cells were spread on regular LB-agar plates supplemented chloramphenicol (34 µg.mL^{-1} final concentration), IPTG (50 µM final concentration) and either enoxacin, aztreonam and spiramycin (at the same concentrations that was present in the library plates (Selection 3.4.5) they were selected off or higher. The plates were incubated at 37 °C overnight.

3.4.8 Curing experiment

Two separate 1.5 mL tubes containing 400 µL of LB only were inoculated with freezer stocks of Mutants 1 and 2 that harboured the pCA24N-*cynT* plasmids. The tubes were incubated at 37 °C for 8 h. A 1 in 20 dilution of the culture (10 µL culture in total of 200 µL LB) was made and 1 µL of this was used to inoculate 400 µL of fresh LB (~ 13 generations to return to saturation). The culture was incubated at 37 °C for 15 h. A 1 in 50 dilution of the overnight culture was prepared (5 µL culture in a total of 250 µL LB) and 1 µL of this was used to inoculate 400 µL of fresh LB and incubated at 37 °C for 7 h (~ 14 generations to return to saturation). This step was repeated three more times with alternating incubation times of 17 h and 7 h at 37 °C. To determine whether the cells had

been cured of the pCA24N-*cynT* plasmid, at ~ 70 generations 1 µL of overnight cells were used to inoculate 400 µL fresh LB supplemented with chloramphenicol (34 µg.mL^{-1} final concentration) and incubated at 37 °C for 9 h. The cells were able to grow in the chloramphenicol-containing medium indicating that the plasmid was still present.

A 1 in 50 dilution of the overnight culture grown in LB only medium was prepared (5 µL culture in a total of 250 µL LB) and 0.5 µL of this was used to inoculate 400 µL of fresh LB and incubated at 37 °C for 7 h (~15 generations to return to saturation). This step was repeated three more times with alternating incubation times of 17 h and 7 h at 37 °C. To determine whether the cells had been cured of the pCA24N-*cynT* plasmid, at ~ 130 generations 1 µL of overnight cells were used to inoculate 400 µL fresh LB supplemented with chloramphenicol (34 µg.mL^{-1} final concentration) and incubated at 37 °C for 9 h. Again, the cells were able to grow indicating that they still harboured the pCA24N-*cynT* plasmid.

CHAPTER 4

Directed evolution of YejG

4.1 Introduction

4.1.1 Directed evolution on *yejG*

The goal in this Chapter was to identify *yejG* variants that improved the host cell's resistance towards two different aminoglycosides (tobramycin and apramycin). The experiments performed on *cynT*, *yeaD* and *ydfW* in Chapter 3 were also performed on *yejG*: random mutations were introduced into *yejG* and the mutated copies were ligated to the pCA24N vector. To eliminate the possibility that improved resistance function was due to YejG-GFP fusion (encoded by pCA24N–*yejG*), the GFP tag was removed (Section 4.5.1). After transforming *E. coli* DH5α-E cells with the pooled plasmids, mutants with improved aminoglycoside resistance were selected for.

Retransformation tests revealed that similar to the previous results, the *yejG* mutants also acquired spontaneous chromosomal mutations. Therefore, I decided to switch from *E. coli* DH5α-E to a strain that minimizes undesired genetic variation (due to the removal of known insertion sequences, IS), the *E. coli* Multiple Deletion Strain 42 (MDS42). Despite using this genetically stable strain, the *yejG* mutants continued to gain beneficial mutations in the chromosome. Nonetheless, a mutant [*E. coli* MDS42 *yejG* (HTD)] was discovered that had acquired a mutation in the chromosomal *fusA* gene that was able to attain maximal resistance to aminoglycosides only when YejG was present (investigated further in Chapters 5 and 6).

4.1.2 YejG

The *yejG* gene encodes a 114 amino acid protein of unknown function (Keseler et al., 2011). The protein shares no significant sequence similarity with any characterised protein in publicly available databases. YejG was selected because cells expressing this protein showed resistance towards three aminoglycoside antibiotics (sisomicin, tobramycin and apramycin) and suggests that YejG may have a specific role in aminoglycoside resistance.

Aminoglycosides are a group of antibiotics derived from soil-dwelling bacteria species of the genera *Streptomyces* and *Micromonospora* (Becker & Cooper, 2013). The drugs work by binding to the 30S bacterial ribosomal subunit leading to misreading of the messenger RNA and synthesis of faulty proteins or premature termination (Bryan & Kwan, 1983). Aminoglycosides are active mainly against aerobic Gram-negative bacteria and are

used to treat urinary tract infections, pneumonia and bacteremia (Gonzalez & Spencer, 1998).

YejG was selected because its over-expression led to enhanced growth in Phenotype Microarray wells that contained three aminoglycosides. Furthermore, this is the first hint of any function for this protein.

4.2 Results

4.2.1 *yejG* error-prone PCR library

The same method used in the *cynT*, *yeaD* and *ydfW* experiments was employed to create the yejG library. A library consisting of 1.1 x 10^6 *yejG* mutants was generated using error-prone PCR (Section 4.5.2). The *yejG* inserts from ten randomly chosen clones were amplified by PCR using plasmid backbone-specific primers (pCA24N.for and pCA24N.rev2; Table I.3, Appendix I.5) and then sequenced (Section 4.5.3). Overall there was a low mutation rate of 1.8 mutations per *yejG* resulting in three out of the ten sequenced genes being wild-type *yejG*. The mutation frequency was lower than predicted from the manufacturer's protocol (Section 4.5.2.1). The mutational spectrum was biased towards A→N and T→N mutations (11 of 18 mutations) with no G→C, C→G mutations observed (Table 10).

PEDEL-AA was used to calculate the mean number of amino acid substitutions at ~ 1 per YejG variant encoded in the library. This meant that ~34 % of the library was likely to encode wild-type YejG. Overall, ~ 2.2 × 10^5 different protein sequences were estimated to be present in the library (Table 11).

Table 10. The mutational spectrum of the *yejG* mutant library

Type(s) of mutations	Frequency	Proportion of total
Transitions		
A→G, T→C	7	38.9%
G→A, C→T	3	16.7%
Transversions		
A→T, T→A	4	22.2%
A→C, T→G	0	0%
G→C, C→G	0	0%
G→T, C→A	4	22.2%
Insertions and deletions		
Insertions	0	0%
Deletions	0	0%
Summary of bias		
Transitions/transversions	1.25	NA[a]
AT→GC/GC→AT	1	NA[a]
A→N, T→N	11	61.1%
G→N, C→N	7	38.9%
Mutation rate		
Mutations per kb	5.26	NA[a]
Mutations per yejG gene[b]	1.8	NA[a]

[a]NA: not applicable.
[b]The cloned *yejG* insert was 342 bp.

Table 11. Composition of the *yejG* mutant library as estimated by PEDEL-AA

Property	Estimate
Total library size	1.1×10^6
Number of variants with no insertions, deletions, or stop codons	9.9×10^5
Mean number of amino acid substitutions per variant	1.2
Unmutated (wild-type) sequences (% of total library)	34.3%
Number of distinct, full-length proteins in the library	2.2×10^5

4.2.2 Selecting *yejG* 'winners' (tobramycin)

The selection experiments were first performed in the presence of tobramycin to identify YejG mutants that conferred increases in resistance to this drug. To do this, first, the lowest tobramycin concentration that the strain expressing wild-type YejG could not grow on was determined. The MIC value of this strain was measured by the agar dilution method (Section 4.5.4) and determined as 10 $\mu g.mL^{-1}$. *yejG* mutants that are able to grow on plates containing tobramycin concentrations of $\geq$ 10 $\mu g.mL^{-1}$, it suggests that the mutant has improved resistance compared to its parent strain. GLUE was used to estimate the number of clones to spread on the tobramycin-containing plates. For my library of 1.1 x 10^6 clones, GLUE estimated that spreading 7.5 x 10^6 cells would ensure that 99.9 % of the cells would be represented in the selection experiment.

The *yejG* mutant library (7.5 x 10^6 cells) was spread on LB agar plates containing tobramycin at concentrations of 12, 18 and 24 $\mu g.mL^{-1}$. After 72 h of incubation, colonies had appeared on all three of the plates. There were 90, 80 and 14 colonies present on the plates containing 12, 18 and 24 $\mu g.mL^{-1}$ of tobramycin, respectively. This corresponded to a low survival rate of 0.0012 % on the plate containing 12 $\mu g.mL^{-1}$ of tobramycin. This means that 1 in every 83,000 clones is resistant to 12 $\mu g.mL^{-1}$ of tobramycin, which equals 13 variants in my library of 1.1 x 10^6 *yejG* mutants.

4.2.3 DNA sequencing of *yejG* mutants

To confirm that the clones on the plates harboured the plasmid-encoded *yejG* gene and were not contaminants, ten colonies were picked at random off each plate for screening by PCR using plasmid backbone-specific primers (pCA24N.for and pCA24N.rev2; Table I.3, Appendix I.5). All 30 of the PCR screens yielded PCR products of the expected band size (550 bp), indicating that all of the colonies contained the plasmid (and were not contaminants). From this set of clones, six were randomly selected for DNA sequencing with one, two and three clones selected off the plates that contained 12, 18 and 24 $\mu g.mL^{-1}$ tobramycin, respectively.

DNA sequencing revealed that all of the six clones harboured mutated versions of *yejG*. One of the clones (mutant 2) had 7 point mutations in the 345 bp *yejG* gene which resulted in 4 amino acid substitutions (Table 12). A glutamine at position 63 (Q63) was mutated in half of the mutants sequenced. Furthermore, a proline at position 35 (P35) and

a glycine at position 51 (G51) were mutated in two mutants each. Mutant 6 contained all of the three aforementioned residues mutated. No C→G or G→C mutations in the DNA were observed, supporting the the predictions made by PEDEL-AA.

Table 12. Mutations in six YejG variants selected in the presence of tobramycin

Mutant	Tobramycin ($\mu g.mL^{-1}$)	Amino Acid mutation	Mutation type
1	12	F22L	Substitution
		L47F	Substitution
2	18	R17H	Substitution
		G51G	*Silent*
		Q63L	Substitution
		C36S	Substitution
		E78D	Substitution
		N86N	*Silent*
		A93A	*Silent*
3	18	V74A	Substitution
4	24	Q63H	Substitution
5	24	I104I	*Silent*
		P35S	Substitution
		L44L	*Silent*
6	24	P35Q	Substitution
		D50E	Substitution
		G51G	*Silent*
		Q63L	Substitution

4.2.4 Retransformation tests and enrichment experiments

Plasmids were isolated from clones 2, 5 and 6 (Table 12) and re-introduced into fresh *E. coli* DH5α-E cells. The transformed cells were grown in the same manner as the mutant library and streaked on plates containing tobramycin at concentrations of 12, 18 and 24 $\mu g.mL^{-1}$. After overnight incubation, only two or three colonies were present on each of the plates for each transformed clone. The plates were incubated for a further two days but no more colonies appeared. When mutants 2, 5 and 6 (originally selected off the plates

containing tobramycin) were streaked on plates containing the same tobramycin concentrations, all three clones grew as a lawn on the plates. These results indicated that the mutants had acquired spontaneous chromosomal mutations. I therefore decided to try a different selection strategy to enrich for *yejG* variants with improved resistance to tobramycin.

The aim was to plate the *yejG* library on a range of tobramycin concentrations, and then to choose the plate on which ~10 % of the population grew. The cells would be scraped off this plate, the plasmid pool prepped from them and used to transform a fresh aliquot of *E. coli* DH5α-E cells. The transformed cells would then be spread on plates containing the same tobramycin concentrations used in the first selection round. Repeating this for three cycles would enrich for true positives (i.e., plasmid-encoded *yejG* variants that increase resistance), while at the same time filtering out the false-positive clones that had acquired chromosomal mutations.

Approximately 7.5 x 10^6 cells were spread on plates containing tobramycin at concentrations of 10, 12 and 16 $\mu g.mL^{-1}$. After 72 h incubation, there were only 314, 100 and 58 colonies on the plates containing 10, 12 and 16 $\mu g.mL^{-1}$ of tobramycin, respectively. This corresponded to a maximum survival rate of 0.003 % (on the 10 $\mu g.mL^{-1}$ plate). PEDEL-AA had estimated that ~ 34 % of the mutant library was wild-type (Table 11), therefore ~1/3 of the library was not expected to survive on the 10 $\mu g.mL^{-1}$ tobramycin plate (since the tobramycin MIC for the wild-type *yejG* expressing strain is 10 $\mu g.mL^{-1}$). However, it appeared that over 65 % of the mutants in the library had not gained increases in tobramycin resistance. This means that the majority of the YejG variants in the library are not any better than wild-type YejG. Since earlier selection experiments (Section 4.2.2) also revealed a low survival rate, this result was expected.

To determine whether there were true positive clones in this population, a further two rounds of enrichment were performed. However, the survival rates did not change much in the successive rounds. There were survival rates of 0.0029 % and 0.0017 % for the second and third selection rounds, respectively, on the plates containing 10 $\mu g.mL^{-1}$ of tobramycin. It was apparent from the results that there was a constant rate of spontaneous mutations (approximately 0.003 %) that kept giving rise to colonies on the 10 $\mu g.mL^{-1}$ tobramycin plate, indicating that performing a retransformation test would not identify any positive mutants. Indeed, when retransformation tests were performed on a total of ten

clones randomly selected off the 10, 12 and 16 μg.mL^{-1} tobramycin plates from the third round of selection, the test was unsuccessful.

4.2.5 Selection of *yejG* 'winners' (apramycin)

Before discarding the library, I decided to select for clones that increased resistance to a second aminoglycoside, apramycin. Using the agar dilution method (Section 4.5.4), the MIC of YejG-expressing cells for this drug was determined as 30 μg.mL^{-1}. As before, a pool of 7.5 x 10^6 cells was spread on plates containing apramycin at concentrations of 32, 36 and 40 μg.mL^{-1}. The next day, there were 500, 146 and 62 colonies on the plates containing 32, 36 and 40 μg.mL^{-1}of apramycin, respectively. To determine whether the colonies contained plasmids with *yejG*-sized inserts, PCR screening was performed ten clones picked at random off each apramycin-containing plate. Using plasmid backbone-specific primers (pCA24N.for and pCA24N.rev2; Table I.3, Appendix I.5), PCR yielded the expected product sizes (~ 550 bp) for each of the ten samples. From this set of clones, six were chosen at random for DNA sequencing.

The sequencing results revealed that there were 1-4 mutations per clone. Mutants 4 and 5 both had mutations that led to substitutions at residue Q88 (Table 13). Mutant 6 on the other hand revealed a truncated protein of 105 amino acids (length of wild-type YejG is 114 amino acids).

Table 13. Mutations in six YejG variants selected in the presence of apramycin

Mutant	Apramycin ($\mu g.mL^{-1}$)	Amino Acid change	Mutation Type
1	32	S48T	Substitution
2	36	P13P	*Silent*
		L47R	Substitution
3	36	L12V	Substitution
4	40	G37G	*Silent*
		Y59C	Substitution
		E78E	*Silent*
		Q88L	Substitution
5	40	P49L	Substitution
		Q88H	Substitution
6	40	E106	NONSENSE

4.2.6 Retransformation tests and investigation of chromosomal *yejG* gene

To test whether the increased resistance to apramycin was due to the plasmid-encoded *yejG* or chromosomal mutations, retransformation tests were performed. Plasmids from three of the mutants, Mutants 3, 4 and 5, were isolated and introduced into fresh *E. coli* DH5α-E cells. The transformed cells were streaked on plates containing apramycin at concentrations of 32, 36 and 40 $\mu g.mL^{-1}$. After 24 h of incubation, only between one and three colonies grew on each plate for each clone. No more colonies grew after 72 h of incubation. This suggests that the mutants were false positives.

The results indicated that the mutants were acquiring beneficial mutations in the chromosome that were increasing their resistance to tobramycin and apramycin. To investigate whether the chromosomal copy of *yejG* (345 bp) or its promoter had acquired a mutation over the course of my experiments, a primer pair was designed to amplify a total region of ~ 800 bp. Mutants 5 and 6 from the tobramycin selection experiments (Table 12) and mutants 4 and 5 from the apramycin selection experiments (Table 13) were selected for DNA sequencing. A wild-type *E. coli* DH5α-E clone was also included to serve as a reference. Colony PCR was performed on the clones and the PCR products were sequenced. The results showed that all of the four mutants and the *E. coli* DH5α-E wild-

type clone had the same DNA sequence. This indicated that the chromosomal copy of *yejG* was not contributing to tobramycin or apramycin resistance. Due to the outcome of these experiments, a different strain that had been constructed to minimize genetic variation was used.

4.3 Investigating yejG in *E. coli* MDS42 cells

The experiments using *E. coli* DH5α-E indicated that the mutants were acquiring a beneficial mutation (or mutations) in their chromosomes (but not in the chromosomal *yejG* gene). I therefore switched to a host strain that would minimise the frequency of spontaneous mutations. *E. coli* strain MDS42 is a reduced-genome derivative of *E. coli* MG1655 (which in turn is an *E. coli* K-12 derivative). The strain has 704 genes (14.3 %) deleted and has a genome size of 3.9 kb. However, this has not impaired the cell's physiology (Pósfai et al., 2006). The genes that have been deleted in this strain include known insertion sequences, virulent genes, and non-essential genes (Posfai et al., 2006). The *E. coli* strain MDS42 is thought to have a reduced evolutionary potential since a large number of the genes that serve to generate mutations (all known mobile DNA sequences) have been disabled (Umenhoffer et al., 2010). Therefore, to minimise the undesired genetic variation in my mutant library, the *E. coli* strain MDS42 was used. This would increase the chance of selecting for mutants with improved antibiotic resistance conferred only by the plasmid-encoded *yejG* insert.

4.3.1 Tobramycin and apramycin resistance of YejG-expressing *E.coli* MDS42 strain

Since a different strain was being used, first, new MIC measurements had to be performed using the agar dilution method (Section 4.5.4). The MIC values of the *E. coli* MDS42 YejG-expressing strain and a control strain (i.e., the clone harbouring an empty pCA24N vector) were determined for tobramycin and apramycin.

The YejG-expressing strain showed 1.5-fold and 2-fold increases in resistance to tobramcyin and apramycin, respectively, compared to the control (Table 14). These MIC values are nearly two times higher than the values obtained with the YejG-expressing *E. coli* DH5α-E strain (Table 1).

Table 14. MICs of *E. coli* MDS42 cells over-expressing YejG

Compound	MIC (μg.mL^{-1}) of YejG-expressing cell	MIC (μg.mL^{-1}) of control	Fold-increase in MIC
Tobramycin	18	12	1.5
Apramycin	55	28	2

4.3.2 Selection of *E. coli* MDS42 *yejG* 'winners'

For this experiment, *yejG* mutants that conferred increases in resistance to tobramycin were selected for. To begin, plasmid DNA was isolated from the *E. coli* DH5α-E *yejG* library and used to transform *E. coli* MDS42 cells. As performed previously, a pool of 7.5 x 10^6 cells was spread on plates containing tobramycin at concentrations of 24, 32 and 48 μg.mL^{-1} (the MIC of YejG-expressing MDS42 cells was 18 μg.mL^{-1}). After 24 h of incubation, no colonies were present on any of the plates. The plates were incubated for a further two days and 91 and 29 colonies appeared on the plates containing 24 and 32 μg.mL^{-1} tobramycin, respectively. There were no colonies on the plate containing 48 μg.mL^{-1} tobramycin.

For the retransformation test, plasmids were isolated from six colonies picked at random off the two plates that had growth. Three colonies were selected off the plate containing 24 μg.mL^{-1} and three were taken from the plate with 32 μg.mL^{-1} tobramycin. The six plasmids were introduced into fresh *E. coli* MDS42 cells, and the transformed cells were grown in the same manner as the mutant library. After overnight incubation, there were no colonies present on plates containing tobramycin at concentrations of 24, 28, 32, 38, 40 or 42 μg.mL^{-1}. There was still no growth after two days of further incubation. To confirm whether the six mutants were able to grow at those tobramycin concentrations because they had acquired a beneficial chromosomal mutation, one out of the three clones originally selected off the 24 μg.mL^{-1} (labeled clone A) and 32 μg.mL^{-1} (labeled clone B) tobramycin plates were used. The two clones were streaked on plates containing the same concentrations of tobramycin as before. Both clones were able to grow at 40 μg.mL^{-1} of tobramycin (MIC = 42 μg.mL^{-1}). The results showed that despite using a more genetically stable *E. coli* strain, the cells were still acquiring spontaneous mutations. It seems like chromosomal mutations appear to be the most likely solution to attaining resistance to the tobramycin concentrations tested.

4.3.3 Investigating the importance of YejG in tobramycin resistant mutants

Experiments were performed to determine whether the *yejG* insert had played any part in the resistance displayed by the clones selected off the plates containing tobramycin. From previous experiments, it was hard to cure the cells of pCA24N (Section 3.2.1.3), however, the pBlueScript plasmid (Table I.2, Appendix I.4) was used to circumvent this problem. The pBlueScript plasmid carries a different antibiotic marker (ampicillin) to the pCA24N plasmid but possesses the same origin of replication. Since the two plasmids share the same origin of replication, they cannot stably co-exist in the same cell because they compete for the same replication factors (Velappan et al., 2007). Therefore, pBlueScript could be used to displace pCA24N. When the pCA24N harbouring cells are transformed with pBlueScript, after successive rounds of cell division, the bacteria will harbour different numbers of each plasmid. In time, some of the bacteria will only carry either pBlueScript or pCA24N. When the cells are grown in the presence of ampicillin, only those carrying pBlueScript will survive. Since pCA24N plasmid carries a chloramphenicol resistance marker, whereas pBlueScript carries an ampicillin resistance marker, it was easy to determine when either plasmid had been replaced.

To confirm that pBlueScript did not contribute to tobramycin resistance, *E. coli* MDS42 cells carrying this plasmid were spread on plates containing varying concentrations of tobramycin. The MIC value of the clone (in the presence and absence of IPTG) was 12 $\mu g.mL^{-1}$. This is the same MIC value of the negative control strain (*E. coli* MDS42 harbouring an empty pCA24N) and *E. coli* MDS42 without any plasmids (in the presence and absence of IPTG) (Table 15). These results show that neither the empty pCA24N vector, nor the pBlueScript plasmid, affect the susceptibility of *E. coli* MDS42 to tobramycin.

Table 15. MIC values of *E. coli* MDS42 cells harbouring either pBlueScript, pCA24N or without plasmids for tobramycin

Clone	MIC (μg.mL^{-1})
E. coli MDS42 (no plasmid)	12
E. coli MDS42 + pBlueScript	12
E. coli MDS42 + empty pCA24N	12

To investigate the involvement of the *yejG* insert in tobramycin resistance, the two clones, A and B (Section 4.3.2) were employed. These clones showed high levels of resistance towards tobramycin (38 μg.mL^{-1}) that appeared to result from acquiring mutations in the chromosome. However, the clones also carried the pCA24N plasmid. To displace pCA24N, the cells were transformed with pBlueScript and grown in the presence of ampicillin (Section 4.5.11). The clones (now carrying pBlueScript) were then spread on plates containing tobramycin. The MIC value for clone A was 24 μg.mL^{-1}, whereas, the MIC value for clone B was 12 μg.mL^{-1} (Table 16). As these strains displayed reduced tobramycin resistance compared to when they over-expressed YejG, it suggested that the *yejG* insert was responsible, in part, for endowing the clones with tobramycin resistance.

To confirm this hypothesis, next, the pBlueScript plasmids from clones A and B were replaced with the same pCA24N plasmid that had previously been replaced. After confirming that both clones had lost the pBlueScript plasmid, clone A and B (now carrying pCA24N) were streaked on plates containing tobramycin. After 24 h of incubation, clone A grew on concentrations up to 42 μg.mL^{-1} of tobramycin, whereas clone B grew on concentrations up to18 μg.mL^{-1} of tobramycin (Table 16). The transformation experiments were repeated twice from the beginning for clone B, however the same results were obtained.

Table 16. MIC values of clones A and B for tobramycin before and after displacing the pCA24N plasmid

Clone	MIC (μg.mL^{-1})
[a]Clone A + pCA24N *yejG*	42
[b]Clone A + pBlueScript	24
[c]Clone A + re-introduced pCA24N *yejG*	42
[1]Clone B + pCA24N *yejG*	42
[2]Clone B + pBlueScript	12
[3]Clone B + re-introduced pCA24N *yejG*	18

[a/1] clone carries chromosomal mutation(s) and is over-expressing either a wild-type YejG or a YejG variant

[b/2] clone carries chromosomal mutation(s) and harbours the pBlueScript plasmid

[c/3] clone carries chromosomal mutation(s) and is over-expressing the same pCA24N-*yejG* (wild-type or mutated version) that was present in [a] (clone A)/ [1] (clone B); i.e., plasmids have been re-introduced

The results suggested that clone A has acquired a beneficial chromosomal mutation that endows the cell with a moderate increase in tobramycin resistance compared to the YejG-expressing *E. coli* MDS42 strain (from 18 to 24 μg.mL^{-1}: 1.3-fold increase in resistance). However, the clone is able to attain maximal tobramycin resistance only when YejG is also present (up to 42 μg.mL^{-1}). This is not the case for clone B. When the pCA24N-*yejG* plasmid is removed from the cell, the clone displays similar tobramycin resistance as the control strain (12 μg.mL^{-1}). When the plasmid is re-introduced inside the cell, the clone now exhibits the same resistance as the YejG-expressing strain (18 μg.mL^{-1}). The results suggest that for clone B, once the pCA24N-*yejG* plasmid has been removed from the cell, it is no longer able to work together with the chromosomal mutation(s) to endow the bacterium with maximal tobramycin resistance when the plasmid is re-introduced. In light of these results, I decided to sequence the *yejG* insert of both clones A and B (to determine whether clone B had different mutations to the insert carried by clone A) but to only investigate clone A further.

4.3.4 DNA sequencing of *yejG* insert from clones A and B

Plasmid DNA was isolated from clone A and sent for DNA sequencing. The results revealed that there were three amino acid substitutions (HTD) and one silent mutation (I)

(Table 17). Therefore, from here on, clone A is referred to as *E. coli* MDS42 *yejG* (HTD) and the protein is referred to as YejG variant [YejG(HTD)]. The *yejG* gene in clone B was different to the one found in clone A. It had acquired three point mutations resulting in two substitution mutations and one silent mutation (Table 17).

Table 17. Mutations in the *yejG* inserts harboured by clones A and B

Clone	Amino acid change	Mutation type
Clone A	Q14H	Substitution
	I30I	*Silent*
	I68T	Substitution
	N110D	Substitution
Clone B	S7S	*Silent*
	S19F	Substitution
	Q88K	Substitution

Next, I wanted to determine whether the mutated *yejG* insert carried by *E. coli* MDS42 *yejG* (HTD) enhanced tobramycin resistance compared to wild-type *yejG*. To do this, the pCA24N-*yejG* (HTD) was replaced with one that encoded wild-type *yejG* in *E. coli* MDS42 *yejG* (HTD) (Section 4.5.11); the clone was first transformed with pBlueScript to make sure that the first pCA24N plasmid was lost. Then, the MIC values of this new clone (expressing wild-type YejG) together with the original *E. coli* MDS42 *yejG* (HTD) clone [expressing the mutated version of YejG, YejG(HTD)] were determined. Both clones displayed MIC values of 42 $\mu g.mL^{-1}$, indicating that the mutations in *yejG* were not essential for maximal tobramycin resistance. Altogether, the results indicated that wild-type or mutated YejG were both able to contribute towards maximal tobramycin resistance. The next step was to investigate the chromosomal mutation(s) present in the *E. coli* MDS42 *yejG* (HTD) clone.

4.3.5 *fusA* gene mutation in *E. coli* MDS42 *yejG* (HTD) clone

To definitely map the chromosomal mutation(s) in the *E. coli* MDS42 *yejG* (HTD) clone, whole genome sequencing was performed. Genomic DNA was extracted from the *E. coli* MDS42 *yejG* (HTD) clone, as well as from its parental strain (*E. coli* MDS42 with no

plasmid). The two genomes were sequenced by Ambry Genetics on an Illumina Genome Analyzer IIx.

The results revealed that the *E. coli* MDS42 *yejG* (HTD) clone had a G→T point mutation (G1478T) in the *fusA* gene, giving rise to the substitution of threonine 493 with asparagine (corresponding to the T493N amino acid substitution). The average coverage (total number of reads at a certain position) was ~ 220 for *E. coli* MDS42 *yejG* (HTD) and ~ 310 for the parental strain. The mutation in *fusA* revealed > 200-fold coverage (i.e., there were more than 200 reads covering this mutation site) with 100 % of the reads revealing a T nucleotide call at the mutation site (G1478T), suggesting that the mutation in *fusA* was true.

There was also a second point mutation, T342G, present in the *E. coli* MDS42 *yejG* (HTD) sequence read that was absent in the parental strain sequence. This mutation was located in between two genes, *ynfB* and *speG*. However, although ~ 150 reads were performed at this position, only 30 % revealed a change to a G nucleotide. The remainder of the reads showed no base change at this position, suggesting that it was a false positive call. To confirm this, primers were designed to amplify a total region of ~1,200 bp that covered *ynfB* and *speG* and the sequence in between the genes (ynfB_speG.for and rev; Table I.3, Appendix I.5). PCR products of the *E. coli* MDS42 *yejG* (HTD) and the parent strains were sequenced. The results revealed that both clones had the same DNA sequences, indicating that the second mutation was false positive. Primers were also designed to amplify the *fusA* gene of *E. coli* MDS42 *yejG* (HTD) and the parent strains (fusA_A.for and rev; Table I.3, Appendix I.5). DNA sequencing revealed that the *E. coli* MDS42 *yejG* (HTD) clone harboured the G1478T mutation.

The *fusA* gene is an essential gene (Baba et al., 2006) that codes for elongation factor G (EF-G), a protein of 704 amino acids that is a member of the GTPase protein superfamily. EF-G catalyses the translocation step of protein synthesis. The protein is folded into five domains. Domain 1 is the head and comprises the GTP-binding G domain and an insertion G domain (Agrawal et al., 1998). Domain 2 is a β-stranded barrel (Liljas & Garber, 1995) and domain 3 affects GTP hydrolysis and translocation (Martemyanov & Gudkov, 2000). Domain 4 is the tail of the molecule and domain 5 is important for ribosome binding of EF-G (Agrawal et al., 1998). The EF-G (T493N) mutation is located in the β-sheet of domain 4 of EF-G. Studies have shown a role for domain 4 in physically displacing

the tRNA in the A-site of the ribosome towards the P-site position of the ribosome (Agrawal et al., 1998; Agrawal et al., 1999).

4.3.6 Resistance of *E. coli* MDS42 *yejG* (HTD) to other aminoglycoside antibiotics

Hou et al., (1994) discovered two *fusA* mutants that were resistant to the aminoglycoside antibiotic, kanamycin. The mutations were at amino acid positions Q495 and G502, which are within a few amino acids of the T493N mutation. The results of Hou et al., (1994) suggested that it may be informative to measure MIC values for kanamycin. The MIC values for the other two aminoglycosides that had been employed in earlier experiments (sisomicin and apramycin) were also measured.

The *E. coli* MDS42 *yejG* (HTD) [harbours *fusA* mutation and over-expressing YejG(HTD)] clone showed increases in resistance towards all three aminoglycosides, compared to the strain expressing YejG with wild-type *fusA* (Table 18). Further experiments showed that the YejG-expressing strain displayed a small increase in kanamycin resistance (1.4-fold) compared to a control strain harbouring pCA24N-GFP (result not shown). The results indicated that both the *E. coli* MDS42 *yejG* (HTD) clone and the YejG-expressing strain displayed broad resistance to aminoglycosides and that when the both the EF-G (T493N) mutation and YejG are present, maximal resistance is attained. Furthermore, the EF-G (T493N) mutation appears to be the biggest contributor to aminoglycoside resistance.

Table 18. MICs of the *E. coli* MDS42 *yejG* (HTD) clone and the YejG-expressing strain for kanamycin, sisomicin, apramycin and tobramycin

Clones	MIC for kanamycin ($\mu g.mL^{-1}$)	MIC for sisomicin ($\mu g.mL^{-1}$)	MIC for apramycin ($\mu g.mL^{-1}$)	MIC for tobramycin ($\mu g.mL^{-1}$)
E. coli MDS42 *yejG* (HTD) clone[a]	80	32	120	42
YejG-expressing clone[b]	26	12	55	18
Fold-increase in MIC	3.1	2.6	2.2	2.3

[a]clone carries chromosomal *fusA*(G1478T) mutation and is over-expressing the YejG variant [YejG(HTD)]

[b]clone harbours the chromosomal wild-type *fusA* gene and is over-expressing wild-type YejG

4.4 DISCUSSION

The goal in this Chapter was to identify *yejG* variants that improved the host cell's resistance towards two different aminoglycosides (tobramycin and apramycin), however, no improved variant was found. The mutation spectrum of the *yejG* mutant library (~2.2 × 10^5 distinct variants; Table 11) was slightly smaller than the *ydfW* library (~3.5 × 10^5 distinct variants; Table 9). It is possible that YejG could not be evolved by directed evolution because of similar reasons described in Section 3.3; that is, there is no improved version of YejG or that major modifications (not accessible by directed evolution) might be required to improve resistance.

When the MIC value of the clone expressing YejG was measured using the agar dilution method, it was more resistant than the negative control (10 versus 6 $\mu g.mL^{-1}$ for tobramycin; 30 versus 12 $\mu g.mL^{-1}$ for apramycin); this result is different to that obtained in the previous experiments which showed that the CA-, YeaD- and YdfW-expressing strains were not more resistant than the control strain using the agar dilution method (section 3.3). Therefore, the selection protocol would not be expected to have negatively affected this study. To further enhance the experimental process, *yejG* mutants were subjected to three consecutive cycles of enrichment before variants with improved function were selected. However, all of the mutants selected had acquired beneficial mutations in the chromosome; that is, the plasmid did not confer increased resistance when it was used to transform a fresh aliquot of *E. coli*.

To minimize the undesired genetic variation, the *E.coli* MDS42 strain was employed. This strain is genetically stable due to the removal of all known mobile elements (Umenhoffer et al., 2010). IS element insertion has been shown to contribute significantly to genetic variation. For example, one study showed the most frequent mutational event that allowed cells resistance to the toxin, colicin B, was IS element insertion in the gene coding for the cytoplasmic protein TonB (Kitamura et al., 1995). One of the advantages of using the MDS42 strain has been demonstrated in a study where the strain improved the maintenance of unstable genetic constructs (Umenhoffer et al., 2010). However, despite using this genetically stable strain, the *yejG* mutants still acquired chromosomal mutations (Section 4.2.6 and 4.3.2).

Indeed spontaneous mutations can arise in cells not only by IS element insertion, but also by other means including point mutations. The MDS42 strain still possesses other

mutation-generating pathways that are accessible under stressful conditions (e.g., in the presence of antibiotics). This includes activation of the SOS response and related error-prone DNA polymerases (Bonner et al., 1988). In a recent article, Csörgő and colleagues (2012) have gone a step further and increased the genetic stability of the MDS42 strain. This was achieved by disabling the stress-induced mutagenesis mechanisms by deleting three genes encoding PolII (*polB*), PolIV (*dinB*), and PolV (*umuDC*) polymerases involved in the stress-induced pathway (Bonner et al., 1988; Napolitano et al., 2000). It is important to note that this additional deletion will help reduce the rate of point mutations but cannot completely eliminate it. In future, it will be interesting to test whether employing the strain with additional genetic stability will help identify improved YejG variants.

Despite the fact that improved YejG variants were not identified, using the MDS42 strain, a clone was discovered that carried a chromosomal *fusA* gene mutation [*fusA*(G1478T)]. This mutation was discovered after sequencing the evolved clone, *E. coli* MDS42 *yejG* (HTD), and the ancestral strain. The genome sequence revealed two mutations that were present in *E. coli* MDS42 *yejG* (HTD) but absent in the ancestor; one in *fusA*(G1478T) and the other in the sequence between *ynfB* and *speG* (T342G). The second mutation however, was confidently ruled out after PCR revealed that both strains had similar DNA sequences at that particular site.

Aminoglycoside resistance can be acquired by several mechanisms: active efflux, decreased permeability, ribosome alteration and drug inactivation (using different enzymes) (Becker & Cooper, 2013; reviewed in Van Hoek et al., 2011). Therefore, it seemed appropriate to perform whole genome sequencing to uncover the gene responsible for resistance in the *E. coli* MDS42 *yejG* (HTD) clone. Since more than one mutation could have given rise to the *E. coli* MDS42 *yejG* (HTD) phenotype, a comprehensive genome search was required. The candidate gene could have been investigated using other approaches, for example microarray-based methods, however, this approach is not as comprehensive and is not always accurate (Dettman et al., 2012). Candidate gene sequencing on the other hand is impractical not to mention time consuming. Other studies have also resorted to whole genome sequencing as a rapid way of determining genetic changes in evolved microbes (Jezequel et al., 2013; Barrick et al., 2009).

A clone carrying the *fusA* mutation alone (and pBlueScript) displayed higher aminoglycoside resistance than the YejG-expressing strain. For example, the MIC value of

tobramycin for a clone carrying the *fusA* mutation alone was 24 $\mu g.mL^{-1}$ (Table 16), whereas an MIC value of 18 $\mu g.mL^{-1}$ was recorded for a clone harbouring wild-type *fusA* and over-expressing YejG (Table 14). To obtain maximal resistance towards the four aminoglycosides (tobramycin, sisomicin, apramycin and kanamycin) however, both the *fusA*(G1478T) mutation and the mutated version of YejG [YejG(HTD)] were required (Table 18). When this YejG variant was present in a clone carrying the *fusA* mutation, resistance towards tobramycin increased to 42 $\mu g.mL^{-1}$ (Table 16). Nonetheless, the *fusA*(G1478T) mutation appeared to be a bigger contributor for tobramycin resistance than YejG. Additional experiments revealed that the YejG variant did not offer any improvement in tobramycin resistance to a clone carrying the *fusA* mutation compared to the wild-type YejG protein (Section 4.3.4). This reinforces the idea that there might not be an improved version of YejG in the library created.

The *fusA* gene codes for EF-G, which catalyses the translocation step of protein synthesis. EF-G binds to both the 30S and 50S subunits of the bacterial ribosome. EF-G binding facilitates the translocation of the aminoacyl (A) and peptidyl (P) site tRNAs to the P and exit (E) sites, respectively. As a result, the ribosome is able to advance along the mRNA by one codon (Agrawal et al., 1998). The charged tRNA binds to the A-site and polypeptide chain elongation occurs when EF-G, in complex with GTP, binds to the ribosome. Hydrolysis of GTP to GDP results in translocation as well as the dissociation of EF-G in a GDP-bound form (Laurberg et al., 2000).

The ribosomal A-site is thought to be the target for aminoglycoside antibiotics, although some aminoglycosides bind close to this site or to other locations (Borovinskaya et al., 2007). Aminoglycoside binding results in an 'on' state of the A-site which allows the binding of other non-cognate tRNAs, thus causing a non-specific increase in A-site affinity for charged tRNAs. This is thought to block translocation of tRNAs to the P-site if there is no way of overcoming the increased affinity of the tRNAs for the A-site. This leads to premature termination or a misreading of mRNA (Becker & Cooper, 2013).

The clone discovered in this work possessed a mutation EF-G (T493N) located in domain 4. This domain comprises the tail of the molecule and is physically involved in displacing the A-site tRNA towards the P-site position (Agrawal et al., 1998; Agrawal et al., 1999). EF-G mutations are commonly associated with resistance to the steroid antibiotic, fusidic acid (Johanson & Hughes, 1994; Macvanin et al., 2000), which stabilizes

the EF-G-GDP complex on the ribosome (Bodley et al., 1969). However, one study revealed that a number of fusidic acid resistance mutants, with mutations in the other EF-G domains were also resistant to the aminoglycoside, kanamycin (Johanson & Hughes, 1994). It therefore seems unlikely that a mutation mapped specifically to domain 4 leads to aminoglycoside resistance. It is possible however, that an EF-G mutant might influence the access of the antibiotic to its ribosomal site if, for example, it leads to a conformational change in the ribosome.

Another possible explanation for the *E. coli* MDS42 *yejG* (HTD) phenotype is pleiotropy associated with mutations in *fusA*. Studies have shown that *fusA* mutations lead to a reduction in cell growth, resulting from a defect in protein synthesis (Hou et al., 1994). As mentioned earlier (above), the binding of EF-G in complex with GTP to the ribosome is followed by hydrolysis of GTP to GDP and then the dissociation of EF-G.GDP. Regeneration of EF-G.GDP is by the spontaneous exchange of GDP for GTP (Macvanin et al., 2003). *fusA* mutants display a reduced GDP-to-GTP exchange leading to a reduction in the rate of protein synthesis (Macvanin et al., 2000). One of the major consequences of this is the perturbation of the global transcription regulator molecule ppGpp (guanosine 3′-biphosphate, 5′-biphospate), which acts as a stress signal (Macvanin et al., 2000).

A role for ppGpp in antibiotic resistance has been recognised (Pomares et al., 2008). For example, Greenway and England (1999) showed that ppGpp production was required for *E. coli* resistance to various antibiotics including the aminoglycoside, gentamicin. Baracchini and colleagues (1988) reported that ppGpp is able to bind directly to RNA polymerase to alter promoter selection by the enzyme. Greenway and England (1999) extended on this work and suggested that other molecules may also be involved. They showed that for certain antibiotics (e.g., anti-folates and cephalosporins), a key sigma factor involved in gene expression during stress conditions and slow growth, σ^s (Fischer et al., 1998) was also required for higher levels of resistance. According to the authors, a defect in protein synthesis increases the levels of ppGpp in the cell. This is thought to affect the expression of certain genes, including the upregulation of intrinsic resistance genes, which in certain cases may involve sigma factors.

It had been observed that the *E. coli* MDS42 *yejG* (HTD) clone revealed a reduction in cell growth (Chapter 6). Because the EF-G (T493N) mutation is likely to reduce the rate of protein synthesis (Hou et al., 1994), it is possible that this may lead to changes in the

levels of ppGpp in the cell. This in turn may result in the down-stream effect of altering the expression of genes associated with aminoglycoside resistance. These may include genes that code for efflux proteins or those that affect cell wall permeability. This hypothesis could be tested by constructing the *fusA*(G1478T) mutation in strains that are unable to produce either ppGpp in a RelA-dependent manner or σ^s (Greenway & England, 1999). The MIC values for aminoglycosides could be measured and then compared to that of control strains. It is possible that a ppGpp deficient strain may have multiple impairements that would make it hard to compare MIC values with that of control strains. Therefore, a second experiment can be considered. Wild-type cells can be treated with the serine analog serine hydroxamate, which increases ppGpp levels inside the cell leading to the stringent response (Shand et al., 1989). The resistance profile of treated cells can then be compared to that of the clone carrying the *fusA*(G1478T) mutation.

To the best of my knowledge, this is the first study to report a link between a T493N mutation in EF-G and aminoglycoside (or any other drug) resistance. This highlights the fact that there are many more routes that can lead to drug resistance that have not yet been recognized. It would be interesting to know whether YejG has a unique link with this novel mutation. However, since YejG endows resistance to cells with or without the *fusA*(G1478T) (Table 18), it is possible that the protein may have no special connection with EF-G. YejG may simply 'buy time' for mutations of larger effect [e.g., *fusA*(G1478T)] to take place. This hypothesis was investigated further in Chapter 5.

4.5 Materials and Methods

Materials and methods specific to this chapter are described in the following sections as well as in Section 3.4 (of Chapter 3). Common materials and molecular biology techniques are described in Appendix I.

4.5.1 Removing GFP

The GFP tag was removed as previously described in Section 3.4.1 (Chapter 3). As expected, a band size that corresponded to a 4,865 bp PCR product was visible on the ethidium bromide-stained agarose gel after running the *Dpn*I digested sample. A 542 bp PCR product was generated after amplifying the pCA24N-*yejG*(-GFP) plasmid with primers pCA24N.for and pCA24N.rev2 (Table I.3, Appendix I.5) and then sequenced.

4.5.2 Creating a mutant library

The error-prone PCR was performed using the GeneMorph II Random Mutagenesis Kit from Agilent technologies that was used to create the *cynT*, *yeaD* and *ydfW* libraries (Section 3.4.2).

4.5.2.1 Error-prone PCR

Random mutations were introduced into *yejG* during PCR using the same method described in section 3.4.2.1 (Chapter 3). The amount of plasmid DNA used in the reaction was 92 ng. The plasmid (pCA24N-*yejG*) was 4,865 bp in size and the amplified *yejG* product (plus flanking sequences) was 542 bp. Therefore, the amount of template DNA in the reaction was 10 ng [(542 bp/4,865 bp) x 92 ng]. For a high mutation rate (9-16 mutations/kb), 0.1-100 ng of template was required. The PCR efficiency (calculated from the total product yield and the amount of starting template) was 0.11 (*yejG* = 92 ng $\rightarrow$ 2,112.5 ng; $d = 4.52$; PCR efficiency = 0.097.

4.5.2.2 Vector and Insert preparation

The pCA24N vector backbone [from pCA24N-*trpD* (4510 bp)] and *yejG* insert (355 bp) were prepared as previously described in 3.4.2.2 (Chapter 3).

4.5.2.3 Preparation of a test library

A test library was prepared according to the protocols described in Section 3.4.2.3 (Chapter 3). The digested *yejG* (12 ng) was ligated to the pCA24N vector backbone (50 ng) at a 3:1 molar excess.

4.5.2.4 Construction of a full-size library

The full-size library was constructed using the same method described in Section 3.4.2.4 (Chapter 3). Three µL aliquots of the 'vector + insert' ligation were used to transform 12 aliquots (50 µL) of electrocompetent *E. coli* DH5α-E cells. A total volume of 7,150 µL transformed cells were spread on a total of two library agar plates supplemented with chloramphenicol (20 $\mu g.mL^{-1}$ final concentration), IPTG (50 µM final concentration) and varying concentrations of either tobramycin (12, 18 and 24 $\mu g.mL^{-1}$) or apramycin (32, 36 and 40 $\mu g.mL^{-1}$).

4.5.3 Analysing the library composition

The composition of the *yejG* library was analysed as described in Section 3.4.3 (Chapter 3). After PCR screening, a product size of ~ 550 bp (542 bp) was observed on an ethidium bromide-stained agarose gel and confirmed that the amplification was successful.

4.5.4 Determining MIC values using agar dilution method

The MIC values of the strains expressing wild-type YejG were measured as previously described in Section 3.4.4 (Chapter 3). This method was also used to determine the MIC values of the YejG-expressing strain, the negative control (expressing GFP) and the *E. coli* MDS42 *yejG* (HTD) clone for various aminoglycosides; tobramycin, kanamycin, sisomicin and apramycin.

4.5.5 Selection of 'winners' from library

Winners from the *yejG* library were selected according to the protocols described in Section 3.4.5 (Chapter 3).

4.5.6 Analysing sequences of 'winners'

Sequences of the 'winners' were analysed as previously described in Section 3.4.6 (Chapter 3). PCR products (542 bp) were sequenced using the pCA24N.for and pCA24N.rev2 primers (Table I.3, Appendix I.5).

4.5.7 Retransformation tests

Retransformation tests were performed using the same method previously described in Section 3.4.7 (Chapter 3).

4.5.8 Transforming *E. coli* MDS42 cells with *yejG* library

One hundred and fifty µL freezer aliquot of mutant *yejG E. coli* DH5α-E library was thawed at room temperature and divided into 10x 15 µL aliquots. The plasmid was extracted from each 15 µL aliquots using the Qiaquick Spin Miniprep Kit (Qiagen). The DNA was eluted in 30 µL EB in each case. The DNA samples were pulled together into x1 tube (300 µL).

E. coli MDS42 cells (Scarab Genomics LLC; Table I.1, Appendix I.3) were made electrocompetent (Appendix I.13). Three µL of the extracted plasmid was used to transform 14 aliquots of *E. coli* MDS42 electrocompetent cells (50 µL). SOC (500 µL) was added to the transformed cells and they were recovered at 37 °C for 1 h. Aliquots (5 µL and 20 µL) of transformed cells were spread on LB-chloramphenicol (20 $\mu g.mL^{-1}$ final concentration) plates. The size of the library was calculated based on the number of colonies present on these plates. The remaining volume (7,675 µL) was evenly spread on three LB agar library plates supplemented with chloramphenicol (20 $\mu g.mL^{-1}$ final concentration). Ten pg of pUC19 was also added to a 50 µL aliquot of MDS42 cells in order to calculate the transformation efficiency. Aliquots (5 µL, 20 µL and 100 µL) of this were spread on LB-ampicillin (100 $\mu g.mL^{-1}$ final concentration). The plates were incubated at 37 °C for 16 h.

4.5.9 Recovering *E. coli* MDS42 cells off library plates

The cells were recovered off the library plates and aliquoted in the same manner as previously described (Section 3.4.3).

4.5.10 Experiments involving the *E. coli* MDS42 *yejG* library

To determine the MIC of strains using the agar dilution method, selection experiments, retransformation tests and analysis of 'winners' were performed as previously described in Sections 3.4.3-3.4.7 (Chapter 3).

4.5.11 Replacing pCA24N with pBlueScript plasmids

Overnight cultures of the strain harbouring the pCA24N plasmid [*E. coli* MDS42 *yejG* (HTD)] were made electrocompetent (Appendix I.13) and transformed (Appendix I.14) with 20 ng of pBlueScript II KS+ plasmid (Fermentas). Five hundred µL of SOC was added to the electroporated cells and they were recovered at 37 °C for 1 h. Aliquots (5 µL and 20 µL) were spread on LB-ampicillin (100 $\mu g.mL^{-1}$ final concentration) plates (to select for clones transformed with pBlueScript) and incubated at 37 °C overnight.

One colony was picked off the LB-ampicillin plate and used to inoculate 2 mL LB-ampicillin (100 $\mu g.mL^{-1}$ final concentration) and the culture was incubated at 37 °C for 6 h. Three µL of this culture was used to inoculate 3 mL LB-ampicillin (100 $\mu g.mL^{-1}$ final concentration) and the culture was incubated at 37 °C for 15 h. Fifty µL of the saturated culture was used to inoculate 5 mL LB-ampicillin (100 $\mu g.mL^{-1}$ final concentration) and the culture was incubated at 37 °C for 4 h. The OD_{600} of the culture was measured and 200 cells were spread on LB-ampicillin (100 $\mu g.mL^{-1}$ final concentration) as well as LB-chloramphenicol (34 $\mu g.mL^{-1}$ final concentration) plates, based on the approximation of $OD_{600} \sim 1 = 2.5 \times 10^8$ cfu/mL (Sambrook & Russell, 2001). The plates were incubated at 37 °C overnight.

The next day, there were no colonies on the LB-chloramphenicol plate even after incubating for 2 days 37 °C. One colony was picked off the LB-ampicillin plate and inoculated in 3 mL LB-ampicillin (100 $\mu g.mL^{-1}$ final concentration) as well as 3 mL LB-chloramphenicol (34 $\mu g.mL^{-1}$ final concentration; to identify clones that still harboured the pCA24N plasmid). The cultures were incubated at 37 °C overnight. There was growth for

the LB-ampicillin culture but no growth for the LB-chloramphenicol culture. This indicated that the pCA24N plasmid in the selected *E. coli* MDS42 *yejG* (HTD) colony had been replaced with pBlueScript. A freezer stock of this culture was prepared [glycerol was added to 300 µL of culture to 12.5 % (v/v) final concentration] and stored at – 80 °C.

The same procedure was employed to replace the pBlueScript plasmid carried in cells with the pCA24N plasmid. However, this time the cells were grown in the presence of chloramphenicol after transforming with 20 ng of pCA24N plasmid until the pBlueScript had been replaced. The cells were passaged for the same amount of time to replace the pBlueScript plasmid.

4.5.12 Preparing DNA for whole genome sequencing

Genomic DNA was isolated from 1 mL overnight culture of *E. coli* MDS42 *yejG* (HTD) and the parental strain (*E. coli* MDS42 with no plasmid) using the Wizard® Genomic DNA Purification Kit (Promega). The DNA was rehydrated using 200 µL of EB (Qiagen) (as directed by the sequencing company; Ambry Genetics) and then measured spectrophotometrically (Appendix I.11); ~ 450 $ng.mL^{-1}$ for *E. coli* MDS42 yejG (HTD) and ~ 440 $ng.mL^{-1}$ for parent strain. One µL of each DNA sample was added to 4 µL water and 1 µL x6 loading dye (Fermentas) and run on a 1 % ethidium-bromide stained gel to confirm the presence of DNA. One hundred µL of the genomic DNA for both strains were sent to Ambry Genetics for whole genome sequencing on the Illumina Genome Analyzer IIx.

CHAPTER 5

Investigating the biochemical basis of aminoglycoside resistance

The ribosome experiments were performed with the help of Dr. Andrew Cridge (University of Otaga). Prof. Greg Cook (University of Otago) performed the membrane potential experiment because our laboratory was not suitably equipped.

Introduction

In the previous chapter, it was shown that YejG over-expression conferred aminoglycoside resistance (Section 4.3.1 & 4.3.6). However, the biochemical basis of resistance is still unclear. To gain insight into this, two possible mechanisms were explored.

In Section 4.1.2, it was mentioned that aminoglycosides bind to the 30S bacterial ribosomal subunit to inhibit protein synthesis. To investigate whether YejG could be binding to the ribosome and somehow reducing the binding of the drug to the ribosome, the interaction between YejG and the ribosome was tested.

The second mechanism that was explored was to assess whether YejG was decreasing uptake of the drug to increase resistance. This can be achieved by regulating the cell's membrane potential. To reach their target, aminoglycosides must cross the bacterial membrane. The transport of the drug inside *E. coli* has been shown to be dependent on the proton motive force (Damper & Epstein, 1981; Fraimow et al., 1991). The positively charged aminoglycoside antibiotic binds to the negatively charged cell surface of the bacteria and is thought to reach the cytoplasm through the disruption of Mg^{2+} bridges between adjacent lipopolysaccharide molecules (Vaara, 1992). Subsequent transfer of the drug across the cytoplasmic (inner) membrane is dependent upon the membrane potential (Bryan & Kwan, 1983).

One of the mechanisms used by bacteria to attain resistance is to decrease the uptake of the toxin by lowering the membrane potential. As a result, access of the antibiotic into the cell is inhibited. The external pH has been demonstrated to affect the membrane potential (Damper & Epstein, 1981; Hirota et al., 1981); when the external pH is reduced, membrane potential falls. A fall in membrane potential inhibits uptake of the drug thereby increasing resistance (Figure 10). Thus, a role for YejG in regulating the membrane potential was investigated.

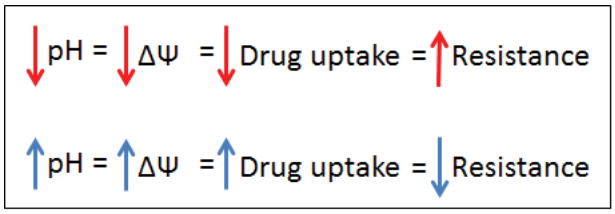


Figure 10. The effect of external pH on the membrane potential (ΔΨ) and drug resistance

For experiments in this chapter, the *E. coli* DH5α-E strain was used because this was the same strain that was employed in the beginning of the project [library-on-library screening experiment (Soo et al., 2011)].

5.2 Results

5.2.1 Testing the interaction between YejG and the ribosome

To investigate whether YejG was binding to the ribosome, the *E. coli* DH5α-E harbouring the pCA24N-*yejG*(-GFP) plasmid (Table I.2; Appendix I.4) was used. A polyclonal antibody against two of the proteins of the large ribosomal subunit (L7/L12) was used to detect the *E.coli* ribosome. Additionally, a polyclonal anti-His_6 antibody was used to probe the His_6-tagged YejG protein. Determining whether YejG associates with the ribosome would indicate that there is likely to be an interaction between the two.

The anti-ribosome antibody was kindly supplied by Prof. Warren Tate (University of Otago). To begin, preliminary dot blots with the anti-ribosome antibody were performed in order to investigate whether it was functional and to determine the appropriate concentration to use in the downstream experiment. The results confirmed that the anti-ribosome antibody was active and that a 1/2000 dilution of primary antibody was the best concentration to use (Figure 11).

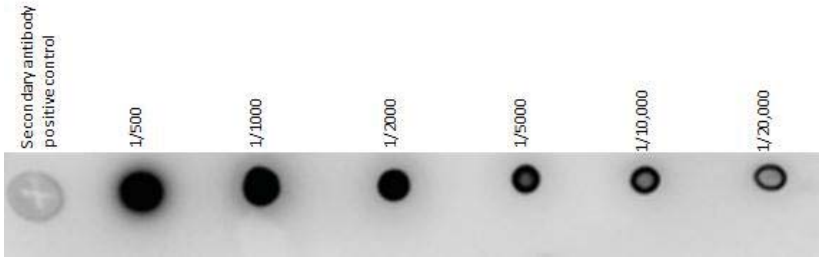


Figure 11. Testing the activity of the anti-ribosome antibody

Dot blot showing the serial dilutions of the *E. coli* anti-ribosomal antibody that were probed with anti-rabbit secondary antibody on a polyvinylidene difluoride (PVDF) membrane. A positive control for the secondary antibody (without primary antibody) was included to confirm that it was also working. The antibodies were detected with a chemiluminescent detection reagent.

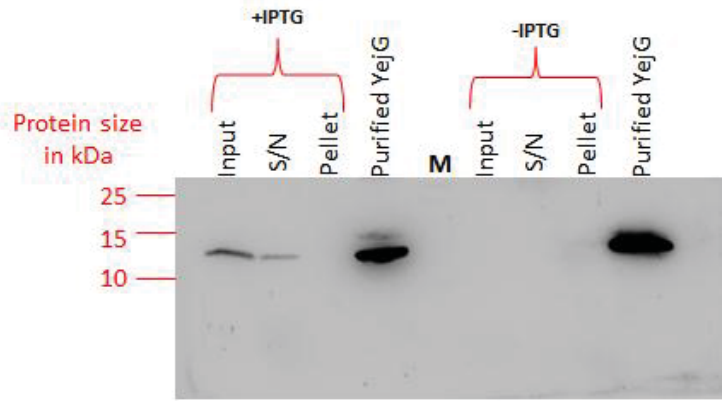
Next, cultures of the *E. coli* DH5α-E strain harbouring pCA24N-*yejG*(-GFP) were grown in the presence and absence of IPTG. The cells were lysed and three fractions were obtained, input (starting sample), pellet (contains ribosome) and supernatant. If YejG is associated with the ribosome, the protein would be detected in the pellet fraction. However, if YejG was detected in the supernatant fraction but absent from the ribosome fraction, it confirmed that the protein did not bind to the ribosome.

Using the anti-His_6 antibody, the YejG protein was detected only in the input and supernatant fractions for the + IPTG culture, at the expected size of 12.5 kDa (Figure 12 A). The purified YejG positive control revealed a band with the same size, confirming the correct size of the over-expressed protein. YejG was not detected in the ribosomal fraction of the + IPTG culture even though it was present in the culture (input fraction). For the uninduced (– IPTG) culture, YejG was not detected in any of the samples. Thus, it is very likely that the detected signal in the + IPTG culture is my protein of interest (YejG) as this signal is absent in the - IPTG culture.

To confirm that the ribosome was pelleted during the experimental procedure, the membrane was stripped and reprobed with the anti-ribosome antibody. The results showed that for the + IPTG culture, the input and pellet fractions revealed two bands, one of which corresponded to the size of the L7/L12 ribosomal subunit (12.3 kDa). The other band (~15 kDa) was probably due to the polyclonal antibody crossreacting with another protein. This indicated that the ribosome was located in the pellet sample (Figure 12 B). No ribosome was detected in the pellet fraction in the – IPTG culture. This may be due to losing the ribosome during sample preparation.

Overall, the experimental results show that YejG was present in the + IPTG culture (input) but was detected only in the supernatant fraction and not in the pellet fraction containing the ribosome. Thus, there is no evidence suggesting that YejG interacts with the ribosome.

A.



B.

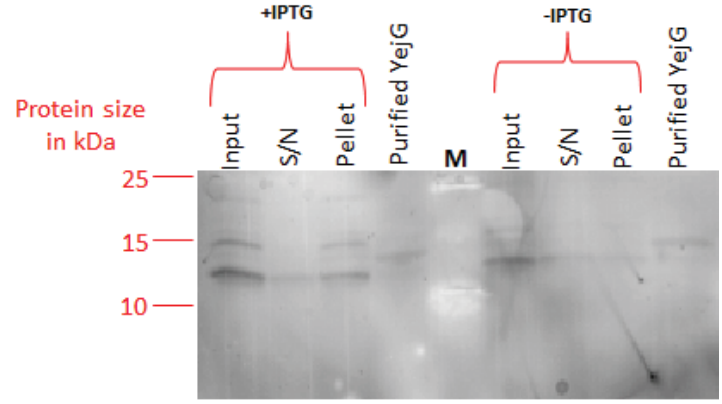


Figure 12. Western blot of YejG and ribosomal proteins from *E. coli* DH5α-E + and – IPTG cultures

The top picture (**A**) is a western blot of YejG from + and – IPTG cultures probed with polyclonal His-antibody on a PVDF membrane. A purified sample of YejG was included as a positive control. The bottom picture (**B**) is the same membrane that has been stripped this time probed with an anti-ribosome antibody. The input, supernatant (S/N) and ribosomal (pellet) fractions and a protein marker (M) are shown.

5.2.2 Measuring tobramycin resistance at varying pH values

As mentioned earlier (Section 5.1), the external pH has been shown to affect the membrane potential, which in turn can alter uptake of a drug (Figure 10). A protein that is able to regulate the membrane potential by lowering it could decrease the transport of the antibiotic into the cell and endow the bacterium with resistance. Hence, the sensitivity of a strain

expressing YejG towards aminoglycosides (represented by tobramycin) at varying pHs was assessed.

The tobramycin sensitivity of an *E. coli* DH5α-E clone carrying the pCA24N-*yejG*(-GFP) plasmid and a control strain carrying the pCA24N-GFP plasmid was tested. The MIC values were measured using the plate-based (E-test strips) method (see Section 2.5.1.1). Cells were spread on minimal agar plates [as performed in the work by Damper and Epstein (1981)] at varying pH values (5.5, 6.5, 7.5 and 9) supplemented with chloramphenicol and IPTG (Section 5.4.2). In this experiment, the MIC values of the YejG-expressing strain were predicted to show less variation with pH, compared to the control, if YejG were regulating the cell's membrane potential.

The results showed that the YejG-expressing clone was more resistant than the control to tobramycin at all of the different pH values (Figure 13 and Table 19). Furthermore, the clones were most resistant to tobramycin at the lowest pH (5.5); reducing the external pH results in a fall in membrane potential and subsequently, a decrease in uptake of the drug (Figure 10). The YejG-expressing clone is nearly three times more resistant than the control strain at this pH, 32 versus 12 $\mu g.mL^{-1}$, respectively. As the pH increases from 5.5 to 9, the YejG-expressing strain's MIC value decreases by at least 2000 fold (from 32 to < 0.016 $\mu g.mL^{-1}$ at pH 9; Table 19). The control strain on the other hand, is unable to grow at this pH.

Initially, the decrease in resistance displayed by the YejG-expressing strain is higher than that displayed by the control strain (~ 11-fold versus 8-fold) (Table 19). However, from pH 6.5 to 7.5, the reduction in tobramycin resistance is less when YejG is expressed (3-fold versus 6-fold for the control). Overall, it is difficult to determine whether YejG is buffering against the changes in membrane potential that result from altering the external pH. Therefore, the membrane potentials of the YejG-expressing strain and the control were measured.

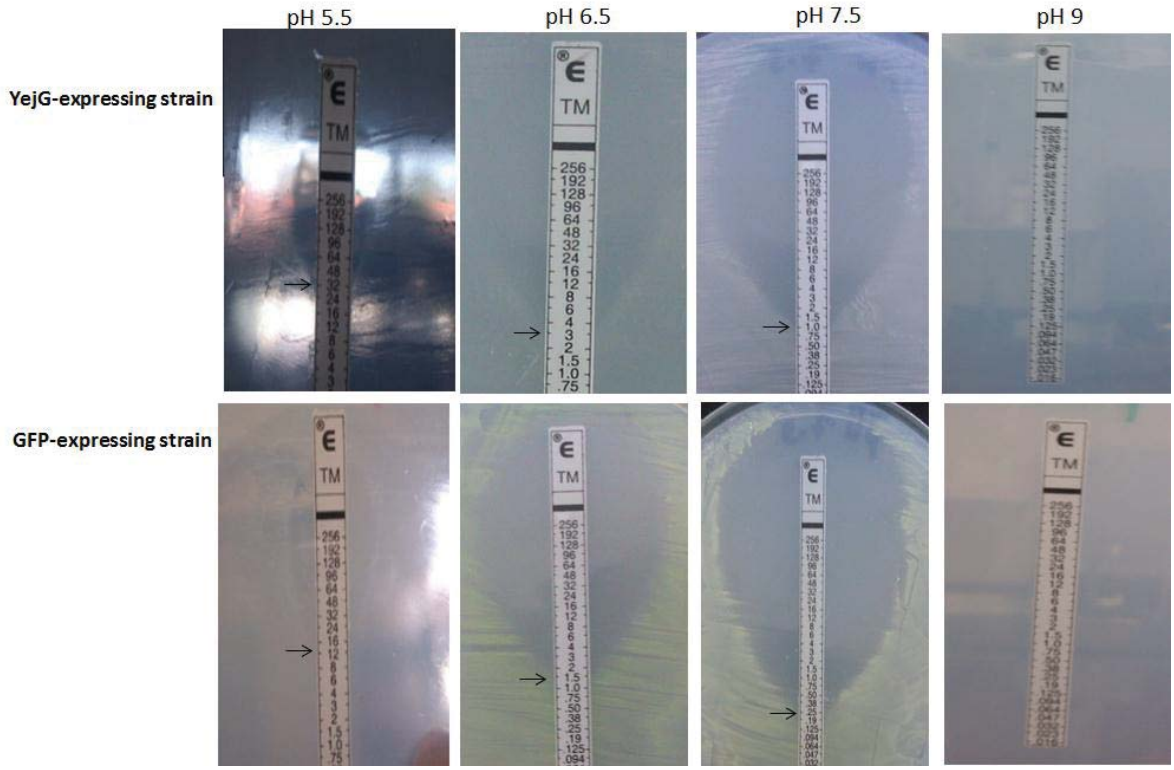


Figure 13. Effect of external pH on tobramycin sensitivity of *E. coli* DH5α-E YejG- and GFP-expressing strains

Plates show the MIC values of the clones at pH 5.5, 6.5, 7.5 and 9 for tobramycin. At pH 9, there was no growth for the control strain and the YejG-expressing strain showed minimal growth. The lowest concentration on the tobramycin E-test was 0.016 µg.mL^{-1}. Black arrows indicate where the MIC values were read off the E-test strip.

Table 19. Tobramycin resistance of YejG- and GFP-expressing strains at varying pH values

The MIC values (µg.mL^{-1}) in of the YejG- and GFP-expressing strains for tobramycin at pH values 5.5, 6.5, 7.5 and 9 were determined using the plate-based E-test method.

	pH 5.5	**pH 6.5**	**pH 7.5**	**pH 9**
DH5α-E pCA24N-*yejG*(-GFP)	32	3	1	<0.016
DH5α-E pCA24N-GFP	12	1.5	0.25	No growth

5.2.3 Measuring the membrane potential of YejG-expressing cells.

In order to confirm whether YejG was regulating the membrane potential, an alternative experiment was carried out. This time, the membrane potentials of YejG- and GFP-expressing *E. coli* DH5α-E strains were measured. This was achieved by determining the distribution of lipophilic charged molecules [such as [^{3}H]tetraphenylphosphonium (TPP+) ions] in the cell. Lipophilic cations distribute across the cytoplasmic membrane in proportion to the membrane potential, thus allowing the membrane potential to be measured (Fraimow et al., 1991). It was hypothesised that if YejG was endowing cells with resistance to aminoglycoside by regulating the membrane potential (to reduce drug uptake), then cells expressing this protein would have a lower membrane potential than the GFP-expressing control.

Initially, a growth assay was performed to assess the effect of protein expression on cell growth (Figure 14). The YejG- and GFP-expressing clones together with a wild-type *E. coli* DH5α-E strain (not carrying a plasmid) were grown in LB broth and induced with IPTG at mid-log phase (OD_{600} ~ 0.6) for ~ 3 h. The absorbance (OD_{600}) of three replicates for each clone was measured over ~ 6 h. The results showed that protein over-expression decreased the growth of the YejG- and GFP-expressing strains compared to the wild-type *E. coli* DH5α-E strain.

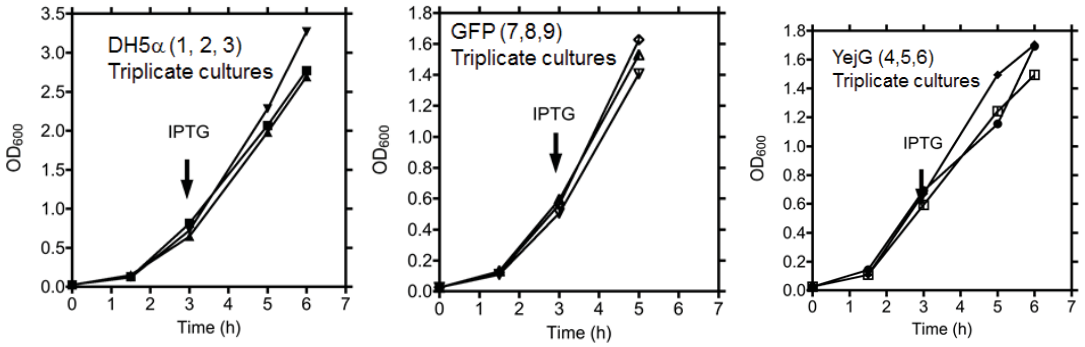


Figure 14. Growth curves of wild-type *E. coli* DH5α-E and YejG- and GFP-expressing strains in the presence of IPTG

Wild-type *E. coli* DH5α-E and YejG- and GFP-expressing strains were grown in triplicates in LB medium and then induced with IPTG after 3 h of incubation. The growth was measured by monitoring OD_{600}.

The membrane potentials of the clones (in Figure 14) were also measured, at two time points after the addition of IPTG (Figure 15). The wild-type clones displayed the lowest membrane potentials (~ 100-120 mV). In contrast, the, the YejG- and GFP-expressing strains had higher membrane potentials and showed similar measurements to each other (~ 160-180 mV). These results fit in well with the growth data in Figure 14. Since the wild-type clones displayed faster growth, their membrane potentials are being drained to drive ATP synthesis and sustain the growth rate. For the strains over-expressing protein (YejG and GFP), the membrane potential is high because the growth rate is slower.

The results showed that the YejG strain has a similar membrane potential as the GFP control strain. Therefore, the YejG protein is unlikely to be endowing cells with aminoglycoside resistance by regulating the membrane potential.

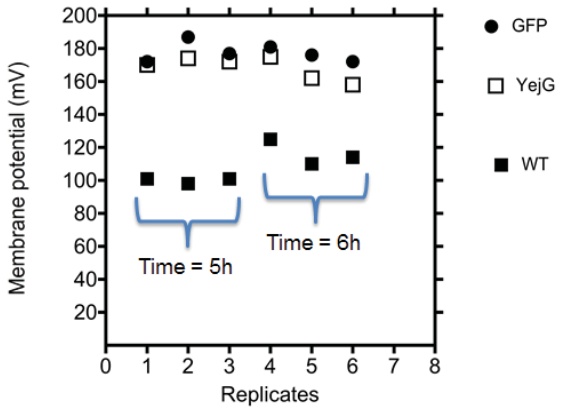


Figure 15. Membrane potential of wild-type *E-coli* DH5α-E and YejG- and GFP-expressing strains

The membrane potentials of *E-coli* DH5α-E cells expressing YejG and GFP and a wild-type DH5α-E clone were measured in triplicates using TPP^+ distribution by membrane filtration. The membrane potential across the cell membrane was calculated from the uptake of TPP^+.

5.3 Discussion

In this chapter the biochemical basis for YejG-mediated aminoglycoside resistance was initially investigated by exploring the protein's association with the ribosome. It was hypothesized that YejG might be binding at or near the A-site of the ribosome (target for aminoglyoside antibiotics) diminishing the ability of aminoglycosides to bind. As a first step towards testing this hypothesis, a western blot was performed to assess whether YejG associated with the ribosome. For this experiment, wild-type *E. coli* DH5α-E cultures expressing or not expressing YejG were lysed and the pellet that contains the ribosome was obtained. Polyclonal anti-His_6 and anti-ribosomal (L7/L12 subunit) antibodies were used to determine whether YejG would be detected in the pellet fraction (containing the ribosome).

When the membrane was probed with the anti-His_6 antibody, a band at the expected size of ~ 12.5 kDa was observed only in the starting sample (input) and the non-ribosomal fraction (supernatant) of the His_6-YejG-expressing culture (Figure 12A). A purified YejG positive control confirmed the correct size of the over-expressing protein. To confirm that the ribosome was present in the pellet sample, the membrane was stripped and re-probed with the polyclonal anti-ribosomal antibody. There were two bands present only in the samples expected to contain the ribosome (input and pellet fractions, Figure 12B). One of the bands (~ 12.5 kDa) corresponded to the expected size of the L7/L12 protein (12.3 kDa). This band was absent from the fractions that contained YejG. The results do not support that YejG binds to the ribosome.

It is possible that YejG might still bind to the ribosome but that this interaction is too weak to be detected using the approach employed in this study. During the course of the experiment, YejG might have been displaced off the ribosome due to the centrifugational forces employed during the experimental procedure. For future experiments, this can be tested by performing formaldehyde cross-linking experiments to stabilize the protein-ribosome (Klockenbusch & Kast, 2010).

To further investigate the function of YejG, a second mechanism concerning the membrane potential was explored. The uptake of aminoglycosides has been shown to be dependent on the membrane potential (Fraimow et al., 1991). If YejG functioned to reduce the cell's membrane potential, this would decrease drug uptake and increase resistance. To test this, MIC values for tobramycin of the clone expressing YejG and the GFP-expressing

strains were measured in conditions that affected the membrane potential. Since external pH alters the membrane potential (Hirota et al., 1981), the strains were grown on agar plates at pH 5.5, 6.5, 7.5 and 9. It was hypothesized that if YejG played a role in regulating the membrane potential, the MIC values of the strain expressing this protein would display less variation with pH, compared to the GFP negative control.

The results showed that the clone expressing YejG was more resistant to tobramycin at all of the different pH values tested when compared to the control (Figure 13 and Table 19). However, the MIC values of the YejG-expressing strain varied considerably with pH. As expected, at pH 5.5 both strains were the most resistant to tobramycin, since a decrease in external pH results in a fall in membrane potential thereby reducing the uptake of the drug. When the pH was increased to 6.5, the YejG-expressing strain decreased in resistance by ~11-fold, whereas resistance dropped by only 8-fold for the control. When the pH was further increased to 7.5 and 9, the clone expressing YejG decreased in resistance by 32-fold and ~ 2000-fold, respectively, from its initial resistance. The control strain on the other hand decreased by 48 fold at pH 7.5 from the initial resistance and was completely susceptible at pH 9. It appeared that YejG was unable to buffer against the changes in membrane potential. Since it was difficult to make solid conclusions from these experiments, for the next step, the YejG- and the GFP-expressing strains were sent to Prof. Greg Cook (University of Otago) for direct measurements of their membrane potential.

For these experiments, it was hypothesized that if YejG were endowing cells with resistance to aminoglycosides by regulating the membrane potential, then the cells expressing YejG would have a lower membrane potential than the GFP-expressing strain. To begin, a growth assay was performed on the YejG- and GFP-expressing strains together with wild-type *E. coli* DH5α-E clones. The results showed that the clones expressing YejG and GFP grew slower than the wild-type clones when IPTG was added to the cultures (Figure 14). This suggested that only the YejG- and GFP-expressing strains were over-expressing protein. Furthermore, the results showed that expression of either protein decreased cell growth, by comparable amounts.

Next, the membrane potential of the YejG- and GFP-expressing strains and also the wild-type clones was measured. The membrane potentials of the strains expressing YejG and GFP were ~160-180 mV (Figure 15). These measurements were higher than the membrane potential of the wild-type clones (~100-120 mV) that showed faster growth.

This can be explained by there being less demand on the membrane potentials of the YejG- and GFP-expressing strains as they showed decreased growth compared to the wild-type clones.

The results revealed that the YejG-expressing strain had the same membrane potential as the control strain. This indicates that YejG does not have a role in regulating the membrane potential. This probably explains why YejG over-expression did not endow the bacterium with resistance to other cationic anti-microbial agents [such as polymyxin B and polylysine (Rivera et al., 1988)] during the library screening experiments, as they also respond to a threshold level of membrane potential.

In this chapter, I was unable to identify a biochemical basis for YejG-mediated aminoglycoside resistance. Nonetheless it was important to investigate whether YejG had a pleiotropic effect on fitness, which may be contributing to resistance.

5.4 Materials and Methods

All reagents were from Sigma, unless otherwise specified. The materials and methods described here are specific to this chapter. Common materials and molecular biology techniques are described in Appendix I.

5.4.1 Testing interaction between YejG and the ribosome

5.4.1.1 Testing the anti-ribosomal antibody

A freeze-dried polyclonal antibody against the large ribosomal subunit L7/L12 that had been raised in rabbit (dating from 1984) was kindly supplied by Prof. Warren Tate (University of Otago). To begin, a preliminary experiment was performed to confirm whether the antibody was active and to determine the concentration of the antibody to use in downstream experiments. The antibody against L7/L12 subunits (~1.5 mg) was rehydrated with 200 µL 1X TBS buffer [4 mM Tris (pH 7.6), 15 mM NaCl and 0.05 % polyethylene sorbitan monolaurate] and then diluted in 1x TBS buffer as follows; 1/500, 1/1000, 1/2000, 1/5000, 1/10,000 and 1/20,000.

Using a pencil, six dots were marked on a 12 cm x 5 cm rectangle of polydivinyl benzene (PDVB) membrane to indicate where each of the six dilutions would be pipetted. The membrane was moistened in methanol (~ 5 s) with slow shaking at room temperature in 1x TBS (~ 10 min). Two µL of each antibody dilution was pipetted onto the marked positions on the membrane and the membrane was placed back in TBS buffer (to avoid drying out). The membrane was blotted in 5 % TBS + milk powder (2.5 g in 50 mL final volume) (low fat milk powder; Pams) with slow shaking for a further 30 min. The membrane was then washed with 1x TBS with slow shaking for 5 min. This washing step was repeated two more times (total of 3 washes).

The membrane was then probed with anti-rabbit IgG 1/100,000 (0.4 mg; Thermo Scientific) + milk (5 % final concentration; 2.5 g in 50 mL final volume) with shaking for 2 h at room temperature. The membrane was washed with 1x TBS with slow shaking for 5 min. This washing step was repeated three times. A solution of chemiluminescence reagent [25 µL of light sensitive Luminol (2.5 mM final concentration), 11 µL of light sensitive p-coumaric acid (40 µM final concentration) and 1.5 µL of 30 % hydrogen peroxide acid (0.009 % final concentration)] was added to 5 mL 0.1 M Tris (pH 8.5) (100 mM final

concentration). Blots were developed with the reagent solution and an image was acquired using a Luminescent Image Analyzer Las4000 (Alphatech).

5.4.1.2 Preparation of ribosomal fractions

This protocol is adapted from a method by Daigle and Brown (2004). Two cultures (150 mL) of *E. coli* DH5α-E harbouring the *yejG*-pCA24N(-GFP) plasmid were grown in LB-chloramphenicol (34 μg.mL^{-1} final concentration) at 37 ºC. At OD_{600} ~ 0.6, IPTG (0.4 mM final concentration) was added to one of the flasks and the cultures were incubated for a further 6 h at 37 ºC. The cultures were centrifuged (4,985 xg, 4 ºC for 15 min) and the pellets were frozen (- 80 ºC).

The next day, the pellets were thawed on ice and resuspended in 10 mL Buffer A (100 mM NH_4Cl, 10 mM $MgCl_2$ and Tris-Cl pH 7.5). The cells were sonicated (45 % amplitude; Ultrasonic processor S-4000, Misonix) and then centrifuged (21,000 xg, 4 ºC for 2.5 h). The supernatant samples (Input) were collected and aliquots (6x 200 μL) from both the +/- IPTG cultures were centrifuged further (150,000 xg, 4 ºC for 2 h). The supernatant was collected (Supernatant) and the pellets (Ribosome) were resuspended in 200 μL of Buffer A. The Input, Supernatant and Ribosome from the +/- IPTG fractions and a pre-stained protein ladder were run on a 12 % SDS -PAGE gel (200 V, 40 min) (Appendix I15).

5.4.1.3 Western blotting

The samples from the SDS-PAGE gel were transferred onto a PDVB membrane. PDVB membrane (15 cm by 11 cm) was moistened in methanol (~ 5 s) and then soaked in Transfer buffer (10 mM $NaHCO_3$, 3 mM Na_2CO_3, 20 % MeOH, pH 9.4) for ~5 min. The sponges and Whatman paper were also soaked in Transfer buffer. For transfer, the following sandwich was made: sponge, Whatman paper, SDS gel, PDVB membrane, Whatman paper and sponge. Air bubbles were removed between the layers by using a flat surface to slide across the sandwich with gentle pressure. The sandwich was clamped and run in a tank containing Transfer buffer for 3 h (0.6 A, ~ 50 V). If the ladder had stained the PDVB membrane, it indicated that the transfer was successful. The ladder sizes, the corners of the gel and the front of the membrane were all marked using a pencil.

The membrane was soaked in 1x TBS + polyethylene sorbitan monolaurate (0.05 % final concentration) + milk powder (1 % final concentration; 1g in 100 mL x1 TBS) and rocked overnight at 4 ºC. The next day, the membrane was probed with anti-His_6 antibody (200 µg; Santa Cruz Biotechnology Inc.) [1:200 dilution in 1x TBS + milk powder (1 % final concentration; 1g in 100 mL x1 TBS), 10 mL final volume] and rocked for 2 h at room temperature. The membrane was washed by rocking in 1x TBS + polyethylene sorbitan monolaurate (0.05 % final concentration) for 5 min, three times. It was then probed with anti-rabbit antibody [1:100,000 dilution in 1x TBS + milk powder (1 % final concentration), 10 mL final volume] and rocked for 2 h at room temperature. Finally, the membrane was washed twice with 1x TBS + polyethylene sorbitan monolaurate (0.05 % final concentration) and once with 1x TBS for 5 min with shaking. The membrane was developed as mentioned previously (Section 5.4.1.1).

5.4.1.4 Protein over-expression, purification and dialysis of YejG

For protein purification, metal affinity chromatography was used since the N-terminus of YejG was fused to a $(His)_6$ tag.

A 4 mL overnight culture was used to inoculate 250 mL LB supplemented with chloramphenicol (34 $\mu g.mL^{-1}$ final concentration) and incubated at 37 °C. When the culture reached OD_{600} ~ 0.6, IPTG (0.4 mM final concentration) was added to induce protein over-expression. The culture was incubated at 28 °C for a further 6 h. The cells were then pelleted at 5,000 xg at 4 °C for 20 min (Heraeus Multifuge 1S-R). The supernatant was discarded and the pellet was stored at – 80 °C overnight.

The pellet was thawed on ice and resuspended in 4 mL Column Buffer [50 mM potassium phosphate buffer (pH 8.0), 300 mM NaCl, 3 mM imidazole, 1 mM β-mercaptoethanol, 10 % (v/v) glycerol] for every 1 g of cells. Lysozyme (0.5 mg/mL, final concentration), DNase I (4 U, final amount) and 50 µL protease inhibitor cocktail (Sigma) were added to the cell suspension and left to incubated on ice for 30 min. To lyse the cells, 10 cycles of sonication (amplitude 45 %; Ultrasonic processor S-4000, Misonix) on ice were performed.

The lysate was centrifuged at 5,000 xg for 40 min, at 4 °C. To separate possible aggregates, the resulting supernatant was filtered through a 0.22-µm filter unit. Meanwhile,

1 mL of Talon® metal affinity resin (bed volume, 0.5 mL; Clontech) washed twice with 8 mL of Column Buffer and centrifuged at 800 xg at 4 °C for 2 min. The supernatant was incubated with the pre-washed resin at 4 °C for 2 h with rocking to allow the $(His)_6$-tagged protein to bind to the resin. The resin was then centrifuged at 800 ×g, 5 min, 4 °C and washed twice with 5 mL of Column Buffer. One mL of Column Buffer was added to the resin before transferring to a gravity flow column (Bio-Rad). The resin was washed twice with 3 mL of Column Buffer. For the last wash, 3 mL of Column Buffer with 15 mM of imidazole was used. The protein was eluted in 3 mL Elution Buffer [50 mM potassium phosphate buffer (pH 8.0), 300 mM NaCl, 150 mM imidazole, 1 mM β-mercaptoethanol, 10 % (v/v) glycerol].

YejG was dialysed overnight against Dialysis Buffer [50 mM potassium phosphate (pH 8.0), 150 mM NaCl, 10 % (v/v) glycerol)] at 4 °C in SpectraPor2 cellulose tubing (MWCO 12–14 kDa; SpectrumLabs) as instructed by the manufacturer. The protein sample recovered from dialysis was concentrated to 1 mL using a centrifugal filtration concentrator (Vivaspin 6; GE Healthcare). Protein aliquots were stored at 4 °C.

5.4.2 Preparing minimal agar plates with varying pH.

5.4.2.1 Preparing buffers

To obtain the desired pH, the following four buffers were required: MES [2-(N-morpholino) ethanesulfonic acid] was used for pHs 5.5 and 6.5; HEPES [4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)] was used for pH 7.5 and; Tris (hydroxymethyl)aminomethane-BASE was used for pH 9. One M of each buffer (100 mL final volume) was prepared and then adjusted to the appropriate pH; 1M HCl was used to adjust the x2 MES buffers to pHs 5.5 and 6.5; 1M NaOH was used to adjust the HEPES buffer and the Tris-BASE buffer to pHs 7.5 and 9, respectively. The buffers were filter-sterilised using a 33 mm 0.22 μm filter (Interlab Ltd) and stored at room temperature.

5.4.2.2 Preparing minimal agar plates with different pHs

To prepare minimal agar plates with the desired pHs, 100 mM (final concentration) of the four buffers (from Section 5.4.2.1) were added to minimal agar (200 mL final volume). To do this, 20 mL of each buffer (1M) was added to 30 mL 5x M9 media (x4 separate media

was prepared). The pH of each M9 medium was checked to confirm that the pHs were still 5.5, 6.5, 7.5 and 9 using a pH meter (Eutech pH Meter 510, Thermo Scientific). If required, they were adjusted with either 1M HCl or 1M NaOH (see Section 5.4.2.1). The media was made up to a final volume of 60 mL using water. Meanwhile, 3 g agar was prepared in a final volume of 136 mL (x4). The M9 media and agar were autoclaved and cooled by stirring at room temperature using magnetic stirrers. Forty µL of each M9 medium was pipetted onto a pH paper test strip (Thermo Fisher Scientific) to check that the pH was still within the range required. Each autoclaved M9 medium was then added to x1 agar and stirred. The remaining components (see below) were added to each agar (to make 200 mL final volume) and stirred once more. The agar was poured into petri-dishes (90 x 15 mm) and left to set at room temperatue.

The M9 agar plates were made up of agar (15g/L), 1× M9 medium, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$ (Merck), 0.4 % (w/v) glucose, 34 µg.mL^{-1} chloramphenicol, 50 µM IPTG (GBT GoldBio) and 2 µg/mL thiamine. Thiamine was added because the *E. coli* DH5α-E (Invitrogen) in this study has a *thi*−1 mutation (Section I.3 in Appendix I). The 1× M9 medium was diluted from a 5× M9 stock (Sigma; 33.9 g/L Na_2HPO_4, 15 g/L KH_2PO_4, 2.5 g/L NaCl and 5 g/L NH_4Cl).

5.4.2.3 Determining MIC values at varying pHs

The MIC of the strains expressing YejG and GFP were determined using the E-test method. The strains were grown and spread on the minimal agar plates at pHs 5.5, 6.5, 7.5 and 9 in the same manner as previously decribed in Section 2.5.1.1. Tobramycin E-test strips were placed in the middle of the plate on top of the bacterial lawn and incubated upside down at 37 ºC overnight. The MIC value was reported as the point the ellipse meets the strip from the zone of inhibition produced.

5.4.3 Measuring membrane potentials

Prof. Greg Cook (University of Otago) provided the following protocol for the experiment that he carried out.

For the determination of membrane potential ($\Delta\Psi$), 3 × 1 mL of log-phase cells were taken directly from the flask and added immediately to a glass tube containing [^{3}H]methyltriphenylphosphonium iodide (TPP^+) (30-60 Ci/mmol, 2.4 nM final concentration). After incubation for 5 min at 37 ºC, the cultures (900 µl in triplicate) were filtered rapidly through a 0.45 µM cellulose-acetate filter (Sartorius) as adapted from Zilberstein et al. (1979). The filters were washed twice with 2 mL of 100 mM LiCl and dried for 60 min at 40 ºC. Filters were resuspended in 2 mL of scintillation liquid and counts per min (CPM) determined using an LKB Wallac 1214 Rackbeta liquid scintillation counter (PerkinElmer Life Sciences). The intracellular volume (2.8 ± 0.5 µl mg of protein^{-1}) was determined previously (Shepherd et al., 2010). The $\Delta\Psi$ across the cell membrane was calculated from the uptake of [^{3}H]TPP^+ according to the Nernst relationship. Non-specific [^{3}H]TPP^+ binding was estimated from cells which had been treated with carbonyl cyanide *m*-chlorophenylhydrazone (CCCP, 250 µM) for 20 min prior to addition of [^{3}H]TPP^+.

CHAPTER 6

Probing the connection between YejG and the mutation in the fusA gene

The phage transduction experiments were performed under the guidance of Dr. Heather Hendrickson (Massey University).

6.1 Introduction

In the previous chapters, it was shown that YejG over-expression conferred resistance to aminoglycosides (Table 1 and 18). In Chapter 4 (Section 4.3.2) the *E. coli* MDS42 *yejG* (HTD) clone was discovered during the directed evolution experiments. Whole genome sequencing revealed that the clone harboured a mutation in the chromosomal *fusA* gene [*fusA*(G1478T), corresponding to the T493N amino acid substitution]. This mutation endowed the cell with a higher level of resistance to the aminoglycoside, tobramycin, than that afforded by YejG over-expression alone (24 $\mu g.mL^{-1}$ versus 18 $\mu g.mL^{-1}$; Table 14 and 16). However, an even greater increase in tobramycin resistance (42 $\mu g.mL^{-1}$) was attained when the *fusA*(G1478T) mutation and YejG were both present (Table 16). A mechanism for resistance conferred by the *fusA*(G1478T) mutation was proposed (Section 4.4). On the other hand, the biochemical basis for YejG-mediated resistance could not be identified (Chapter 5). Therefore, the next step was to investigate whether YejG had a pleiotropic effect on fitness in the presence of aminoglycosides and/or the *fusA*(G1478T) mutation.

It was mentioned earlier (Section 4.3.6) that Hou et al., (1994) discovered two *fusA* mutants that were also resistant to an aminoglycoside (kanamycin). The mutants in their study harboured substitutions that were at amino acid positions close to the T493N substitution found in the *E. coli* MDS42 *yejG* (HTD) clone. The *fusA* mutants displayed measurable defects in growth that were attributed to slower rates of protein synthesis compared to a control clone. Similarly, throughout the course of my experiments, a reduction in the growth rate for the *E. coli* MDS42 *yejG* (HTD) clone had been observed. However, this was true only when YejG was not over-expressed. Therefore, the aim of this Chapter was first, to determine whether carrying the *fusA*(G1478T) mutation was associated with a defect in cell growth and second, to investigate whether YejG might be improving the cell's fitness.

For these experiments, the competitive abilitity of selected clones were measured by performing fitness assays. To gain a better understanding of the fitness assay results, growth assays were also performed on the clones. Fitness experiments measure the competitive abilitity of a cell during 24 h. Performing growth assays allows us to gain a better insight into the fitness assay results by identifying potential changes in the growth cycle (lag, exponential and stationary phase), since an apparent difference in fitness can be caused by changes at any of these stages.

For the fitness experiments, attempts were made to construct an *E. coli* MDS42 strain that carried the *fusA*(G1478T) mutation. It was desirable not to use the *E. coli* MDS42 *yejG* (HTD) clone because although it carried the *fusA*(G1478T) mutation, it also harboured the pCA24N-*yejG*(HTD) plasmid. Although this plasmid could be replaced with the pBlueScript plasmid (Section 4.3.3), it was still preferable to construct a clone carrying the chromosomal *fusA* mutation but lacking plasmids. To begin, I attempted to construct such a clone by using two techniques, the Red recombinase system and P1 phage transduction. It had already been anticipated that it might be difficult to create a mutation in *fusA* since it is an essential gene.

6.1.1 Red recombinase system and generalized transduction

The Red recombinase system of bacteriophage λ enables homologous recombination between a chromosomal region of interest and a linear PCR product with arms that are homologous to the target DNA (Datsenko & Wanner, 2000). Disruption of a target gene is achieved by generating a PCR product containing an antibiotic marker flanked by homology arms that can be as short as 50 bp. To create a point mutation (such as in this chapter), a DNA sequence containing the desired base mutation is used. Recombination is promoted by the proteins encoded by the Red operon, which consists primarily of three genes encoding a nuclease inhibitor (Redγ) and two proteins that mediate site-specific recombination (Redα and β). Redγ binds to the bacterial RecBCD enzyme to inhibit its activities, Redα digests the 5´-ended strand of dsDNA end and Redβ binds to the exposed 3´ single-stranded overhangs promoting strand annealing and recombination (Datsenko & Wanner, 2000). Interaction between Redα and Redβ is required to catalyse the homologous recombination reaction (Stahl et al., 1997; Muyrers et al., 2000).

The Red/ET system is another way of disrupting and mutating genes in bacteria and is similar to λ Red. While the λ Red system is mediated by bacteriophage λ proteins (Redαβδ), the Red/ET recombination relies on the Rac prophage coded proteins, RecE and RecT. The Redα/β and RecE/T protein pairs are functionally and operationally equivalent (Muyrers et al., 2000).

On the other hand, generalised transduction is a process that allows the packaging of any section of bacterial DNA into infectious phage particles (without the presence of

phage DNA), which can then be transferred from one bacterium to another (Sandulache et al., 1984). P1 phage can be used to mediate generalised transduction in *E. coli* and can transduce any marker on the *E. coli* chromosome at a frequency of 10^{-4}-10^{-5} (Miller, 1972). P1 can package up to 115 kb of DNA into the phage head (Fineran et al., 2009), which is ~ 2.5 % of the *E. coli* genome. To inject the DNA into the host, P1 binds to the host via a terminal glucose of the lipopolysaccharide (LPS) core of the bacterial outer membrane (Sandulache et al., 1984). The injected DNA undergoes homologous recombination with the chromosome of the recipient bacterium within one hour of infection. This applies only if the injected DNA is not degraded by nucleases or does not remain extrachromosomal within the cytoplasm (unaffected by recombination and degradation). Using this technique, it is usual for ~2 % of transduced DNA to be successfully recombined into the recipient genome (Fineran et al., 2009).

To begin, the recombination experiments were performed with the Red recombinase system using PCR products and a single-stranded oligonucleotide that both included the *fusA*(G1478T) mutation.

6.2 Results

6.2.1 Attempts at mutagenesis using the Red recombinase system

The λ Red recombinase system was employed, in an attempt to introduce the *fusA*(G1478T) mutation into a naïve *E. coli* MDS42 background. *E. coli* MDS42 is a stable strain with a low propensity for recombination (Chakiath & Esposito, 2007); therefore, the first test was to determine whether the λ Red system would work in this strain. To do this, the *metC* gene in *E. coli* MDS42 and in another strain, *E. coli* BW25113, which served as a positive control was knocked out (Section 6.4.1.1). *E. coli* strain BW25113 was selected as the positive control because a previous study had successively inactivated genes using the λ Red system in this strain (Baba et al., 2006). To disrupt *metC*, a 1.4 kb kanamycin resistance cassette derived from the *E. coli* BW25115 Δ*metC* knockout strain from the Keio collection (Baba et al., 2006) that had been amplified by PCR was used. The kanamycin resistance marker was flanked by 54 bp and 47 bp sequences that were homologous to the 5′ and 3′ ends of the *metC* gene, respectively.

The *metC* gene knockout experiments were performed and the Red recombination protocol yielded 41 and ~ 500 colonies of *E. coli* MDS42 and BW25113, respectively. To confirm that the colonies were true positives, colony PCR was performed on three colonies randomly selected off each plate using primers that bound upstream of the *metC* gene and to the kanamycin resistance marker. PCR yielded a 750 bp PCR product confirming that the *metC* gene had successfully been replaced with the kanamycin marker (Figure 16).

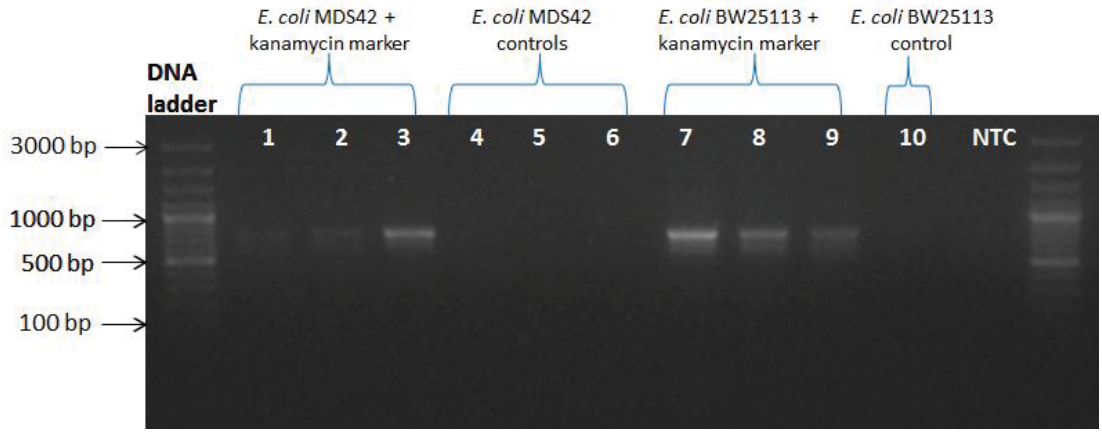


Figure 16. Successful recombination using the kanamycin resistance marker

The λ Red recombinase system was used to introduce the kanamycin resistance marker into *E. coli* MDS42 (lanes 1-3) and BW25113 cells (lanes 7-9). A successful recombination event yielded a 750 bp product after performing PCR. Control samples (lanes 4-6 and 10) represent cells that have undergone the same procedure but in the absence of the kanamycin resistance marker. A no template control (NTC) was included in the PCR reaction.

Since recombination was successful in *E. coli* MDS42 using the λ Red system, the next step was to proceed with constructing the *fusA*(G1478T) mutation in this strain. In this approach, using two primer pairs specific for the *fusA* gene, fusA_A and fusA_B (Table I.3, Appendix I.5), two regions of different sizes were amplified by PCR, 584 bp and 816 bp, respectively, from the *E. coli* MDS42 *yejG* (HTD) clone. Two regions of different sizes were amplified in order to investigate whether the length of the DNA would affect recombination efficiency. Both PCR products contained the *fusA*(G1478T) mutation.

The PCR products were introduced into *E. coli* MDS42 cells expressing the λ Red proteins. For the negative control, *E. coli* MDS42 strain expressing the Red proteins were electroporated in the absence of DNA. It was possible to select for clones that had acquired the mutation because *E. coli* MDS42 cells that harbour the *fusA*(G1478T) mutation are resistant to higher concentrations of tobramycin than cells that do not carry the mutation (MIC values of 24 $\mu g.mL^{-1}$ and 12 $\mu g.mL^{-1}$, respectively; Section 4.3.3). The next day, 2 candidate colonies were obtained for primer fusA_A and 17 colonies were obtained when primer fusA_B was used. However, the negative control plate contained 7 colonies. Therefore, the next step was to check whether any of the colonies contained the *fusA*(G1478T) mutation.

To check whether recombination had been successful, a PCR discrimination assay was used (Figure 17). Two primer pairs were designed that were specific for the mutated *fusA* allele (fusA_m) and wild-type *fusA* (fusA_w). For the fusA_m primer pair, the 3´ end of the reverse primer ended with base A, at a position that matched in sequence to the G1478T mutation (Figure 17). If the clone carried the mutation in the *fusA* gene, then the PCR would yield a 477 bp PCR product. For the fusA_w primer pair, the 3´ end of the forward primer ended with G. A 399 bp PCR product was amplified if the gene was wild-type. All of the colonies that grew on the non-control plates were tested by PCR discrimination using the fusA_m and fusA_w primers. The results showed that only the primer pair specific for the wild-type *fusA* gene yielded bands for all of the clones. For the positive control, the PCR discrimination assay was also performed on the *E. coli* MDS42 *yejG* (HTD) clone whose DNA was used as a template to generate the PCR products used in the recombination reaction. This PCR yielded a band with the expected size (~ 500 bp). This suggested that recombination had not been successful in any of the 19 colonies tested.

The experiment was repeated but this time using fragments of *fusA* that were similar or longer in DNA size to the kanamycin resistance marker (1.4 kb). Another set of primers specific to *fusA*, fusA_C and fusA_D (Table I.3, Appendix I.5), were used to amplify 1,482 bp and 2,211 bp PCR products, respectively from the *E. coli* MDS42 *yejG* (HTD) clone. When the recombination experiment was performed, two colonies were present on one (out of two) of the control plates but there was no growth on the sample plates.

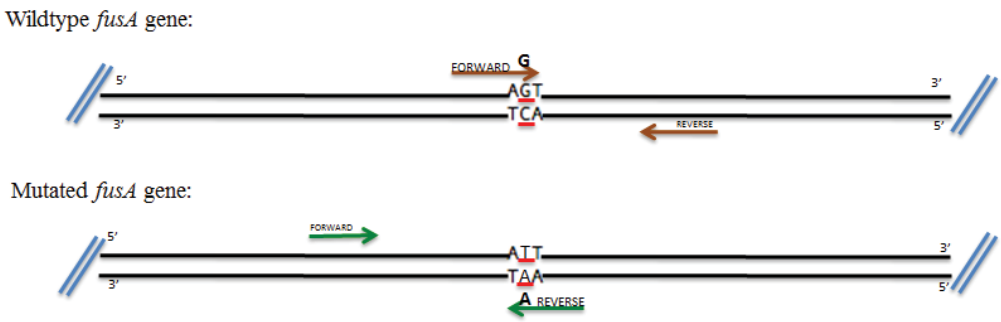


Figure 17. Designing primers for PCR discrimination to detect the *fusA*(G1478T) mutation

The diagram shows the design of the fusA_w (19 bp) and fusA_m (19bp) primer pairs that are specific for the wild-type and mutated *fusA* gene (2115 bp), respectively. The fusA_m primer pairs generated a 477 bp PCR product only when the *fusA*(G1478T) was present.

As a result of the negative outcome, attempts were made to optimise the Red recombination protocol. One such change was decreasing the recombination time from 3 h to 2.5 h, 40 min and 20 min. The reason for this step was to decrease the time available for cells that had not recombined from outgrowing the minority that had successfully recombined. This way, it was more likely to select for the desired clone from the aliquots of cells spread on plates. Another modification was to try a different tobramycin concentration (20 $\mu g.mL^{-1}$) for the selection stage. This was to inhibit the growth of cells that had acquired a chromosomal mutation that conferred a lower level of tobramycin resistance than that conferred by the *fusA*(G1478T) mutation (24 $\mu g.mL^{-1}$). Finally, clones were selected in the presence of the different aminoglycosides (sisomicin and apramycin). Despite these efforts, none of a total of 76 clones analysed had successfully incorporated the G1478T mutation.

As a last resort, the 'Quick & Easy *E. coli* Gene Deletion Kit' (Gene Bridges GmbH), which uses the Red/ET system, was employed. As mentioned earlier, the Red/ET system is similar to the λ Red system. This kit was used because according to the manufacturers, it provided an optimized system for disrupting and modifying genes on the *E. coli* chromosome. As before, a test was carried out to confirm that the Red/ET system could work in *E. coli* MDS42. Again, the *metC* gene was disrupted in the *E. coli* MDS42 and BW25113 strains with the 1.4 kb kanamycin resistance marker. PCR screening of three randomly picked colonies of each strain confirmed that all six colonies had successfully recombined.

A different strategy was used for this experiment to see whether this would yield the desired result. Ellis and colleagues (2001) showed that it was possible to create point mutations with the Red recombinase system using single-stranded DNA. They and others (Wang et al., 2009; Costantino & Court, 2003) have also shown that higher recombination efficiencies can be obtained with the oligonucleotide that corresponds in sequence to the lagging strand. The model for single-stranded recombination is that upon entering the cell, the RecT protein binds to the oligonucleotide to protect it from nuclease degradation (Karakousis et al., 1998). The binding of RecT also mediates the annealing of the oligonucleotide to the lagging strand at transiently single-stranded regions at the replication fork to generate the recombinant. Single-stranded regions are formed during processes

including DNA replication (Ellis et al., 2001). The oligonucleotide is then joined to the surrounding Okazaki fragments by DNA ligase.

For this experiment, a 71 nucleotide synthetic single-stranded oligonucleotide, including the G1478T mutation at base 36 flanked by 35 nucleotides of wild-type *fusA* was designed as the substrate for recombination. A 71-base DNA sequence was chosen because oligonucleotides of similar length have been successfully used for recombination in other studies (Costantino & Court, 2003; Ellis et al., 2001). To increase the recombination efficiency, the oligonucleotide included four phosphorothioate bonds located at the 5′ terminus to prevent *in vivo* degradation of the synthetic DNA by endogenous exonucleases (Wang et al., 2009). To determine the 'lagging strand' sequence, the direction of transcription of the *fusA* gene relative to the direction of replication in *E. coli* from the origin towards the terminus was identified (Figure 18).

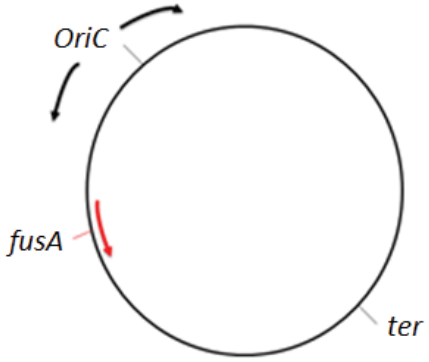


Figure 18. Designing single-stranded oligonucleotides that correspond to the lagging strand

E. coli chromosome (represented as a circle) is replicated bi-directionally from the origin of replication (*oriC*) toward the terminus (*ter*), as indicated by black arrows. The direction of transcription of the *fusA* gene (red arrow) is shown. Thus the leading and lagging strand sequences can be inferred from the orientation of *fusA* relative to *OriC*.

E. coli MDS42 cells expressing the Red proteins were transformed with the phosphorothioate-containing single-stranded oligonucleotide. A control expressing the Red proteins was electroporated in the absence of the phosphorothioate-containing DNA. Following recombination and overnight incubation on tobramycin-containing plates, there were 4, 0 and 1 colony on the sample plates and 6 colonies on the control plate. Despite the

low number of colonies that grew, all of the 5 colonies that grew on the sample plates were screened using the PCR discrimination assay. However, recombination had not been successful in any of the clones. The experiment was repeated several times with many of the steps altered but none of the clones analysed had successfully incorporated the mutant sequence. Attempts were also made to introduce the mutation into *E. coli* strains DH5α-E and BW25113 using this system. Over 200 colonies were analysed, however, none of clones investigated had acquired the *fusA*(G1478T) mutation. The PCR discrimination assay was performed successfully on clone *E. coli* MDS42 *yejG* (HTD), indicating that the assay was working (Figure 19).

Finally, the experiment was repeated using the λ Red system with the 584 PCR product [that included the *fusA*(G1478T) mutation] in naïve *E. coli* MDS42 cells that were expressing YejG. This was to test whether YejG might facilitate acquiring the *fusA* mutation. However, the clones that appeared on the plates containing tobramycin were false positives.

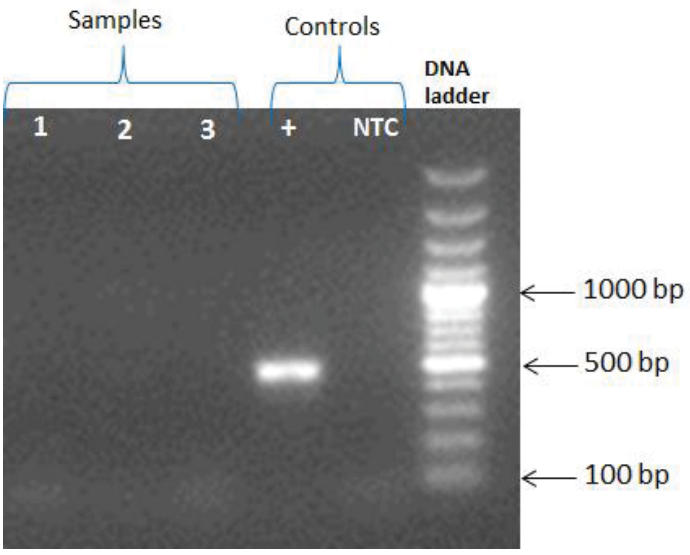


Figure 19. Gel of PCR discrimination assay

PCR discrimination assay was performed on *E. coli* MDS42 cells after attempted recombination with the single-stranded oligonucleotide, using the Red/ET system (lanes 1-3). A 477 bp product was present after PCR if the *fusA*(G1478T) mutation was present in the chromosome, indicating that recombination was successful. The *E. coli* MDS42 *yejG* (HTD) clone carrying the *fusA*(G1478T) mutation (positive control, +) and a no template control (NTC) were included as controls.

6.2.2 P1 general phage transduction

The next attempt to introduce the *fusA*(G1478T) mutation was by P1 phage transduction. The plan for this experiment was to infect *E. coli* MDS42 *yejG* (HTD) cells that carried the *fusA*(G1478T) mutation with P1 lysates prepared from *E. coli* CSH128 (lysogen for P1*cml*,*clr100*) by heat induction (at 42 °C). P1*cml*,*clr100* is a mutant of P1 that harbours the gene for chloramphenicol resistance (*cml*) and carries a mutation that makes P1 thermo-inducible (*clr*100) (Miller, 1972). Infection with the P1 lysate would allow different regions of the host chromosome (including the mutated *fusA* gene) to be packaged inside phage heads. The phage could then be incubated with wild-type *E. coli* MDS42 cells, allowing the packaged DNA to be injected into the wild-type cells and undergo homologous recombination with the recipient chromosome.

To begin, the methodology was tested by transducing *E. coli* strain BW25113 with P1 lysate that had been incubated with an *E. coli* BW25113 Δ*yejG* knockout strain, which carried a kanamycin resistance marker. Kanamycin resistant clones were then selected for. Two controls were included in the experiment: a 'no cell' control that did not contain the recipient *E. coli* BW25113 cells, and a 'no phage' control that did not contain P1 lysate. After performing the experiment, the next day, no colonies had appeared on any of the control plates. There were 14 and 104 colonies on the plates spread with 20 and 130 μL non-control sample cells, respectively. To confirm that the clones were true positives, PCR screening was performed on four randomly chosen colonies. Primers that bound upstream and downstream of *yejG* were used. If phage transduction were successful, a ~ 1500 bp PCR product (kanamycin marker (*KanR*) flanked by segments of the *yejG* gene) was expected. A product of ~ 450 bp indicated that phage transduction had not worked and that the *yejG* gene had not been disrupted. PCR yielded the expected size for the *KanR* product (Figure 20).

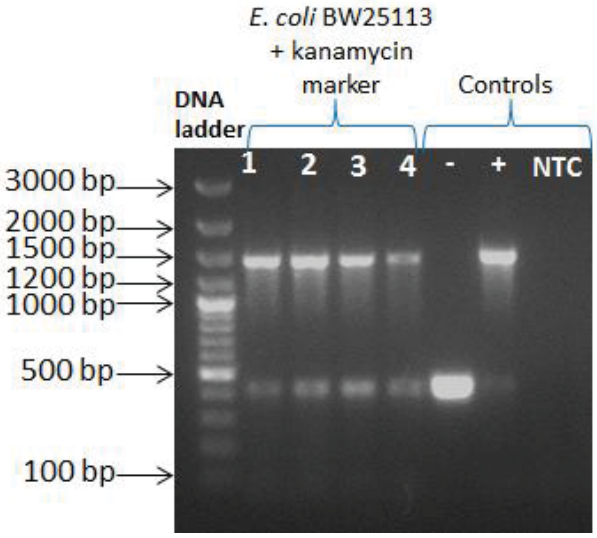


Figure 20. Successful P1 phage transduction of kanamycin marker

The kanamycin resistance marker from the *E. coli* BW25113 Δ*yejG* knockout strain was introduced into *E. coli* strain BW25113 by phage transduction (lanes 1-4). PCR yielded a ~ 1500 bp product if the kanamycin marker was present, and a ~ 450 bp product if the marker was absent. *E. coli* BW25113 before transduction (negative control, -) and the *E. coli* BW25113 Δ*yejG* knockout strain (positive control, +) were included in the PCR reaction. The no template control (NTC) is also shown.

Having shown that P1 transduction could work, an attempt was made to construct the *E. coli* MDS42 strain carrying the *fusA*(G1478T) mutation. The *E. coli* MDS42 *yejG* (HTD) clone was infected with the P1 phage and the P1 lysates were incubated with an overnight culture of *E. coli* MDS42 cells not harbouring the mutation. Two similar control samples as above were included in the experiment. After 24 h of incubation, there were a total of 133 and 150 colonies on the non-control and the 'no phage' control plates (contains the *E. coli* MDS42 cells recipient cells only), respectively, and no colonies on the 'no cell' control plates (contains the P1 phage only). This suggested that the naïve *E. coli* MDS42 cells were acquiring beneficial chromosomal mutations and that the clones on the non-control plates were probably false positives. To confirm this, 20 colonies randomly selected off the non-control plate were screened by the PCR discrimination assay. The results revealed that all of the colonies were wild-type *E. coli* MDS42.

The infection and transduction steps were varied in numerous repeats of the experiment (summarised in Section 6.4.1.3.3). However, it was not possible to prevent the growth of naïve *E. coli* MDS42 on the 'no phage' plates. A total of 204 clones were analysed and none carried the *fusA* mutation. These results suggested that expressing FusA

(T493N) was highly deleterious to the growth of cells. To test this hypothesis, fitness assays were performed using an *E. coli* MDS42 strain that carried the *fusA*(G1478T) mutation and the pBlueScript plasmid.

6.2.3 Fitness assays involving the *fusA*(G1478T) mutation and YejG

This section explains the rationale behind the construction of specific clones that were used in the fitness assays, to investigate whether the *fusA*(G1478T) mutation imparts a fitness cost and to assess the role of YejG in cell fitness. Specifically, the following questions were investigated:

i) Does harbouring the *fusA*(G1478T) mutation incur a fitness cost?
ii) Is there a benefit to over-expressing YejG in a clone with the *fusA*(G1478T) mutation?
iii) Does over-expressing YejG alleviate any fitness cost the *fusA*(G1478T) mutation may impart?
iv) Does over-expressing the mutated version of YejG [YejG(HTD)] improve or alleviate any fitness cost the *fusA*(G1478T) mutation may impart?

i) Does harbouring the *fusA*(G1478T) mutation incur a fitness cost?

To address these questions, a clone with the *fusA*(G1478T) mutation and another with the wild-type *fusA* gene were required. As it was not possible to construct a clone with the *fusA* mutation (Section 6.2), fitness assays were performed with the *E. coli* MDS42 *yejG* (HTD) clone [carries the *fusA*(G1478T) mutation and the pCA24N-*yejG*(HTD) plasmid]. However, the aim was to investigate whether the *fusA*(G1478T) mutation incurred a fitness cost in the following conditions; (i) in the absence of aminoglycoside and (ii) without the presence of plasmid-encoded YejG or the mutated version of YejG. Therefore, experiments were carried out to replace pCA24N-*yejG*(HTD) with pBlueScript, since it had also proven impossible to cure cells of the pCA24N plasmid (Section 6.4.3.3).

The pBlueScript plasmid (Table I.2, Appendix I.4) was useful for replacing pCA24N for several reasons: it did not contribute to aminoglycoside resistance; it carried a different

antibiotic marker (ampicillin) to the pCA24N plasmid; and it turned colonies blue in the presence of Xgal, allowing for blue/white screening. If pBlueScript were to be used in the fitness experiment, it would have to be carried by both the strains with wild-type *fusA*, and the *fusA*(G1478T) allele. However, this would mean that all of the cells would form blue colonies on Xgal plates, making it impossible to distinguish between the two clones. To solve this problem a pBlueScript plasmid with a non-functional *lacZ* gene (pWhite) was constructed.

An arbitrarily chosen non-functional fragment of the *tdcD* gene (535 bp) was cloned into the middle of *lacZ*α located on the pBlueScript plasmid (Section 6.4.3.2) to create pWhite (Table I.2, Appendix I.4). As a result, the *lacZ*α fragment was disrupted and became out of frame, no longer enabling the formation of blue colonies in the presence of Xgal. I now had a plasmid that did not contribute to aminoglycoside resistance, had a different resistant marker to pCA24N, did not have a functional *lacZ* and could displace pCA24N-*yejG*(HTD). Now it was trivial to mark one strain with a functional copy of *lacZ*. As described above, if both the clones were transformed with pWhite, then all of the cells would form white colonies on Xgal plates, again making it impossible to differentiate between the two clones. To resolve this issue, one of the clones (the *E. coli* MDS42 strain with wild-type *fusA*) was marked with a functional chromosomal copy of *lacZ* (Section 6.4.3.1) before being transformed with the pWhite plasmid. In the end, to investigate whether the *fusA*(G1478T) mutation incurred a fitness cost, the following clones were used: *E. coli* MDS42 with *fusA*(G1478T) mutation + pWhite versus *E. coli* MDS42 with wild-type *fusA* + *lacZ* + pWhite.

The results revealed that the *fusA*(G1478T) mutation imparts a statistically significant reduction in fitness ($W = 0.93 \pm 0.01$; $P < 0.01$). This implies that harbouring this mutated *fusA* allele in the absence of aminoglycoside is detrimental to the cell's growth. Although it was not surprising that this mutation incurred a fitness cost, given the difficulty in creating a *fusA* mutated clone, a larger fitness decrease was expected.

ii) Is there a benefit to over-expressing YejG in a clone with the *fusA*(G1478T) mutation?

To determine whether there was a benefit to over-expressing YejG in a clone with the *fusA*(G1478T) mutation, the *E. coli* MDS42 *yejG* (HTD) clone [with the *fusA*(G1478T) mutation + pCA24N-*yejG*(HTD)] was first transformed with pBlueScript, to displace the existing plasmid. After confirming that the cells had lost pCA24N-*yejG*(HTD) (the cells were ampicillin resistant and chloramphenicol sensitive), they were transformed with either pCA24N-*yejG* or pCA24N-GFP. To check whether the clones had lost the pBlueScript plasmid, the cells were spread on LB agar containing ampicillin to confirm that they were susceptible. The strain expressing GFP was marked with a functional copy of *lacZ* so that the two clones could be distinguished.

The fitness assays showed that there was a significant fitness benefit to over-expressing YejG in a strain with the *fusA*(G1478T) mutation ($W = 1.18 \pm 0.03$; $P < 0.01$). To follow on from this, it was investigated whether this benefit might include alleviating the fitness cost imparted by the *fusA*(G1478T) mutation.

iii) Does over-expressing YejG alleviate any fitness cost the *fusA*(G1478T) mutation may impart?

The developing hypothesis was that YejG could alleviate the fitness cost associated with the *fusA*(G1478T) mutation. This was tested by performing a competitive fitness assay between cells with the *fusA*(G1478T) mutation + pCA24N-*yejG* (the same clone used in (ii) above) and a strain with the wild-type *fusA* gene + pCA24N-GFP. The strain with the wild-type *fusA* gene was marked with a functional copy of *lacZ*.

Surprisingly, the results revealed that the strain harbouring the *fusA*(G1478T) mutation and over-expressing YejG was significantly fitter than the strain with the wild-type *fusA* gene and no YejG over-expression ($W = 1.46 \pm 0.05$; $P < 0.01$). It would appear that YejG not only alleviates the fitness cost incurred by the *fusA*(G1478T) mutation, but it also improves the fitness. However, since these results relate to wild-type YejG, but the *E. coli* MDS42 *yejG* (HTD) clone [with the *fusA*(G1478T) mutation] was discovered carrying a

mutated version of YejG (Section 4.3.4), I investigated whether harbouring this YejG variant would change the fitness.

iv) Does over-expressing the mutated version of YejG [YejG(HTD)] improve or alleviate any fitness cost that the *fusA*(G1478T) mutation may impart?

Next, it was investigated whether a clone with the *fusA*(G1478T) mutation combined with the mutated version of YejG [which was the plasmid carried by the *E. coli* MDS42 *yejG* (HTD) clone] was fitter than when the mutation is combined with wild-type YejG. A similar experiment to that performed in (iii) was carried out, but this time the *E. coli* MDS42 with the *fusA*(G1478T) mutation harboured pCA24N-*yejG*(HTD), instead of wild-type *yejG*.

Over-expressing the mutated YejG variant in a strain with the *fusA*(G1478T) mutation resulted in no reduction in fitness ($W = 1.00 \pm 0.01$; $P > 0.01$). The results suggest that although this YejG variant may alleviate the fitness cost incurred by the chromosomal mutation, it is not able to enhance the cell's fitness to the extent that wild-type YejG can.

Finally, a fitness assay was performed to determine whether YejG would also improve the fitness of a strain with the wild-type *fusA*. To examine this, a YejG-expressing strain and a clone marked with a functional copy of *lacZ* over-expressing GFP were employed. The results revealed that in the absence of aminoglycoside, the YejG-expressing strain was significantly fitter ($W = 1.16 \pm 0.03$; $P < 0.01$) than the control strain.

The fitness experiments revealed that YejG was able to improve the fitness of cells whether or not they carried the *fusA* mutation. In the next set of experiments, growth assays were performed in order to investigate whether YejG expression was affecting a particular stage of cell growth.

6.2.4 Growth assays

In the fitness experiments, it was shown that there was a fitness cost associated with harbouring the *fusA*(G1478T) mutation and that YejG was able to alleviate this cost. To further understand this, growth assays were performed to investigate how YejG was improving cell fitness.

Throughout my experiments, I observed that cells expressing YejG grew faster than cells that were not expressing the protein. This improvement in cell growth was detected within the first ~ 3 h of incubation and was assessed by measuring the optical density of the growing culture. However, when the cultures were incubated overnight, the cultures expressing YejG were less turbid than those not expressing the protein. This led me to hypothesise that YejG might be promoting an early transition from the lag phase into the log phase and that this could be observed by performing a growth assay. The same clones used in the fitness experiments were employed for these experiments.

For the first growth assays, *E. coli* MDS42 cells over-expressing either YejG or GFP (control) were compared (Figure 21). It was apparent from the results that when YejG was over-expressed, the bacteria entered the log phase earlier than strains that did not express the protein; the YejG-expressing strain reached the log phase ~3 h before the control strain. YejG also appeared to drive the cells into an early stationary phase. The YejG-expressing cells reached a lower final density (OD_{600} ~ 0.7) than the control strain (OD_{600} ~ 1.2). This confirmed and quantified my previous observation that YejG-expressing cultures were less turbid after overnight incubation than GFP-expressing cultures.

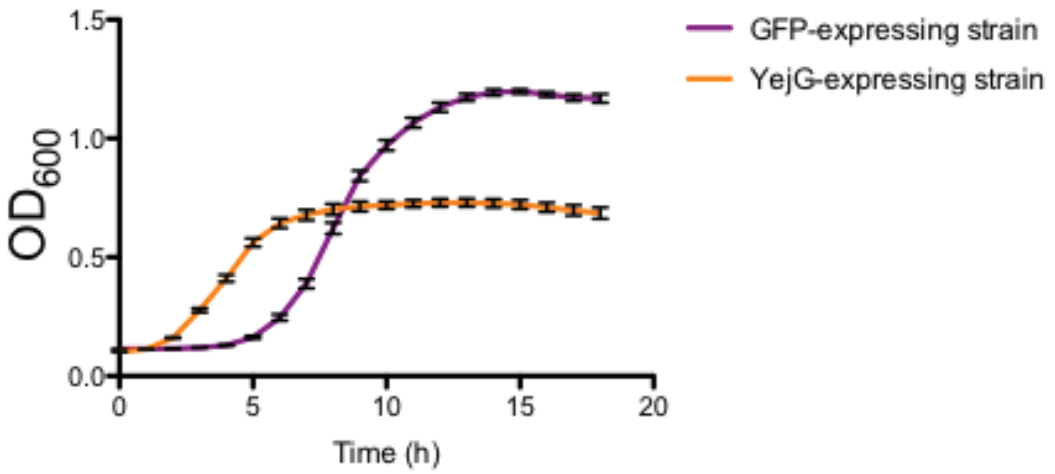


Figure 21. YejG-expressing cells show early entry into the log phase

The YejG- and GFP-expressing strains were grown separately in LB supplemented with chloramphenicol (to maintain the pCA24N plasmid) and IPTG and monitored for ~ 20 h. The graph is the average of six replicates; error bars are shown. Results show that over-expression of YejG drives the cell into an early log phase as well as an early stationary phase.

Next, the growth of a strain carrying the *fusA*(G1478T) mutation and one harbouring the wild-type *fusA*, both carrying the pWhite plasmid were compared (Figure 22). The results revealed that carrying this mutation might slightly delay the clone entering into the log phase by (< 1 h). Both clones showed similar log and stationary growth phases. A slight delay in entering the log phase might be a possible explanation for the fitness decrease observed for the *E. coli* MDS42 *yejG* (HTD) clone during the fitness experiment [$W = 0.93 \pm 0.01$) (Section 6.2.3 (i)]. To assess how YejG alleviates the fitness cost associated with the *fusA*(G1478T) mutation, growth assays of the strain carrying the *fusA* mutation and over-expressing either YejG or GFP were performed.

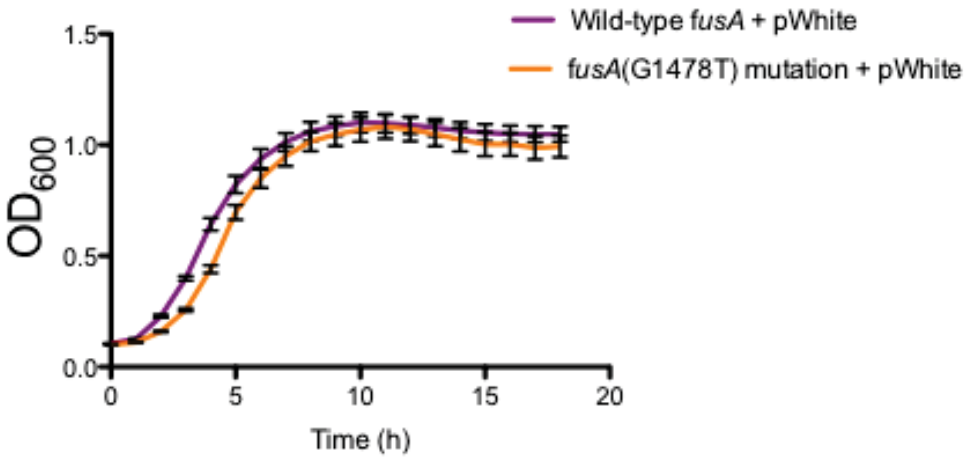


Figure 22. Cells carrying the *fusA*(G1478T) mutation show slightly delayed entry into the log phase

Cells carrying the *fusA*(G1478T) mutation and cells harbouring wild-type *fusA* (both carrying the pWhite plasmid) were grown separately in LB supplemented with ampicillin (to maintain the pWhite plasmid) and monitored for ~ 20 h. The graph is the average of six replicates; error bars are shown. The graph reveals that the *fusA*(G1478T) mutation may lead to a slight delay in growth.

When YejG was over-expressed in a strain carrying the *fusA*(G1478T) mutation, the cells reached the log phase ~ 1 h earlier than cells harbouring same mutation but over-expressing GFP (Figure 23). Similar to Figure 21, the YejG-expressing cells also entered the stationary phase earlier.

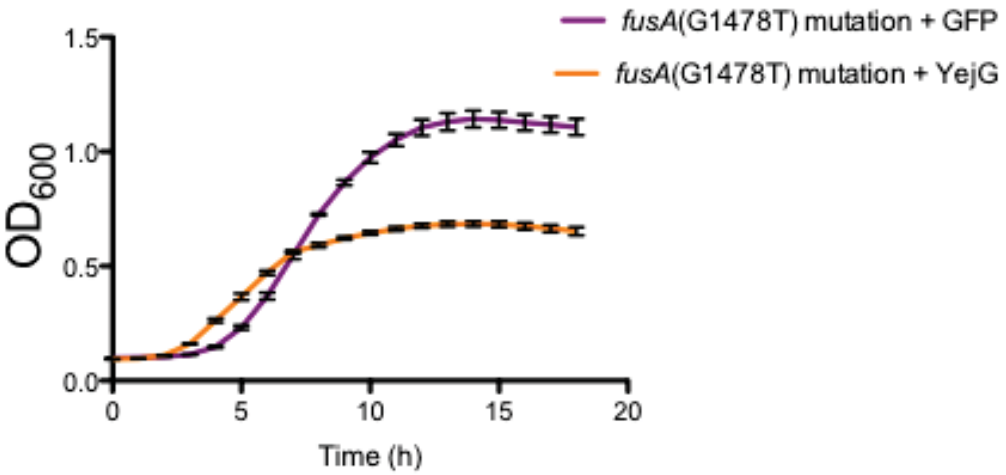


Figure 23. Cells carrying the *fusA*(G1478T) mutation and over-expressing YejG also show early entry into the log phase

Cells carrying the *fusA*(G1478T) mutation and over-expressing either YejG or GFP were grown separately in LB supplemented with chloramphenicol (to maintain the pCA24N plasmid) and IPTG and monitored for ~ 20 h. The graph is the average of six replicates; error bars are shown. YejG over-expression also promotes an early log phase in cells that carry the *fusA*(G1478T) mutation.

To determine how YejG might be alleviating the fitness cost conferred by the *fusA*(G1478T) mutation [as revealed in the fitness experiments Section 6.2.3, (iii)], a growth assay was carried out on the strain harbouring the *fusA*(G1478T) mutation and over-expressing YejG. This growth curve was compared to that of the strain carrying the wild-type *fusA* and over-expressing GFP from Figure 21 (Figure 24). The same growth pattern seen in Figure 21 and 23 was observed. YejG-over-expression allows the cell to reach the log phase ~ 3 h earlier than the GFP-expressing strain. An early entry into the log phase in a pool of limited resources would ensure a massive fitness advantage for the cell. In the competitive fitness assays, the resources are limited. A strain expressing YejG will have a 3 h early access to this limited resource before its competitor enters the log phase. The nutrients that remain after the 3 h will then be shared by both the YejG-expressing strain and the competing strain, until the YejG-expressing strain enters the stationary phase. This may well explain the high fitness increase observed for the strain carrying the *fusA*(G1478T) mutation and over-expressing YejG during the fitness experiment (W = 1.46 ± 0.05) [Section 6.3, (iii)].

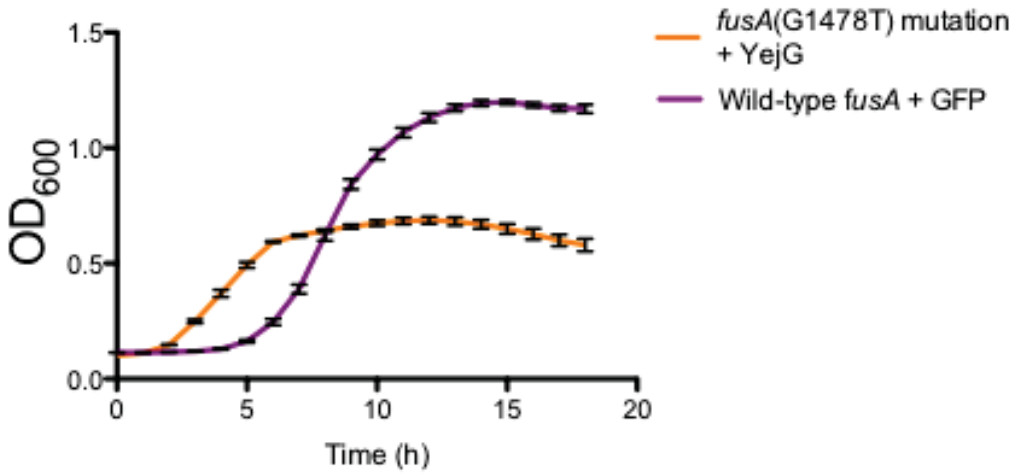


Figure 24. YejG over-expression allows cells carrying the *fusA*(G1478T) mutation an early growth compared to cells with the wild-type *fusA* over-expressing GFP

Cells carrying the *fusA*(G1478T) mutation and over-expressing YejG were grown in LB supplemented with chloramphenicol (to maintain the pCA24N plasmid) and IPTG and monitored for ~ 20 h. The graph is the average of six replicates; error bars are shown. This growth curve is compared to that of the cells harbouring wild-type *fusA* and over-expressing GFP (in Figure 21). The results show that YejG over-expression allows cells that carry the *fusA*(G1478T) mutation to reach an earlier log phase compared to cells that harbour a wild-type *fusA* and over-expressing GFP.

Finally, the effect of over-expressing the YejG variant [YejG(HTD)] on the growth of cells carrying the *fusA* mutation was assessed (Figure 25). The experiment was performed with the strain carrying the *fusA*(G1478T) mutation and over-expressing YejG(HTD) and the strain carrying the wild-type *fusA* and over-expressing GFP. From the results, it appears that the YejG(HTD)-expressing strain displays a similar log phase to the GFP-expressing strain. Furthermore, the YejG(HTD)-expressing strain does not reveal a delayed entry into the log phase in the presence of the *fusA* mutation (see Figure 22). Over-expressing the YejG variant also lead to a lower final cell density similar to the wild-type protein (OD_{600} ~ 0.7).

Overall, the results suggested that YejG enhanced fitness by driving the cells into an early log phase, effectively giving them a head start in the competitive fitness assays. However, the protein also promotes an early entry into stationary phase.

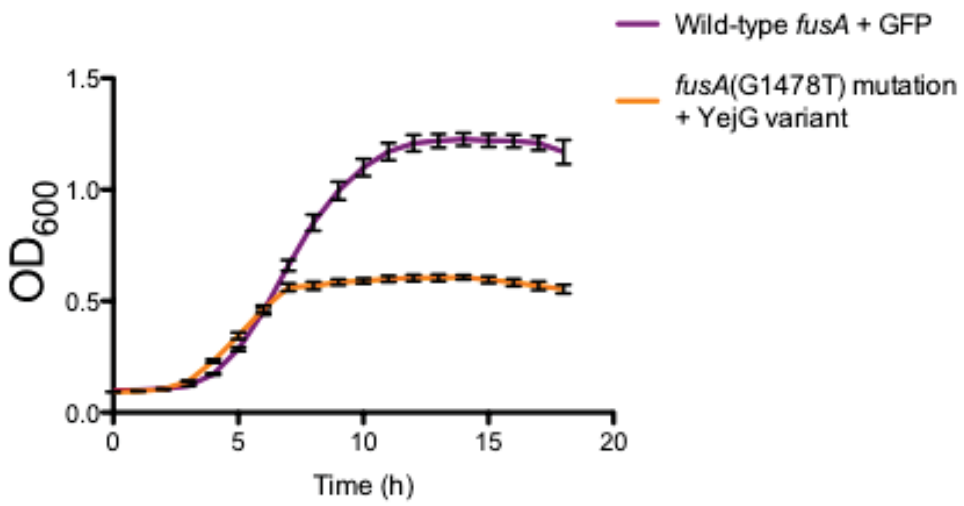


Figure 25. Effect of YejG variant in the presence of the *fusA*(G1478T)

Cells carrying the *fusA*(G1478T) mutation and over-expressing the YejG variant and cells harbouring wild-type *fusA* and over-expressing GFP were grown separately in LB supplemented with chloramphenicol (to maintain the pCA24N plasmid) and IPTG and monitored for ~ 20 h. The graph is the average of six replicates; error bars are shown. When the YejG variant is over-expressed, the cell does not display a delayed entry into the log phase in the presence of the *fusA*(G1478T) mutation.

6.3 Discussion

Previously, it was shown that YejG-expressing cells show moderate increases in tobramycin resistance (MIC = 18 μg.mL^{-1}) compared to a suitable negative control (MIC = 12 μg.mL^{-1}) (Table 14). The *fusA* mutation [*fusA*(G1478T)], discovered in the *E. coli* MDS42 *yejG* (HTD) clone, conferred even higher levels of resistance to tobramycin (24 μg.mL^{-1}; Table 16). Together, the *fusA*(G1478T) mutation and YejG were able to endow the bacterium with maximal tobramycin resistance (42 μg.mL^{-1}; Table 16). In this Chapter, I investigated whether YejG was mediating pleiotropic effects on fitness and assessed the impact of YejG over-expression on a clone carrying the *fusA*(G1478T) mutation. To achieve this, fitness experiments and growth assays were performed.

In order to conduct competitive fitness assays, the first aim was to construct an *E. coli* MDS42 clone harbouring the *fusA* mutation in the chromosome but lacking plasmids. Two methods were used to try and create this clone, the Red recombinase system (both lambda Red and Red/ET versions) and P1 phage transduction. However, despite many attempts, the clone could not be constructed (Section 6.2.1 and 6.2.2). To confirm that both systems worked in *E. coli* strain MDS42, the methodology was tested and shown to be successful (Figure 16 and 20). The results suggested that a clone with the *fusA*(G1478T) mutation could not be created because the mutation was deleterious to the growth of the cells. This could be rationalized since *fusA* is an essential gene that codes for elongation factor G (EF-G), a protein involved in protein synthesis. However, this mutation is unlikely to be completely harmful to the cell since the *E. coli* MDS42 *yejG* (HTD) clone carried it.

Bielecki and colleagues (2012) were able to introduce T1370C and T1988A mutations into *Pseudomonas aeruginosa* by homologous recombination using a 2042 bp fragment of *fusA* cloned into a suicide plasmid (pEX18AP). In this thesis, fragments of *fusA* of different sizes (584 bp, 816 bp, 1,482 bp and 2,211 bp; full length *fusA* is 2115 bp) were used for the recombination experiment. Since both the Red recombinase system (lambda Red and Red/ET) and transduction have been shown to work (Section 6.2.1 and 6.2.2) and a similar DNA fragment size (2,211 bp) to that used in the Bielecki et al., (2012) study was employed (2,042 bp), it is unlikely that any of these factors contributed to the failure of the experiment. Perhaps the position of the mutation in *fusA* may have a role to play, although this is hard to rationalise. This can be investigated experimentally by testing

whether the T1370C and T1988A mutations can be introduced into *E. coli* MDS42 using the Red recombinase system or by transduction.

To determine whether the *E. coli* MDS42 *yejG* (HTD) clone survived gaining the *fusA*(G1478T) mutation because the cell was also over-expressing YejG, a second set of experiments was performed. This time, the Red recombinase system was used to try and create the *fusA*(G1478T) mutation in *E. coli* MDS42 cells that were over-expressing YejG (Section 6.2.1). However, this did not work either. The *E. coli* MDS42 *yejG* (HTD) clone carried a plasmid with a mutated *yejG* insert [*yejG* (HTD); Table 17]. It may be possible that attempts to construct the clone were unsuccessful because over-expression of this mutated version of YejG alone allows the cells to survive the *fusA*(G1478T) mutation. However, this seems unlikely, in light of the results that were obtained subsequently. In the end, fitness experiments were therefore performed with *E. coli* MDS42 clones harbouring a variety of different plasmids.

Competitive fitness assays revealed that there is a fitness defect associated with the *fusA*(G1478T) mutation ($W = 0.93 \pm 0.01$; $P < 0.01$) but that over-expressing YejG is advantageous to this strain (Figure 26A and B; Table 20). Not only is YejG able to alleviate the cost imparted by the *fusA* mutation (Figure 26C) but more generally, the protein also increases cell growth (i.e., the number of divisions in 24 h) (Figure 26D; Table 20). In Section 4.4, it was proposed that the *fusA* mutation causes perturbation of the global transcription regulator ppGpp. As a first step to gain more insight into YejG's role, experiments exploring how the protein is improving growth in a bacterium with reduced EF-G function would be informative. The effect of YejG over-expression in cells carrying chromosomal mutations that are known to have an effect on ppGpp levels (both high and low) inside the cell could be investigated. This would inform whether there may be a connection between YejG and ppGpp levels, that is, genes that are up-regulated or down-regulated in reponse to ppGpp levels.

While it appears that for the fitness results, the expression of the mutated YejG protein helped to alleviate the fitness cost of the *fusA*(G1478T) mutation, the effect was less than when wild-type YejG was expressed (compare 26C and E; Table 20). This result confirms that the mutated *yejG* clone, identified from an error-prone PCR library (Section 4.3.2) was a false positive. Instead, the tobramycin-resistant colony was selected because

it contained the rare *fusA*(G1478T) chromosomal mutation, combined with a *yejG* allele that was less ideal than wild-type *yejG* (but nevertheless, not deleterious).

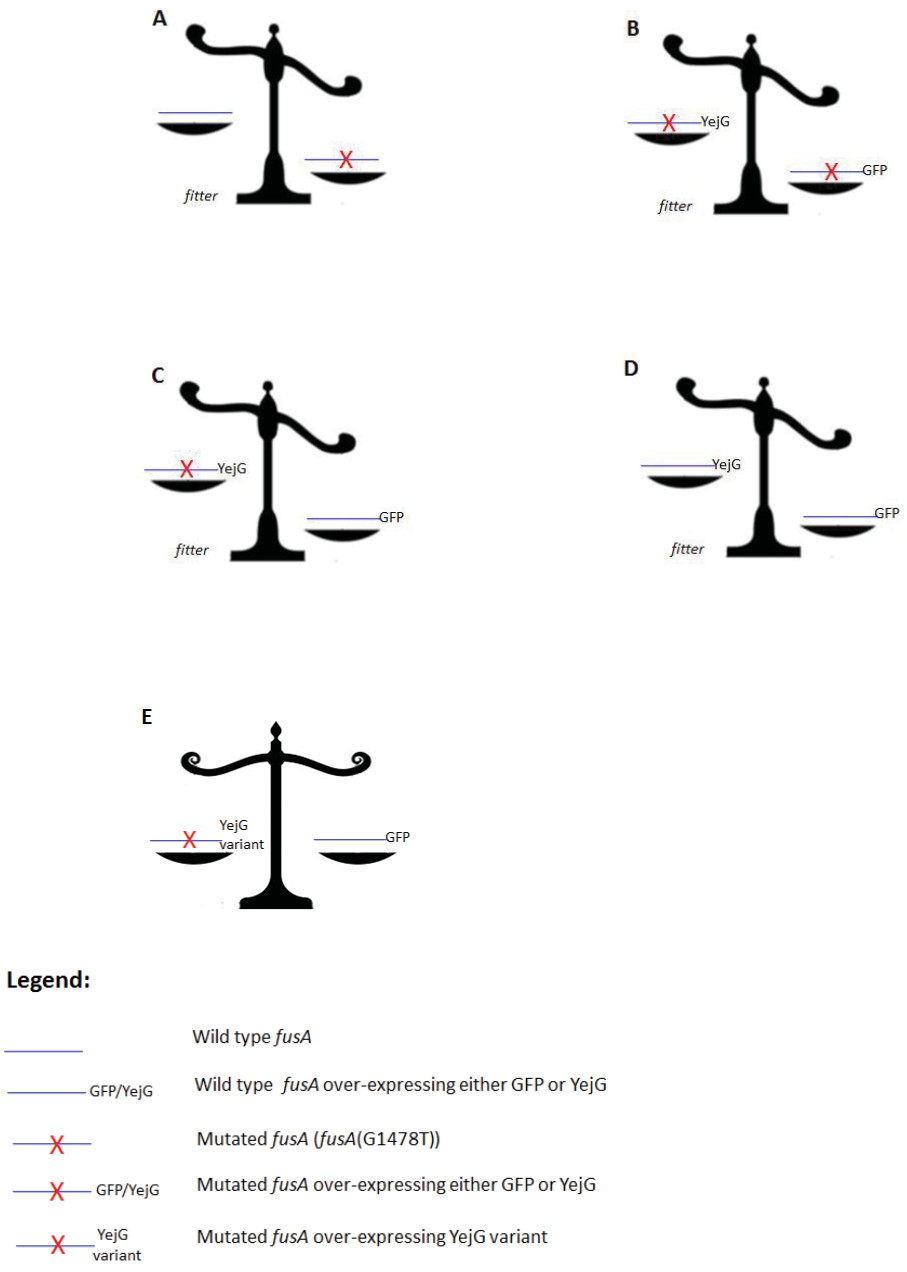


Figure 26. Diagram showing qualitative outcomes of fitness experiments involving the *fusA*(G1478T) mutation and YejG over-expression

The diagram shows outcomes of the fitness assays performed in Section 6.2.3. The results show that cells over-expressing YejG are fitter than cells over-expressing GFP, even in the presence of the *fusA*(G1478T) mutation.

Table 20. Summary of fitness experiments involving *fusA*(G1478T) mutation and YejG over- expression

Summary of the fitness experiments performed in Section 6.2.3. Fitness experiments were performed in the absence of toxin but in the presence of IPTG. Fitness values are from eight replicates (mean ± SE). The fittest clone out of the two competitors is shown with a tick (√). The order in which the fitness assays are shown in the table below corresponds to the order of the diagrams A through to E in Figure 26.

Figure 26 diagram	Strain	Fittest clone (√)	Fitness (W ± SE)
A	*E. coli* MDS42 with wild-type *fusA* *E. coli* MDS42 with *fusA*(G1478T) mutation	√	0.93 ± 0.01
B	*E. coli* MDS42 with *fusA*(G1478T) mutation + YejG *E. coli* MDS42 with *fusA*(G1478T) mutation + GFP	√	1.18 ± 0.03
C	*E. coli* MDS42 with *fusA*(G1478T) mutation + YejG *E. coli* MDS42 with wild-type *fusA* + GFP	√	1.46 ± 0.05
D	*E. coli* MDS42 with wild-type *fusA* + YejG *E. coli* MDS42 with wild-type *fusA* + GFP	√	1.16 ± 0.03
E	*E. coli* MDS42 with *fusA*(G1478T) mutation + YejG variant *E. coli* MDS42 with wild-type *fusA* + GFP		1.00 ± 0.01

It was noteworthy that a clone harbouring the *fusA* mutation and over-expressing the YejG(HTD) variant was found to have the same fitness as a clone carrying the wild-type *fusA* and over-expressing GFP [Section 6.2.3 (iv)]. However, the two strains have very different resistance profiles. The MIC values for tobramycin, kanamycin and apramycin of a strain harbouring the *fusA* mutation and over-expressing YejG(HTD) were 42, 80 and 120 μg.mL^{-1}, respectively. The MIC values for the same antibiotics of a clone carrying the wild-type *fusA* and over-expressing GFP were 12, 18 and 28 μg.mL^{-1}, respectively. A strain carrying the *fusA* mutation and over-expressing the wild-type YejG was able to improve the fitness of cells compared to when this strain was over-expressing YejG(HTD) (W = 1.46 versus W = 1.00). Curiously, both strains showed the same resistance to tobramycin (MIC = 42 μg.mL^{-1}; Table 16 and Section 4.3.3). This implies that a bacterium over-expressing YejG(HTD) only loses the benefit of an increased fitness and not resistance. Extrapolating, YejG over-expression combined with the *fusA*(G1478T) mutation is likely to yield a similar resistance profile for the other aminogycosides (kanamycin, apramycin and sisomicin; Table 18). Overall, this means that in the presence of an aminoglycoside,

having the *fusA*(G1478T) mutation and over-expressing YejG is highly advantageous. Not only would the bacterium be highly resistant to the antibiotic but it would also grow faster than the neighbouring cells.

Surprisingly, growth curves revealed that YejG over-expression promoted an early entry into stationary phase, at a lower cell density (Figure 21, 23 and 24). Thus, YejG over-expression affords a massive advantage; a shorter lag phase in a population of competitors allows the faster-growing bacterium to be the first to have access to the available resources in the environment.

Together, these results suggest a possible pathway involving YejG and the *fusA*(G1478T) mutation for evolving high-level aminoglycoside resistance. In the presence of an aminoglycoside, tobramycin, the bacterium acquires a mutation that leads to the over-expression of YejG. Mutations of this type are comparatively common and are estimated to occur at a rate of 10^{-2} to 10^{-5} per cell generation (Sandegren & Andersson, 2009). Increased expression can be acquired in a number of ways including, promoter mutations, increasing mRNA stability and ribosome binding site mutations (Bergthorsson et al., 2007; Ohki & Tateno, 2004; Matten et al., 1998). This provides the cell not only with a moderate level of resistance to the drug (from 12 to 18 $\mu g.mL^{-1}$; Table 14), it also affords the cell with a fitness advantage. The bacterium then acquires a second mutation, *fusA*(G1478T), that confers a higher level of resistance although this mutation imparts a fitness cost. This mutation will be rarer, because it's a specific amino acid substitution, however, more work is required to elucidate whether other EF-G mutations might have a similar effect. The combination of the *fusA* mutation and YejG not only alleviates the fitness cost but improves it. In addition, the bacterium is able to attain maximal aminoglycoside resistance (from 12 to 42 $\mu g.mL^{-1}$; Table 14 and 18). Finally, additional mutations that are able to restore the cell's fitness may relieve the selection pressure for over-expressing YejG. Again, this remains to be tested experimentally.

In this pathway, although the *fusA* mutation endows the cells with a higher level of resistance (24 versus 18 $\mu g.mL^{-1}$; Table 14 and 16) than YejG, YejG-over-expression is likely to come first. This is because a cell that acquires the *fusA*(G1478T) mutation as a first step will have impaired growth and will be outcompeted by neighbouring cells at times when tobramycin levels fall (e.g., between doses in a clinical setting). When YejG over-expression precedes the *fusA*(G1478T) mutation however, fitness is no longer an issue and

resistance is also acquired by the cell. Thus, YejG serves as an initial solution in the presence of the toxin and paves the way for mutations, such as *fusA*(G1478T), that confer higher levels of resistance to occur. After the cells have acquired the *fusA*(G1478T) mutation, selection for YejG's role is a little relaxed but is still required to enhance the cell's growth. This is evidenced by the directed evolution experiments, in which a mutated version of YejG (that is still able to alleviate the cost of the *fusA* mutation) was discovered in a clone carrying the *fusA*(G1478T) mutation (Table 17).

Overall, the results show the importance and usefulness of genes that have yet to be studied in the war against resistance. This study has shown that, YejG can endow cells with resistance against more than one aminoglycoside; and that the protein can also improve cell fitness even when the bacterium's growth is reduced through second-site mutations. How this is achieved is not quite clear but preliminary results from this study indicate that YejG does not associate with the ribosome, nor does it regulate the membrane potential. We are hopeful that solving the structure of this protein will give clues about its function. Structure determination is currently under way in the lab of Prof. Joel Mackay (University of Sydney).

6.4 Materials and Methods

All reagents were from Sigma, unless otherwise specified. The materials and methods described here are specific to this chapter. Common materials and molecular biology techniques are described in Appendix I.

The Quick & Easy *E. coli* Gene Deletion kit (Gene Bridges) was used to attempt mutagenesis.

6.4.1 Attempts at mutagenesis

All of the *fusA* PCR products used in the recombination experiments included the G1478T mutation and were amplified from the *E. coli* MDS42 *yejG* (HTD) clones. The following primers were used to amplify the different fragments of *fusA*: fusA_A.for and rev generated a 584 bp product, fusA_B.for and rev generated a 816 bp product, fusA_C.for and rev generated a 1,482 bp product and fusA_D.for and rev generated a 2,211 bp product. All the primers can be found in Table I.3, Appendix I.5. PCR was performed as described in Appendix I.8. The PCR products were purified using a commercial spin column and eluted in 30 µL EB. The single-stranded oligonucleotide (71 nucleotides) with phosphorothioate bonds located at the 5′ terminus also included the G1478T mutation at base 36 flanked by 35 bases on both sides (Table I.3, Appendix I.5).

The 1.4 kb kanamycin resistance marker was derived from the *E. coli* Δ*metC* strain in the Keio collection (Baba et al., 2006). The resistance marker is flanked by 54 nucleotides at the 5′ end that is homologous to the 5′ end of the genomic *metC* gene and 47 nucleotides at 3′ end that is homologous to the 3′ end of the genomic *metC* gene. It was obtained from our lab stock (Dr. Wayne Patrick). The 1.4 kb kanamycin resistance cassette was used in these experiments to test the methodology.

6.4.1.1 λ Red recombinase system

Forty ng of the helper plasmid, pKD46 (Appendix I.4), was used to transform naïve *E. coli* BW25113 or MDS42. The electroporated cells were recovered in 500 µl SOC and incubated at 28 °C for 90 min. Twenty-five µL of the recovered cells were spread on LB-ampicillin (100 µg.mL^{-1} final concentration) plates to select for clones transformed with

pKD46. The plates were incubated overnight at 28 °C. A colony was selected off the LB-ampicillin plate and inoculated in 1 mL LB-ampicillin (100 μg.mL^{-1} final concentration) and grown overnight at 28 °C. A freezer stock of the culture was made and stored at −80 °C.

The freezer stock (*E. coli* BW25113 or MDS42 cells harbouring the pKD46) were used to inoculate 3 mL LB-ampicillin (100 μg.mL^{-1} final concentration) and incubated overnight at 30 °C. The next day, 150 μL of saturated culture was used to inoculate 3 mL LB-ampicillin (100 μg.mL^{-1} final concentration) and the culture was incubated at 30 °C for 4 h. To induce the expression of the recombinases from pKD46, L-Arabinose (final concentration, 5 mM) was added to the culture and incubated for a further 1.5 h at 37 °C. The culture was evenly divided into two aliquots (1.5 mL each) which were pelleted at 17,000 ×g for 1 min. The pellets were washed a total of three times with 1 mL of pre-chilled 10 % (v/v) glycerol and pelleted as previously mentioned after each wash.

The cell pellets were pooled together and resuspended in 50 μL of the pre-chilled 10 % (v/v) glycerol. The cells were electroporated at 2.5 V (0.2 cm cuvette gap) with 100 ng of the purified *KanR* cassette (1.4 kb; Section 6.4.1) for *E. coli* BW25113 and 100 ng of *fusA* PCR products (584 bp, 816 bp, 1,482 bp or 2,211 bp; Section 6.4.1) for *E. coli* MDS42. The amount of the *fusA* PCR product used for electroporation was increased to 600 ng in a separate experiment. A negative control, *E. coli* BW25113 or MDS42 cells expressing the Red proteins electroporated in the absence of DNA, was always included. SOC (500 μL) supplemented with L-Arabinose (1mM) and the culture was incubated at 37 °C for 3.5 h. Aliquots (50 μL, 150 μL and 250 μL) of the recovered cells were spread on LB agar plates supplemented with kanamycin (30 μg.mL^{-1} final concentration; for *E. coli* BW25113) or tobramycin (18 μg.mL^{-1} final concentration; for *E. coli* MDS42). The plates were incubated overnight at 37 °C.

Colonies that appeared on the selection plates were picked for further screening. To determine whether recombination had been successful, the colonies were tested by PCR screening using KeioKanROut.for and metC-dwnstr.rev primers (Appendix I.3) to amplify the kanamycin resistance marker (750 bp) (*E. coli* BW25113). PCR discrimination (Section 6.4.2) was performed on *E. coli* MDS42 using fusA_m primers (Appendix I.3).

6.4.1.2 Quick and Easy *E. coli* Gene deletion kit (Gene Bridges GmbH)

The Red recombinase system was used in an attempt to introduce the *fusA*(G1478T) mutation and to knock out the *met C* gene (with a kanamycin resistance marker) in *E. coli* strains MDS42 and BW25113 (target strains), respectively. This protocol is as recommended in the Quick & Easy *E. coli* Gene Deletion kit (Gene Bridges).

A single colony of the target strain (naïve *E. coli* BW25113 or *E. coli* MDS42) was used to inoculate LB medium (1 mL) and incubated overnight at 37 °C. Two separate microfuge tubes containing 1.4 mL LB medium were inoculated with 30 µL of the overnight culture and incubated for 2-3 h at 37 °C. The cells were centrifuged in a cooled bench top centrifuge (11,600 xg, 2°C for 30 sec) and the supernatant was discarded. The cells were placed on ice and the pellet was resuspended with 1 mL of chilled 10 % glycerol. The centrifugation and resuspension steps were repeated. The cells were centrifuged and the supernatant was discarded once more. The pellets were resuspended in the residual glycerol (~ 20 µL) and kept on ice. Twenty ng of the Red/ET Recombination expression plasmid (pRedET) was added to the cell pellet, mixed and transferred to a pre-chilled electroporation cuvette (0.1 cm cuvette gap). The plasmid was introduced into the cells by electroporation (1,350 V, 5 ms pulse) and then the cells were recovered with 1 mL LB medium without antibiotics. The cells were transferred back to the microfuge tube and incubated at 30 °C for 70 min. One hundred µL of the culture was plated on LB agar containing tetracycline (3 $\mu g.mL^{-1}$ final concentration). The plate was incubated at 30 °C overnight.

The next day, one colony was picked off the plate and used to inoculate 1 mL LB medium supplemented with tetracycline (3 $\mu g.mL^{-1}$ final concentration) in a 1.5 mL tube. A hole was punctured in the lid of the tubes for air before incubating at 30 °C overnight (as instructed by the manufacturer). Thirty µL of the overnight culture transformed with the pRedET plasmid was added to three separate microfuge tubes (with punctured lids) containing 1.4 mL fresh LB plus tetracycline (3 $\mu g.mL^{-1}$ final concentration). The tubes were incubated at 30 °C for ~ 2 h (OD_{600} ~ 0.3) and 50 µL of 10 % L-arabinose was added to two of the cultures (final concentration 0.3 %-0.4 %). This step was to induce the expression of the Red/ET Recombination proteins. The third culture was not induced and served as the 'No L-arabinose control'. The tubes were incubated at 37 °C for 45 min to 1 h.

The cultures were centrifuged in a cooled microfuge bench top centrifuge (11,600 xg, 2 °C for 30 sec). The supernatant was discarded and the pellets were placed on ice. The cells were resuspended in 1 mL chilled 10 % glycerol. The centrifugation step was repeated and the cells were resuspended and then centrifuged once more. The supernatant was discarded and the pellet was resuspended in residual glycerol (~ 20 µL) and kept on ice. Approximately 800 ng of the PCR product (or 71-base single-stranded oligonucleotide; Appendix I.5) containing the *fusA*(G1478T) mutation was added to one of the pellets of induced cells and the uninduced cells. The other induced cells served as a 'No PCR product control'. The cells were pipetted into pre-chilled electroporation cuvettes. The samples were electroporated ([a]1,350 V, 5 ms pulse, 0.1 cm cuvette gap) and recovered with 1mL LB medium before transferring back into the microfuge tubes. The cultures were incubated at 37 °C for ~ [b]3 h (recombination should occur). The cultures were centrifuged (11,600 xg, room temperature for 30 sec) and 900 µL of the supernatant was removed. The cells were resuspended in the remaining medium and aliquots (30 and 70 µL) of each culture were spread on LB agar plates supplemented with [c]18 µg.mL^{-1} of [d]tobramycin (for *E. coli* MDS42) or 30 µg.mL^{-1} of kanamycin (for *E. coli* BW25113).

In an attempt to optimise the Red recombination protocol a few of the steps were altered:

[a]The samples were also electroporated at 1,800 V (0.1 cm cuvette gap) and 2,500 V (0.2 cm cuvette gap) in separate experiments.

[b]The recombination time was changed from 3 h to 2.5 h, 40 min and 20 min.

[c]The culture was spread on plates containing 20 µg.mL^{-1} instead of 18 µg.mL^{-1} of tobramycin.

[d]The culture was spread on plates containing 60 µg.mL^{-1} of apramycin and 30 µg.mL^{-1} of kanamycin instead of tobramycin. Although the MIC of *E. coli* MDS42 harbouring the *fusA*(G1478T) mutation alone had not been measured for these antibiotics, I assumed that the MICs will at least be higher than those of the strain expressing YejG alone, but lower than the strain harbouring the *fusA* mutation and over-expressingYejG (Table 18). The MIC for the YejG-expressing strain for apramycin and kanamycin is 55 µg.mL^{-1} and 26 µg.mL^{-1}, respectively (Table 18).

6.4.1.3 Phage transduction

For phage transduction, lysates were first prepared from the P1*cml*,*clr*100 lysogen (*E. coli* strain that harbours the virus in its chromosome), followed by phage titration and then phage transduction. P1*cml*,*clr*100 is a mutant of P1 that harbours the gene for chloramphenicol resistance (*cml*) and carries a mutation that makes P1 thermo-inducible (*clr*100) (Miller, 1972).

6.4.1.3.1 Preparation of P1 lysates

A freezer stock of *E. coli* CSH128 (lysogen for P1*cml*,*clr*100) was used to inoculate 3 mL LB-chloramphenicol medium (12.5 μg.mL^{-1} final concentration) and incubated at 28 ºC overnight. Three aliquots of the overnight culture (2x 200 μL and 1x 500 μL) were added to three separate sterile glass flasks containing 10 mL LB broth supplemented with $MgSO_4$ (100 mM), and then incubated at 28 ºC (shaking at 190 rpm) to OD_{600} ~ 0.7. To make lysates, 1x 200 μL and 500 μL cultures were transferred to a 40 ºC shaking water bath for 5 min, shifted to 40 ºC for 45 min (with shaking) and then moved to 37 ºC for an additional hour (with shaking). The second 200 μL culture served as a control and was left to incubate at 28 ºC during this period. To collect the P1 lysate, all of the cultures were centrifuged (2,400 xg, room temperature for 10 min) and the supernatant was filtered into sterilised glass bottles using a syringe (0.22 μm, 33 mm filters; Merck). To store the samples, 600 μL and 900 μL of chloroform was added to the 2x 200 μL and 500 μL supernatant samples, respectively, mixed and placed at 4 ºC.

6.4.1.3.2 Phage titration

To decide which dilution of P1 lysates to use for phage transduction, phage titering was performed. The lysate from the 200 μL culture (not the control) was serially diluted (10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}) in LB broth. One hundred μL aliquots of a fresh *E. coli* DH5α-E overnight culture were transferred into six separate microcentrifuge tubes containing 50 μL of $CaCl_2$ (100 mM final volume). One hundred μL of the 10^{-2} to 10^{-6} lysate dilutions were added to five of the tubes containing the *E. coli* DH5α-E cells. The sixth tube served as a control; therefore no phage was added. All of the tubes were incubated at 37 ºC for 15 min (no shaking).

Meanwhile, 'LB hard agar' (2 %) supplemented with $CaCl_2$ (5mM final concentration) was poured into six regular petri dishes to fill ¼ of the plate and left to set at room temperature. The cultures (250 µL total volume) were pipetted onto the centre of six separately prepared agar plates. Three mL of 'LB soft agar' (0.6 %) was then pipetted on top of the cells (and covered the centre of the 'LB hard agar') and left to set at room temperature. The plates were incubated at 37 ºC overnight. The lysate dilution used for P1 phage transduction was selected based on the plate that contained the highest phage dilution with the most plaques (indicative of cell lysis).

6.4.1.3.3 Phage transduction

To package the DNA of interest into P1 phage, an overnight culture (5 mL) of the strain carrying the desired DNA [*E. coli* BW25113 Δ*yejG* knockout strain or *E. coli* MDS42 *yejG* (HTD)] was prepared. $CaCl_2$ (5 mM final concentration) was added to the culture before pipetting 300 µL of culture into four individual microcentrifuge tubes. For phage adsorption, 10 µL, 30 µL and 100 µL of [a]P1 lysate (10^{-2} dilution) was added only to three of the tubes. No phage was added to the fourth culture as it served as a control. The tubes were incubated at 37 ºC for 20 min (no shaking).

The cultures were transferred into four separate glass flasks containing 10 mL LB broth supplemented with $CaCl_2$ (5mM final concentration). The cultures were incubated at 40 ºC (with shaking) for [b]4 h; this step allows for growth to occur after ~ 1 h followed by lysis at 2-3 h. To select which sample to use for downstream experiments, the cultures containing phage were compared to the control culture. The culture with the smallest volume of P1 phage that shows lysis (i.e., less turbidity) was selected. The culture was centrifuged (2,325 xg room temperature for 10 min). The supernatant (containing the phage with packaged target DNA) was filtered into glass bottles using a syringe. To kill cells that may have filtered through, 600 µL of chloroform was added to the filtered supernatant and stored at 4 ºC.

For the P1 phage transduction step, the following experiments were performed. Three 15 mL Falcon tubes containing $CaCl_2$ (5 mM final concentration) were labelled as sample, cell control and phage control. To the sample and cell control tubes, 1.3 mL of the recipient overnight culture (naïve *E. coli* BW25113 or *E. coli* MDS42) was added. To the sample and phage control tubes, 100 µL of P1 lysate [DNA from donor cells *E. coli*

BW25113 Δ*yejG* knockout strain or *E. coli* MDS42 *yejG* (HTD)] was added. The tubes were incubated at 37 ºC for [c]15 min with no shaking (infection step).

The cells were centrifuged (2,325 xg room temperature for 3 min) and the pellets were washed three times with 1mL LB, supplemented with sodium citrate (25 mM). The pellets were resuspended in 1 mL LB-sodium citrate and incubated at 37 ºC for 2 h (with shaking) (expression step). The cells were centrifuged once again (2,325 xg, room temperature for 10 min) and resuspended in 150 µL LB broth. Aliquots (20 µL and 130 µL) were spread on LB agar selection plates supplemented with sodium citrate (5 mM final concentration) and incubated at 37 ºC overnight. For the experiments involving the *fusA*(G1478T) mutation and the *kanR* marker, agar plates containing 18 $\mu g.mL^{-1}$ tobramycin and 30 $\mu g.mL^{-1}$ kanamycin were used for selection, respectively.

If colonies appeared on the agar selection plates, PCR (Appendix I.8) was performed using the yejG.for and rev primers (Table I.3, Appendix I.5) to confirm whether recombination had been successful: PCR screening (Appendix I.8) using yejG.for and rev primers (Appendix I.3) was performed on *E. coli* BW25113 colonies and PCR discrimination (Section 6.4.2) using fusA_m.for and rev primers (Appendix I.3) was performed on *E. coli* MDS42 colonies.

In an attempt to optimise this experiment a few of the steps were altered:

[a]For phage adsorption 5 µL, 50 µL and 80 µL of P1 lysate (10^{-2} dilution) were also used in separate experiments.

[b]For the lysis step, the cultures were incubated at 40 ºC (with shaking) for 6 and 8 h.

[c]The infection step was increased from 15 min to 20 and 30 min.

6.4.2 PCR discrimination assay

This experiment was to determine whether *E. coli* MDS42 Colonies that appeared on the selection plates containing tobramycin, apramycin or sisomicin had incorporated the *fusA*(G1478) mutation. Colonies were picked off the plates and transferred into individual thin-walled, 0.2 mL tubes containing 5 µL of sterile water. The cells were lysed by incubating the tubes at 95 °C for 5 min in a thermocycler (MJ Mini Bio-Rad). One µL of lysed cell was used for PCR. The reaction consisted of 1× Green GoTaq® Reaction buffer (Promega), 0.5 µM of the fusA_m.for and fusA_m.rev primers (Table I.3, Appendix I.5),

0.25 mM dNTP mix (iNtRON Biotechnology), 1.25 U i-Taq™ DNA polymerase (iNtRON Biotechnology) and DNA template. The PCR was run at the conditions: 95 °C for 1 min, 30 cycles of 95 °C for 30 sec, 95 °C for 30 s (denaturing), 44 °C for 30 s (annealing), 72 °C for 30 s (extension) and finally, 72 °C for 1 min.

A 477 bp PCR product was generated only when the *fusA* mutation was present. The fusA_w.for and fusA_w.rev primers (Table I.3, Appendix I.5) were used to identify colonies with the wild-type chromosomal *fusA* gene; a 399 bp PCR product was amplified in this case. The PCR was run at the same conditions as previously described but this time the annealing temperature was changed to 60 °C. The PCR products were run on an agarose gel (Appendix I.9).

6.4.3 Fitness experiments

6.4.3.1 Creating lacZ-marked strains

The same method as previously described in Section 2.5.2.1 (Chapter 2) was used to mark the strains with a functional copy of *lacZ*.

6.4.3.2 Creating the pWhite plasmid

The pWhite plasmid was constructed by cloning an arbitrarily chosen non-functional fragment of the *tdcD* gene (535 bp) into the middle of *lacZα* located on the pBlueScript plasmid. This plasmid was used to replace the pCA24N-*yejG* (HTD) plasmid carried by the *E. coli* MDS42 *yejG* (HTD) strain and allowed for blue/white screening in the fitness experiments (Section 6.2.3).

The pCA24N-*tdcD* plasmid was extracted from a saturated culture of the strain harbouring pCA24N-*tdcD* (ASKA collection; Kitagawa et al., 2005) (Table I.2, Appendix I.4) using the QIAprep Spin Miniprep kit (Qiagen). The purified pCA24N-*tdcD* plasmid (~ 3000 ng) and the pBlueScript II KS+ plasmid (Fermentas) (Table I.2, Appendix I.4) (3600 ng) were both digested with *Hind*III-HF restriction enzyme (New England Biolabs) in separate reactions (30 µL final volume) to generate compatible sticky ends. The digestion contained 20 U of *Hind*III-HF restriction enzyme, x1 reaction buffer and the plasmids. The reactions were incubated at 37 °C for ~ 3.5 h.

The samples were run on a 1 % SYBR-Safe™-stained gel (Appendix I.9). One band was obtained for the pBlueScript plasmid (2961 bp) and three bands for pCA24N-*tdcD* (4,955 bp, 970 bp and 535 bp). The pBlueScript DNA band (2961 bp) and the DNA band of the 535 bp fragment of the pCA24N-*tdcD* plasmid were excised from the gel and purified using the QIAquick purification kit series (Qiagen) (Appendix I.10). The DNA was eluted in 14 µL EB for the *tdcD* gene fragment and 30 µL EB for pBlueScript. The *tdcD* gene fragment (27 ng) was ligated to the linearised pBlueScript vector (50 ng) in a 10 µL reaction (3-fold molar excess than vector) (Appendix I.12).

One µL of the ligated DNA (3,496 bp) was used to transform electrocompetent MDS42 strains (Appendix I.3). The cells were covered by adding 500 µL SOC and incubating at 37 °C for 1 h. Aliquots (5 µL and 50 µL) were spread on LB agar supplemented with ampicillin (100 µg.mL^{-1} final concentration) and X-gal (40 µg.mL^{-1} final concentration). The presence of white colonies on the plates indicated that the *tdcD* insert had successfully ligated to the pBlueScript vector and dirupted the *lacZ*α fragment.

6.4.3.3 Replacing pCA24N-yejG(HTD) with the pWhite plasmid

This experiment was performed as previously described in Section 4.5.10 (Chapter 4).

6.4.3.4 Fitness assays

The fitness experiments (from Section 6.2.3) involving *E. coli* MDS42 harbouring the *fusA*(G1478T) mutation and carrying the pWhite plasmid; *E. coli* MDS42 harbouring the wild-type *fusA* and carrying pWhite; *E. coli* MDS42 harbouring the *fusA*(G1478T) mutation alone and over-expressing either YejG or GFP; *E. coli* MDS42 harbouring the wild-type *fusA* and over-expressing GFP and; *E. coli* MDS42 harbouring the *fusA*(G1478T) mutation and over-expressing YejG(HTD) were performed using the same protocol previously described in Section 2.5.2.2 (Chapter 2).

For the strains carrying the pWhite plasmid, individual colonies were picked off LB-ampicillin (100 µg.mL^{-1} final concentration) plates and the fitness experiments were performed in LB-ampicillin (100 µg.mL^{-1} final concentration) medium. The cells were plated on LB plates supplemented with ampicillin (100 µg.mL^{-1} final concentration) and X-gal (40 µg.mL^{-1} final concentration) for the T_0 and T_{24} plates.

6.4.4 Growth assays

One mL of the appropriate overnight culture was centrifuged at full speed in a benchtop centrifuge (17,000 xg, room temperature for 2 min). The supernatant was removed and the pellet was washed once with sterile water and then centrifuged as before. The cells were resuspended in 1 mL sterile water and incubated at 37 °C for ~ 2 h. The cells were diluted 1000-fold in LB supplemented with the appropriate antibiotic and IPTG (50 μM final concentration), where necessary. Two hundred μL of the culture was aliquotted into 96-well microtiter plates in six replicates. A 'media only control' sample was also aliquotted into the wells in duplicates. The plates were read in a Synergy 2 plate reader. Growth was monitored at OD_{600} for ~ 20 h, with continuous slow shaking and a reading was taken every 15 min. The final growth measurements were the average of six replicates performed in two or more independent experiments.

CHAPTER 7

Summary and outlook

7.1 The bacterial resistome

Antibiotic resistance is threatening the successful management of bacterial infections, both in humans and in animals. One way to better understand this problem is to investigate pathways by which resistance can occur. Bacteria possess many genes that contribute directly or indirectly to inherent resistance (the bacterial 'resistome') (Wright, 2012). Evidence for this is shown in a study where ~ 500 species of soil bacteria revealed resistance to a number of antibiotics, including natural and completely synthetic drugs (D'Costa et al., 2006). The majority of bacteria do not cause diseases in humans or animals. Thus, the findings of D'Costa et al., (2006) reinforced the idea that an understanding of antibiotic resistance would require a focus on non-pathogenic bacteria, as well as pathogens. Building on this insight, previous work in our laboratory showed that proteins in non-pathogenic bacteria (laboratory strains of *E. coli*) were able to endow the cell with resistance to many antibiotics, provided they were over-expressed (Soo et al., 2011). In the first part of this thesis, the level of resistance imparted by a selection of these proteins was quantified. The effect of resistance on cell fitness was also assessed.

In Chapter 2, I showed that over-expression of pre-existing gene products was able to endow non-pathogenic bacteria with moderate levels of resistance. This evolutionary step (protein over-expression) may play an important role in the emergence of high-level resistance, by serving as an initial solution to the problem of surviving in an antibiotic-containing environment (Martinez & Baquero, 2000). In some of the cases studied, clinically-relevant levels of resistance were attained (e.g., through *folA* and *ycgZ* over-expression). This emphasises how non-pathogenic bacteria are able to evade the inhibitory effects of drugs with ease, by making use of intrinsic resistance elements. This study uncovered novel secondary functions of previously-characterised proteins, as well as shedding light on the roles of proteins with previously unknown functions. It has shown that antibiotic resistance can be used as a model system to gain fundamental insights into protein function.

Since the connection between fitness and resistance may play an important role in a clinical setting, it was investigated. The results revealed that the relationship between fitness and resistance is complex; combinations of a high- or a low-level resistance with either an increase or a decrease in fitness were observed. Resistance mutations that confer a marginal growth improvement may play an important role during the initial stages of

adaptation; such mutations will allow the bacterium to produce more descendants than its competitors and increase the chances of it acquiring beneficial mutations to further enhance fitness.

7.2 Evolution of antibiotic resistance

Jensen's model for the evolution of new functions posits that primordial cells possessed a small number of genes that encoded enzymes with numerous functions (promiscuity) (Jensen, 1976). An appealing model that connects genetic processes with enzyme promiscuity in the evolution of new function is the IAD model (Bergthorsson et al., 2007). It proposes that through duplication, mutation and selection of pre-existing genes that encode proteins with promiscuous activities, enzymes with specialized functions (modern enzymes) have emerged. Not only that, but Bergthorssen et al. (2007) suggested that the IAD model continues to drive the emergence of new functions, such as antibiotic resistance, today. In this thesis, I explored the idea that promiscuous activities can serve as the starting points for the evolution of new functions.

In Chapters 3 and 4, I investigated whether the proteins carbonic anhydrase (CA), YeaD, YdfW and YejG, each of which imparted low to moderate-levels of resistance, could be evolved to improve their resistance functions. In all four cases, error-prone PCR and selection failed to identify variants that improved the antibiotic resistance of the host cell. In the case of YeaD, YdfW and YejG, it was difficult to provide a rationale, because the native functions of these proteins are not known. If their native functions are to confer antibiotic resistance, then it is possible that the resistance functions they possess can not be improved further. In the case of CA, which was displaying promiscuous activity towards enoxacin, it is possible that major modifications to the active site (through multiple mutations) may be required to improve the promiscuous resistance activity.

This work has hinted that directed evolution can be used to explore the limits of protein function [this has been argued previously (Miller et al., 2006)]. Ultimately, alternative directed evolution strategies, combined with investigations into the structures and functions of YeaD, YdfW and YejG will be necessary to confirm this.

In Chapters 4, 5 and 6, a novel route to tobramycin resistance involving YejG was discovered. Over-expression of the protein, together with the chromosomal *fusA*(G1478T)

mutation, endowed *E. coli* with maximal resistance to tobramycin. A role for YejG that was previously unknown in promoting an early entry into log phase (as well as an early exit into stationary phase) was uncovered. Initially, it was observed that over-expression of YejG endowed *E. coli* with low-level tobramycin resistance and a small increase in fitness. The chromosomal *fusA*(G1478T) mutation imparted higher levels of tobramycin resistance but reduced fitness. YejG alleviated this fitness cost. This emphasizes the importance of mutations that confer low-levels of resistance with the effect of improving fitness in facilitating the emergence of mutations that confer higher level resistance. The ability of resistance mutations with different fitness effects to be mixed and matched to improve survival of the bacterium has been made evident.

This thesis has identified a function of YejG in promoting an early entry into log phase of growth, although the molecular mechanism of its growth effects remain to be determined. From the results, there is no evidence to suggest that YejG is binding to the ribosome or regulating the membrane potential. Sequence analysis revealed the presence of the same protein in the bacterial pathogens *Shigella flexneri* (99 % identical) and *Salmonella enterica* serovar Choleraesuis (92 % identical). This extremely high level of conservation suggests an important role for YejG in these organisms. It is plausible that YejG might be found to be involved in the development of clinically-relevant resistance in future.

7.3 Concluding remarks

The work presented in this thesis supports the idea that bacteria possess surprisingly many intrinsic resistance elements that can be easily accessed through over-expression mutations. The proteins that confer these weak resistance phenotypes may be difficult to improve through random mutagenesis and selection; however, they are not insignificant. Bacteria can capitalise on such activities to facilitate the emergence of higher levels of resistance. A good understanding and an early identification of genes that make up the bacterial resistome will allow for downstream opportunities in drug development. The screening and directed evolution methods that I have explored should be useful tools for identifying the ways in which pathogens can evolve resistance to new drugs, and therefore, whether these new drugs are likely to be useful in the clinic.

[Blank page]

APPENDIX I

General Materials and Methods

I.1 REAGENTS

The purity of the reagents used were analytical grade or higher. Autoclaved MilliQ® water was used to prepare the solutions. The chemicals were purchased from Sigma-Aldrich, unless otherwise mentioned. Solutions and media were autoclaved by heating up to 121 °C and holding this temperature for 20 min (Autoclave; J. Mercer & Sons Ltd).

I.2 GROWTH MEDIA and ANTIBIOTICS

LB medium (ForMedium; 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) was used for the growth of *E. coli*. The culture was grown at 37 °C with agitation at 180 rpm. To prepare agar plates, bacteriological agar (ForMedium) was added to LB medium at a final concentration of 1.5 % (w/v). The appropriate antibiotic was added to all growth media.

All general antibiotics were purchased from Melford and dissolved in water to make stock solutions. Chloramphenicol on the other hand was bought from Duchefa Biochemie and dissolved in ethanol to prepare the stock solution. The final concentrations for general antibiotics are as follows: chloramphenicol, 34 $\mu g.mL^{-1}$; ampicillin, 100 $\mu g.mL^{-1}$; kanamycin, 30 $\mu g.mL^{-1}$; and gentamicin, 15 $\mu g.mL^{-1}$. The antibiotics were stored at – 20 °C.

I.3 BACTERIAL STRAINS

A single bacterial colony was grown overnight in selective media. Glycerol was added to a 300 µL aliquot of the overnight culture to 12.5 % (v/v) final concentration and the culture was stored at – 80 °C for long term. Agar plates containing bacterial colonies were stored at 4 °C and used within five days.

Table I.1 Characteristics of *E. coli* strains used in this thesis

Strain	Genotype	Reference
BW25113 Δ*metC*	F− Δ(*araD*–*araB*)567 Δ*lacZ*4787(::*rrnB*−3) λ⁻ *rph*−1 Δ(*rhaD*–r*haB*)568 *hsdR*514 Δ*metC*::*Kan*^{R}	(Baba et al., 2006)
BW25113 ΔyejG	F− Δ(*araD*–*araB*)567 Δ*lacZ*4787(::*rrnB*−3) λ⁻ *rph*−1 Δ(*rhaD*–*rhaB*)568 *hsdR*514 Δ*yejG*::*Kan*^{R}	(Baba et al., 2006)
DH5α-E	F⁻ φ80*lacZ*ΔM15 Δ(*lacZYA*–*argF*)U169 *recA*1 *endA*1 *hsdR*17(r_k^-, m_k^+) *gal*⁻ *phoA supE*44 λ⁻ *thi*−1 *gyrA*96 *relA*1	Invitrogen
DH5α-E::*lacZ*	DH5α-E with *Gm*^{R}-*lacZ* marker inserted at the *att*Tn7 site	Lab stock; Dr. Wayne Patrick
E. coli MDS42 *yejG* (HTD)	MDS42 strain harbouring the *fusA*(G1478T) mutation and carrying pCA24N-*yejG* plasmid with Q14H, I68T and N110D mutations	Present study
K12	Wild-type *E.coli*; $F^+ \lambda^+$	American Type Culture Collection (ATCC) 10798
MDS42	MG1655 multiple-deletion strain (Pósfai et al., 2006), *recA*-1819, *lacZ*ΔM15	Scarab Genomics LLC
MDS42::*lacZ*	MDS42 with *Gm*^{R}-*lacZ* marker inserted at the *att*Tn7 site	Present study
E. coli CSH128	P1*cml*,*clr*100	Dr. Heather Hendrickson

I.4 PLASMID

Purified plasmids were kept at −20 °C.

Table I.2 Features of plasmids used in this thesis

Plasmid name	Relevant characteristics	Reference
pCA24N-*cynT*(-GFP)	*Cam*^{R}; ASKA plasmid carrying the *cynT* gene; no GFP tag present	Present study
pCA24N-GFP	*Cam*^{R}; pCA24N plasmid with GFP tag (with no fusion partner) used for negative control	Lab stock; Dr. Wayne Patrick
pCA24N-*tdcD*	CamR; ASKA plasmid carrying the tdcD gene	(Kitagawa et al., 2005)
pCA24N-*trpD*(-GFP)	*Cam*^{R}; ASKA plasmid carrying the *trpD* gene; no GFP tag present	Present study
pCA24N-*yeaD*(-GFP)	*Cam*^{R}; ASKA plasmid carrying the *yeaD* gene; no GFP tag present	Present study
pCA24N-*ydfW*(-GFP)	*Cam*^{R}; ASKA plasmid carrying the *ydfW* gene; no GFP tag present	Present study
pCA24N-*yejG*(-GFP)	*Cam*^{R}; ASKA plasmid carrying the *yejG* gene; no GFP tag present	Present study
pCA24N- *yejG* (HTD)	*Cam*^{R}; ASKA plasmid carrying a mutated version of *yejG* gene with Q14H, I68T and N110D mutations; 'winner' from *yejG* library; no GFP tag present	Present study
pCP20	*Amp*^{R}; plasmid used to remove gentamicin selection marker which is flanked by FRT sites	(Datsenko & Wanner, 2000)
pKD46	*Amp*^{R}; temperature-sensitive plasmid expressing the arabinose-induced λ Red recombinase	(Datsenko & Wanner, 2000)
pRedET	*Tet*^{R}; temperature-sensitive plasmid expressing the arabinose-induced Red/ET recombination proteins	Gene Bridges GmbH
pUC19	*Amp*^{R}; 2.7 kb control plasmid used in transformation	Invitrogen
pUC18R6K-mini-Tn7T-*Gm-lacZ*	*Amp*^{R}, *Gm*^{R} on mini-Tn7T suicide vector; *ori*$_{R6K}$; *lacZ* transcriptional fusion vector	Lab stock; Dr. Wayne Patrick
pUX-BF13	*Amp*^{R}; *ori*$_{R6K}$; helper plasmid encoding the TnsABC+D	Dr. Monica Gerth
pBlueScript II KS+	*Amp*^{R}; 2,961 bp plasmid containing α-complementing fragment of beta-galactosidase	Fermentas
pWhite	*Amp*^{R}; pBlueScript II KS+ plasmid with *lacZ*α fragment disrupted with *tdcD* fragment (3,496 bp)	Present study

I.5 OLIGONUCLEOTIDES

Integrated DNA Technologies (Coralville, Iowa) custom synthesized all the primers. The lyophilized primers were resuspended in TE buffer (10 mM Tris-HCl, pH 8.0; 1 mM EDTA) to a final concentration of 100 μM. A working stock was prepared by diluting with sterile MilliQ® water to obtain a concentration of 10 μM. The primer stocks were stored at − 20 °C.

Table I.3 Oligonucleotides used in this thesis

Primer name	Primer sequence (5´→3´)	Reference
ecGlmS.rev	CAGTTAACTGTGCTGTTGATGC	Lab stock; Dr. Wayne Patrick
ecTn7R.for	CAGCATAACTGGACTGATTTCAG	Lab stock; Dr. Wayne Patrick
fusA_m.for	GTAACTTTCTGGCGGATAG	Present study
fusA_m.rev	CGGTCTGAAAGACGTAACC	Present study
fusA_w.for	CCTGACCTTTGAGCATACC	Present study
fusA_m.rev	AGGTTGCTTACCGTGAAAA	Present study
fusA_A.for	CCGAAGTGCAGACGAATACCCATGTCTACTACCGGGTAGC	Present study
fusA_A.rev	GGTCTGAAAGACGTAACCACTGGTGACACCCTGTGTGACC	Present study
fusA_B.for	CCCTTCAGGAGAGAGCACGGGACTTTGG	Present study
fusA_B.rev	GGACTGACGAAGAATCTAACCAGACCATCATCGC	Present study
fusA_C.for	CCCTTCAGGAGAGAGCACGGGACTTTGG	Present study
fusA_C.rev	CGAATGGCACCAGAACCTGATCGAATCCG	Present study
fusA_D.for	CCCTTCAGGAGAGAGCACGGGACTTTGG	Present study
fusA_D.rev	GGAGTTTTAGTCACCAGGCGGGCGCTTCCAGTAAGC	Present study
gyrA_setA.for	GCAGCCTGGACTTTTCAATTCAAACAAGGG	Present study
gyrA_setA.rev	GGTCGTCCGATCGTCAACCTGCTGC	Present study

gyrA_setB.for	GAGGACAGTTTTCTTCACGGTACCGTTAGCGG	Present study
gyrA_setB.rev	GTGGCATTGCACCATGGTCAGCCG	Present study
gyrA_setC.for	GAGCTTTACGCAGTTCGAAAATAGTACGAC	Present study
gyrA_setC.rev	CAGGCATTGGATGTGAATAAAGCG	Present study
gyrB_setA.for	CCGCCCTTTTCTTATAGCTTCTTGCCGGATGCGGCG	Present study
gyrB_setA.rev	CTCTATCGCGCTGGACGGCGCAACGC	Present study
gyrB_setB.for	CGACGCTCCATACGATTGATCATTTTCTGCG	Present study
gyrB_setB.rev	GTCAGCGCCACCGGTGACGATGC	Present study
gyrB_setC.for	CCGCCGATTTCACCTCAGAAGAAACC	Present study
gyrB_setC.rev	GTCAGCGCGATCAGTGCTGAACACG	Present study
KeioKanROut.for	CGGTGCCCTGAATGAACTGC	Lab stock; Dr. Wayne Patrick
metC-dwnstr.rev	GACTTTTCACAATAAAATGTCTGCAAAATTGTCCAAAAG	Lab stock; Dr. Wayne Patrick
parC_setA.for	CGCTGCCGGATGACGACTTAACG	Present study
parC_setA.rev	CTGAACCTGTCACCATTGTGCTGTCGC	Present study
parC_setB.for	GCTCTTACCTTTCACCGCCGC	Present study
parC_setB.rev	CGATCTGCGCGATGAATCTGACC	Present study
parC_setC.for	CCAGATCGGTGGTAGCGAAGAGG	Present study
parC_setC.rev	CGCGGCAGATAATGTAGTATCTCCGG	Present study

parE_setA.for	GCCGTAAAAGGTGACGTGCTGGAGATGAATATCCG	Present study
parE_setA.rev	GGAAGATATCTGGGATCGCTGCGCCTATGTGC	Present study
parE_setB.for	CGCCATCTCCGCCAGCAGTTCAGC	Present study
parE_setB.rev	CGCATTCACGGGCTTCGAAGATTATTTCAACC	Present study
parE_setC.for	GCGTTTACCGCAAGTGCCGACAACCTGTAAATCCTGC	Present study
parE_setC.rev	CGCATTCACGGGCTTCGAAGATTATTTCAACC	Present study
pCA24N.for	GATAACAATTTCACACAGAATTCATTAAAGAG	(Patrick et al., 2007)
pCA24N.rev2	CAAATCCAGATGGAGTTCTGAGG	(Patrick et al., 2007)
pCA24N_NoGFP.for*	*Phos*-TAAGGGTCGACCTGCAGCCAAGC	Present study
cynT_NoGFP.for*	*Phos*-GCGGCCGCATAGGCCTTACGCTGCGGTCGGTTGGC	Present study
yeaD_NoGFP.for*	*Phos*-GCGGCCGCATAGGCCTTAACGTTTCGCAACGCGAATG	Present study
ydfW_NoGFP.for*	*Phos*-GCGGCCGCATAGGCCTTAATCTTTTTTATCGGGATCAGGCTTC	Present study
yejG_NoGFP.for*	*Phos*-GCGGCCGCATAGGCCTTAAAAAGGATTGTAGTTTGAAAAGGG	Present study
ynfB_speG.for	CCAGGGCAGAAAGCATCCAATAATTTCACACAGCGC	Present study
ynfB_speG.rev	CATTGTTGATGAAGGAGTAAAACCCGCCGTCGAACGG	Present study
yejG. for	ACCGCAGCAGGAGTATACGCGTCAGC	Present study
yejG.rev	GTAAACAAATTGTGAAGTGAATGTGC	Present study

*Primers were phosphorylated at 5´end

The 71 base single-stranded oligonucleotide included the G1478T mutation (underlined below) at base 36 flanked by 35 nucleotides of wild-type *fusA* and four phosphorothioate bonds (*PT*) located at the 5´ terminus:

PT-PT-PT-PT-
GTTTACCTTCAACATCGGTAACTTTCTGGCGGATA**T**TTTCACGGTAAGCAACCTGCGGTTTACCTACGTTC

I.6 DNA SEQUENCING

Massey Genome Service (Palmerston North) or Macrogen Inc. (South Korea) carried out all the general DNA sequencing in this study. An automated capillary-based ABI3730 DNA Analyzers (Applied Biosystems) was employed by both companies for routine dye-terminator sequencing. Whole genome sequencing was performed by AMBRY Genetics (America). The genome was sequenced on the Illumina Genome Analyzer IIx.

I.7 ANALYTICAL SOFTWARE

DNA and protein sequences were viewed, designed, aligned to references and modified using MacVector (MacVector, Inc.). Chromatograms from DNA sequences were imported using this software. GraphPad Prism 5.0d (GraphPad Software) was used to analyse and draw graphs.

I.8 PCR SCREENING

Routinely, a 20-µL reaction in 0.2 mL thin-walled tube was performed for PCR screening in an MJ Mini Thermal Cycler (Bio-Rad). The reaction consisted of 1× Green GoTaq® Reaction buffer (pre-mixed with two tracking dyes; Promega), 0.5 µM of the forward and reverse primers, 0.25 mM dNTP mix (iNtRON Biotechnology), 1.25 U i-Taq™ DNA polymerase (iNtRON Biotechnology), and DNA template. The template was either 1 µL plasmid DNA, saturated bacterial culture or one colony from a cultured agar plate. The PCR was run at the conditions: 95 °C for 4 min, 30 cycles of 95 °C for 30 sec, annealing temperature for 30–60 s, 72°C for 1 min/kb, and finally, 72 °C for 1 min. The annealing temperature was usually 3°C below the lower Tm of the two primers.

I.9 AGAROSE GEL ELECTROPHORESIS

Different gel concentrations were prepared depending on the on the expected size of the DNA. For a 1–10 kb DNA, a 0.8 % (w/v) agarose was prepared and a 1.5 % (w/v) agarose was made for a 100–1000 bp DNA. Agarose powder (Axygen) was dissolved in 1× TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA, pH 8.0). To stain the nucleic acid, ethidium bromide (Bio-Rad) was added to a final concentration 0.5 µg/mL. Before loading

the DNA sample into the well, one volume of 6× DNA Loading Dye (Fermentas) was mixed with five volumes of the sample.

For each gel run, an appropriate DNA ladder was loaded alongside the samples and run at 100 V until the bottom tracking dye had migrated through $^{3}/_{4}$ of the gel. A UV transilluminator (part of the Gel Doc™ 2000 Gel Documentation System, Bio-Rad) was used to view the gel.

A different DNA stain (1× SYBR Safe™ Gel DNA stain; supplied as 10,000× concentrate by Invitrogen) was used instead of ethidium bromide when the DNA was to be purified from the gel. In this case, a Safe Imager™ Blue Light Transilluminator (Invitrogen) was used. This avoided DNA exposure to UV light.

I.10 DNA EXTRACTION and PURIFICATION

To purify plasmid DNA the QIAprep Spin Miniprep kit (Qiagen) was used. To clean up DNA from PCRs, ligation reactions and agarose gels the QIAquick purification kit series (Qiagen) was employed. MinElute Spin columns (Qiagen) were used in place of the usual QIAprep or QIAquick Spin columns to obtain a concentrated DNA sample (e.g., elution of DNA in 12-µL EB). Genomic DNA was extracted using the Wizard® Genomic DNA Purification Kit (Promega), according to the manufacturer's instructions.

I.11 DNA QUANTIFICATION

To determine the concentration of a DNA sample, usually 1 µL of DNA was added to 49 µL in a UVette (Eppendorf) measured spectrophotometrically in an Eppendorf Biophotometer. Alternatively, the DNA sample was run alongside a suitable DNA ladder [GeneRuler^{TM} 100 bp or GeneRuler^{TM} 1 kb DNA ladder (Thermo Scientific)] and the intensity of the samples with the intensity of the bands in the ladder was compared.

I.12 RESTRICTION DIGEST AND LIGATION

Typically, a digestion reaction (10 µL) consisted of 1× reaction buffer, 0.1–2 µg of DNA and 2-10 U restriction endonuclease. Information for incubation temperature, bovine serum

albumin supplementation, buffer compatibilities (especially important for a double digestion) and heat inactivation were provided by the manufacturer. DNA was digested for 1–16 h unless otherwise stated. All restriction endonucleases were obtained from New England Biolabs.

For ligation reactions, a 10 μL reaction was prepared in a 0.2 mL thin-walled tube that contained 1× T4 DNA Ligase Buffer, 2.5 Weiss units of T4 DNA Ligase (Fermentas) and linearized vector and digested gene insert (in 3-fold molar excess than vector). The reaction was incubated at 16 °C for 16 h and then heat-inactivated at 65 °C for 10 min. If ligated products were used to transform *E. coli*, they were first diluted (at least 5-fold) using sterile MilliQ® water or cleaned up (Appendix I.10).

I.13 PREPARING ELECTROCOMPETENT CELLS

The protocol for preparing electrocompetent cells is adapted from that described in Sambrook and Russell, (2001). A freezer stock or a single colony of *E. coli* was grown overnight in 3 mL LB supplemented with the appropriate antibiotics. The overnight culture was used to inoculate 250 mL SOB medium (20 g/L tryptone, 5 g/L yeast extract, 10 mM NaCl, 2.5 mM KCl) supplemented with the appropriate antibiotics. The culture was incubated at 37 °C until an OD_{600} ~ 0.35-0.4 was reached. Volumes of the culture were evenly transferred into 6 × 50 mL pre-chilled centrifuge tubes. The tubes were placed on ice for another 30 min-1 h. The following steps were carried out at 4 °C.

The cells were centrifuged at 3,200 ×g for 15 min, at 4 °C (Heraeus Multifuge 1S-R) and the pellet was resuspended in 35 mL of pre-chilled 10 % (v/v) glycerol and re-pelleted as described earlier. The six cell pellets were resuspended with 10 % glycerol and pooled into 4 × 35 mL. The cells were centrifuged once more. The four pellets were resuspended using 10 % glycerol and this time pooled into 2 × 35 mL. The cells were pelleted, and the two pellets were resuspended with 10 % glycerol and pooled into 1 × 35 mL. A final centrifugation step was performed.

The weight of the final cell pellet was determined and resuspended with 0.1 mL of 10 % glycerol per 0.1 g of wet cell weight. The OD_{600} of the resulting suspension was usually in the range of 100–150. The cells were aliquoted into 50 μL volume and used immediately for electroporation (see Appendix I.14) or stored at −80 °C until it was ready to be used.

I.14 TRANSFORMATION

Electrocompetent cells that had been prepared earlier and frozen were thawed on ice and usedfor regular cloning experiments. To construct error-prone libraries, electrocompetent cells were prepared and transformed on the same day.

The standard protocol according to Sambrook and Russel (2001) was followed. DNA was added to an aliquot of electrocompetent cells (50 µL) and allowed to incubate on ice for 1 min. The cells were transferred to a sterile, pre-chilled Gene Pulser cuvette 0.1 cm gap (Bio-Rad). The sample was electroporated at 1.8 kV, 200 Ω and 25 µF in a Gene Pulser unit with Pulse controller (Bio-Rad Gene Pulser II) or at 2.5 kV, 200 Ω and 25 µF in an electroporation cuvette with 0.2 cm gap (Bio-Rad). Time constants of 4.5–5.5 ms were generally observed for all electroporations. After electroporation, 0.5 mL of pre-warmed SOC medium (20 g/L tryptone, 5 g/L yeast extract, 10 mM NaCl, 2.5 mM KCl and 20 mM glucose) was added to the cells before being transferred to a sterile 15 mL Falcon tube. The cells were recovered at 37 °C for 1 h and then spread on selective LB agar plates containing the appropriate antibiotics. The plates were incubated overnight at 37 °C. To determine the transformation efficiency, 10 pg pUC19 (Invitrogen) was added to 50 µL of electrocompetent *E. coli* cells and electroporated and recovered as described. Aliquots of the cells (5 µL, 20 µL and 100 µL) were spread on LB-ampicillin plates (100 $\mu g.mL^{-1}$ final concentration) and incubated overnight at 37 °C. The transformation efficiency was calculated based on the colony counts.

I.15 SDS-PAGE

The buffer system used for SDS-PAGE was the discontinuous system. To make polyacrylamide gels, the stock solutions were as follows: 40 % acrylamide: *N,N′*-methylene-bis-acrylamide (29:1) solution (Bio-Rad), 1.5 M Tris-HCl (pH 8.8), 0.5 M Tris-HCl (pH 6.8), 10 % (w/v) sodium dodecyl sulfate, 10 % (w/v) ammonium persulfate and *N,N,N′,N′*-tetramethylethylenediamine (~ 6.67 M).

The resolving gel mixtures consisted of the following components: 12 % acrylamide: *N,N'*-methylene-bisacrylamide (29:1) solution, 375 mM Tris-HCl (pH 8.8), 0.1 % sodium dodecyl sulfate, 0.0 5 % ammonium persulfate and 6.67 mM *N,N,N′,N′*-tetramethylethylenediamine.

All stacking gel mixtures consisted of 4 % acrylamide: *N,N'*-methylene-bis-acrylamide (29:1) solution, 125 mM Tris-HCl (pH 6.8), 0.1 % sodium dodecyl sulfate, 0.05 % ammonium persulfate and 6.67 mM *N,N,N′,N′*-tetramethylethylenediamine.

The Mini-PROTEAN gel casting assembly (Bio-Rad) was used to prepare polyacrylamide gels of 1 mm thickness. The resolving gel mixture was pipetted in between the glass plates until it was ~ 3.5 cm from the top of the plates. Seventy % (v/v) ethanol was added on top of the gel mixture up to 1 cm and left for ~ 30 min for gel polymerization to occur. The ethanol was removed by pouring off the polymerized gel before rinsing with running water. The stacking gel mixture was pipetted on top of the resolving gel and a 10-well comb was inserted. The stacking gel was also left for ~ 30 min to polymerize. If the gel was not used immediately, it was wrapped in moistened paper towels and a plastic film, stored at 4 ºC and used within two weeks.

The comb was removed from the gel under running water and the polyacrylamide gel was assembled into the Mini-PROTEAN Tetra Cell electrophoresis tank (Bio-Rad)., the heated samples were loaded into each well. The protein samples to be run were mixed with an equal volume of 2× Loading Buffer [100 mM Tris-HCl (pH 8.8), 4 % (w/v) sodium dodecyl sulfate, 20 % (v/v) glycerol, 0.2 % (w/v) bromophenol blue, 200 mM β-mercaptoethanol]. The mixtures were heated at 95 °C for 5 min in a thermocycler (MJ Mini Bio-Rad). The gel was run in 1× SDS PAGE buffer [25 mM Tris, 250 mM glycine (pH 8.3), 0.1 % (w/v) sodium dodecyl sulfate], at 200 V. The run was stopped when the tracking dye was 0.5 cm from the bottom of the gel.

To stain the gel, it was removed from the glass plates and submerged in Coomassie Blue stain [2.5 g/L (w/v) Coomassie R250 (Bio-Rad) in 4 volumes of water: 5 volumes of methanol: 1 volume of acetic acid] with gentle rocking for 30 min. The stained gel was washed repeatedly with water before destaining with Destaining Solution (6 volumes of water: 3 volumes of methanol: 1 volume of acetic acid) for 1h or more with gentle rocking.

APPENDIX II

Statement of Contributions

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION TO DOCTORAL THESIS CONTAINING PUBLICATIONS

(To appear at the end of each thesis chapter/section/appendix submitted as an article/paper or collected as an appendix at the end of the thesis)

We, the candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of Candidate: PAULINA HANSON-MANFUL

Name/Title of Principal Supervisor: AUSTEN GANLEY

Name of Published Research Output and full reference:
Hanson-Manful, P and Patrick, W M (2013) Construction and analysis of randomized protein-encoding libraries using error-prone PCR. Methods in molecular biology (Clifton, N.J.). 996, 251–67.

In which Chapter is the Published Work: Chapters 3 and 4

Please indicate either:

- The percentage of the Published Work that was contributed by the candidate: 60%

  and / or

- Describe the contribution that the candidate has made to the Published Work:
  The candidate conducted the experiments that are used as examples in this book chapter (which describes experimental protocols). The candidate provided her lab notes as the basis of the chapter draft; however, the chapter was written by Dr. Patrick.

Candidate's Signature

06 June 2013
Date

Principal Supervisor's signature

06 June 2013
Date

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION TO DOCTORAL THESIS CONTAINING PUBLICATIONS

(To appear at the end of each thesis chapter/section/appendix submitted as an article/paper or collected as an appendix at the end of the thesis)

We, the candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of Candidate: PAULINA HANSON-MANFUL

Name/Title of Principal Supervisor: AUSTEN GANLEY

Name of Published Research Output and full reference:
Soo VWC, Hanson-Manful P and Patrick WM (2011) Artificial gene amplification reveals an abundance of promiscuous resistance determinants in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America. 108 (4), 1484–9.

In which Chapter is the Published Work: Chapter 2

Please indicate either:

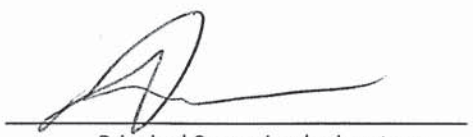
- The percentage of the Published Work that was contributed by the candidate: 10%

  and / or

- Describe the contribution that the candidate has made to the Published Work:
  The candidate performed some of the experiments in this paper, and assisted with some of the data analysis.

Candidate's Signature

06 June 2013
Date



Principal Supervisor's signature

06 June 2013
Date

Appendix III

Publications arising from this work

Chapter 15

Construction and Analysis of Randomized Protein-Encoding Libraries Using Error-Prone PCR

Paulina Hanson-Manful and Wayne M. Patrick

Abstract

In contrast to site-directed mutagenesis and rational design, directed evolution harnesses Darwinian principles to identify proteins with new or improved properties. The critical first steps in a directed evolution experiment are as follows: (a) to introduce random diversity into the gene of interest and (b) to capture that diversity by cloning the resulting population of molecules into a suitable expression vector, en bloc. Error-prone PCR (epPCR) is a common method for introducing random mutations into a gene. In this chapter, we describe detailed protocols for epPCR and for the construction of large, maximally diverse libraries of cloned variants. We also describe the utility of an online program, PEDEL-AA, for analyzing the compositions of epPCR libraries. The methods described here were used to construct several libraries in our laboratory. A side-by-side comparison of the results is used to show that, ultimately, epPCR is a highly stochastic process.

Key words Directed evolution, Random mutagenesis, Error-prone PCR, GeneMorph II, Mutazyme II DNA polymerase, Library, Mutation spectrum, Mutational bias, PEDEL-AA

1 Introduction

In the past two decades, directed evolution has emerged as a powerful method for altering the properties of proteins. It involves mimicking the process of Darwinian evolution on a single gene, on a laboratory timescale. In the first step of a directed evolution experiment, mutations are introduced at random into copies of the target gene, resulting in a large and diverse library of variants (typically 10^3–10^9 clones). The members of this library are subjected to a suitably high-throughput screen or genetic selection, in order to identify rare variants with improvements in the desired property. Multiple rounds of mutagenesis and screening/selection enable the accumulation of beneficial mutations that may have been impossible to predict, a priori. Directed evolution has been adopted widely for tailoring industrially relevant biocatalysts with improvements in properties such as substrate specificity, enantioselectivity, and

Juliet A. Gerrard (ed.), *Protein Nanotechnology: Protocols, Instrumentation, and Applications*, Methods in Molecular Biology, vol. 996, DOI 10.1007/978-1-62703-354-1_15,

thermostability (1, 2). It has also been used to address fundamental questions about protein structure, function, and evolution (3).

Many methods have been developed for introducing molecular diversity into parent sequences (4, 5). One of these methods, error-prone polymerase chain reaction (epPCR), remains a particularly common means of generating random mutations at any position in the target gene. Conceptually, epPCR is simple: the target gene is amplified exponentially, under conditions in which the fidelity of the polymerase is reduced. Cloning the randomly mutagenized PCR product into an appropriate expression vector yields a library of variants that can be used in downstream screening or selection. While outside the scope of this chapter, it is worth noting that these downstream steps are also critically important for the success of any directed evolution experiment (6).

The original—and still the cheapest—way to carry out an epPCR is to reduce the fidelity of *Taq* DNA polymerase, by adding Mn^{2+} ions and unbalanced ratios of dNTPs (7, 8). Detailed protocols for using *Taq* polymerase to construct epPCR libraries have been described previously (9, 10). However, *Taq*-generated epPCR libraries suffer from biases in the types of mutations that are observed; in particular, mutations at A:T base pairs are massively overrepresented (11).

We have argued that an unbiased and maximally diverse library has the highest probability of containing variants with the desired function (12). The GeneMorph II Random Mutagenesis Kit from Agilent Technologies (http://tinyurl.com/3mb6x66) is designed specifically for the construction of unbiased epPCR libraries. In the protocols below, we describe the use of the GeneMorph II kit, as well as the subsequent steps that are required to construct a library. We illustrate the protocols with data from a library we have recently constructed. We also describe the analyses that we routinely perform to assess the compositions of our libraries. Our target for randomization was the *Escherichia coli cynT* gene, which was identified in a previous experiment because of its contribution to antibiotic resistance when it was over-expressed (13). We conclude the chapter by comparing the *cynT* epPCR library with two other libraries and with the results that are predicted by the manufacturer of the GeneMorph kit (Agilent). This analysis highlights the stochastic nature of epPCR.

2 Materials

2.1 Error-Prone PCR

1. Plasmid containing the gene that is to be amplified by epPCR.
2. Spectrophotometer and cuvettes for measuring DNA concentration, e.g., an Eppendorf Biophotometer and UVettes.
3. Oligonucleotide primers for the epPCR amplification (see Note 1).

4. GeneMorph II Random Mutagenesis Kit (Agilent). The kit contains Mutazyme II DNA polymerase (2.5 U/μL), 10× Mutazyme II reaction buffer, and a dNTP mix (10 mM each dNTP).
5. Thermocycler with a heated lid.
6. Agarose gels, stained with ethidium bromide at 0.5 μg/mL.
7. DNA ladder with bands that contain known amounts of DNA.
8. Apparatus for agarose gel electrophoresis.
9. QiaQuick PCR Purification Kit (Qiagen). Equivalent kits from other manufacturers are also suitable.

2.2 Vector and Insert Preparation

1. Protein expression vector, into which the epPCR product will be cloned (see Note 2).
2. QiaPrep Spin Miniprep Kit (Qiagen). Equivalent kits from other manufacturers are also suitable.
3. Restriction enzyme(s) that facilitate directional, sticky-ended cloning (e.g., SfiI from New England Biolabs).
4. Restriction enzyme that cuts within the expression vector's stuffer fragment (see Note 3).
5. Restriction enzyme DpnI (New England Biolabs).
6. Agarose gels stained with 1× SYBR Safe DNA gel stain (Invitrogen).
7. Safe Imager 2.0 Blue-Light Transilluminator (Invitrogen).
8. Clean razor blades for excising bands from gels.
9. MinElute Gel Extraction Kit (Qiagen).

2.3 Preparation of a Test Library

1. T4 DNA ligase and ligation buffer. We obtain comparable results with the T4 DNA ligases from New England Biolabs, Fermentas, and Enzymatics Inc.
2. Aliquots (50 μL) of electrocompetent *E. coli* cells (see Note 4).
3. Gene Pulser electroporation cuvettes (BioRad, 0.2 cm electrode gap).
4. Gene Pulser electroporation unit with Pulse Controller (BioRad).
5. Sterile SOC medium: 20 g/L tryptone; 5 g/L yeast extract; 10 mM NaCl; 2.5 mM KCl; 20 mM glucose.
6. LB-agar plates containing the correct antibiotic for selecting plasmid-containing cells.

2.4 Analysis of Library Composition

1. Thermocycler with a heated lid (e.g., an MJ Mini from BioRad).
2. Primers for amplifying cloned inserts from the epPCR library (see Note 1).

3. Reagents for a standard PCR screen. While there are many alternate (and equally good) suppliers, we routinely use 5× Green GoTaq Reaction Buffer (Promega), i-Taq DNA polymerase (iNtRON Biotechnology), and the dNTP mix that is supplied with the polymerase (which contains 2.5 mM of each dNTP).
4. Ethidium bromide-stained agarose gels and electrophoresis apparatus.
5. QiaQuick PCR Purification Kit (Qiagen), or equivalent.

2.5 Construction and Storage of the Full-Sized Library

1. T4 DNA ligase and ligation buffer, as listed in Subheading 2.3.
2. QiaQuick PCR Purification Kit (Qiagen), or equivalent.
3. Standard-sized LB-agar plates (circular, 85–90 mm diameter) containing the correct antibiotic for selecting plasmid-containing cells.
4. Two square bioassay dishes (245 mm × 245 mm) from Corning or Nunc. Each bioassay dish holds 200 mL of LB agar, supplemented with the appropriate antibiotic for maintaining the library vector.
5. A fresh batch of electrocompetent *E. coli* cells (see Note 4).
6. Electroporation cuvettes and apparatus, as described in Subheading 2.3.
7. SOC medium, as described in Subheading 2.3.
8. Supercoiled pUC19 control plasmid (10 pg/μL; Invitrogen).
9. Sterile 50 mL tube (Falcon or similar).
10. LB medium supplemented with the appropriate antibiotic (~20 mL, total).
11. Refrigerated centrifuge with a rotor that takes the 50 mL tube listed above (item 9).
12. Spectrophotometer and cuvettes for measuring cell density (OD_{600}), e.g., an Eppendorf Biophotometer and UVettes.
13. Sterile glycerol (50% v/v).
14. Cryogenic vials, suitable for storage at −80°C.

3 Methods

3.1 Error-Prone PCR

1. Prepare purified plasmid DNA, containing the gene that is to be mutagenized. Measure its concentration spectrophotometrically.
2. The amount of template used in the epPCR affects the mutation rate (see Note 5). Calculate the amount of plasmid DNA that is required for the desired mutation rate (see Notes 6 and 7).

3. Prepare the epPCR reagents in a thin-walled, 0.2 mL tube:

x μL	Plasmid DNA template (see Note 6)
y μL	Water to a total volume of 50 μL
5 μL	10× Mutazyme II reaction buffer
1 μL	dNTP mix (gives 200 μM of each dNTP, final concentration)
2 μl	Forward primer (from 10 μM stock solution)
2 μL	Reverse primer (from 10 μM stock solution)
1 μL	Mutazyme II DNA polymerase (2.5 U)

4. Mix the sample and place the tube in the thermocycler.
5. Run the epPCR program:

Step 1:	1 min	95°C
Step 2:	20 s	94°C
Step 3:	20 s	Annealing temperature for primers (see Note 1)
Step 4:	1 min	72°C (for a ~1 kb gene; see Note 8)
Step 5:	Repeat steps 2–4 for an additional 29 cycles	
Step 6:	2 min	72°C
Step 7:	Hold	4°C (for product storage, if necessary)

6. Run 2 μL of the product on an ethidium bromide-stained agarose gel, alongside a DNA ladder. Determine the total yield of the epPCR product by comparing the intensity of the epPCR sample with the intensity of the bands in the ladder. The total yield of epPCR product is required to calculate the PCR efficiency (see Note 9).
7. Purify the remainder of the epPCR product using the PCR Purification Kit. Elute the purified DNA from the spin column in 30 μL elution buffer (EB).
8. Estimate the concentration of the purified sample by running 1 μL on an agarose gel, alongside a suitable DNA ladder.

3.2 Vector and Insert Preparation

1. Prepare 3–4 μg of the plasmid that will be used for cloning and expression of the epPCR library (see Note 2). In the example discussed below, we used vector pCA24N (14), which was purified from a saturated overnight culture using the QiaPrep Spin Miniprep Kit (Qiagen).

2. Digest the vector (~3 μg) and epPCR product (~1 μg) to completion, with restriction enzyme(s) that introduce sticky ends and facilitate directional cloning. We typically digest pCA24N and the epPCR product with 20 U of SfiI, in total reaction volumes of 30 μL. Under these conditions, incubating for 5 h at the enzyme's optimal temperature (50°C) is generally sufficient for complete digestion.
3. For additional improvements in the quality of the final library, add fresh restriction enzymes, as follows:
   (a) Vector preparation—add 10 U of an enzyme that cuts within the stuffer fragment (see Note 3).
   (b) Insert preparation—add 10 U of DpnI, to eliminate any of the methylated, unmutated template that may have carried over from the epPCR.

   Incubate the reactions for a further 2 h at 37°C, then heat inactivate the enzymes (where possible), according to the manufacturer's guidelines.
4. Run the two reactions on separate agarose gels, with agarose concentrations that are appropriate for the fragments being resolved (e.g., 0.8% agarose for the vector and 1.2% agarose for the insert).
5. Excise the bands that correspond to the digested vector and insert. We strongly recommend the use of a blue-light transilluminator and a compatible stain (SYBR Safe), rather than ethidium bromide and a UV transilluminator, for this step. See Note 10.
6. Purify the vector and insert DNA from the excised gel bands. We use the MinElute Gel Extraction Kit (Qiagen) and elute the DNA from each spin column in 12 μL EB.
7. Determine the concentrations of the purified vector and insert DNA by running 2 μL aliquots of each on an ethidium bromide-stained agarose gel, as described above (Subheading 3.1, step 6).
8. Store the purified DNA at –20°C, as necessary.

3.3 Preparation of a Test Library (See Note 11)

1. Prepare two ligation reactions: one with vector DNA only and one with the vector and a threefold molar excess of the insert DNA. Each reaction should contain the following: 1× ligation buffer; 50 ng of vector DNA (see Subheading 3.2, step 6); T4 DNA ligase (1 U); plus or minus the insert DNA; and water to a final volume of 10 μL. Add the T4 DNA ligase last and mix gently.
2. Incubate the ligation reactions at 16°C for 16 h.
3. Use a 1 μL aliquot of each ligation reaction to transform 50 μL aliquots of *E. coli*, by electroporation. See Note 12.
4. Immediately after electroporation, add 500 μL of SOC medium to the cuvette and transfer the cells to a sterilized, capped test

tube or a 15 mL tube (Falcon or similar). Allow the cells to recover by incubating them at 37°C, with shaking, for 1 h.

5. Store the remaining 9 μL of each ligation reaction (leftover from step 3, above) at −20°C.
6. Spread aliquots (10 and 50 μL) of the two recovery cultures on LB-agar plates. Incubate the plates at 37°C for 12–16 h.
7. Count the number of colonies on each plate. Use the results from the "vector only" plates to calculate the fraction of the library (as represented on the "vector + insert" plates) that contains recircularized vector. This background must be minimized, to avoid wasting time on futile library screens. If the "vector only" background is >1% of the total library, we recommend preparing a fresh batch of the vector (Subheading 3.2, above), and lengthening the incubation time with each restriction enzyme.
8. The number of colonies on the "vector + insert" plates also allows the size of the final, scaled-up library to be estimated. The final library is likely to be ~10^3 times larger than the total number of colonies on the "vector + insert" test plates (see Note 13).

3.4 Analysis of Library Composition (See Note 14)

1. Use 2 μL pipette tips (or sterile toothpicks) to pick 10–20 colonies at random from the "vector + insert" test plates (Subheading 3.3, step 7).
2. Transfer each colony into a thin-walled, 0.2 mL tube containing 5 μL of sterile water.
3. Lyse the cells by incubating the tubes at 95°C for 5 min, in a thermocycler.
4. Amplify the randomly mutagenized gene inserts from each colony by PCR. We have listed our routine protocol, for guidance. However, many variations are possible; the goal here is merely to generate enough of the amplified product for DNA sequencing. We typically set up 25 μL PCRs in thin-walled 0.2 mL tubes, as follows:

14.75 μL	Water
5 μL	5× Green GoTaq buffer
2 μL	dNTP mix (gives 200 μM of each dNTP, final concentration)
1 μl	Forward primer (from 10 μM stock solution)
1 μL	Reverse primer (from 10 μM stock solution)
0.25 μL	*Taq* DNA polymerase (1.25 U)
1 μL	Cell lysate (from step 2, above)

Table 1
Matrix of point mutations identified in 16 *cynT* variants

		Mutation To			
		T	**C**	**A**	**G**
Mutation From	**T**	–	15	16	6
	C	10	–	8	0
	A	18	3	–	13
	G	10	4	18	–

4. Mix each sample and place the tubes in the thermocycler.
5. Run an appropriate PCR program, such as the one listed in Subheading 3.1, step 5.
6. Run a 2 μL aliquot of each PCR product on an agarose gel, to confirm successful amplification.
7. Purify the remainder of each PCR product using the QiaQuick PCR Purification Kit (or equivalent). Elute the purified DNA from each spin column in 30 μL EB.
8. Sequence each PCR product. Use the forward and/or reverse primers from the PCR as the sequencing primer(s), as necessary.
9. Align the sequence of each PCR product with the known sequence of the unmutated parental gene. Computer programs such as MacVector are useful for this analysis.
10. Tabulate all of the point mutations in the sequenced samples. Also note any insertions or deletions that may have arisen during the epPCR. The point mutations should be grouped by type. For example, we randomized the *E. coli cynT* gene and sequenced 16 clones from the resulting test library. The 121 point mutations that we identified in the 16 variants are summarized in Table 1. There were also two deletions and one insertion in the data set; in total, the sequencing revealed 124 mutations.
11. Use the tabulated data to calculate the overall mutation rate and to assess biases in the mutation spectrum of the epPCR library. There are three key indicators of bias (see Note 15): (a) the ratio of transition (Ts) to transversion (Tv) mutations; (b) the ratio of AT→GC transitions to GC→AT transitions; and (c) the frequency of mutations at A:T base pairs, to mutations at G:C base pairs. The mutation rate and bias measures for our *cynT* epPCR library are shown in Table 2.

Table 2
Mutational spectrum of the *cynT* epPCR library

Type(s) of mutations	Frequency	Proportion of total
Transitions		
A→G, T→C	28	22.6%
G→A, C→T	28	22.6%
Transversions		
A→T, T→A	34	27.4%
A→C, T→G	9	7.3%
G→C, C→G	4	3.2%
G→T, C→A	18	14.5%
Insertions and deletions		
Insertions	1	0.8%
Deletions	2	1.6%
Summary of bias		
Transitions/transversions	0.86	NA[a]
AT→GC/GC→AT	1	NA[a]
A→N, T→N	71	57.3%
G→N, C→N	50	40.3%
Mutation rate		
Mutations per kb	11.8	NA[a]
Mutations per *cynT* gene[b]	7.8	NA[a]

[a]NA: not applicable
[b]The cloned *cynT* insert was 657 bp

12. The library analysis program PEDEL-AA (15), available online at http://guinevere.otago.ac.nz/stats.html, should now be used to predict the utility of the final epPCR library. PEDEL-AA has an easy-to-use web interface and takes the following parameters as its inputs:
    (a) The sequence of the gene that was randomized
    (b) The estimated size of the scaled-up library (Subheading 3.3, step 8)
    (c) The nucleotide mutation matrix (Subheading 3.4, step 10; see Table 1 for an example)
    (d) The mean number of mutations per gene in the library (Subheading 3.4, step 11; see Table 2 for an example)

Table 3
PEDEL-AA outputs for the *cynT* epPCR library

Property	Estimate
Total library size	1.4×10^7
Number of variants with no insertions, deletions, or stop codons	9.0×10^6
Mean number of amino acid substitutions per variant	5.5
Unmutated (wild-type) sequences (% of total library)	3.0%
Number of distinct, full-length proteins in the library[a]	7.4×10^6

[a]Calculated using the PCR efficiency parameter. PEDEL-AA also calculates a less accurate estimate the number of distinct, full-length proteins in the library by using the simplifying assumption of Poisson statistics

(e) The number of cycles in the epPCR (Subheading 3.1, step 3)

(f) The PCR efficiency parameter for the epPCR (see Note 9)

(g) The mean number of insertions per gene in the library (Subheading 3.4, step 10)

(h) The mean number of deletions per gene in the library (Subheading 3.4, step 10)

The program outputs a variety of statistics about the protein variants that are encoded by the epPCR library. A selection of these statistics, calculated for our *cynT* library, is shown in Table 3. Together, the data in Tables 2 and 3 allow an informed decision to be made about whether to scale up the library (or whether to start over, with different epPCR and ligation conditions).

3.5 Construction and Storage of the Full-Sized Library

1. Prepare "vector only" and "vector + insert" ligation reactions that are tenfold larger than those described in Subheading 3.3, step 1 (see Note 16). Each reaction should contain the following: 1× ligation buffer; 500 ng of vector DNA; T4 DNA ligase (10 U); plus or minus insert DNA (3-fold molar excess over vector); and water to a final volume of 100 μL.
2. Incubate the ligation reactions at 16°C for 16 h.
3. Add the remaining 9 μL of each test ligation (Subheading 3.3, step 5) to the scaled-up "vector only" and "vector + insert" ligation reactions.
4. Purify the products from each ligation reaction using the QiaQuick PCR Purification Kit (or equivalent). Elute the purified DNA from each spin column in 42 μL EB.
5. Prepare the LB-agar plates on which the transformed cells of the library will be spread. For each library, we typically use two 245 mm × 245 mm square bioassay dishes (see Note 17).

6. Prepare a fresh batch of electrocompetent *E. coli* cells (see Note 4). In our hands, the transformation efficiencies of freshly prepared cells are four to fivefold higher than cells that have undergone a freeze/thaw cycle.
7. Add 3 μL aliquots of the "vector + insert" ligation to 14 × 50 μL aliquots of electrocompetent *E. coli* cells. Transform each aliquot, and recover the transformed cells, as described previously (Subheading 3.3, steps 3 and 4).
8. Transform a single 50 μL aliquot of cells with 3 μL of the "vector only" ligation.
9. Transform one more 50 μL aliquot of the electrocompetent *E. coli* with an appropriate plasmid for determining the transformation efficiency of the cells. We routinely use 10 pg of supercoiled pUC19.
10. Pool all of the cells that were transformed with the "vector + insert" ligation, in a sterile 15 mL tube. The total volume should be 7.7 mL (14 electroporations; 550 μL per recovery culture).
11. Mix the cells briefly, by inverting the tube 2–3 times.
12. Spread 1, 5, and 25 μL aliquots on regular LB-agar plates (diluting as necessary to obtain a spreadable volume).
13. Spread the remainder of the library on the two large plates (see step 5, above); ~3.85 mL per plate.
14. Spread 1, 5, and 25 μL aliquots of the "vector only" control on regular LB-agar plates.
15. Spread suitable aliquots (typically 2 and 10 μL) of the cells transformed with the pUC19 control on LB-agar plates that contain ampicillin (100 μg/mL).
16. Incubate all of the dilution and control plates at 37°C for 16 h. Incubate the two large library plates at 30°C, to avoid the formation of a confluent lawn.
17. Count the number of colonies on each plate, except for the large library plates (which should be covered in dense lawns of small colonies).
18. Use the pUC19 control to calculate the transformation efficiency of your electrocompetent cells. The easiest way to increase the size of an epPCR library is to improve the transformation efficiency of the cells. A good batch of *E. coli* cells should yield $>10^9$ colonies per microgram of pUC19 used in the transformation.
19. The regular LB-agar plates with aliquots of the library and the "vector only" control should be used to estimate the final size of the library, and to verify that the "vector only" background is <1% of the total library (see Subheading 3.3, step 7).

20. The large library plates should each be covered in thousands (or millions) of small colonies. To recover the library from one of the plates, pipette 4 mL of LB medium (supplemented with the appropriate antibiotic) into the center of the plate.
21. Use a glass spreader to scrape cells off the surface of the plate; pool them in one corner of the plate.
22. Remove the resuspended cells with a P1000 pipettor, and transfer them to a sterile 50 mL tube.
23. Pipette a second 4 mL aliquot of LB onto the same plate, and repeat the scraping step.
24. Use another 2 × 4 mL aliquots of LB to recover the cells from the second library plate.
25. All of the recovered cells should be pooled in the same 50 mL tube.
26. Break up cell clumps and mix well, by pipetting up and down repeatedly with a P1000 pipettor.
27. Pellet the cells by centrifugation at 3,000 × *g*, 4°C, for 15 min. Use a P1000 pipettor to remove the supernatant.
28. Resuspend the cell pellet in 1 mL of LB medium, plus antibiotic. The pellet is likely to be large; resuspension is likely to involve vigorous pipetting and/or gentle vortexing.
29. Knowing the cell density is likely to be useful for planning downstream screening/selection experiments. Mix 1 μL of the cell suspension with 999 μL of sterile water, and measure the OD_{600} (against a water blank). The final OD_{600} of the resuspended library is usually >100, corresponding to a cell density of $>2.5 \times 10^{10}$ cells per milliliter (see Note 18).
30. Split the library into 100 μL aliquots and transfer each aliquot to a cryogenic vial.
31. Add 50 μL of sterile glycerol (50% v/v) to each aliquot and mix well by pipetting.
32. Store the aliquots at −80°C until you are ready to proceed with screening/selection to identify improved variants in the library.

3.6 Summarizing Stochasticity in epPCR Library Construction

1. Tables 1 and 2 show that the *cynT* epPCR library, constructed with the GeneMorph II kit, is not free of mutational bias. For example, G→C and C→G mutations occur much less frequently than A→T and T→A mutations (Table 2)—indicating that the Mutazyme II polymerase retains some of the bias of *Taq* polymerase. For reference, we have included a side-by-side comparison of the *cynT* data with two other epPCR libraries that were constructed using the same protocol, together with the guidelines from the manufacturer (Table 4). All three randomized genes (*cynT*, *ydfW*, and *yeaD*) are latent contributors to antibiotic resistance (13).

Table 4
Comparison of three epPCR libraries with Agilent's product guidelines

	cynT	*ydfW*	*yeaD*	[Agilent][a]
epPCR details				
Gene length	657 bp	147 bp	882 bp	NA[b]
Plasmid in the epPCR	90 ng	90 ng	90 ng	NA[b]
Amount of target DNA[c]	15 ng	7 ng	18 ng	0.1–1,000 ng
Library bias indicators				
Transitions/transversions	0.86	0.56	0.56	0.90
AT→GC/GC→AT	1	1.50	1.17	0.60
A→N, T→N	57.3%	71.4%	61.1%	50.7%
G→N, C→N	40.3%	28.6%	38.9%	43.8%
Mutation rates				
Mutations per kb	11.8	9.5	4.1	0–16
Mutations per gene	7.8	1.4	3.6	NA[b]

[a]Source: Tables 1 and 2 of the GeneMorph II Random Mutagenesis Kit manual, available for download from http://tinyurl.com/3mb6x66
[b]NA: not applicable
[c]Calculated as described in Note 6

2. The data in Table 15.4 highlight the stochasticity that is inherent in epPCR. No two libraries are identical. While the GeneMorph II kit introduces less mutational bias than other epPCR methods, variation in mutational spectra is still to be expected. Mutation rates are also variable, and the recommendations given here (see Note 6) should be considered a rough guide only.
3. Bearing these facts in mind, an epPCR practitioner should be prepared to construct several libraries, combining the analyses that we have described here with considerable trial and error!

4 Notes

1. Primers should be noncomplementary and should have melting temperatures that are within 5°C of each other. Melting temperatures can be estimated accurately using the OligoAnalyzer tool from Integrated DNA Technologies: http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/. The optimal annealing temperature to use in a PCR is typically 3–5°C cooler than the lowest primer melting

temperature. We routinely resuspend lyophilized primers in TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0), to a concentration of 100 μM. Working stocks (10 μM) are made by tenfold dilution of these master stocks, using sterile water.

2. A high-quality vector preparation is critical for constructing a large library. We find it useful to use a plasmid with a stuffer fragment in the cloning cassette. Excision of this stuffer fragment allows the progress of the restriction digestion to be monitored. It also ensures that the doubly digested vector (with stuffer fragment removed) can be resolved from undigested and singly digested material on an agarose gel.
3. Using a restriction enzyme that cuts within the stuffer fragment minimizes the number of "vector only" clones in the final library. We routinely use BglII (New England Biolabs) for this purpose.
4. The choice of *E. coli* strain will depend on the downstream selection or screen that is being employed. In general, the final size of the epPCR library is directly proportional to the transformation efficiency of the host strain. Therefore, strains with high transformation efficiencies (such as *E. coli* DH5α-E) are preferable. We prepare electrocompetent cells according to the method of Hanahan (16).
5. The overall mutation frequency depends on the error rate of the polymerase and also the number of times that each template is duplicated in the reaction. If the initial amount of template is high, it will undergo few duplications in the epPCR. On the other hand, a low amount of template will result in a greater number of duplications, and more mutations will be introduced. This is discussed further in the GeneMorph II Random Mutagenesis Kit manual, available for download from http://tinyurl.com/3mb6x66.
6. Agilent recommends 500–1,000 ng of template DNA for a low mutation rate (0–4.5 mutations/kb); 100–500 ng of template for a medium mutation rate (4.5–9 mutations/kb), and 0.1–100 ng of template for a high mutation rate (9–16 mutations/kb). Note that the amount of template is not the amount of purified plasmid DNA. Instead, it is the amount of target DNA to be amplified. For example, we used 90 ng of plasmid pCA24N-*cynT* as the starting point for one of our epPCR libraries. In total, this plasmid was 5,180 bp in size. However, the amplified product (i.e., the *cynT* gene, plus flanking sequences) was 857 bp. Therefore, the amount of template DNA in the reaction was given by

$$\left(\frac{857\text{ bp}}{5{,}180\text{ bp}}\right)\times 90\text{ ng} = 15\text{ng}$$

7. We typically aim for a medium-to-high mutation rate, because this generates libraries that contain minimal numbers of "wasted" variants (i.e., unmutated copies of the template, or multiple copies of variants with any one point mutation). This strategy is discussed in more detail elsewhere (12, 15).
8. We routinely use extension times that are calculated at a rate of 1 min/kb. For example, a 30 s extension time is used for a 500 bp product, and a 90 s extension time is used for a 1,500 bp product.
9. Calculating the PCR efficiency parameter allows robust statistical analysis of library composition (*see* Subheading 3.4, step 12). When the total product yield and the amount of starting template are known, the number of doublings in the PCR, d, can be calculated as follows:

$$d = \frac{\log(\text{Product} / \text{Template})}{\log 2}$$

The PCR efficiency (i.e., the probability that any particular sequence is duplicated in any one cycle of the PCR, *eff*) is then given by

$$eff = 2^{(d/n)} - 1$$

where n is the number of PCR cycles ($n = 30$ in our protocol). An online tool for calculating *eff*, given d and n, can be found at http://guinevere.otago.ac.nz/cgi-bin/aef/PCReff.pl.

10. It is well known that UV transillumination of ethidium bromide-stained DNA can induce damage, resulting in lower cloning and transformation efficiencies (17). Even short exposures to UV (<60 s) can have dramatic and deleterious effects. Constructing a large epPCR library (>10^6 variants) requires the highest possible quality of DNA. Therefore, we use SYBR Safe stain and a blue-light transilluminator for preparation of our library vector and epPCR insert. In our hands, this results in libraries that are 5–10 times larger than equivalent libraries prepared with ethidium bromide-stained DNA.
11. Before scaling up to a full-sized library, we find it useful and expedient to construct a test library. This allows the epPCR mutation spectrum to be determined. It also ensures that the ligation protocol is optimized for constructing a full-sized library with low "vector only" background and the maximum number of insert-containing clones.
12. Aliquots of cells should be thawed on ice. DNA is added to each 50 µl aliquot of cells and chilled on ice in a sterile Gene Pulser cuvette. Samples are electroporated at 2.5 kV, 200 Ω, and 25 µF in a Gene Pulser unit with Pulse Controller.

13. This is a rough estimate, based on the following: (a) Tenfold scale-up of the ligation reaction, (b) transforming 10–20 aliquots of electrocompetent cells, and (c) spreading all 550 μL of each recovery culture (instead of 10–50 μL aliquots).
14. Clones from the test library (Subheading 3.3) should be sequenced in order to analyze the mutation rate and the spectrum of mutations that arose in the epPCR. As discussed in Notes 5 and 6, some control over the mutation rate is possible. However, in our experience, there is considerable experiment-to-experiment variation in the outcomes of the epPCR process (see Subheading 3.6). Therefore, we recommend conducting the analyses described in Subheading 3.4, to avoid wasting time and resources on a scaled-up library that contains little molecular diversity.
15. A library with an unbiased spectrum of mutations will be maximally diverse; that is, it will have the lowest probability of duplicated variants. Therefore, it is more likely to contain at least one improved variant (12). One indicator of bias is the ratio of transitions (i.e., purine-to-purine and pyrimidine-to-pyrimidine mutations) to transversions (purine-to-pyrimidine and pyrimidine-to-purine mutations). There are four possible transitions and eight possible transversions (listed in Table 15.2). Therefore, a completely unbiased error-prone polymerase would generate libraries with transition/transversion (Ts/Tv) ratios of 0.5. Provided that the GC content of the gene is ~50%, the ratio of AT → GC transitions to GC → AT transitions (i.e., AT → GC/GC → AT) in an unbiased epPCR library should also be 1. Similarly, the number of mutations at A:T base pairs (A → N, T → N) should also be the same as the number of mutations at G:C base pairs (G → N, C → N). The effects of mutational bias on overall library composition can be assessed by altering the input parameters for PEDEL-AA analysis (Subheading 3.4, step 12).
16. When constructing the full-sized library, the focus should be on scaling everything up by as much as possible. All of the remaining epPCR insert (Subheading 3.2, step 7) should be used in a scaled-up ligation, and as many aliquots of electrocompetent *E. coli* as possible should be transformed with the ligated products. The protocol that we describe is a typical example from our laboratory.
17. The large volume and high surface area of the square bioassay dishes make them prone to "sweating" when they are incubated at 30–37°C (particularly if the LB agar is too hot when the plates are poured). They may need to be pre-warmed at 37°C for 4–6 h and/or dried in a laminar flow hood or a class II biosafety cabinet (10–15 min), before they are dry enough to use.
18. For *E. coli* strain DH5α-E, we find that $OD_{600} = 1$ corresponds to ~2.5×10^8 cells/mL.

Acknowledgment

The authors gratefully acknowledge financial support for this work from the New Zealand Marsden Fund.

References

1. Turner NJ (2009) Directed evolution drives the next generation of biocatalysts. Nat Chem Biol 5:567–573
2. Jäckel C, Hilvert D (2010) Biocatalysts by evolution. Curr Opin Biotechnol 21:753–759
3. Peisajovich SG, Tawfik DS (2007) Protein engineers turned evolutionists. Nat Methods 4:991–994
4. Lutz S, Patrick WM (2004) Novel methods for directed evolution of enzymes: quality, not quantity. Curr Opin Biotechnol 15:291–297
5. Otten LG, Quax WJ (2005) Directed evolution: selecting today's biocatalysts. Biomol Eng 22:1–9
6. Arnold FH, Georgiou G (eds) (2003) Directed enzyme evolution: screening and selection methods. *Methods in Molecular Biology*, vol 230, Humana Press, Totowa, New Jersey.
7. Leung DW, Chen E, Goeddel DV (1989) A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. Technique 1:11–15
8. Cadwell RC, Joyce GF (1992) Randomization of genes by PCR mutagenesis. PCR Methods Appl 2:28–33
9. Cirino PC, Mayer KM, Umeno D (2003) Generating mutant libraries using error-prone PCR. Methods Mol Biol 231:3–9
10. McCullum EO, Williams BA, Zhang J, Chaput JC (2010) Random mutagenesis by error-prone PCR. Methods Mol Biol 634:103–109
11. Shafikhani S, Siegel RA, Ferrari E, Schellenberger V (1997) Generation of large libraries of random mutants in *Bacillus subtilis* by PCR-based plasmid multimerization. Biotechniques 23:304–310
12. Patrick WM, Firth AE, Blackburn JM (2003) User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. Protein Eng 16:451–457
13. Soo VWC, Hanson-Manful P, Patrick WM (2011) Artificial gene amplification reveals an abundance of promiscuous resistance determinants in *Escherichia coli*. Proc Natl Acad Sci USA 108:1484–1489
14. Kitagawa M, Ara T, Arifuzzaman M, Ioka-Nakamichi T, Inamoto E, Toyonaga H, Mori H (2005) Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. DNA Res 12:291–299
15. Firth AE, Patrick WM (2008) GLUE-IT and PEDEL-AA: new programmes for analyzing protein diversity in randomized libraries. Nucleic Acids Res 36:W281–W285
16. Hanahan D, Jessee J, Bloom FR (1991) Plasmid transformation of *Escherichia coli* and other bacteria. Methods Enzymol 204:63–113
17. Hartman PS (1991) Transillumination can profoundly reduce transformation frequencies. Biotechniques 11:747–748

[Blank page]

REFERENCES

Agrawal RK, Heagle AB, Penczek P, Grassucci RA and Frank J (1999) EF-G-dependent GTP hydrolysis induces translocation accompanied by large conformational changes in the 70S ribosome. *Nature Structural Biology*. 6 (7), 643–647.

Agrawal RK, Penczek P, Grassucci RA and Frank J (1998) Visualization of elongation factor G on the *Escherichia coli* 70S ribosome: the mechanism of translocation. *Proceedings of the National Academy of Sciences of the United States of America*. 95 (11), 6134–6138.

Agrawal RK, Penczek P, Grassucci RA and Frank J (1998) Visualization of elongation factor G on the Escherichia coli 70S ribosome: The mechanism of translocation. *Proceedings of the National Academy of Sciences of the United States of America*. 95 (11), 6134–6138.

Aharoni A, Gaidukov L, Khersonsky O, McQ Gould S, Roodveldt C and Tawfik DS (2005) The "evolvability" of promiscuous protein functions. *Nature Genetics*. 37 (1), 73–76.

Alekshun MN and Levy SB (2007) Molecular mechanisms of antibacterial multidrug resistance. *Cell*. 128 (6), 1037–1050.

Allwood AC, Walter MR, Kamber BS, Marshall CP and Burch IW (2006) Stromatolite reef from the Early Archaean era of Australia. *Nature*. 441 (7094), 714–718.

Andersson DI (2003) Persistence of antibiotic resistant bacteria. *Current Opinion in Microbiology*. 6 (5), 452–456.

Andersson DI and Levin BR (1999) The biological cost of antibiotic resistance. *Current Opinion in Microbiology*. 2 (5), 489–493.

Andrews JM (2001) Determination of minimum inhibitory concentrations. *Journal of Antimicrobial Chemotherapy*. 48 (suppl 1), 5–16.

Arnold FH (1998) When blind is better: protein design by evolution. *Nature Biotechnology*. 16 (7), 617–618.

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko Kirill A, Tomita M, Wanner Barry L and Mori H (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*. 2, 2006.0008.

Babtie A, Tokuriki N and Hollfelder F (2010) What makes an enzyme promiscuous? *Current Opinion in Chemical Biology*. 14 (2), 200–207.

Baquero F, Negri MC, Morosini MI and Blázquez J (1997) The antibiotic selective process: concentration-specific amplification of low-level resistant populations. *Ciba Foundation Symposium*. 207, 93–105; discussion 105–111.

Baracchini E, Glass R and Bremer H (1988) Studies *in vivo* on *Escherichia coli* RNA polymerase mutants altered in the stringent response. *Molecular & General Genetics*. 213 (2-3), 379–387.

Barlow M and Hall BG (2002) Predicting Evolutionary Potential: *In Vitro* Evolution Accurately Reproduces Natural Evolution of the TEM {beta}-Lactamase. *Genetics*. 160 (3), 823–832.

Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE and Kim JF (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature*. 461 (7268), 1243–1247.

Becker B and Cooper MA (2013) Aminoglycoside antibiotics in the 21st century. *ACS Chemical Biology*. 8 (1), 105–115.

Bergthorsson U, Andersson DI and Roth JR (2007) Ohno's dilemma: evolution of new genes under continuous selection. *Proceedings of the National Academy of Sciences of the United States of America*. 104 (43), 17004–17009.

Bershtein S and Tawfik DS (2008) Advances in laboratory evolution of enzymes. *Current Opinion in Chemical Biology*. 12 (2), 151–158.

Bielecki P, Lukat P, Hüsecken K, Dötsch A, Steinmetz H, Hartmann RW, Müller R and Häussler S (2012) Mutation in elongation factor G confers resistance to the antibiotic argyrin in the opportunistic pathogen *Pseudomonas aeruginosa*. *ChemBioChem*. 13 (16), 2339–2345.

Bisht R, Katiyar A, Singh R and Mittal P (2009) Antibiotic resistance – A global issue of concern. *Asian Journal of Pharmaceutical and Clinical Research*. 2 (2), 34–39.

Björkman J, Nagaev I, Berg OG, Hughes D and Andersson DI (2000) Effects of environment on compensatory mutations to ameliorate costs of antibiotic resistance. *Science*. 287 (5457), 1479–1482.

Björkman J and Andersson DI (2000) The cost of antibiotic resistance from a bacterial perspective. *Drug Resistance Updates*. 3 (4), 237–245.

Blondeau JM (2004) Fluoroquinolones: mechanism of action, classification, and development of resistance. *Survey of Ophthalmology*. 49 (Suppl 2), S73–S78.

Bloom JD, Labthavikul ST, Otey CR and Arnold FH (2006) Protein stability promotes evolvability. *Proceedings of the National Academy of Sciences of the United States of America*. 103 (15), 5869–5874.

Bochner BR, Gadzinski P and Panomitros E (2001) Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Research*. 11 (7), 1246–1255.

Bodley JW, Zieve FJ, Lin L and Zieve ST (1969) Formation of the ribosome-G factor-GDP complex in the presence of fusidic acid. *Biochemical and Biophysical Research Communications*. 37 (3), 437–443.

Bonner CA, Randall SK, Rayssiguier C, Radman M, Eritja R, Kaplan BE, McEntee K and Goodman MF (1988) Purification and characterization of an inducible *Escherichia coli* DNA polymerase capable of insertion and bypass at abasic lesions in DNA. *The Journal of Biological Chemistry*. 263 (35), 18946–18952.

Borovinskaya MA, Pai RD, Zhang W, Schuwirth BS, Holton JM, Hirokawa G, Kaji H, Kaji A and Cate JHD (2007) Structural basis for aminoglycoside inhibition of bacterial ribosome recycling. *Nature Structural & Molecular Biology*. 14 (8), 727–732.

Brustad E, Bushey ML, Brock A, Chittuluru J and Schultz PG (2008) A promiscuous aminoacyl-tRNA synthetase that incorporates cysteine, methionine, and alanine homologs into proteins. *Bioorganic & Medicinal Chemistry Letters*. 18 (22), 6004–6006.

Bryan LE and Kwan S (1983) Roles of ribosomal binding, membrane potential, and electron transport in bacterial uptake of streptomycin and gentamicin. *Antimicrobial Agents and Chemotherapy*. 23 (6), 835–845.

Cadwell RC and Joyce GF (1992) Randomization of genes by PCR mutagenesis. *Genome Research*. 2 (1), 28–33.

Cartron ML, Maddocks S, Gillingham P, Craven CJ and Andrews Simon C (2006) Feo-transport of ferrous iron into bacteria. *BioMetals*. 19 (2), 143–157.

Catrina I, O'Brien PJ, Purcell J, Nikolic-Hughes I, Zalatan JG, Hengge AC and Herschlag D (2007) Probing the origin of the compromised catalysis of *E. coli* alkaline phosphatase in its promiscuous sulfatase reaction. *Journal of the American Chemical Society*. 129 (17), 5760–5765.

Chakiath CS and Esposito D (2007) Improved recombinational stability of lentiviral expression vectors using reduced-genome *Escherichia coli*. *BioTechniques*. 43 (4), 466-470.

Chittori S, Simanshu DK, Savithri HS and Murthy MRN (2007) Structure of the putative mutarotase YeaD from *Salmonella typhimurium*: structural comparison with galactose mutarotases. *Acta Crystallographica. Section D, Biological Crystallography*. 63 (Pt 2), 197–205.

Choi KH, Gaynor JB, White KG, Lopez C, Bosio CM, Karkhoff-Schweizer RR and Schweizer HP (2005) A Tn7-based broad-range bacterial cloning and expression system. *Nature Methods*. 2 (6), 443–448.

Cizman M (2003) The use and resistance to antibiotics in the community. *International Journal of Antimicrobial Agents*. 21 (4), 297–307.

Clardy J, Fischbach MA and Walsh CT (2006) New antibiotics from bacterial natural products. *Nature Biotechnology*. 24 (12), 1541–1550.

Cohen SP, Hächler H and Levy SB (1993) Genetic and functional analysis of the multiple antibiotic resistance (*mar*) locus in *Escherichia coli*. *Journal of Bacteriology*. 175 (5), 1484–1492.

Cole FE, Kalyanpur MG and Stevens CM (1973) Absolute configuration of α-isopropylmalate and the mechanism of its conversion to β-isopropylmalate in the biosynthesis of leucine. *Biochemistry*. 12 (17), 3346–3350.

Colombo ML and Bosisio E (1996) Pharmacological activities of *Chelidonium majus* L. (Papaveraceae). *Pharmacological Research*. 33 (2), 127–134.

Cooper MA and Shlaes D (2011) Fix the antibiotics pipeline. *Nature*. 472 (7341), 32.

Costantino N and Court DL (2003) Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proceedings of the National Academy of Sciences of the United States of America*. 100 (26), 15748–15753.

Counts GW, Gregory D, Zeleznik D and Turch M (1977) Cefazaflur, a new parenteral cephalosporin: *in vitro* studies. *Antimicrobial Agents and Chemotherapy*. 11 (4), 708–711.

Csörgo B, Fehér T, Tímár E, Blattner FR and Pósfai G (2012) Low-mutation-rate, reduced-genome *Escherichia coli*: an improved host for faithful maintenance of engineered genetic constructs. *Microbial Cell Factories*. 11, 11.

D'Anna F, La Marca S and Noto R (2008) Kemp elimination: a probe reaction to study ionic liquids properties. *The Journal of Organic Chemistry*. 73 (9), 3397–403.

D'Costa VM, King CE, Kalan L, Morar M, Sung WWL, Schwarz C, Froese D, Zazula G, Calmels F, Debruyne R, Golding GB, Poinar HN and Wright GD (2011) Antibiotic resistance is ancient. *Nature*. 477 (7365), 457–461.

D'Costa VM, McGrann KM, Hughes DW and Wright GD (2006) Sampling the antibiotic resistome. *Science*. 311 (5759), 374–377.

Daigle DM and Brown ED (2004) Studies of the interaction of Escherichia coli YjeQ with the ribosome *in vitro*. *Journal of Bacteriology*. 186 (5), 1381–1387.

Damper PD and Epstein W (1981) Role of the membrane potential in bacterial resistance to aminoglycoside antibiotics. *Antimicrobial Agents and Chemotherapy*. 20 (6), 803–808.

Dantas G, Sommer MOA, Oluwasegun RD and Church GM (2008) Bacteria subsisting on antibiotics. *Science*. 320 (5872), 100–103.

da Silva NA and Bailey JE (1986) Theoretical growth yield estimates for recombinant cells. *Biotechnology and Bioengineering*. 28 (5), 741–746.

Datsenko KA and Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America*. 97 (12), 6640–6645.

Davies J (2007) Microbes have the last word. A drastic re-evaluation of antimicrobial treatment is needed to overcome the threat of antibiotic-resistant bacteria. *The European Molecular Biology Organisation Reports*. 8 (7), 616–621.

Davies J and Davies D (2010) Origins and evolution of antibiotic resistance. *Microbiology and Molecular Biology Reviews*. 74 (3), 417–433.

Dettman JR, Rodrigue N, Melnyk AH, Wong A, Bailey SF and Kassen R (2012) Evolutionary insight from whole-genome sequencing of experimentally evolved microbes. *Molecular Ecology*. 21 (9), 2058–2077.

Drevland RM, Jia Y, Palmer DRJ and Graham DE (2008) Methanogen homoaconitase catalyzes both hydrolyase reactions in coenzyme B biosynthesis. *The Journal of Biological Chemistry*. 283 (43), 28888–28896.

Edgar R and Bibi E (1997) MdfA, an *Escherichia coli* multidrug resistance protein with an extraordinarily broad spectrum of drug recognition. *Journal of Bacteriology*. 179 (7), 2274–2280.

Ellis HM, Yu D, DiTizio T and Court DL (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proceedings of the National Academy of Sciences of the United States of America*. 98 (12), 6742–6746.

Engberg J, Neimann J, Nielsen EM, Aerestrup FM and Fussing V (2004) Quinolone-resistant Campylobacter infections: risk factors and clinical consequences. *Emerging Infectious Diseases*. 10 (6), 1056–1063.

Enne VI, Bennett PM, Livermore DM and Hall LMC (2004) Enhancement of host fitness by the sul2-coding plasmid p9123 in the absence of selective pressure. *The Journal of Antimicrobial Chemotherapy*. 53 (6), 958–963.

Fernández-Gacio A, Codina A, Fastrez J, Riant O and Soumillion P (2006) Transforming carbonic anhydrase into epoxide synthase by metal exchange. *ChemBioChem*. 7 (7), 1013–1016.

Fineran PC, Petty NK and Salmond GPC (2009) Transduction: Host DNA transfer by bacteriophages. *Encyclopedia of Microbiology* (3rd edition). 666-679.

Firth AE and Patrick WM (2008) GLUE-IT and PEDEL-AA: new programmes for analyzing protein diversity in randomized libraries. *Nucleic Acids Research*. 36 (Web Server issue), W281–W285.

Fischer D, Teich A, Neubauer P and Hengge-Aronis R (1998) The general stress sigma factor sigmaS of *Escherichia coli* is induced during diauxic shift from glucose to lactose. *Journal of Bacteriology*. 180 (23), 6203–6206.

Fitzhugh AL (1998) Antibiotic inhibitors of the peptidyl transferase center. 1. clindamycin as a composite analogue of the transfer RNA fragments L-Pro-Met and the D-ribosyl ring of adenosine. *Bioorganic & Medicinal Chemistry Letters*. 8 (1), 87–92.

Flensburg J and Skold O (1987) Massive overproduction of dihydrofolate reductase in bacteria as a response to the use of trimethoprim. *European Journal of Biochemistry*. 162 (3), 473–476.

Fraimow HS, Greenman JB, Leviton IM, Dougherty TJ and Miller MH (1991) Tobramycin uptake in *Escherichia coli* is driven by either electrical potential or ATP. *Journal of Bacteriology*. 173 (9), 2800–2808.

Francez-Charlot A, Castanié-Cornet MP, Gutierrez C and Cam K (2005) Osmotic regulation of the *Escherichia coli bdm* (biofilm-dependent modulation) gene by the RcsCDB His-Asp phosphorelay. *Journal of Bacteriology*. 187 (11), 3873–3877.

Frey PA (1996) The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose. *The Journal of the Federation of American Societies for Experimental Biology*. 10 (4), 461–470.

Gabelli SB, Bianchet MA, Xu W, Dunn CA, Niu ZD, Amzel LM and Bessman MJ (2007) Structure and function of the *E. coli* dihydroneopterin triphosphate pyrophosphatase: a Nudix enzyme involved in folate biosynthesis. *Structure*. 15 (8), 1014–1022.

Gagneux S, Long CD, Small PM, Van T, Schoolnik GK and Bohannan BJM (2006) The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. *Science*. 312 (5782), 1944–1946.

Galperin MY, Bairoch A and Koonin EV (1998) A superfamily of metalloenzymes unifies phosphopentomutase and cofactor-independent phosphoglycerate mutase with alkaline phosphatases and sulfatases. *Protein Science*. 7 (8), 1829–1835.

Gerken H, Charlson ES, Cicirelli EM, Kenney LJ and Misra R (2009) MzrA: a novel modulator of the EnvZ/OmpR two-component regulon. *Molecular Microbiology*. 72 (6), 1408–1422.

Gilliver MA, Bennett M, Begon M, Hazel SM and Hart CA (1999) Antibiotic resistance found in wild rodents. *Nature*. 401 (6750), 233–234.

Gomez MJ and Neyfakh AA (2006) Genes involved in intrinsic antibiotic resistance of *Acinetobacter baylyi*. *Antimicrobial Agents and Chemotherapy*. 50 (11), 3562–3567.

Gonzalez III, Luis S and Spencer JP (1998) Aminoglycosides: A Practical Review. *American Family Physician*. 58 (8), 1811–1820.

Gould SM and Tawfik DS (2005) Directed evolution of the promiscuous esterase activity of carbonic anhydrase II. *Biochemistry*. 44 (14), 5444–5452.

Greenway DL and England RR (1999) ppGpp accumulation in *Pseudomonas aeruginosa* and *Pseudomonas fluorescens* subjected to nutrient limitation and biocide exposure. *Letters in Applied Microbiology*. 29 (5), 298–302.

Guilloton MB, Korte JJ, Lamblin AF, Fuchs JA and Anderson PM (1992) Carbonic anhydrase in *Escherichia coli*. A product of the cyn operon. *The Journal of Biological Chemistry*. 267 (6), 3731–3734.

Hamed RB, Gomez-Castellanos JR, Henry L, Ducho C, McDonough MA and Schofield CJ (2013) The enzymes of β-lactam biosynthesis. *Natural Product Reports*. 30 (1), 21–107.

Hansen LH, Mauvais P and Douthwaite S (1999) The macrolide-ketolide antibiotic binding site is formed by structures in domains II and V of 23S ribosomal RNA. *Molecular Microbiology*. 31 (2), 623–631.

Hanson-Manful P and Patrick WM (2013) Construction and analysis of randomized protein-encoding libraries using error-prone PCR. *Methods in Molecular Biology*. 996, 251–267.

Hawkey PM (1998) The origins and molecular basis of antibiotic resistance. *British Medical Journal*. 317 (7159), 657–660.

Hiramatsu K, Hanaki H, Ino T, Yabuta K, Oguri T and Tenover FC (1997) Methicillin-resistant *Staphylococcus aureus* clinical strain with reduced vancomycin susceptibility. *The Journal of Antimicrobial Chemotherapy*. 40 (1), 135–136.

Hirota N, Matsuura S, Mochizuki N, Mutoh N and Imae Y (1981) Use of lipophilic cation-permeable mutants for measurement of transmembrane electrical potential in metabolizing cells of *Escherichia coli*. *Journal of Bacteriology*. 148 (2), 399–405.

Holden HM, Rayment I and Thoden JB (2003) Structure and function of enzymes of the Leloir pathway for galactose metabolism. *The Journal of Biological Chemistry*. 278 (45), 43885–43888.

Hooper DC (1999) Mechanisms of fluoroquinolone resistance. *Drug Resistance Updates*. 2 (1), 38–55.

Hou Y, Lin YP, Sharer JD and March PE (1994) *In vivo* selection of conditional-lethal mutations in the gene encoding elongation factor G of *Escherichia coli*. *Journal of Bacteriology*. 176 (1), 123–129.

Huovinen P, Sundström L, Swedberg G and Sköld O (1995) Trimethoprim and sulfonamide resistance. *Antimicrobial Agents and Chemotherapy*. 39 (2), 279–289.

James LC and Tawfik DS (2001) Catalytic and binding poly-reactivities shared by two unrelated proteins: The potential role of promiscuity in enzyme evolution. *Protein Science*. 10 (12), 2600–2607.

Jensen RA (1976) Enzyme recruitment in evolution of new function. *Annual Review of Microbiology*. 30, 409–425.

Jezequel N, Lagomarsino MC, Heslot F and Thomen P (2013) Long-term diversity and genome adaptation of Acinetobacter baylyi in a minimal-medium chemostat. *Genome Biology and Evolution*. 5 (1), 87–97.

Johanson U and Hughes D (1994) Fusidic acid-resistant mutants define three regions in elongation factor G of *Salmonella typhimurium*. *Gene*. 143 (1), 55–59.

Kado CI (1998) Origin and evolution of plasmids. *Antonie van Leeuwenhoek*. 73 (1), 117–126.

Karakousis G, Ye N, Li Z, Chiu SK, Reddy G and Radding CM (1998) The beta protein of phage lambda binds preferentially to an intermediate in DNA renaturation. *Journal of Molecular Biology*. 276 (4), 721–731.

Keseler IM, Collado-Vides J, Santos-Zavaleta A, Peralta-Gil M, Gama-Castro S, Muñiz-Rascado L, Bonavides-Martinez C, Paley S, Krummenacker M, Altman T, Kaipa P, Spaulding A, Pacheco J, Latendresse M, Fulcher C, Sarker M, Shearer AG, Mackie A, Paulsen I, Gunsalus RP and Karp PD (2011) EcoCyc: a comprehensive database of *Escherichia coli* biology. *Nucleic Acids Research*. 39 (Database issue), D583–D590.

Khersonsky O, Roodveldt C and Tawfik DS (2006) Enzyme promiscuity: evolutionary and mechanistic aspects. *Current Opinion in Chemical Biology*. 10 (5), 498–508.

Khersonsky O and Tawfik DS (2010) Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annual Review of Biochemistry*. 79, 471–505.

Kimber MS and Pai EF (2000) The active site architecture of *Pisum sativum* beta-carbonic anhydrase is a mirror image of that of alpha-carbonic anhydrases. *The European Molecular Biology Organisation Journal*. 19 (7), 1407–1418.

Kimura M and Ota T (1974) On some principles governing molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America*. 71 (7), 2848–2852.

Kitagawa M, Ara T, Arifuzzaman M, Ioka-Nakamichi T, Inamoto E, Toyonaga H and Mori H (2005) Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA Research*. 12 (5), 291–299.

Kitamura K, Torii Y, Matsuoka C and Yamamoto K (1995) DNA sequence changes in mutations in the *tonB* gene on the chromosome of *Escherichia coli* K12: insertion

elements dominate the spontaneous spectra. *The Japanese Journal of Genetics*. 70 (1), 35–46.

Klockenbusch C and Kast J (2010) Optimization of formaldehyde cross-linking for protein interaction analysis of non-tagged integrin beta1. *Journal of Biomedicine & Biotechnology*. 2010, 927585.

Krzewinski F, Brassart C, Gavini F and Bouquelet S (1996) Characterization of the Lactose Transport System in the Strain Bifidobacterium bifidum DSM 20082. *Current Microbiology*. 32 (6), 301–307.

Künzler DE, Sasso S, Gamper M, Hilvert D and Kast P (2005) Mechanistic insights into the isochorismate pyruvate lyase activity of the catalytically promiscuous PchB from combinatorial mutagenesis and selection. *The Journal of Biological Chemistry*. 280 (38), 32827–32834.

Kurihara S, Oda S, Kato K, Kim HG, Koyanagi T, Kumagai H and Suzuki H (2005) A novel putrescine utilization pathway involves gamma-glutamylated intermediates of *Escherichia coli* K-12. *The Journal of Biological Chemistry*. 280 (6), 4602–4608.

Lamblin AF and Fuchs JA (1994) Functional analysis of the *Escherichia coli* K-12 *cyn* operon transcriptional regulation. *Journal of Bacteriology*. 176 (21), 6613–6622.

Lampe MF and Bott KF (1985) Genetic and physical organization of the cloned *gyrA* and *gyrB* genes of *Bacillus subtilis*. *Journal of Bacteriology*. 162 (1), 78–84.

Laurberg M, Kristensen O, Martemyanov K, Gudkov AT, Nagaev I, Hughes D and Liljas A (2000) Structure of a mutant EF-G reveals domain III and possibly the fusidic acid binding site. *Journal of Molecular Biology*. 303 (4), 593–603.

Leroi AM, Bennett AF and Lenski RE (1994) Temperature acclimation and competitive fitness: an experimental test of the beneficial acclimation assumption. *Proceedings of the National Academy of Sciences of the United States of America*. 91 (5), 1917–1921.

Levin BR, Perrot V and Walker N (2000) Compensatory Mutations, Antibiotic Resistance and the Population Genetics of Adaptive Evolution in Bacteria. *Genetics*. 154 (3), 985–997.

Levy SB (2002) Factors impacting on the problem of antibiotic resistance. *Journal of Antimicrobial Chemotherapy*. 49 (1), 25–30.

Levy SB and Marshall B (2004) Antibacterial resistance worldwide: causes, challenges and responses. *Nature Medicine*. 10 (12 Suppl), S122-S129.

Li J, Attila C, Wang L, Wood TK, Valdes JJ and Bentley WE (2007) Quorum sensing in *Escherichia coli* is signaled by AI-2/LsrR: effects on small RNA and biofilm architecture. *Journal of Bacteriology*. 189 (16), 6011–6020.

Liljas A and Garber M (1995) Ribosomal proteins and elongation factors. *Current Opinion in Structural Biology*. 5 (6), 721–727.

Lipsitch M, Singer RS and Levin BR (2002) Antibiotics in agriculture: when is it time to close the barn door? *Proceedings of the National Academy of Sciences of the United States of America*. 99 (9), 5752–5754.

Macinga DR and Rather PN (1999) The chromosomal 2'-N-acetyltransferase of *Providencia stuartii*: physiological functions and genetic regulation. *Frontiers in Bioscience*. 4, D132–D140.

Macvanin M, Björkman J, Eriksson S, Rhen M, Andersson DI and Hughes D (2003) Fusidic acid-resistant mutants of *Salmonella enterica* Serovar Typhimurium with low fitness *in vivo* are defective in RpoS induction. *Antimicrobial Agents and Chemotherapy*. 47 (12), 3743–3749.

Macvanin M, Johanson U, Ehrenberg M and Hughes D (2000) Fusidic acid-resistant EF-G perturbs the accumulation of ppGpp. *Molecular Microbiology*. 37 (1), 98–107.

Malakar P and Venkatesh K V (2012) Effect of substrate and IPTG concentrations on the burden to growth of *Escherichia coli* on glycerol due to the expression of Lac proteins. *Applied Microbiology and Biotechnology*. 93 (6), 2543–2549.

Marcusson LL, Frimodt-Møller N and Hughes D (2009) Interplay in the selection of fluoroquinolone resistance and bacterial fitness. *PLoS Pathogens*. 5 (8), e1000541.

Marinelli F and Tomasz A (2010) Antimicrobials. *Current Opinion in Microbiology*. 13 (5), 547–550.

Marshall BM, Ochieng DJ and Levy SB (2009) Commensals: Underappreciated Reservoir of Antibiotic Resistance. *Micobiology*. 4 (5), 231–238.

Martemyanov KA and Gudkov AT (2000) Domain III of elongation factor G from *Thermus thermophilus* is essential for induction of GTP hydrolysis on the ribosome. *The Journal of Biological Chemistry*. 275 (46), 35820–25824.

Martinez JL and Baquero F (2000) Mutation Frequencies and Antibiotic Resistance. *Antimicrobial Agents and Chemotherapy*. 44 (7), 1771–1777.

Mastrobattista E, Taly V, Chanudet E, Treacy P, Kelly BT and Griffiths AD (2005) High-throughput screening of enzyme libraries: *in vitro* evolution of a beta-galactosidase by fluorescence-activated sorting of double emulsions. *Chemistry & Biology*. 12 (12), 1291–1300.

Matten SR, Schneider TD, Ringquist S and Brusilow WSA (1998) Identification of an Intragenic Ribosome Binding Site That Affects Expression of the *uncB* Gene of the *Escherichia coli* Proton-Translocating ATPase (*unc*) Operon. *Journal of Bacteriology.* 180 (15), 3940–3945.

Mauzy CA and Hermodson MA (1992) Structural and functional analyses of the repressor, RbsR, of the ribose operon of *Escherichia coli*. *Protein Science*. 1 (7), 831–842.

Miller BG and Raines RT (2004) Identifying latent enzyme activities: substrate ambiguity within modern bacterial sugar kinases. *Biochemistry*. 43 (21), 6387–6392.

Miller J (1972) Experiment in Molecular Genetics*. Cold Spring Harbor Lab., New York, USA*.

Miller SP, Lunzer M and Dean AM (2006) Direct demonstration of an adaptive constraint. *Science*. 314 (5798), 458–461.

Muyrers JPP, Zhang Y, Buchholz F and Stewart AF (2000) RecE/RecT and Redalpha/Redbeta initiate double-stranded break repair by specifically interacting with their respective partners. *Genes & Development.* 14 (15), 1971–1982.

Nagaev I, Bjorkman J, Andersson DI and Hughes D (2001) Biological cost and compensatory evolution in fusidic acid-resistant *Staphylococcus aureus*. *Molecular Microbiology*. 40 (2), 433–439.

Napolitano R, Janel-Bintz R, Wagner J and Fuchs RP (2000) All three SOS-inducible DNA polymerases (Pol II, Pol IV and Pol V) are involved in induced mutagenesis. *The European Molecular Biology Organisation Journal*. 19 (22), 6259–6265.

Nathan C and Goldberg FM (2005) Outlook: the profit problem in antibiotic R&D. *Nature Reviews Drug Discovery*. 4 (11), 887–891.

Neu HC (1992) The crisis in antibiotic resistance. *Science*. 257 (5073), 1064–1073.

Norberg P, Bergström M, Jethava V, Dubhashi D and Hermansson M (2011) The IncP-1 plasmid backbone adapts to different host bacterial species and evolves through homologous recombination. *Nature Communications*. 2, 268.

Normark S, Edlund T, Grundström T, Bergström S and Wolf-Watz H (1977) *Escherichia coli* K-12 mutants hyperproducing chromosomal beta-lactamase by gene repetitions. *Journal of Bacteriology*. 132 (3), 912–922.

Nyquist AC (1998) Antibiotic Prescribing for Children With Colds, Upper Respiratory Tract Infections, and Bronchitis. *The Journal of the American Medical Association*. 279 (11), 875–877.

O'Brie PJ and Herschlag D (2001) Functional Interrelationships in the Alkaline Phosphatase Superfamily: Phosphodiesterase Activity of *Escherichia coli* Alkaline Phosphatase. *Biochemistry*. 40 (19), 5691–5699.

O'Brien PJ and Herschlag D (1999) Catalytic promiscuity and the evolution of new enzymatic activities. *Chemistry & Biology*. 6 (4), R91–R105.

O'Loughlin TL, Patrick WM and Matsumura I (2006) Natural history as a predictor of protein evolvability. *Protein Engineering, Design & Selection*. 19 (10), 439–442.

Ochman H, Lawrence JG and Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature*. 405 (6784), 299–304.

Ohki R and Tateno K (2004) Increased stability of bmr3 mRNA results in a multidrug-resistant phenotype in *Bacillus subtilis*. *Journal of Bacteriology*. 186 (21), 7450–7455.

Ohno S (1970) Evolution by gene duplication. *New York: Springer-Verlag*.

Okeke IN and Edelman R (2001) Dissemination of antibiotic-resistant bacteria across geographic borders. *Clinical Infectious Diseases*. 33 (3), 364–369.

Olguin LF, Askew SE, O'Donoghue AC and Hollfelder F (2008) Efficient catalytic promiscuity in an enzyme superfamily: an arylsulfatase shows a rate acceleration of 10(13) for phosphate monoester hydrolysis. *Journal of the American Chemical Society*. 130 (49), 16547–16555.

Patrick WM, Firth AE and Blackburn JM (2003) User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. *Protein Engineering Design and Selection*. 16 (6), 451–457.

Patrick WM and Matsumura I (2008) A study in molecular contingency: glutamine phosphoribosylpyrophosphate amidotransferase is a promiscuous and evolvable phosphoribosylanthranilate isomerase. *Journal of Molecular Biology*. 377 (2), 323–336.

Patrick WM, Quandt EM, Swartzlander DB and Matsumura I (2007) Multicopy suppression underpins metabolic evolvability. *Molecular Biology and Evolution*. 24 (12), 2716–2722.

Paulsen IT, Nguyen L, Sliwinski MK, Rabus R and Saier MH (2000) Microbial genome analyses: comparative transport capabilities in eighteen prokaryotes. *Journal of Molecular Biology*. 301 (1), 75–100.

Pocker Y and Stone JT (1965) The catalytic versatility of erythrocyte carbonic anhydrase. The enzyme-catalyzed hydrolysis of rho-nitrophenyl acetate. *Journal of the American Chemical Society*. 87 (23), 5497–5498.

Poeta P, Radhouani H, Pinto L, Martinho A, Rego V, Rodrigues R, Gonçalves A, Rodrigues J, Estepa V, Torres C and Igrejas G (2009) Wild boars as reservoirs of extended-spectrum beta-lactamase (ESBL) producing *Escherichia coli* of different phylogenetic groups. *Journal of Basic Microbiology*. 49 (6), 584–588.

Pósfai G, Plunkett G, Fehér T, Frisch D, Keil GM, Umenhoffer K, Kolisnychenko V, Stahl B, Sharma SS, De Arruda M, Burland V, Harcum SW and Blattner FR (2006) Emergent properties of reduced-genome *Escherichia coli*. *Science*. 312 (5776), 1044–1046.

Quail MA, Jordan P, Grogan JM, Butt JN, Lutz M, Thomson AJ, Andrews SC and Guest JR (1996) Spectroscopic and voltammetric characterisation of the bacterioferritin-associated ferredoxin of *Escherichia coli*. *Biochemical and Biophysical Research Communications*. 229 (2), 635–642.

Rivera M, Hancock RE, Sawyer JG, Haug A and McGroarty EJ (1988) Enhanced binding of polycationic antibiotics to lipopolysaccharide from an aminoglycoside-supersusceptible, *tolA* mutant strain of *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy*. 32 (5), 649–655.

Romero PA and Arnold FH (2009) Exploring protein fitness landscapes by directed evolution. *Nature Reviews. Molecular cell Biology*. 10 (12), 866–876.

Roth JR, Kugelberg E, Reams AB, Kofoid E and Andersson DI (2006) Origin of mutations under selection: the adaptive mutation controversy. *Annual Review of Microbiology*. 60, 477–501.

Rownd R and Mickel S (1971) Dissociation and Reassociation of RTF and r-Determinants of the R-factor NR1 in Proteus mirabilis. *Nature*. 234 (45), 40–43.

Rozen DE, McGee L, Levin BR and Klugman KP (2007) Fitness costs of fluoroquinolone resistance in Streptococcus pneumoniae. *Antimicrobial Agents and Chemotherapy*. 51 (2), 412–416.

Ruiz J (2003) Mechanisms of resistance to quinolones: target alterations, decreased accumulation and DNA gyrase protection. *The Journal of Antimicrobial Chemotherapy*. 51 (5), 1109–1117.

Sambrook J and Russell DW (2001) *Molecular cloning. A laboratory manual* (3rd edition). New York: CSHL Press.

Sandegren L and Andersson DI (2009) Bacterial gene amplification: implications for the evolution of antibiotic resistance. *Nature Reviews Microbiology*. 7 (8), 578–588.

Sander P, Springer B, Prammananan T, Sturmfels A, Kappler M, Pletschette M and Bottger EC (2002) Fitness Cost of Chromosomal Drug Resistance-Conferring Mutations. *Antimicrobial Agents and Chemotherapy*. 46 (5), 1204–1211.

Sandulache R, Prehm P and Kamp D (1984) Cell wall receptor for bacteriophage Mu G(+). *Journal of Bacteriology*. 160 (1), 299–303.

Schulz zur Wiesch P, Engelstädter J and Bonhoeffer S (2010) Compensation of fitness costs and reversibility of antibiotic resistance mutations. *Antimicrobial Agents and Chemotherapy*. 54 (5), 2085–2095.

Shachrai I, Zaslaver A, Alon U and Dekel E (2010) Cost of unneeded proteins in *E. coli* is reduced after several generations in exponential growth. *Molecular Cell*. 38 (5), 758–767.

Shafikhani S, Siegel RA, Ferrari E, Schellenberger V (1997) Generation of large libraries of random mutants in *Bacillus subtilis* by PCR-based plasmid multimerization. *BioTechniques*. 23 (2), 304 – 310.

Shand RF, Blum PH, Mueller RD, Riggs DL and Artz SW (1989) Correlation between histidine operon expression and guanosine 5'-diphosphate-3'-diphosphate levels during amino acid downshift in stringent and relaxed strains of *Salmonella typhimurium*. *Journal of Bacteriology*. 171 (2), 737–743.

Shepherd M, Sanguinetti G, Cook GM and Poole RK (2010) Compensations for diminished terminal oxidase activity in *Escherichia coli*: cytochrome *bd*-II-mediated respiration and glutamate metabolism. *The Journal of Biological Chemistry*. 285:18464-18472.

Singh S, Kate BN and Banerjee UC (2008) Bioactive Compounds from Cyanobacteria and Microalgae: An Overview. *Critical Reviews in Biotechnology*. 25 (3), 73-95.

Smith DL, Dushoff J and Morris JG (2005) Agricultural antibiotics and human health. *PLoS Medicine*. 2 (8), e232.

Smith DR and Calvo JM (1982) Nucleotide sequence of dihydrofolate reductase genes from trimethoprim-resistant mutants of *Escherichia coli*. *Molecular & General Genetics*. 187 (1), 72–78.

Smith KS and Ferry JG (2000) Prokaryotic carbonic anhydrases. *Federation of European Microbiological Societies Microbiology Reviews*. 24 (4), 335–366.

Sommer MOA, Dantas G and Church GM (2009) Functional Characterization of the Antibiotic Resistance Reservoir in the Human Microflora. *Science*. 324 (5944), 1128-1131.

Soo VWC, Hanson-Manful P and Patrick WM (2011) Artificial gene amplification reveals an abundance of promiscuous resistance determinants in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*. 108 (4), 1484–1489.

Spellberg B, Powers JH, Brass EP, Miller LG and Edwards JE (2004) Trends in antimicrobial drug development: implications for the future. *Clinical Infectious Diseases*. 38 (9), 1279–1286.

Stahl MM, Thomason L, Poteete AR, Tarkowski T, Kuzminov A and Stahl FW (1997) Annealing vs. invasion in phage lambda recombination. *Genetics*. 147 (3), 961–977.

Sykes RB and Bonner DP (1985) Aztreonam: the first monobactam. *The American Journal of Medicine*. 78 (2A), 2–10.

Thati V, Shivannavar CT and Gaddad SM (2011) Vancomycin resistance among methicillin resistant Staphylococcus aureus isolates from intensive care units of tertiary care hospitals in Hyderabad. *The Indian Journal of Medical Research*. 134 (5), 704–708.

The European Committee on Antimicrobial Susceptibility Testing [EUCAST] (2012) Clinical breakpoints. http://www.eucast.org/clinical_breakpoints

The UniProt Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt) (2012). *Nucleic Acids Research*. 40, D71–D75.

Tracewell CA and Arnold FH (2009) Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Current Opinion in Chemical Biology*. 13 (1), 3–9.

Tripp BC, Smith K and Ferry JG (2001) Carbonic anhydrase: new insights for an ancient enzyme. *The Journal of Biological Chemistry*. 276 (52), 48615–48618.

Tschowri N, Busse S and Hengge R (2009) The BLUF-EAL protein YcgF acts as a direct anti-repressor in a blue-light response of *Escherichia coli*. *Genes & Development*. 23 (4), 522–534.

Umenhoffer K, Fehér T, Balikó G, Ayaydin F, Pósfai J, Blattner FR and Pósfai G (2010) Reduced evolvability of *Escherichia coli* MDS42, an IS-less cellular chassis for molecular and synthetic biology applications. *Microbial Cell Factories*. 9 (1), 38.

Vaara M (1992) Agents that increase the permeability of the outer membrane. *Microbiological Reviews*. 56 (3), 395–411.

van de Sande-Bruinsma N, Grundmann H, Verloo D, Tiemersma E, Monen J, Goossens H and Ferech M (2008) Antimicrobial drug use and resistance in Europe. *Emerging Infectious Diseases*. 14 (11), 1722–1730.

van Hoek AHAM, Mevius D, Guerra B, Mullany P, Roberts AP and Aarts HJM (2011) Acquired antibiotic resistance genes: an overview. *Frontiers in Microbiology*. 2, 203.

Velappan N, Sblattero D, Chasteen L, Pavlik P and Bradbury ARM (2007) Plasmid incompatibility: more compatible than previously thought? *Protein Engineering, Design & Selection*. 20 (7), 309–313.

Walsh C (2000) Molecular mechanisms that confer antibacterial drug resistance. *Nature*. 406 (6797), 775–781.

Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR and Church GM (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature*. 460 (7257), 894–898.

Wang SC, Johnson WH and Whitman CP (2003) The 4-oxalocrotonate tautomerase- and YwhB-catalyzed hydration of 3E-haloacrylates: implications for the evolution of new enzymatic activities. *Journal of the American Chemical Society*. 125 (47), 14282–14283.

Warnes SL, Highmore CJ and Keevil CW (2012) Horizontal transfer of antibiotic resistance genes on abiotic touch surfaces: implications for public health. *mBio*. 3 (6), e00489.

Whitman WB (1998) Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences of the United States of America*. 95 (12), 6578–6583.

Wiegand I, Hilpert K and Hancock REW (2008) Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nature Protocols*. 3 (2), 163–175.

Wise R (2002) Antimicrobial resistance: priorities for action. *Journal of Antimicrobial Chemotherapy*. 49 (4), 585–586.

Wise R, Hart T, Cars O, Streulens M, Helmuth R, Huovinen P and Sprenger M (1998) Antimicrobial resistance. *British Medical Journal*. 317 (7159), 609–610.

Wright GD (2010) Antibiotic resistance in the environment: a link to the clinic? *Current Opinion in Microbiology*. 13 (5), 589–594.

Wright GD (2007) The antibiotic resistome: the nexus of chemical and genetic diversity. *Nature Reviews Microbiology*. 5 (3), 175–186.

Wright GD (2012) The origins of antibiotic resistance. *Handbook of Experimental Pharmacology*. (211), 13–30.

Yasutake Y, Yao M, Sakai N, Kirita T and Tanaka I (2004) Crystal structure of the *Pyrococcus horikoshii* isopropylmalate isomerase small subunit provides insight into the dual substrate specificity of the enzyme. *Journal of Molecular Biology*. 344 (2), 325–333.

You W, Qiu X, Zhang Y, Ma J, Gao Y, Zhang X, Niu L and Teng M (2010) Crystallization and preliminary X-ray diffraction analysis of the putative aldose 1-epimerase YeaD from *Escherichia coli*. *Acta Crystallographica. Section F, Structural Biology and Crystallization Communications*. 66 (Pt 8), 951–953.

Zhang J (2003) Evolution by gene duplication: an update. *Trends in Ecology and Evolution*. 18 (6), 292-298.

Zilberstein D, Schuldiner S, and Padan E (1979) Proton electrochemical gradient in *Escherichia coli* cells and its relation to active transport of lactose. *Biochemistry*. 18 (4), 669-673.

Zuckerman JM (2004) Macrolides and ketolides: azithromycin, clarithromycin, telithromycin. *Infectious Disease Clinics of North America*. 18 (3), 621–649.