Enzyme promiscuity and the origins of cellular innovations

A dissertation presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy
in
Biochemistry

at Massey University, Albany, New Zealand.

Valerie Wooi Chee Soo

2012

ABSTRACT

Biochemistry textbooks define enzymes as being efficient and highly specific. However, these characteristics are usually associated with a lack of versatility, and therefore, an inability to evolve new functions. In spite of this, it is known that new enzymes can arise rapidly (such as when bacteria evolve antibiotic resistance). One hypothesis proposes that enzymes are actually promiscuous (Jensen, 1976); that is, they are able to carry out secondary reactions, in addition to the one they evolved to catalyze. The goal of this research was to explore the role that promiscuity plays in the origins and evolution of enzyme functions, using *Escherichia coli* as a model organism.

In the first part of this thesis, I report the discovery of two enzymes (alanine racemase and cystathionine β-lyase) that are reciprocally promiscuous, and are dependent on the cofactor pyridoxal 5'-phosphate (PLP) for activity. *In vivo*, the cofactor-mediated promiscuous activities of alanine racemase and cystathionine β-lyase were each successfully improved to near wildtype levels using directed evolution experiments. These results extend Jensen's hypothesis, and led me to propose that PLP played a significant role in the evolution of new enzymes, in the primordial world.

In the second part of the thesis, I developed a comprehensive library-on-library screen to search for *E. coli* proteins that could mediate improved growth in environments containing either a foreign nutrient or a toxin. Proteins were over-expressed in an attempt to increase their weak, promiscuous activities, and to mimic the common genetic phenomenon of gene amplification. Over-expression of individual proteins conferred improved growth to the host cell in 35% of ~2,000 environments. The findings have important implications for understanding bacterial adaptation to new environments, such as when antibiotic resistance emerges. The ability of promiscuous proteins to drive the emergence of new phenotypes, when their expression is increased, validates the feasibility of the Innovation, Amplification and Divergence (IAD) model for the evolution of new genes (Bergthorsson *et al*., 2007).

Overall, the work described in this thesis demonstrates that protein promiscuity is common, though difficult to predict *a priori*. My experimental results are consistent with the work of others, in suggesting that promiscuous activities are evolvable. Together, the high frequency and evolvability of promiscuous proteins appear to underpin many different cellular innovations.

[Blank page]

ACKNOWLEDGEMENTS

Biochemistry was one of my least liked subjects when I was an undergraduate.

But this has changed in less than four years. For introducing me to the coolness of Biochemistry, I thank Dr. Wayne Patrick. As a PhD supervisor, he has done everything he could to make sure I didn't become a burnt pancake — as they say, the first one is usually a tricky one. Above all, Wayne has been a source of inspiration. I particularly appreciate his constructive feedback, which (usually) manifests in the form of red scribbles all over the papers. Perhaps the most meaningful gift I could present to him is not this thesis, but a stash of red pens (which can last until his retirement)!

I am grateful for the all-round advice from Dr. Monica Gerth. Her patience for me is amazing, despite my endless questions about science, life, the universe… She just seems to have a knack of setting me straight, and telling me not to sway when in doubt.

My secondary supervisor, Prof. Paul Rainey, has been generous with his time and evolutionary insights whenever I needed them. Dr. Xue-Xian Zhang has been kind to share his knowledge on microbial genetics. Both have my thanks.

I extend my gratitude to Prof. Dan Andersson (Uppsala University) for his helpful discussions, which have partly shaped the work presented in Chapters 2 and 3. Dan also paid for my meals and lodging during my short trip to Uppsala in 2010. Special thanks to Prof. John Roth (University of California, Davis) for his invaluable comments on the work described in Chapters 4 and 5.

I also wish to thank Dr. Gill Norris for her help in interpreting the mass spec data; Dr. Christopher Squire (University of Auckland) for taking an interest in solving the structure of MetC(P113S); Dr. Tim Clausen (Research Institute of Molecular Pathology, Vienna) for his willingness to share one of his old articles and the structure coordinates of an inhibitor-bound MetC.

Financial support for my PhD came from the Institute of Molecular Biosciences/Institute of Natural Sciences (INS) Doctoral Scholarship, and the New Zealand (NZ) Institute of Advanced Study/INS Postgraduate Research Excellence Award. For sorting out my financial (and other administrative) issues, I thank the Institute

Administrator, Muharram Khoussainova. Travel grants from the NZ Society for Biochemistry and Molecular Biology, and the Federation of European Biochemical Societies, allowed me to attend conferences without financial worries.

A big thank-you goes to the people in the Patrick Lab, especially Paulina Hanson-Manful, Ilana Gerber, Susan Morton, Laura Nigon and Natasha le Roux. In addition to their technical support, the girls have also been excellent lunch buddies. I can't thank Matteo Ferla enough for his help with Illustrator and Photoshop. Mack Saraswat has also been great in taking my mind off research during our chats over coffee.

Colleagues and friends from Building 11 have been very supportive, especially towards the end of my PhD. Moral support from Saumya Agrawal and Martina Dautel was very much appreciated on days when I seemed to be at the end of my tether. Ralph Grand, Jarod Young and Matilda Newton have been offering heavy doses of encouragement and optimism.

I am also indebted to the *ad hoc* proofreading team: Monica, Ralph, Saumya and Matilda.

A large chunk of this thesis was written at home. For keeping the noise volume to a minimum in our house, I thank my flatmates: Kataraina Hohua, Nurul Syaza Abdul Latif and Noriza Ahmad.

I am thankful to have Jingjing Khoo as a best friend. We have supported each other when our experiments didn't go well. I hope she will overcome her last hurdle soon in obtaining a PhD in Biochemistry (from Monash University).

To Mum, Dad and Sis, thank you for believing in me, and my dreams. Your sheer love is unparalleled. 感謝家人一直給慧琪的愛護，支持，鼓勵和包容。 爹哋（蘇芳先生），媽咪（彭水珍女士）和小妹（慧琳），謝謝你們！

Last but not least, I owe various *E. coli* strains (see Section I.3 in Appendix I) great thanks. Without them, this thesis would turn out to be very different.

TABLE OF CONTENTS

ABBREVIATIONS

Alr	Alanine racemase
ASKA	A complete set of *E. coli* K-12 ORF Archive
A_{xxx}	Absorbance measured at xxx nm
cfu	Colony-forming unit
CHES	*N*-cyclohexyl-2-aminoethanesulfonic acid
DAAO	D-amino acid oxidase
DMSO	Dimethyl sulfoxide
DTNB	5,5′-dithiobis(2-nitrobenzoic acid)
EC	Enzyme Commission number
EDTA	Ethylenediaminetetraacetic acid
epPCR	Error-prone PCR
GFP	Green Fluorescent Protein
$(His)_6$	Hexa-histidine
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
lacZ	β-galactosidase
LB	Lysogeny broth
LDH	Lactate dehydrogenase
m/*z*	Mass-to-charge ratio
MetC	Cystathionine β-lyase
MIC	Minimum inhibitory concentration
NAD^+	Oxidized form of β-nicotinamide adenine dinucleotide
NADH	Reduced form of β-nicotinamide adenine dinucleotide
NoIns (or NoInsert)	The negative control plasmid, pCA24N-NoIns (or *E. coli* harbouring the pCA24N-NoIns plasmid)

OD_{600}	Optical density measured at 600 nm
ORF	Open reading frame
PDB	The RCSB Protein Data Bank
PLP	Pyridoxal 5′-phosphate
PM	Phenotype Microarray
RMSD	Root Mean Square Deviation
S.E.M.	Standard error of mean
SOB	Super Optimal Broth
SOC	SOB with Catabolite repression
Tris	Tris(hydroxymethyl)aminomethane
W	Relative fitness
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

CHAPTER 1

Introduction

Life is shaped by networks of chemical reactions. Protein molecules called enzymes are responsible for accelerating thousands of these reactions, under biologically relevant conditions (*e.g.* in an aqueous environment with a near-neutral pH, at atmospheric pressure and temperatures below 100°C). In the absence of enzymes, these chemical reactions are unlikely to happen due to their unfavourable thermodynamics and/or kinetics. For example, orotidine 5'-phosphate (OMP) decarboxylase enhances the uncatalyzed rate of OMP decarboxylation by a factor of 10^{17} (Radzicka & Wolfenden, 1995; Callahan & Miller, 2007). However, the evolutionary processes that give rise to, and optimize, such remarkable catalysts remain poorly understood.

New enzyme activities must arise as species adapt to changing environments. At the molecular level, the presence of novel nutrients or toxins can drive the evolution of new functions by imparting selection for enzymes that can recognize or metabolize these compounds (Hegeman & Rosenberg, 1970; Mortlock, 1984). Understanding the emergence of enzymes and metabolic pathways is important for learning about the past; that is, how extant organisms evolved to be what they are. Further, lessons from the past can allow predictions to be made about the future, such as predicting the routes by which antibiotic resistance may emerge in disease-causing bacteria. These "looking back" and "looking forward" perspectives on enzyme evolution will be explored in this thesis.

1.1 ORIGINS OF NEW FUNCTIONS AND METABOLIC PATHWAYS

An early attempt to explain how new enzymes could evolve was provided by Norman Horowitz. He proposed a model of retrograde evolution, in which the depletion of a metabolite is the selection pressure for evolution of a new enzyme (Horowitz, 1945). Selection for this new enzyme is based on its ability to use an available precursor(s) to catalyze the formation of the required metabolite. For instance, once product *A* is exhausted (as the organism grows), it can be replenished from a closely related compound (precursor *B*) if a suitable enzyme *E1* evolves (Figure 1.1). The depletion of precursor *B* will further drive the evolution of enzyme *E2*, which can synthesize compound *B* from precursor *C*. Hence, metabolic pathways evolve backwards. This evolutionary model surmises that both new and old enzymes must have similar binding specificities, as chemically related compounds are continuously being recruited into the pathway. Horowitz and Metzenberg (1965) further postulated that all enzymes involved in the same multi-step biosynthetic pathway are homologous — each enzyme located upstream in the pathway is most likely to have arisen by duplication and divergence from the immediate progenitor downstream in the pathway [enzyme *E2* from *E1*; enzyme *E3* from *E2* in Figure 1.1].

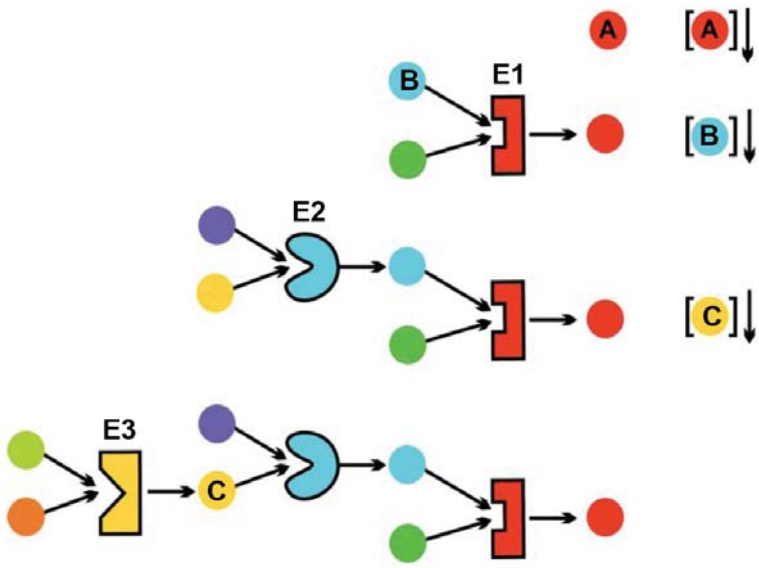


Figure 1.1: Horowitz's model of retrogade evolution.

A new enzyme (E1) emerges when a metabolite (product A) is exhausted by cellular consumption. The enzyme E1 catalyzes the formation of the metabolite by recruiting a chemically-related compound (precursor B) as a substrate. The iterative process further suggests that the last enzyme (E1) in a metabolic pathway is the first to appear, and the first enzyme (E3) is the last to appear. Figure adapted from Holliday *et al*. (2007).

An example that appears to follow the retrogade evolution model is the enzymes from tryptophan biosynthesis. In bacteria, the enzymes phosphoribosylanthranilate isomerase (TrpF), indole-3-glycerol phosphate synthase (TrpC) and the α-subunit of tryptophan synthase (TrpA) catalyze three sequential reactions to yield indole, the precursor for synthesizing tryptophan. The $(\beta/\alpha)_8$-barrel structures of the three enzymes are very similar to one another (Wilmanns *et al.*, 1992), with an RMSD of ~1.0 Å for their superimposed main chains (Wilmanns *et al.*, 1991). However, neither the enzymes before TrpF in the pathway (TrpE, TrpG and TrpD), nor the final enzyme (TrpB) are homologous $(\beta/\alpha)_8$-barrels. This suggests that Horowitz's model may be an over-simplification.

In 1976, Roy Jensen pointed out that Horowitz's hypothesis depended on the availability of chemically related metabolites in primordial cells (Jensen, 1976). Hypothetically, the cell would have to obtain each metabolic intermediate from its environment. This seemed highly unlikely, given the "extreme chemical lability of many intermediary metabolites, as well as by the barriers to their transport that may have existed in the absence of specialized transport systems" (Jensen, 1976). Instead, an alternative hypothesis of enzyme evolution was proposed, after taking the minimal gene contents in primordial cells into account (Yčas, 1974; Jensen, 1976). This hypothesis assumes that primitive enzymes possessed broad specificities for multiple substrates. Many pathways for synthesizing key metabolites could therefore exist in a patchwork fashion, albeit at a low level (Figure 1.2A). Genes for advantageous enzymes could be duplicated and then accumulate mutations that enabled them to provide increased levels of key metabolites (Figure 1.2B). The combination of broad-specificity enzymes and fortuitous chemical reactions could allow the production of key metabolites from novel intermediates, even if several enzymatic steps were required simultaneously (Jensen, 1976; Rison & Thornton, 2002). This model is appealing, because such a "multitasking" strategy would allow the cell to cope with as many fluctuations as possible with a minimal number of enzymes (and therefore the smallest possible genome).

With its emphasis on enzymes with secondary activities, this model of patchwork evolution has also proven useful for explaining the on-going evolution of modern enzyme activities. A classic example is the emergence of a phosphotriesterase (PTE) enzyme in *Pseudomonas diminuta* (Lewis *et al.*, 1988). PTE catalyzes the hydrolysis of synthetic organophosphate pesticides, particularly paraoxon. Although the bacterium first encountered these human-made compounds in the mid-20th century, PTE has evolved to have a catalytic efficiency (k_{cat}/K_M) for paraoxon hydrolysis that approaches diffusion-limited 'catalytic perfection' [$\geq 10^7$ $s^{-1}.M^{-1}$] (Omburo *et al.*, 1992). The evolution of such an efficient catalyst took less than fifty years, as organophosphates were synthesized for the first time in the 1930s and the paraoxonase activity was discovered in the 1980s (Lewis *et al.*, 1988). Further studies suggested that PTE has evolved from an ancestor with lactonase activity, which might have been involved in quorum quenching (Afriat *et al.*, 2006; Chow *et*

al., 2009). Consistent with Jensen's model, it seems likely that this ancestor possessed a weak affinity for phosphotriesters such as paraoxon, even before this activity was selected for (Roodveldt & Tawfik, 2005; Afriat *et al.*, 2006). Optimization of the enzyme's affinity and activity towards a new substrate — paraoxon — became rapid after organophosphates were introduced into the environment.

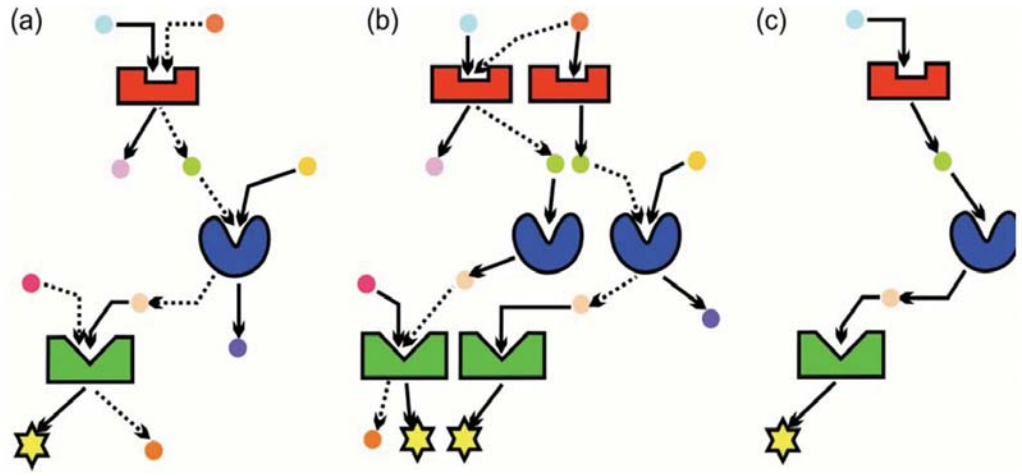


Figure 1.2: Jensen's model of patchwork evolution.

(A) Primordial enzymes may favour one substrate and/or catalytic mechanism (solid arrows), but are capable of promiscuously catalyzing other reactions (dotted arrows). Therefore, many metabolite chains synthesizing key metabolites may have existed. (B) Duplication of any of these primordial enzymes would increase the production of key metabolites, thereby increasing the fitness of the cell. (C) Further enzyme specialization would account for extant pathways. Figure taken from Holliday *et al*. (2007).

Similar to primordial enzymes, including the recent lactonase ancestor of PTE, many extant enzymes are now known to be promiscuous (see Section 1.3). O'Brien and Herschlag (1999) were the first to coin the term "catalytic promiscuity" to refer to enzymes that are able to catalyze different reactions. Another term, "substrate ambiguity" is used to describe the ability of enzymes to catalyze similar reactions on a range of different substrates (Khersonsky *et al.*, 2006). However, for the sake of convenience, in this thesis I broadly define "promiscuity" as the ability of proteins (including enzymes) to perform any secondary functions that have not been selected for during their evolution to date. The key feature of promiscuous proteins is that they are not completely specific. Instead, they possess "hidden talents" (Copley, 2003), that manifest in the form of broad substrate specificities or weak secondary activities. As explained in Jensen's model, these secondary activities or affinities are thought to act as starting points for the evolution of new functions (O'Brien & Herschlag, 1999; Copley, 2003; Khersonsky *et al.*, 2006; Khersonsky & Tawfik, 2010).

1.2 EVOLUTION OF NEW FUNCTIONS, FROM THE GENETIC PERSPECTIVE

Jensen's explanation for the evolution of new functions appears to be biochemically plausible. Yet, how can this model of enzyme evolution be reconciled with genetic mechanisms for the birth of new genes?

Fundamentally, new enzymes are the products of new genes. The primary source of novel genes is gene duplication and divergence from pre-existing ones (Lewis, 1951; Ohno, 1970). Duplication supplies a redundant copy of the gene that is free from selective pressure for its original function. This allows previously "forbidden" mutations to accumulate, potentially leading to a new function. Kimura and Ohta (1974) went as far as to state that gene duplication must always precede the emergence of a new gene with a new function. This evolutionary route to functionally novel proteins is referred to as the classic neo-functionalization model, or the model of Mutation During Non-functionality (MDN), and has been proposed as a fundamental process in adaptive evolution (Hughes, 1994).

The MDN model assumes that the newly duplicated gene is neutral until it acquires beneficial mutations or compensatory mutations that allow it to be selected (Ohta, 1988; Wagner, 2008). However, in the absence of selection, loss of the duplicate gene's function is expected to be much more frequent than gain of a new function (Hughes, 1994; Wagner, 1998). Common fates for the extra duplicate are loss via drift, inactivating mutations or recombinational segregation, and gene conversion, thereby resulting in the formation of pseudogenes (Walsh, 1995; Wagner, 1998). Hence, Force *et al*. (1999) proposed a subfunctionalization model that offers an alternative explanation for the maintenance of gene duplicates during the off-selection period. In this model, duplicate copies accumulate degenerate mutations, which partially damage the ancestral function of both genes to a point that neither could perform the original function well. This creates a selection pressure to maintain both subfunctionalized copies. While this strategy can buy time for the duplicated genes to remain in the population until selection starts to act on them, the likelihood of evolving a novel function is not clear. The reason is that the subfunctionalized (ancestral) function of the duplicates is still being selected, and therefore, neither of the genes is entirely free to acquire a new function.

Without selection, it seems unlikely that a duplicate gene could be maintained intact, at a high frequency in a population, until mutations accumulate and give it a new and selectable function. One way to maintain the duplicated gene is to assume that it remains under selection for its original function. However, by definition, this would limit its ability to lose its old function, as it gained a new one. This contradiction of the MDN model has been referred to as "Ohno's dilemma" (Bergthorsson *et al.*, 2007).

To address Ohno's dilemma, Bergthorsson and his colleagues (2007) proposed the Innovation, Amplification and Divergence (IAD) model for the evolution of new genes. This

IAD model emphasizes a role for secondary functions (*i.e.*, protein promiscuity), and invokes continuous selection during all processes from gene duplication and higher-order amplifications, to maintenance of the gene copies, to divergence of their functions. The model also predicts that gene amplification (as a means to increase gene dosage) is common, and this has been observed experimentally (Anderson & Roth, 1981; Andersson *et al.*, 1998; Hendrickson *et al.*, 2002). The three steps described by the model are as follows:

(i) *Innovation*

The parent gene encodes a primary activity that is being selected as well as a handful of minor activities that are neither under selection nor deleterious (Figure 1.3A). When the environment changes (*e.g.* by the introduction of a novel nutrient or toxin), one of the secondary activities becomes important and selection will favour an increase in its levels.

(ii) *Amplification*

Gene duplication and amplification events are at least four orders of magnitude more likely than point mutations (Bergthorsson *et al.*, 2007). Amplification increases gene dosage, and therefore, the level of the secondary activity that is under selection. Since the secondary activity is valuable under the selective condition, there is selection pressure to maintain the extra gene copies in the population (Figure 1.3B). Under continuous selection, gene amplification relieves the new gene copies from having to retain their original function (provided at least one copy retains it). This amplification step solves the basic problem of Ohno's dilemma.

(iii) *Divergence*

An increase in gene copies provides a higher number of targets for mutation and recombinational reassortment (Figure 1.3B). This increases the probability of improving the secondary activity of at least one gene copy, thereby relaxing the selection pressure on other copies. These extra copies can now be lost from the population via drift and inactivating mutations. On the other hand, the improved copy remains under selection for increasing its new activity. The original activity of the parental gene may (or may not) be maintained if it does not incur a fitness cost.

This IAD model is a testable hypothesis for the mechanism by which promiscuous activities drive the evolution of new genes. Its plausibility remains to be seen; yet, evidence that supports the different steps of the IAD model is emerging. This will be discussed in the following section.

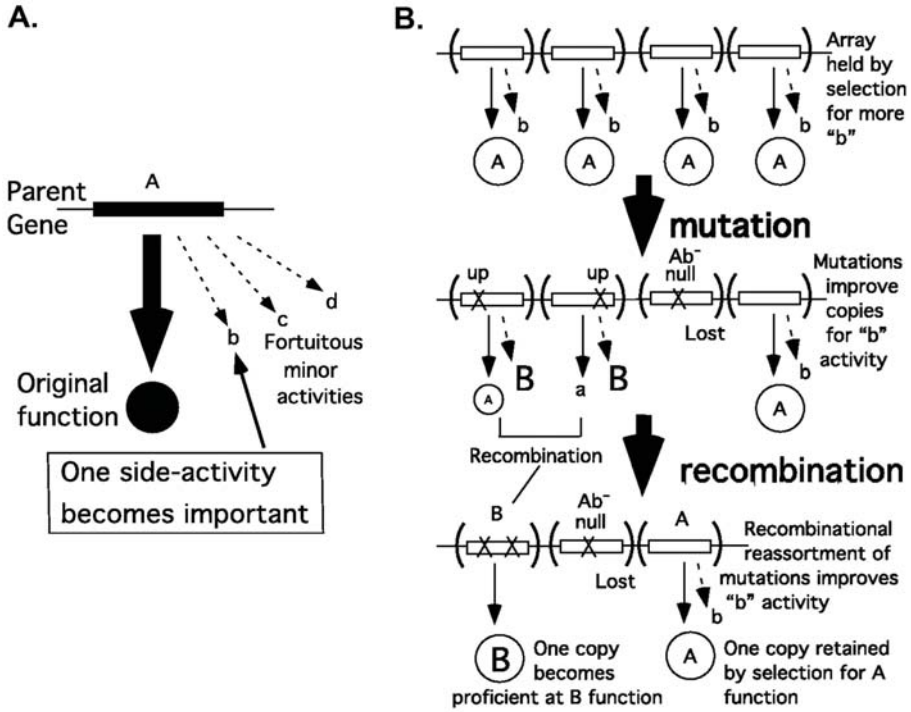


Figure 1.3: The IAD model of the evolution of new gene.

See text for details. (A) The Innovation step. (B) The Amplification and Divergence steps. Figure adapted from Bergthorsson *et al*. (2007).

1.3 EXPERIMENTAL VALIDATION OF THE IAD MODEL

Consistent with both Jensen's model and the IAD model, a growing number of studies suggest that secondary activities are common in extant enzymes. Multicopy suppression studies are a powerful and tractable way of uncovering these promiscuous activities (Berg *et al.*, 1988), and further demonstrating that adaptation can begin with the increased expression of promiscuous proteins. The latter is analogous to the gene amplification step described in the IAD model. Under selective conditions, elevated levels of promiscuous activities caused by increased gene dosage can provide a selective advantage to the cells. In multicopy suppression experiments, conditionally auxotrophic bacterial strains are typically transformed with plasmid-encoded open reading frames (ORFs) of interest, or plasmid-encoded genomic libraries. The resulting clones, with each harbouring different plasmid constructs, are screened for genes that can suppress the phenotype of the original mutation. At the outset of this thesis, a handful of promiscuous proteins had been identified from these suppression screens, including four transaminases with overlapping substrate specificities (Berg *et al.*, 1988), and an uncharacterized YjbQ protein that showed thiamine phosphate synthase activity (Morett *et al.*, 2008).

Many promiscuous activities are latent, and thus, they are not detectable unless their expression has been increased by using a strong promoter. To search for these latent activities, Miller and Raines (2004) constructed an *E. coli* genomic library in a plasmid with the T7 promoter. The authors then over-expressed the ORFs of the library in an auxotrophic *E. coli* strain that lacked the enzyme glucokinase, and was therefore a glucose auxotroph. Three suppressors were found: *nanK*, *yajF* and *ycfX*. Their gene products possessed weak glucokinase activity, although their activities were more than 10^4-fold lower than the endogenous bacterial glucokinase. Despite their ability to catalyze the glucokinase reaction, NanK, YajF and YcfX show only insignificant sequence identity (~7%) with *E. coli* glucokinase. This suggests that extant proteins with overlapping functions are not necessarily descendants of a single ancestral protein.

Non-homology-based promiscuity was further demonstrated by Patrick and his colleagues (2007). In their suppression screens, the authors made use of the *E. coli* ASKA and Keio libraries in an attempt to assess the prevalence of promiscuity, on a proteome-wide scale. The ASKA library consists of a plasmid collection all *E. coli* ORFs cloned into an over-expression vector [pCA24N] (Kitagawa *et al.*, 2005), whereas the Keio collection represents an *E. coli* mutant library consisting of all viable single-gene deletions (Baba *et al.*, 2006). The plasmids of the ASKA library were pooled, and used to transform 104 conditionally auxotrophic, single-gene deletion strains from the Keio collection. Twenty-one of the auxotrophies were specifically suppressed by the over-expression of non-cognate *E. coli* proteins (Patrick *et al.*, 2007). Among these, eight strains were rescued by non-homologous suppressor enzymes that appeared to have weak activity, or affinities for the substrate, of the deleted enzyme. The overall implication from this study and that of Miller and Raines (2004) is that protein promiscuity is a feature of not only primordial enzymes, but also extant enzymes. This fulfils the innovation step of the IAD model, for the evolution of new functions from pre-existing activities.

In some cases, promiscuous enzymes do not simply replace the functions of deleted enzymes. They can catalyze unrelated reactions, provided these reactions bypass the step that is blocked by gene deletion (Desai & Miller, 2008). A recent suppression study showed that several proteins with promiscuous activities could be patched together to form a new bypass pathway for synthesizing the vitamin B6 cofactor (Kim *et al.*, 2010). Promiscuous activities from three enzymes (and a non-enzymatic activity) diverted an intermediate from the serine biosynthetic pathway to a product that was subsequently incorporated in the cofactor biosynthesis pathway. The elegance of this work is two-fold: (i) promiscuous activities from extant enzymes are clearly contributing to metabolic and phenotypic robustness by providing an alternative pathway bypassing a blocked reaction; and more importantly, (ii) the patching of promiscuous activities typifies the evolution of new pathway as described by Jensen (1976).

Finally, the IAD model assumes that secondary activities are evolvable, *i.e.*, they can be readily improved by point mutations. *In vitro* evolution studies reveal that this is often the case. For example, the promiscuous lactonase and esterase activities of a PTE variant (also described previously in Section 1.1) could be improved with as few as one critical mutation [H254R] (Roodveldt & Tawfik, 2005). Two additional mutations (F306C and P342A) improved the esterase activity of PTE by two orders of magnitude, while decreasing the enzyme's native paraoxonase activity only by a factor of three (Aharoni *et al.*, 2005).

In a separate study, the *in vivo* TrpF (also described previously in Section 1.1) activity of *E. coli* PurF (glutamine phosphoribosylpyrophosphate amidotransferase) could be improved to almost wildtype levels using random mutagenesis (Patrick & Matsumura, 2008). Yet, the *in vitro* TrpF activity of the evolved PurF variant was $>10^7$-fold less than wildtype TrpF. These results implied that even very weak promiscuous activities could provide immediate selective advantages for the cells.

The emerging consensus from *in vitro* evolution studies is that promiscuity is readily improved by incorporating a few mutations (and often, without compensating the native activity). Along with Jensen's model, the IAD model appears to be a plausible route for the evolution of new enzymes. The models also suggest that protein promiscuity continues to play critical roles in the evolution of new enzymes. A comprehension of where and how new enzyme activities emerge will provide us with a better understanding of adaptive processes at the organismal level, as well as insights for biotechnological applications.

1.4 AIMS OF THIS STUDY

The goals of this research were to investigate how enzymes evolve, and where new enzymes originate from. The specific aims were:

1. To explore patterns of overlapping promiscuity in non-homologous enzymes.

Jensen's model predicts that two proteins with overlapping patterns of promiscuous activities are likely to be homologues (*i.e.*, descended from a common ancestor). However, there is an increasing number of examples in which a specialist enzyme (*e.g.* Glk, TrpF) and its promiscuous suppressor (Nan/YajF/YcfX, PurF) are not homologues (Miller & Raines, 2004; Patrick & Matsumura, 2008). In Chapters 2 and 3, I will describe a novel case in which two non-homologous enzymes are reciprocally promiscuous. Biochemical studies and directed evolution experiments extend Jensen's model, and confirm that cofactors can play a critical role in driving enzyme evolution.

2. To screen for promiscuous proteins that allow cells to adapt to novel environments.

Current efforts are focused on discovering promiscuous proteins that can complement auxotrophies when other pre-existing enzymes are deleted (see Section 1.3). It has been hypothesized that promiscuous activities can also serve as the evolutionary starting points for the emergence of entirely new biochemical functions (Bergthorsson *et al.*, 2007). However, the plausibility of this model has not been demonstrated experimentally, on a proteome-wide scale. In Chapters 4 and 5, the development of a comprehensive screen for promiscuous proteins is described. This screen was used to identify proteins that can enable improved growth when *E. coli* cells are exposed to novel nutrients and toxins. The results are presented in the context of the IAD model (Bergthorsson *et al.*, 2007).

On the whole, the first part of this thesis (Chapters 2 and 3) is analogous to "looking backwards in enzyme evolution", where I unravel the way evolution has shaped and optimized extant enzyme activities. The second part (Chapters 4 and 5) is about "looking forwards in enzyme evolution" — exploring the possible routes by which new functions may emerge in the future. These two themes (looking backwards and forwards in enzyme evolution) sum up this thesis.

CHAPTER 2

Cofactor-mediated promiscuity in alanine racemase

Acknowledgement: Trevor Loo (from the Norris Lab, Institute of Molecular Biosciences, Massey University, Palmerston North) performed the mass spectrometry analyses, as described in Section 2.4.8.

2.1 INTRODUCTION

Jensen (1976) has proposed that new catalytic functions can evolve from the secondary activities of pre-existing proteins. Enzymes that perform similar catalytic transformations therefore share significant sequence or structural homology because they have evolved from a shared ancestral scaffold. Examples include enzymes belonging to the same enzyme superfamily, such as those from the enolase and amidohydrolase superfamilies (Babbitt & Gerlt, 1997; Gerlt & Babbitt, 2001; Glasner *et al.*, 2006). Despite catalyzing different overall reactions, members from the same superfamily share a common structural fold, and a common partial reaction, intermediate or transition state.

However, several studies have also shown that proteins with no significant homology can catalyze the same reaction. For example, James and Tawfik (2001) found that human serum albumin and a catalytic antibody (that had been raised against a diketone hapten) were able to catalyze the Kemp elimination of benzisoxazole. The affinity and activity for benzisoxazole appeared to stem from a reactive lysine and apolar patches in the active sites of both non-homologous proteins. In a separate study, the affinity for similarly phosphoribosylated substrates allowed *E. coli* PurF (glutamine phosphoribosylpyrophosphate amidotransferase) to catalyze the isomerization of phosphoribosylanthranilate, which is the reaction ordinarily catalyzed by TrpF (phosphoribosylanthranilate isomerase) (Patrick & Matsumura, 2008). In spite of the sequence and structural dissimilarities, *E. coli* Δ*trpF* was rescued by the over-expression of PurF. This indicates that, unlike the Kemp elimination activity shown by the serum albumin and antibody (James & Tawfik, 2001), the promiscuous activity of PurF can be physiologically relevant in the absence of TrpF.

The TrpF activity of PurF is not a one-off example of molecular promiscuity by non-homologous enzymes. A proteome-wide screen for promiscuous enzymes revealed that 21 out of 104 single-gene deletion strains of *E. coli* were successfully rescued by one or more non-cognate *E. coli* genes (Patrick *et al*. 2007). In 30 of 41 cases, the suppressor gene and the deleted gene were unrelated. One particularly interesting case was the Alr/MetC combination — over-expression of Alr rescued *E. coli* Δ*metC* on minimal medium supplemented with glucose. The expression of MetC, or cystathionine β-lyase, is essential for *E. coli* growth on minimal media (Kim & Copley, 2007; Patrick *et al.*, 2007), due to its role in cleaving Cβ-S bonds during methionine biosynthesis. On the other hand, Alr (alanine racemase) is involved in alanine biosynthesis and is not known to possess any Cβ-S bond cleavage activity. Alr is not a sequence or structural homolog of MetC, although both MetC and Alr are dependent upon the cofactor, pyridoxal 5′-phosphate (PLP), to be catalytically active. The implication was that the promiscuous MetC activity of Alr was mediated by PLP. However, at the outset of this study, it was unclear how this PLP-mediated chemistry

gave rise to promiscuous activity in non-homologous scaffolds. The aim of this study was to investigate the basis for the PLP-mediated promiscuity in Alr.

2.1.1 The pyridoxal 5'-phosphate-dependent enzymes

Pyridoxal 5'-phosphate (PLP; Figure 2.1) is the biologically active form of the vitamin B6 cofactor. It has been dubbed the most versatile organic cofactor due to the catalytic and structural diversities associated with PLP-dependent enzymes (Schneider *et al.*, 2000; Percudani & Peracchi, 2003). Generally, PLP-dependent enzymes catalyze reactions involving the Cα, Cβ and Cγ atoms of amino acids, or compounds with an amino group. Examples of reactions catalyzed include transamination, α,β-elimination, β,γ-elimination, β-decarboxylation, β-replacement, γ-replacement, racemization and retro aldol cleavage. Since PLP-dependent enzymes are heavily involved in amino acid metabolism, it is not surprising that these enzymes are widespread among bacteria, archaea and eukaryotes. In prokaryotes alone, Percudani and Peracchi (2003) estimated that as many as 1.5% of all genes code for PLP-dependent enzymes. The same authors also reported that there were about 140 distinct PLP-dependent activities documented by the Enzyme Commission (EC), corresponding to ~4% of all classified catalytic activities at that time.

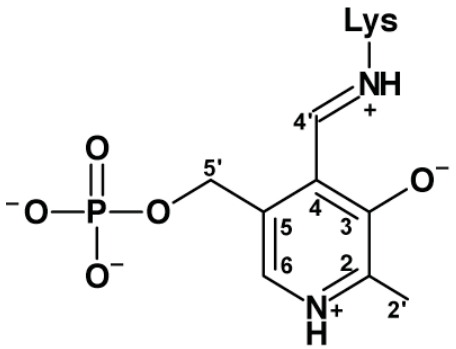


Figure 2.1: The structure of PLP, and the carbon numbering used throughout the thesis.
In the resting state, the PLP is bound to an active site lysine through an imine bond, forming an internal aldimine.

All PLP enzymes are classified according to their fold. It is widely accepted that there are five fold groups (Figure 2.2; Schneider *et al.*, 2000), although a recent study has proposed two additional groups, as new structures have been reported that are unlike any other PLP-dependent enzymes (Percudani & Peracchi, 2009). In this thesis, fold types I (MetC) and III (Alr) are of particular interest, and they will be discussed further throughout Chapters 2 and 3. Overall, there is no correlation between the fold type and the reaction catalyzed (Eliot & Kirsch, 2004). The prevalence of different reaction types within each fold group suggests

that extant PLP-dependent enzymes have evolved multiple times over history (Christen & Mehta, 2001). However, despite the structural dissimilarities, all five scaffolds harbour a similar "phosphate-binding cup" subdomain for anchoring the PLP molecule in the active site (Denesyuk *et al*., 2002).

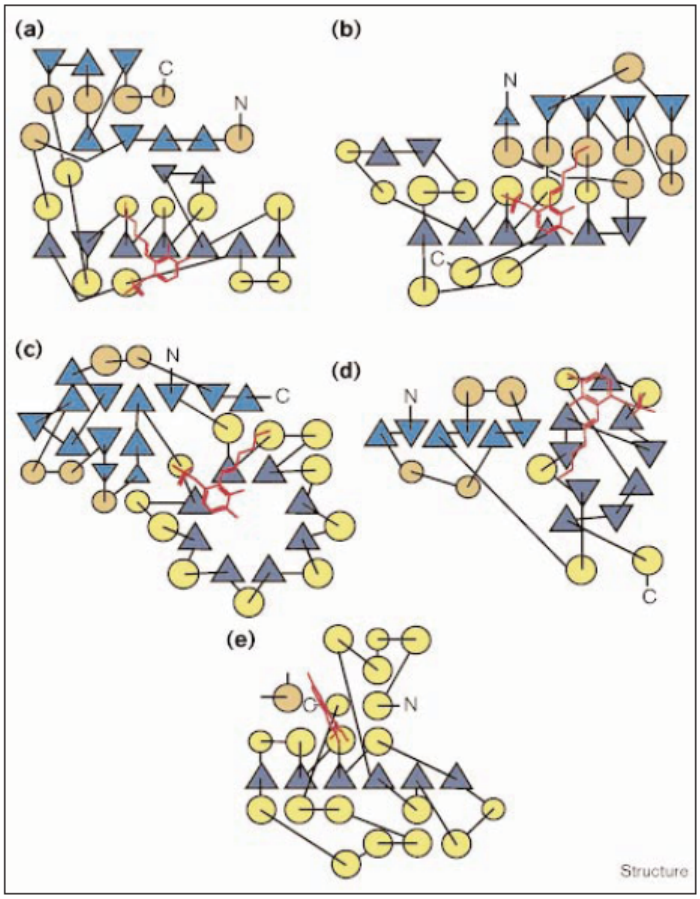


Figure 2.2: The five commonly accepted PLP-binding folds.

Circles represented α-helices, whereas triangles represented β-strands. The β-strands in indigo and α-helices in yellow depicted the PLP-interacting domains. The β-strands in blue and α-helices in orange depicted other domains. The PLP cofactor bound to the catalytic lysine was shown in red. (A) Fold type I. (B) Fold type II. (C) Fold type III. (D) Fold type IV. (E) Fold type V. Figure taken from Schneider *et al*. (2000).

The mechanistic features of catalysis are very similar in most PLP-dependent enzymes (except for glycogen phosphorylase from fold type V). This similarity is important in accounting for the diversity of PLP-dependent activities. Using serine as an example of a typical substrate, the possible outcomes catalyzed by PLP-dependent enzymes are illustrated in Figure 2.3. During the resting state, a PLP molecule is bound covalently to the ε-amino group of a catalytic lysine in the active site as an internal aldimine (Figures 2.1 and

2.3). The amine group of an incoming substrate then replaces lysine to form an external aldimine. From this point onwards, the reaction routes diverge depending on the conformation of the imine complex that is formed (Figure 2.3). Generally, the bond perpendicular to the planar PLP-imine-π-system will be broken (Dunathan, 1966). For instance, the substrate is deprotonated if the C-H bond (coloured in red, Figure 2.3) is perpendicular to the PLP-imine-π-system. The resulting negative charge on the carbanion intermediate is delocalized throughout the conjugated π-bonding system of PLP (Richard *et al.*, 2009). Following this, two outcomes are possible for the deprotonated carbanion intermediate. In a racemization step, the intermediate is reprotonated at Cα, on the opposite plane from which the initial proton was removed. Otherwise, a transamination reaction results from the protonation of the C4′ of the intermediate.

The substrate and reaction outcome for a PLP-dependent enzyme is largely dependent on the types of catalytic residues in the active site. As all PLP-dependent enzymes (except glycogen phosphorylase from fold type V) share the initial mechanistic steps (*i.e.*, the formation of internal and external aldimines), the imperfect control of the reaction and substrate selectivities occasionally result in side reactions (Eliot & Kirsch, 2004).

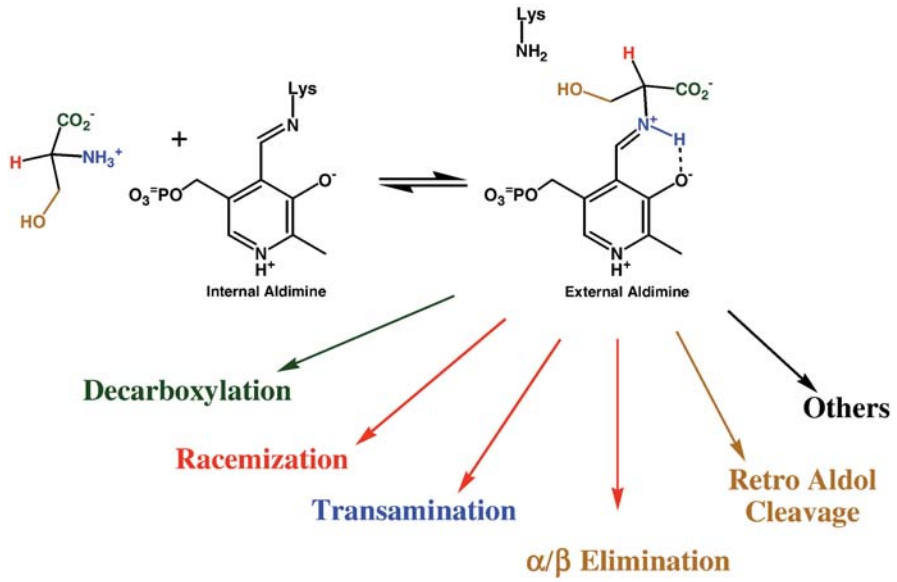


Figure 2.3: The possible catalytic routes taken by PLP-dependent enzymes.

A substrate example is serine, which is depicted on the left. In the first step of catalysis, the substrate displaces the catalytic lysine to form an external aldimine with PLP. The formation of the external aldimine is common in all PLP-dependent catalyses. Depending on the bond conformation and leaving groups (which are coloured differently), the possible catalytic routes are shown (in matching colours). Figure taken from Toney (2011).

2.1.2 Cystathionine β-lyase (MetC)

MetC (EC 4.4.1.8) produces L-homocysteine from L-cystathionine in the penultimate step of methionine biosynthesis (Figure 2.4A). MetC has been identified as a potential antibiotic target, as its inhibition impacts overall protein synthesis and DNA synthesis (via the methyl donor, *S*-adenosylmethionine) in bacteria (Ejim *et al.*, 2007). Unlike human MetC, prokaryotic MetC (and MetB, cystathionine γ-synthase) are parts of the reverse transsulfuration pathway that synthesizes methionine from cysteine (Wada & Takagi, 2006). In contrast, the transsulfuration pathway in humans converts methionine to cysteine. In *E. coli*, MetC is dispensable for growth on rich medium such as LB. In contrast, MetC is essential for growth on minimal medium supplemented with either glycerol or glucose as the carbon source (Joyce *et al.*, 2006; Kim & Copley, 2007; Patrick *et al.*, 2007). Similar to other methionine biosynthetic genes, the MetC expression in *E. coli* is negatively co-regulated by the MetJ repressor and *S*-adenosylmethionine (Weissbach & Brot, 1991).

E. coli MetC is a tetrameric protein, with each monomer consisting of 395 amino acids (Figure 2.4B). Each monomer adopts a type I PLP-binding fold (Clausen *et al.*, 1996) [also refer to Figure 2.2 in Section 2.1.1 for PLP-binding folds]. Within the active site of each monomer, one PLP molecule is bound to Lys210 (Belfaiza *et al.*, 1986; Clausen *et al.*, 1996). As with other enzymes of fold type I, the MetC enzyme has three functional domains. The N-terminal domain contributes to tetramerization, whereas the C-terminal domain consists of an anti-parallel β-sheet with α-helices on the solvent side. The middle domain harbours the PLP cofactor, which is found in a parallel seven-stranded β-sheet covered by adjacent α-helices.

In addition to L-cystathionine, MetC also recognizes alternative substrates such as L-cystine, L-cysteine, L-djenkolic acid, L-homolanthionine and L-*meso*-lanthionine (Uren, 1987). Degradation of L-cysteine to pyruvate, ammonia and hydrogen sulfite by MetC has been demonstrated using a crude cell extract (Awano *et al.*, 2003), but kinetic measurements of L-cysteine by purified MetC have remained a challenge so far. This is because the free thiol group of L-cysteine reacts with the coupling agent of the MetC assay (the assay details will be described in Sections 2.2.4.1 and 2.4.7.2). In the presence of a mild reducing agent (dithioerythritol), MetC was inhibited by excess L-cystine, and the maximal velocity with L-cystine is only 1% of that with L-cystathionine (Uren, 1987). Although the intracellular level of cysteine must be carefully regulated to prevent cytotoxicity (Sørensen & Pedersen, 1991; Ohtsu *et al.*, 2010), cysteine degradation in *E. coli* is also catalyzed by four other enzymes in addition to MetC, namely CysK (cysteine synthase A), CysM (cysteine synthase B), MalY (β-cystathionase/negative regulator of maltose regulon) and TnaA (L-cysteine desulfhydrase/tryptophanase) (Awano *et al.*, 2003; Awano *et al.*, 2005). Thus, cysteine degradation by MetC is much less likely to be physiologically relevant than its role in methionine biosynthesis.

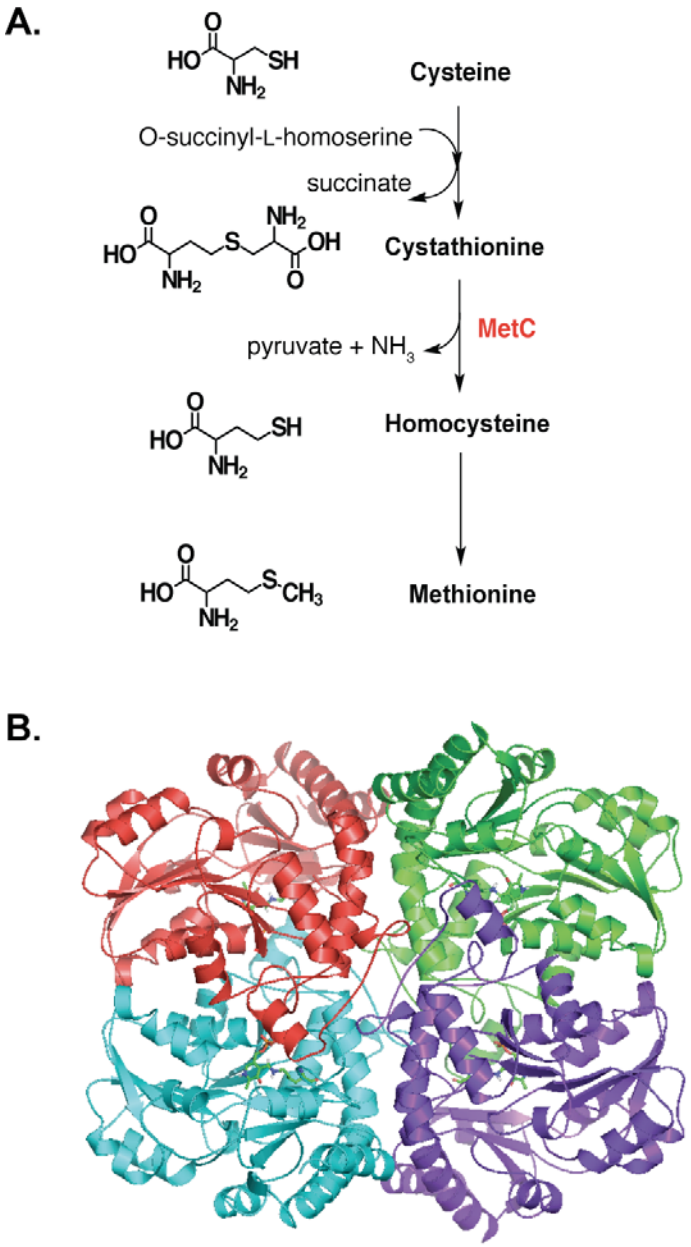


Figure 2.4: The function and structure of MetC.

(A) A simplified version of the methionine biosynthetic pathway in *E. coli*. MetC catalyzes the β-elimination of cystathionine, producing homocysteine, pyruvate and ammonia. Homocysteine is then methylated to form methionine. (B) The ribbon diagram of MetC's structure (PDB code: 1cl1), as reported by Clausen *et al*. (1996). Each monomer is shown in a different colour. Within each monomer, the PLP molecule bound to Lys210 is shown as sticks. Loops have been smoothed for clarity.

2.1.3 Alanine racemase (Alr)

Alr (EC 5.1.1.1) is an important enzyme in a bacterial cell, due to its role in synthesizing cell wall components. Alr mainly isomerizes L-alanine to D-alanine [Figure 2.6A; (Lambert & Neuhaus, 1972)], although the enzyme also catalyzes the reverse reaction. Both alanine isomers are constituents of the tetrapeptide linker found in the peptidoglycan layer of bacterial cell wall (Figure 2.5). D-alanine is a critical component of the tetrapeptide chain due to its role in interpeptide crosslinking within the peptidoglycan layer. The absence of D-alanine leads to a defective bacterial cell wall.

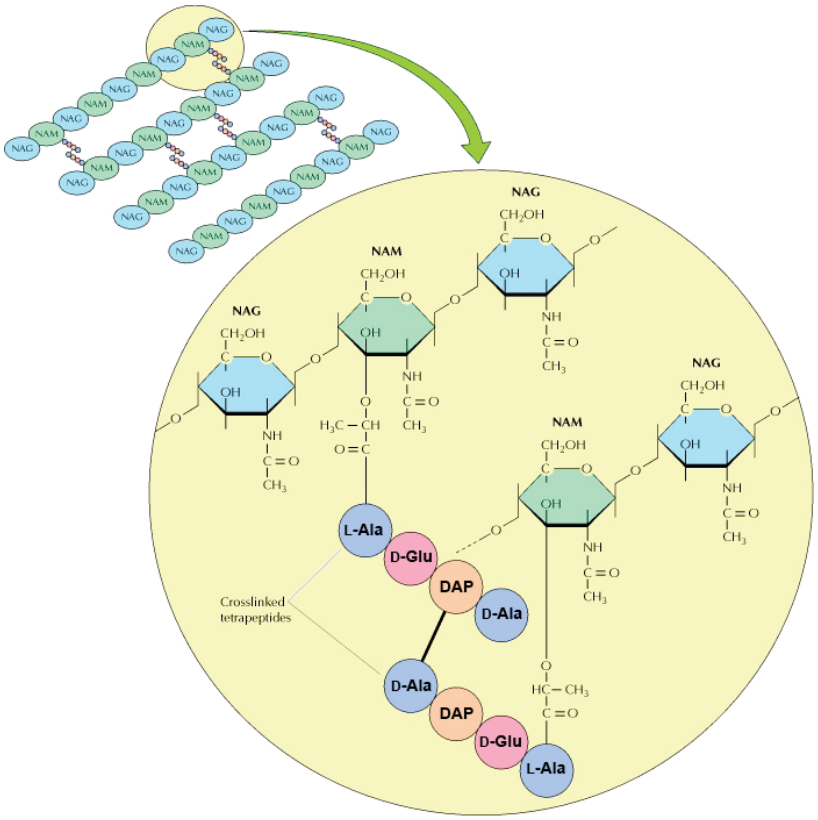


Figure 2.5: The peptidoglycan layer of *E. coli*.

The sugar backbone consists of repeating units of *N*-acetylglucosamine (NAG) and *N*-acetylmuramic acid (NAM), which are linked by β(1,4)-glycosidic bonds. Attached to every NAM residue is a tetrapeptide, which varies in different bacterial species. In *E. coli*, the tetrapeptide is L-Ala, D-Glu, *meso*-diaminopimelic acid (DAP) and D-Ala. An interpeptide bond links a DAP of one chain to D-Ala of the other chain. Figure adapted from Cooper (2000).

Unlike L-alanine, the only biosynthetic source of D-alanine in bacteria comes from the catalytic activity of alanine racemase(s). In *E. coli*, there is an extra alanine racemase, DadX, in addition to Alr. While both Alr and DadX catalyze the same reaction, the regulation of

their expression is different. Alr is expressed constitutively in the cell, whereas DadX is repressible by glucose. *E. coli* cells with both their *alr* and *dadX* genes deleted can not survive, unless D-alanine is supplied exogenously (Strych *et al.*, 2001). Similarly, Alr is essential for growth of mycobacterial species such as *Mycobacterium tuberculosis*, because they possess only one alanine racemase (Strych *et al.*, 2001; Milligan *et al.*, 2007). The essentiality of Alr in the biosynthesis of the bacterial cell wall renders it an attractive antibiotic target (Lambert & Neuhaus, 1972; LeMagueres *et al.*, 2005). The absence of Alr homologues in humans also indicates that Alr inhibitors should have minimal clinical side effects (Strych *et al.*, 2001). One of the clinical antibiotics used to treat tuberculosis is an Alr inhibitor, D-cycloserine (Koul *et al.*, 2011). However, the use of D-cycloserine causes neurotoxicities, which appear to stem from non-specific binding of the drug to other PLP-dependent enzymes in humans (Burman *et al.*, 1999; Di Perri & Bonora, 2004). This has led to the call for new Alr inhibitors (Anthony *et al.*, 2011), which are more specific to the enzyme but less toxic to patients.

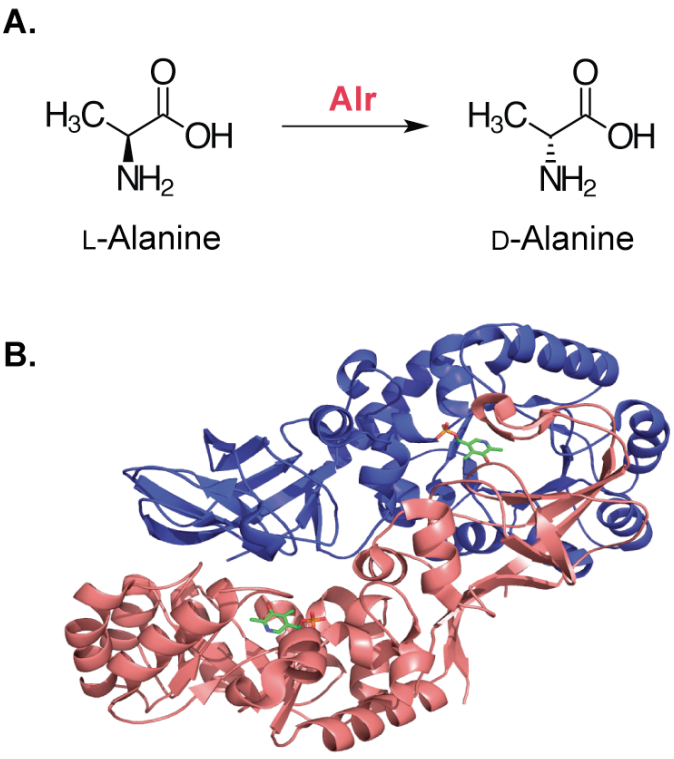


Figure 2.6: The function and structure of Alr.

(A) Alr catalyzes the interconversion of L-alanine and D-alanine. (B) The ribbon diagram of Alr's structure (PDB code: 2rjg), as reported by Wu *et al.* (2008). The monomers are shown in blue and red. Within each monomer, the PLP molecule bound to Lys34 is shown as sticks. Loops have been smoothed for clarity.

The native *E. coli* Alr protein is a dimer [Figure 2.6B; (Wu *et al.*, 2008)]. Each monomer is composed of 359 amino acids, and its overall structure adopts a type III PLP-binding fold (refer to Figure 2.2 in Section 2.1.1 for PLP-binding folds). Two domains are found in each monomer: a N-terminal domain consisting of a $(\beta/\alpha)_8$-barrel, and a C-terminal domain consisting mainly of β-strands. The core of the $(\beta/\alpha)_8$-barrel houses the active site, where the PLP cofactor binds covalently to Lys34. Proton transfer during alanine racemization is typically achieved by a catalytic lysine and a catalytic tyrosine (Watanabe *et al.*, 1999). In *E. coli* Alr, Lys34 and Tyr255′ (primed number denotes a residue from the other monomer) were identified as the catalytic pair (Wu *et al.*, 2008).

Similar to MetC, Alr is also known to catalyze secondary reactions. *E. coli* Alr can catalyze α,β-elimination, as evident from the elimination of halide from suicide substrates such as β-fluoroalanine, β,β-difluoroalanine, β,β,β-trifluoroalanine, β-chloroalanine, β,β-dichloroalanine and 3-halovinylglycine (Wang & Walsh, 1978; Wang & Walsh, 1981; Thornberry *et al.*, 1991). Alr from *Geobacillus stearothermophilus* was found to catalyze the transamination of both alanine isomers, albeit only once every 10^7 turnovers (Kurokawa *et al.*, 1998). The rate of transamination activity was increased by ~7-fold when two conserved residues (Tyr265′ and Arg219) were mutated (Yow *et al.*, 2003). In another study, Seebeck and Hilvert (2003) mutated Tyr265′ to Ala, and reported an increase of 10^5-fold increase in phenylserine aldolase activity. Taken together, these studies underline the intrinsic promiscuity of Alr, which can be easily accessed by non-synonymous mutations. However, the physiological and evolutionary implications of these promiscuous activities remain to be studied.

2.1.4 FOCUS OF THIS CHAPTER

I hypothesized that the promiscuous activity of Alr is responsible for rescuing the Δ*metC* strain of *E. coli*. This promiscuous activity has not been demonstrated previously. Alr and MetC enzymes are unrelated in terms of their sequence, structure, reaction type, substrate choice and participation in biosynthetic pathways. In this chapter, I offer genetic and biochemical evidence to confirm the promiscuous MetC activity of Alr. I also show that this promiscuous activity of Alr can be improved by directed evolution. This substantiates the notion that promiscuous activities from non-homologous protein scaffolds are evolvable (Patrick & Matsumura, 2008).

2.2 RESULTS

2.2.1 ALR OVER-EXPESSION RESCUED *E. COLI* Δ*metC*

All the results described in this chapter were built on the results reported by Patrick and his colleagues (2007). According to their results, the growth of *E. coli* Δ*metC* mutant on glucose-supplemented M9 medium was complemented by the over-expression of Alr. As this experiment was performed a while ago, and in a different laboratory, I set out to verify this phenotype. The growth complementation assays were set up based on the described methods (Patrick *et al*., 2007; Section 2.4.4). Instead of using their negative control plasmid (pCA24N-*lacZ*), a new negative control plasmid was constructed for this assay (pCA24N-NoIns; Section 2.4.1). The pCA24N-NoIns plasmid does not carry any gene insert, and its vector backbone is essentially identical to all other ASKA plasmids (pCA24N).

First, *E. coli* Δ*metC* cells were transformed with pCA24N-*metC*, pCA24N-*alr* and pCA24N-NoIns to form the positive control clone, test clone and negative control clone, respectively. It has been known that *E. coli* Δ*metC* cells alone were unable to grow on glucose-supplemented M9 medium (Kim & Copley, 2007; Patrick *et al*., 2007). Likewise, no growth was observed for the negative control clone (*E. coli* Δ*metC* harbouring pCA24N-NoIns) plated on the IPTG-supplemented M9/glucose agar, even after 3 weeks of incubation at 28°C. The positive control clone (*E. coli* Δ*metC* harbouring pCA24N-*metC*) took 2 days to form full-sized colonies on the IPTG-supplemented M9/glucose agar. Full-sized colonies were defined as colonies with a diameter of at least 1 mm. Similar to the reported results (Patrick *et al*., 2007), the Alr-over-expressing clone took 5–6 days to form full-sized colonies. Although the cells were incubated at a lowered temperature [28°C instead of 30°C as reported by Patrick and co-workers (2007)], the growth patterns of the MetC- and Alr-over-expressing clones, as well as the negative control clones were comparable to those previously reported (Patrick *et al*., 2007). All clones were grown on separate M9 agar plates to ensure that there was no cross-feeding. For every complementation assay, approximately 10^3–10^4 cfu of each clone were plated on the selective M9 agar. The complementation assays were individually repeated to ascertain that the growth results were reproducible. An example of growth complementation assay is illustrated in Figure 2.7.

Every gene insert in the ASKA plasmids is fused to a $(His)_6$ tag at the 5′ end, and a GFP tag at the 3′ end. As the GFP is a considerably large fusion protein (243 amino acids), it was not known if the promiscuous activity shown by the Alr-expressing clone (during growth complementation assays) was due to the Alr-GFP fusion. For this reason, the GFP tag was eliminated from pCA24N-*alr*, and the resulting plasmid was termed as pCA24N-*alr*(GFP−) [Section 2.4.2]. *E. coli* Δ*metC* cells over-expressing the Alr(GFP−) protein took 3 days to form colonies, and by the fourth day of incubation, full-sized colonies were formed. The reduced amount of time for the formation of colonies suggested that the plasmid-encoded GFP fused

to the C-terminus of Alr had reduced the latter's activity, possibly because the GFP hinders the access of the substrate to the active site. It was concluded that the GFP fusion tag did not endow promiscuous activity to Alr.

Figure 2.7: The growth of *E. coli* Δ*metC* clones on IPTG-supplemented M9/glucose agar after 4 days of incubation at 28°C.

Different clones were normally plated or streaked on separate agar plates. In this figure, different clones were streaked on the same agar plate for a comparative purpose. Each quadrant of the agar plate was streaked with the clone over-expressing the following plasmids (clockwise, from top left): (1) pCA24N-*metC*; (2) pCA24N-*alr*; (3) pCA24N-*alr*(GFP−); (4) pCA24N-NoIns.

For comparative purposes with the pCA24N-*alr*(GFP−) in subsequent growth complementation and enzyme activity assays, the GFP tag of the pCA24N-*metC* was eventually removed (Section 2.4.3). The resulting plasmid, pCA24N-*metC*(GFP−), was electroporated into *E. coli* Δ*metC* to form the new positive control clone. Cells over-expressing the MetC(GFP−) protein took only 1 day to form full-sized colonies on the IPTG-supplemented M9/glucose agar. Similar to Alr, the removal of GFP tag from the MetC protein also reduced the incubation period required (from 2 days to 1 day) for colony formation. To facilitate subsequent protein purification via the immobilized metal affinity purification approach (Section 2.4.5), the $(His)_6$ tags of both MetC and Alr proteins were retained.

No mutations were found in the *alr* inserts after sequencing the pCA24N-*alr* and pCA24N-*alr*(GFP−) plamids. This confirmed that the native Alr was sufficient to complement *E. coli* Δ*metC* on M9/glucose agar. A summary of the growth patterns of different *E. coli* Δ*metC* clones is presented in Table 2.1.

Table 2.1: Summary of the growth complementation of *E. coli* Δ*metC* and Δ*metE* clones on M9/glucose supplemented with IPTG (50 μM).

Over-expressed protein	Δ*metC*	Δ*metC* + L-methionine	Δ*metE*	Δ*metE* + L-methionine
MetC	2	NT	NT	NT
MetC(GFP−)	1–2	1–2	NT	NT
NoIns	No growth	1–2	No growth	2
Alr	5–6	NT	NT	NT
Alr(GFP−)	3–4	1–2	No growth	2
MetE	NT	NT	2	2

Each value represents the number of days that each clone took to form colonies on the agar medium. NT, Not tested.

2.2.2 *E. COLI* Δ*metC* CELLS WERE STARVED FOR METHIONINE

E. coli MetC has been shown to catalyze two reactions *in vitro*, depending on the substrates (Figure 2.8). MetC can catalyze either the degradation of cysteine (Awano *et al.*, 2003), or the Cβ-S bond breakage of cystathionine to generate homocysteine, a precursor for synthesizing methionine (Uren, 1987). However, the physiological importance of both activities was unclear. In other words, it was not known whether the growth of *E. coli* Δ*metC* on M9/glucose was limited by the accumulation of cysteine, or the lack of methionine, in the cells. My hypothesis was that the MetC deletion had impacted the methionine biosynthesis more than cysteine accumulation in *E. coli*. Therefore, the growth of *E. coli* Δ*metC* on M9/glucose was likely to be limited by the lack of methionine in the cells. Understanding the growth limitation of the Δ*metC* strain was important for determining whether Alr is involved in the methionine biosynthetic pathway, or the degradation pathway of L-cysteine.

To test the methionine auxotrophic state of the strain, the growth complementation assay was repeated with the M9/glucose agar that had been supplemented with L-methionine (100 μM). Three clones of *E. coli* Δ*metC* were tested: the positive control clone [pCA24N-*metC*(GFP−)], the test clone [pCA24N-*alr*(GFP−)] and the negative control clone (pCA24N-NoIns). As shown in Figure 2.9 (and also summarized in Table 2.1 in Section 2.2.1), all three Δ*metC* clones were able to grow on the methionine-supplemented M9/glucose agar. The rate of colony formation appeared to be identical among the clones, with full-sized colonies appearing after 1–2 days of incubation at 28°C. The confluent growth of the *E. coli* Δ*metC* clones, especially the negative control (Plate 2 in Figure 2.9), appeared to be consistent with the initial assumption that methionine starvation was the limiting factor in the growth of the Δ*metC* strain. In contrast, if cysteine accumulation were the factor that limited the growth of *E. coli* Δ*metC*, the negative control clone would not have

been able to grow, even in the presence of exogenous methionine. These results confirmed my hypothesis that the *in vivo* promiscuous activity of Alr is likely to be involved in the methionine biosynthesis (*e.g.* the formation of homocysteine from cystathionine), instead of the breakdown of cysteine.

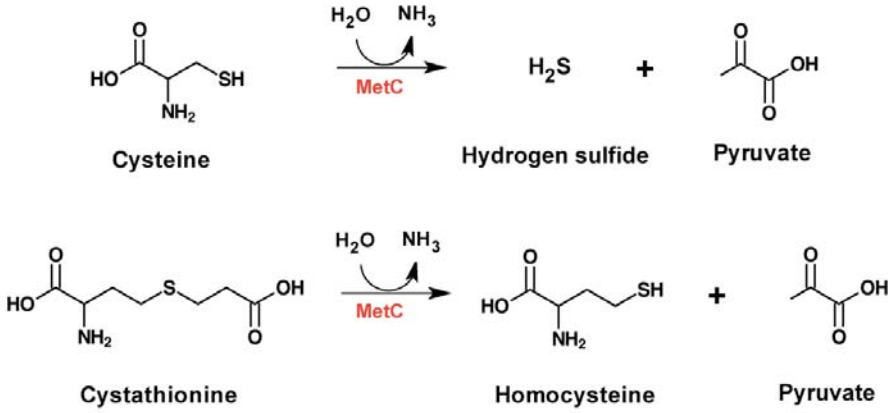


Figure 2.8: The catalytic activities of MetC with two different substrates, cysteine and cystathionine.
In the first reaction, the MetC-catalyzed cysteine degradation is shown. In the second reaction, MetC converts cystathionine into homocysteine during the biosynthesis of methionine.

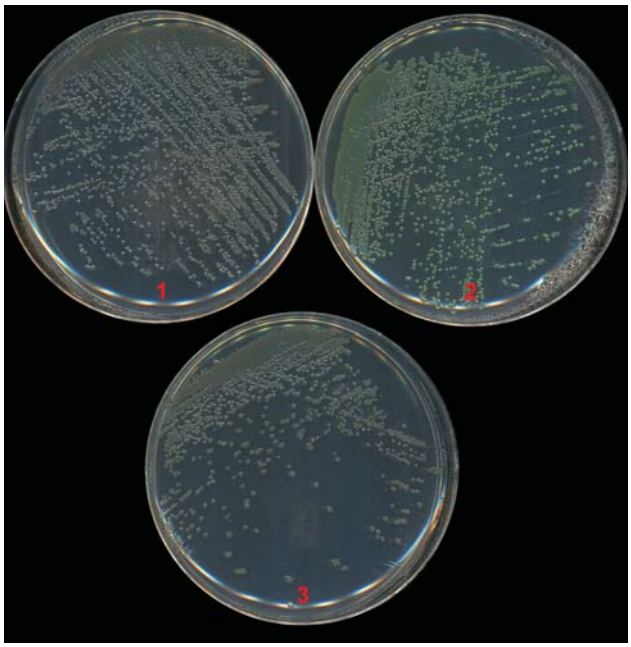


Figure 2.9: The growth of *E. coli* Δ*metC* clones on M9/glucose agar containing 100 µM L-methionine and 50 µM IPTG after 2 days of incubation at 28°C.
The numbered agar plates were streaked with *E. coli* Δ*metC* harbouring the following plasmids: (1) pCA24N-*metC*(GFP−); (2) pCA24N-NoIns; (3) pCA24N-*alr*(GFP−).

2.2.3 ALR OVER-EXPRESSION DID NOT RESCUE *E. COLI* Δ*metE*

I also considered another possibility, in which the promiscuous activity of Alr might produce the methionine via a novel metabolic bypass. In this case, over-expression of Alr would still be able to rescue the Δ*metC* auxotroph on M9/glucose medium. In the final step of *de novo* methionine biosynthesis, a methyl group is transferred to the MetC-generated homocysteine via the activity of either the cobalamin-dependent MetH (homocysteine transmethylase) or the cobalamin-independent MetE (methionine synthase) [Figure 2.10]. If *alr* had encoded a bypass pathway for methionine biosynthesis, over-expression of the Alr would also rescue *E. coli* deleted for enzymes that catalyze the final step of the biosynthetic pathway.

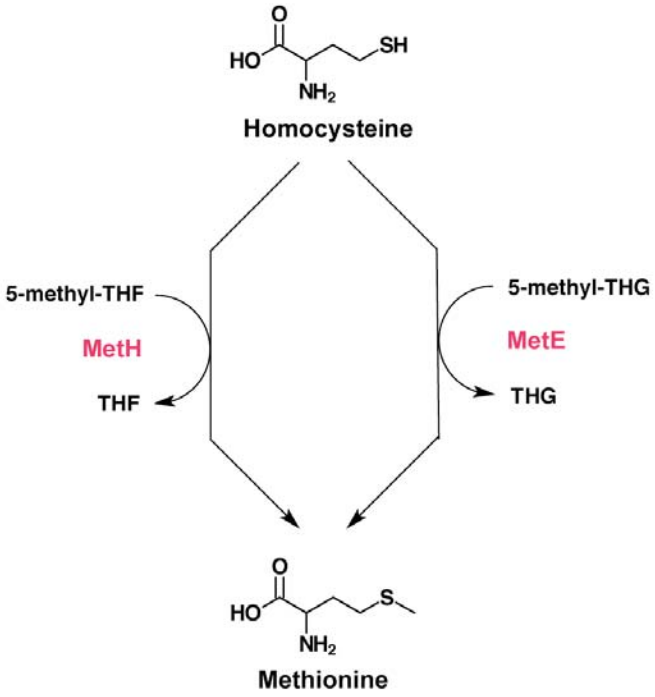


Figure 2.10: The final step of *de novo* methionine biosynthesis in *E. coli*.
The transfer of a methyl group to homocysteine can be catalyzed by either MetH or MetE. Both 5-methyl-THF and 5-methyl-THG are methyl donors. THF, tetrahydrofolate; THG, tetrahydropteroyltri-L-glutamate.

Previous studies have shown that *E. coli* Δ*metH* is able to grow on glucose (Kim & Copley, 2007; Patrick *et al.*, 2007), due to a functional copy of *metE* in the cells. In contrast, the cobalamin-dependent MetH in the Δ*metE* strain is unlikely to be functional due to the absence of cobalamin in the minimal medium. Hence, *E. coli* Δ*metE* is unable to grow on the

M9/glucose medium, unless a plasmid-encoded MetE protein was provided (Patrick *et al*., 2007).

Therefore, the Δ*metE* strain (from the Keio collection) was transformed with pCA24N-*metE*, pCA24N-NoIns and pCA24N-*alr*(GFP−) to form the positive control clone, the negative control clone and the test clone, respectively. The growth of the various Δ*metE* clones was assessed on M9/glucose medium supplemented with IPTG. Figure 2.11A depicted the growth of different *E. coli* Δ*metE* clones on M9/glucose medium containing IPTG. The Δ*metE* strain was rescued by over-expressing the plasmid-encoded MetE, but not by the plasmid-encoded Alr or the empty vector. The agar plates were incubated at 28°C for up to one week, but no growth was observed for the Alr-expressing and the negative control clone by the end of incubation. The lack of growth for the Δ*metE* clone harbouring the pCA24N-*alr*(GFP−) was also consistent with the previous observation — the Δ*metE* mutation was not suppressed by other gene, except *metE* (Patrick *et al*., 2007). When the same set of the Δ*metE* clones were plated on M9/glucose agar supplemented with methionine (and IPTG), full-sized colonies were observed for all clones after 2 days of incubation at 28°C (Figure 2.11B). Overall, the growth complementation results (tabulated in Table 2.1 of Section 2.2.1) supported the hypothesis that the *in vivo* promiscuous activity of Alr is the cystathionine β-lyase activity that is normally catalyzed by MetC.

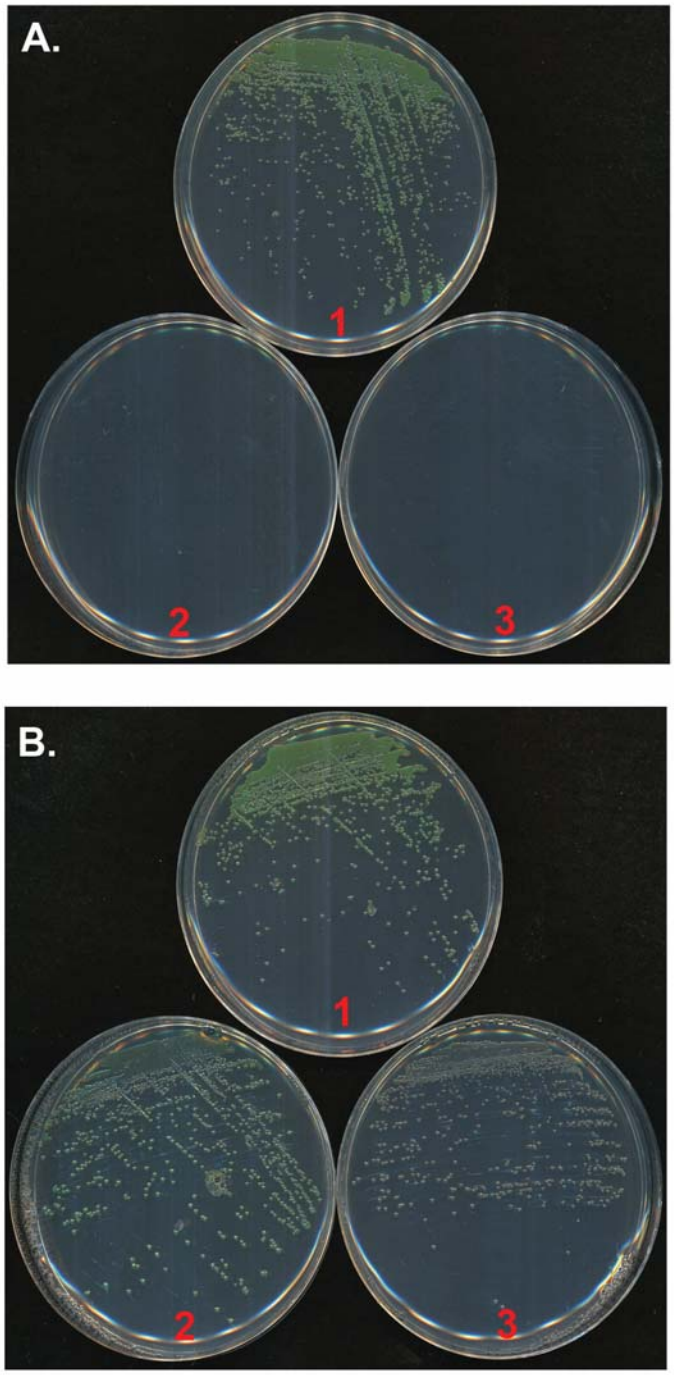


Figure 2.11: The growth of *E. coli* Δ*metE* clones on (A) M9/glucose alone; and (B) M9/glucose with L-methionine (100 μM) after 2 days of incubation at 28°C.

All M9/glucose agar plates were supplemented with 50 μM IPTG. The numbered agar plates were streaked with *E. coli* Δ*metE* harbouring the following plasmids: (1) pCA24N-*metE*; (2) pCA24N-NoIns; (3) pCA24N-*alr*(GFP−).

2.2.4 Cystathionine β-lyase activities of MetC and Alr

2.2.4.1 Kinetic activities of cystathionine β-lyase

Throughout Sections 2.2.1 to 2.2.3, I present genetic evidence for the *in vivo* cystathionine β-lyase activity in *E. coli* Alr. The *in vitro* cystathionine β-lyase activities of MetC and Alr are tested in this section.

E. coli Δ*metC* was used as the host strain to over-express the MetC enzyme from pCA24N-*metC*(GFP−). For the Alr enzyme, *E. coli* MB2795 (Δ*alr* Δ*dadX*; see Section I.3 in Appendix I) was used as the host strain. Each MetC subunit was fused to a 17-amino acids linker, including the $(His)_6$ tag, at the N-terminus. Similarly, each Alr subunit was fused to the same linker at the N-terminus. Both MetC and Alr were largely soluble, and were $(His)_6$-tagged purified to >95% purity (see Section 2.4.5 for more details), as judged by SDS-PAGE (Figure 2.12). The molecular sizes of both MetC and Alr subunits were also verified by SDS-PAGE (MetC, 45 kDa; Alr, 41 kDa).

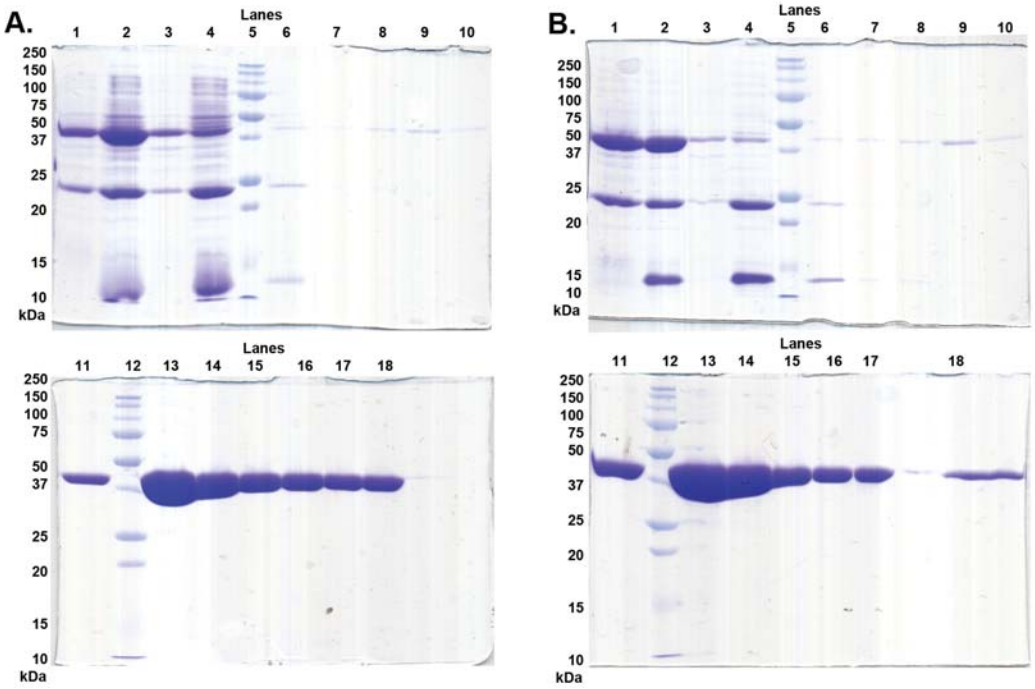


Figure 2.12: SDS-PAGE analyses of the purified (A) MetC and (B) Alr.

Each lane was loaded with a 6-μL aliquot from the corresponding fractions. Lane 1, Crude cell lysate after 6 hours of protein over-expression. Lane 2, Soluble lysate. Lane 3, Insoluble fraction. Lane 4, Proteins not bound to the Talon® metal affinity resin. Lanes 5 and 12, Bio-Rad Precision Plus protein All Blue standard. Lanes 6–7, Batch washes. Lane 8, Flow-through from the enzyme-bound resin in the gravity flow column. Lanes 9–10, Column washes with low imidazole concentration (3 mM). Lane 11, Column wash with high imidazole concentration (15 mM). Lanes 13–18, Elution fractions 1–6.

Both enzymes were dialyzed extensively to remove the reducing agent, β-mercaptoethanol, from the elution buffer (see Section 2.4.5 for the recipe of elution buffer). Thiol-containing compounds (such as β-mercaptoethanol) interfere with the cystathionine β-lyase activity assay (see Figure 2.13 and Section 2.4.7.2 for details). After dialysis, the enzymes were quantified as described in Section 2.4.6. The yield of purified MetC always exceeded 18 mg/L of culture, whereas purified Alr always yielded more than 20 mg/L of culture. Throughout this study, at least 70% of either MetC or Alr were catalytically active, as defined by the ratio of the concentration of PLP-bound enzymes to the total enzyme concentrations (see Section 2.4.6 for details).

The continuous assay system that was described by Uren (1987) was modified and used to measure the cystathionine β-lyase activity (refer to Section 2.4.7.2 for details). In this assay, a coupling agent [DTNB; 5,5′-dithiobis(2-nitrobenzoic acid); (Ellman, 1959)] is added to react with the product of the reaction, homocysteine [Figure 2.13]. One mole of TNB^{2-} (2-nitro-5-thiobenzoate dianion) is generated for every mole of homocysteine reacted with DTNB. Since the TNB^{2-} ion absorbs strongly at A_{412}, the increase of A_{412} can be used as a proxy to measure the rate of homocysteine formation.

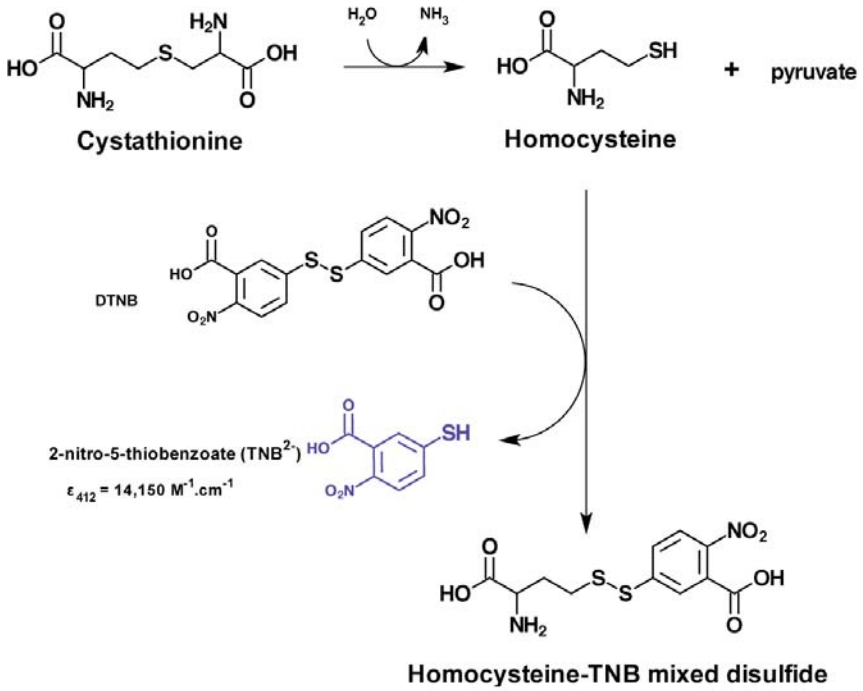


Figure 2.13: The DTNB assay used to measure the *in vitro* cystathionine β-lyase activity.

The homocysteine produced by the cystathionine β-lyase catalysis is reacted with DTNB, to form TNB^{2-} (in blue). The TNB^{2-} anion was measured spectrophotometrically at 412 nm, using an extinction coefficient of 14,150 $M^{-1}.cm^{-1}$. DTNB, 5,5′-dithiobis(2-nitrobenzoic acid); TNB, nitro-thiobenzoate.

First, the kinetic parameters were determined for *E. coli* MetC. The K_M for cystathionine was 39 ± 7 μM, whereas the k_{cat} was 90 ± 14 s^{-1}. The final values were the mean of three repeats, and each repeat was obtained from the average of duplicate assays. While the K_M value was very close to that obtained by Uren (1987) [39 μM *versus* 40 μM], the k_{cat} was approximately 2.5-fold higher than that reported by Lodha and Aitken (2011). Overall, the catalytic efficiency (k_{cat}/K_M) of MetC was 2.3 × 10^6 $s^{-1}.M^{-1}$, which is about 10-fold higher than the catalytic efficiency reported by Lodha and Aitken (2011). The slight difference in the k_{cat} and k_{cat}/K_M values between this study and Lodha and Aitken (2011) is likely due to the the higher temperature (37°C, instead of 25°C) used to assay the activity in this study.

No cystathionine β-lyase activity was detected for Alr up to 10 mM L-cystathionine. The use of higher L-cystathionine concentration was not possible due to low solubility of L-cystathionine (~5.6 mg/mL in 0.04 N HCl). The activity assays of Alr were also monitored at different temperatures (25°C, 30°C, and 37°C) for prolonged duration (30 min), but there was still no detectable cystathionine β-lyase activity. The native activity of Alr (alanine racemization) was detectable (refer to Table 2.7 in Section 2.2.6.5), hence confirming that the purified Alr was not inactive. Throughout this study, the *in vitro* cystathionine β-lyase activity of Alr could not be detected even with Alr aliquots purified from different batches. Overall, the kinetic parameters of the cystathionine β-lyase activity in Alr could not be established using the steady state enzyme kinetics approach.

2.2.4.2 ENDPOINT ASSAYS AND MASS SPECTROMETRY

The undetectable cystathionine β-lyase activity of Alr in the spectrophotometric assay was possibly due to an overall low catalytic turnover of L-cystathionine. The low substrate turnover could be explained by two scenarios: (i) the enzyme was assayed under non-physiological conditions [Tris-HCl buffer (pH 8.8), lack of intracellular reducing agents, diluted enzyme and substrate concentrations, *etc.*]; and (ii) the enzyme's activity and/or affinity for L-cystathionine was very low. Instead of measuring the rate of the cystathionine β-lyase activity in Alr within a short period of time (~30 min), I attempted to demonstrate the catalytic turnover of cystathionine by Alr using endpoint assays.

In the first instance, MetC (0.8 μM, final concentration) and Alr (5.1 μM, final concentration) were reacted with 5 mM L-cystathionine in 50 mM Tris-HCl assay buffer (pH 8.8) for 16 h. A blank control containing no enzyme was also included. All reactions contained DTNB for detection of the product, homocysteine. The reactions were incubated in the dark, at 37°C. At the end of the incubation, MetC and Alr were removed from the reactions (see Section 2.4.8 for details).

Figure 2.14A shows the absorbance spectra of the enzyme-free reaction products. After a 16-h incubation, it was evident that the homocysteine produced by MetC had formed mixed disulfides with DTNB, and hence, generated TNB^{2-} ions in the reaction. The TNB^{2-}

ions gave a strong absorbance peak at 410 nm (purple line, Figure 2.14A). In contrast, no peaks at 410 nm were observed for Alr and the blank control. Instead, both reactions displayed an absorbance peak at 325 nm, which was attributed to unreacted DTNB (Figure 2.14B). The *in vitro* cystathionine β-lyase activity of Alr was still undetectable by endpoint assays alone.

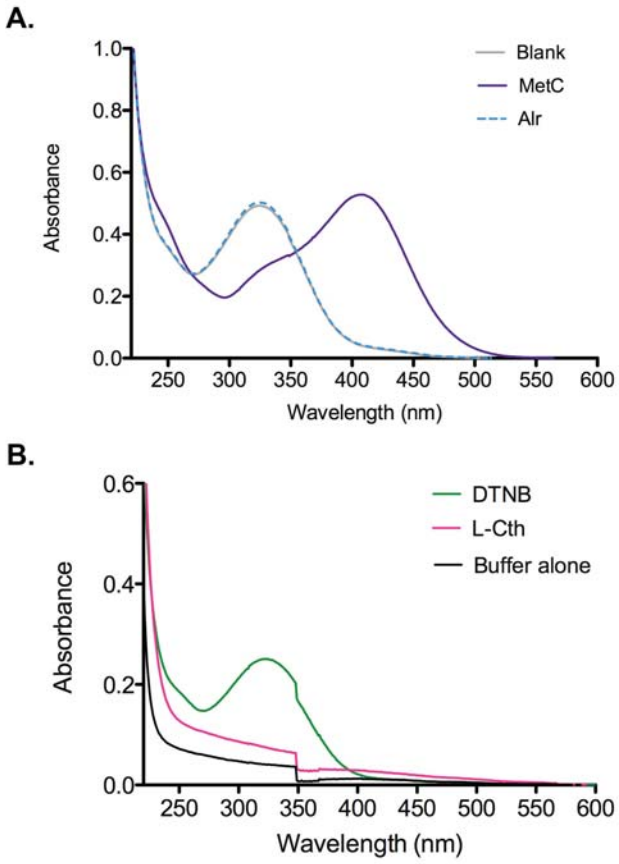


Figure 2.14: The absorbance spectra of the endpoint reactions, and different components of the reactions.
(A) The peak at 410 nm in the MetC reaction (purple line) was ascribed to the TNB^{2-} ions, which had been generated from the sulfide interchange of DTNB by homocysteine, the product of the cystathionine β-lyase activity. The peak at 325 nm in the Blank and Alr reactions was attributed to unreacted DTNB. (B) Freshly prepared DTNB generated a peak at 325 nm, which was similar to those observed previously in the Blank and Alr reactions. The assay buffer and L-cystathionine did not yield any absorbance peak from 220–600 nm wavelengths.

To determine whether Alr could turnover cystathionine in endpoint assays, the next logical step was to employ a more sensitive method of detecting the reaction products. Electrospray mass spectrometry was chosen for this purpose. For mass spectrometry analyses, the endpoint assays were repeated without DTNB. The Tris-HCl assay buffer was replaced with a NH_4HCO_3 buffer (see Section 2.4.8 for details). The reason was that the

NH_4HCO_3 buffer would evaporate as NH_3 and CO_2 when the samples were concentrated under vacuum, and hence, no buffer components would be detected in the mass spectrometer. Preliminary assays indicated that the use of NH_4HCO_3 buffer had no effect on the cystathionine β-elimination activity — the initial rates (V_0) of MetC activity with various concentrations of L-cystathionine in the NH_4HCO_3 buffer were similar to those performed in the Tris-HCl buffer (results not shown).

The mass spectra of the reaction products are shown in Figure 2.15. For the positive control in which MetC was added, homocysteine (observed mass, 135.04 Da; predicted mass, 135.18 g/mol) was observed as a dominant species (Figure 2.15B). No cystathionine (222.07 Da) was detected, thereby confirming that all cystathionine had been turned over by MetC. Zoomed-in areas of the spectra revealed that there was a small peak at 135 Da that corresponded to the mass of homocysteine, in the Blank, Alr and Alr+PLP reactions (Gill Norris, personal communication). However, the size of this peak in these reactions was similar, implying that no significant amount of homocysteine was formed in the reactions containing Alr and Alr+PLP. As the detection threshold of the mass spectrometer is in the range of femtomoles (Gill Norris, personal communication), and the amount of Alr was ~1 nmole per reaction, the upper limit for Alr's turnover of L-cystathionine was estimated to be $<6.1 \times 10^{-8}$ per h per Alr molecule. This turnover rate seems too low to be physiologically relevant. It was possible that the *in vitro* assay conditions were too far from the *in vivo* conditions to accurately estimate the activity of Alr *in vivo*.

For the Alr reaction, two peaks were observed (Figure 2.15C). The peak at 222.07 Da corresponded to cystathionine (predicted mass, 222.26 g/mol), while the identity of the other peak at 133.02 Da remains unknown. These two peaks were also present in the Blank control and Alr reaction supplemented with 50 μM PLP (Figures 2.15A and D). The implications were: (i) the 133-Da species was related to the substrate (L-cystathionine) used in the reaction; and (ii) that adding excess PLP to the reaction did not increase the cystathionine β-lyase activity of Alr. The appearance of the 133-Da species is intriguing, because no biological compound involved in either methionine or cysteine metabolism has a molecular mass of 133 Da. Several compounds related to methionine and cysteine metabolism in *E. coli* are O-succinyl-L-homocysteine (218 g/mol), succinate (118 g/mol), pyruvate (88 g/mol), methionine (149 g/mol), cysteine (121 g/mol) and homoserine (119 g/mol). Due to the fact that the 133-Da compound was also present in the Blank, the 133-Da compound was likely a decomposed product, or an impurity, of L-cystathionine. The L-cystathionine stock in our lab was of TLC-grade, and a reported purity of ~90% (Sigma). The 133-Da compound might have originated from the remaining 10% impurity. Curiously, this peak (at 133.02 Da) was reduced by ⅞ in the MetC reaction, thereby suggesting that MetC could turnover this 133-Da impurity (Gill Norris, personal communication).

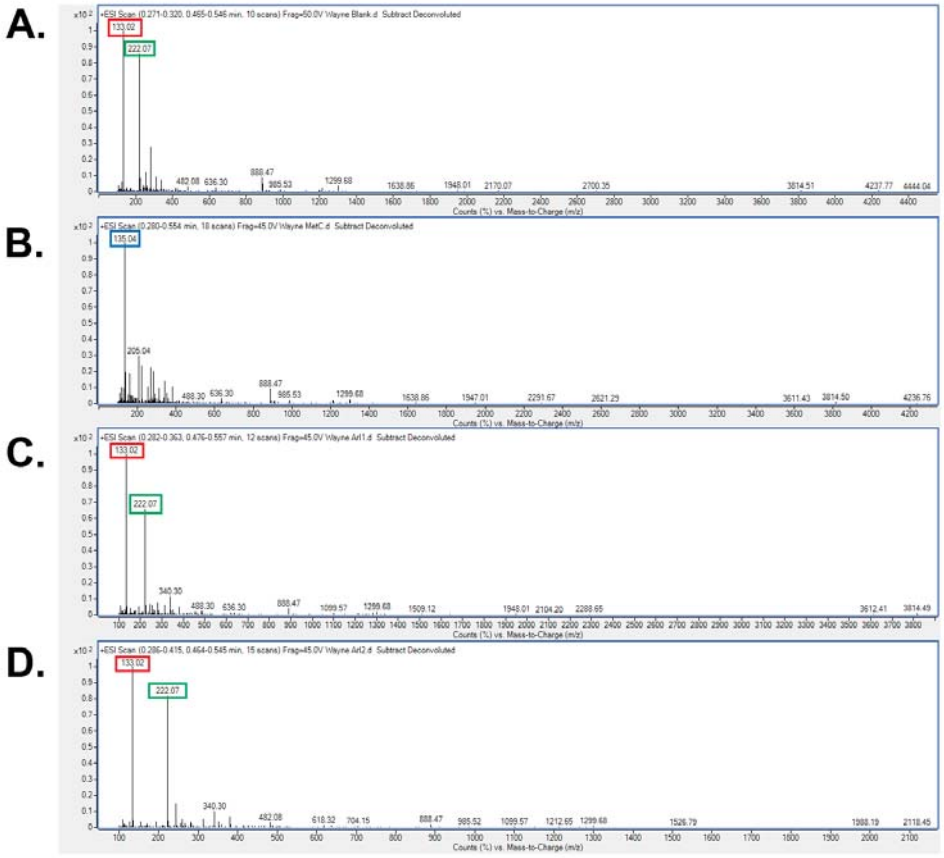


Figure 2.15: Mass spectra of the reaction products from endpoint assays.

(A) Blank control. (B) MetC. (C) Alr. (D) Alr + 50 μM PLP. The species with a m/z value of 135.04 corresponded to homocysteine (blue-boxed; panel B). The species with a m/z value of 222.07 corresponded to cystathionine (green-boxed; panels A, C and D). The identity of the species with a m/z value of 133.02 (red-boxed; panels A, C and D) is not known.

Two compounds with a mass of 133 Da were inferred (Figure 2.16), based on logic and species searches from the National Institute of Standards and Technology (NIST) database (2011). The compounds were α-keto-γ-thiobutyric acid and aspartic acid, although it is not known if the former is of biological origin. Efforts in determining the identity of the 133-Da compound, and confirming whether it has been originated from L-cystathionine, are under way (Gill Norris, personal communication).

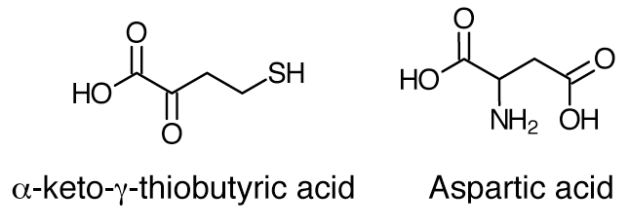


Figure 2.16: Inferred compounds with a molecular mass of 133 Da.

2.2.4.3 pH EFFECT ON THE CYSTATHIONINE β-LYASE ACTIVITY OF ALR

The *in vitro* cystathionine β-lyase activity was assayed at alkaline pH, because the coupling reagent (DTNB) is highly stable in mildly alkaline conditions (Ellman, 1959; Riddles *et al.*, 1979). Furthermore, MetC and Alr are known to function optimally between pH 8–9 (Dwivedi *et al.*, 1982), and 8.5–11 (Esaki & Walsh, 1986; Kurokawa *et al.*, 1998), respectively. Yet, the pH conditions where the promiscuous activity of Alr was observed during the growth complementation assays (Section 2.2.1) were likely to be near neutral or slightly acidic (due to accumulation of metabolic waste from cell growth). The disparity of the pH conditions between *in vivo* and *in vitro* assays prompted a preliminary study on the effect of pH on the promiscuous activity of Alr.

Figure 2.17 shows the effect of low pH on the *in vitro* activity assays for Alr. Each reaction was initiated by the addition of L-cystathionine (5 mM, final concentration) at t = 0.5–0.8 min. Although the cystathionine β-lyase activities were low (as depicted by the increasing slope of the activity curves), the overall trend suggested that they were detectable at low pHs. The biphasic curves also hinted at burst phase kinetics for the enzymes under these conditions. The amount of product that was produced during the putative burst phase was calculated from the total increase in absorbance over the range t = 0.8–1.2 min (pH 6.0), and t = 0.85–1.4 min (pH 5.5). A total of 1.4 μM of product formed at pH 6.0, and 1.0 μM of product formed at pH 5.5. As expected for a burst phase, this is approximately the same as the total of Alr concentration (0.8 μM) in each assay; that is, $\Delta c \approx [E]_T$. The implication was that release of the spectrophotometrically active product (homocysteine) from the Alr active site was more rapid than the turnover of the enzyme. While preliminary, this interpretation is consistent with the hypothesis that the second reaction product, pyruvate, is released from the MetC active site after homocysteine (Clausen *et al.*, 1996). However, the kinetics of Alr catalyzing the MetC reaction at low pH need to be investigated further, due to the instability of the enzyme in acidic conditions.

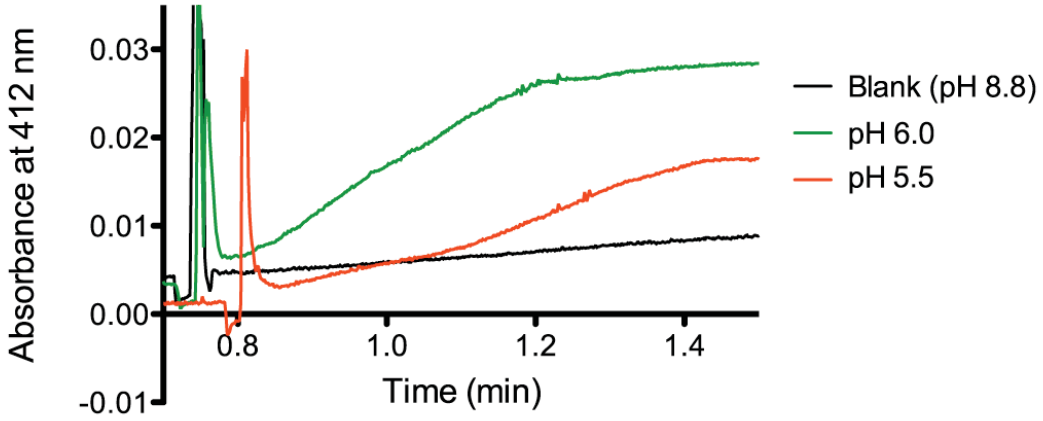


Figure 2.17: The cystathionine β-lyase activity of Alr assayed at pH 5.5 and 6.0.

2.2.5 Enzyme activities of Alr(R209E)

The preliminary results in the previous section (see Figure 2.17) suggested that a low pH environment might be favourable to the cystathionine β-lyase activity of Alr. This begged the question whether the protonation state of PLP was affected by pH changes. Therefore, I used site-directed mutagenesis to change a key residue in Alr that might alter the protonation state of PLP.

In Alr, the pyridine nitrogen of PLP has been regarded as a mechanistic determinant for the alanine racemase activity (Sun & Toney, 1999; Yow *et al.*, 2003; Rubinstein & Major, 2010; Griswold & Toney, 2011). Under physiological conditions, the L- to D-alanine racemization in *E. coli* Alr is achieved by the actions of a pair of catalytic base (Lys34 and Tyr255′) located at the opposite sides of the PLP-conjugated plane (Figure 2.18A). In principle, the substrate (L-Ala) displaces Lys34 to form an external aldimine with the PLP, and Tyr255′ abstracts a proton from the substrate. The resulting carbanion intermediate was then reprotonated by Lys34 on the opposite side to form D-Ala. The D-Ala product was subsequently released after being displaced by Lys34 to reform the internal aldimine with PLP.

In MetC, the negatively charged aspartic acid within the vicinity of the pyridine nitrogen ensures that the PLP is fully protonated (Figure 2.18C). The protonated pyridine ring plays a key role in stabilizing the negative charge developed at the Cα after the formation of external aldimine (Figure 2.18B). However, the aspartic acid was replaced by a positively charged arginine in Alr. The arginine residue prevents full protonation of the PLP (Figure 2.18C), thereby negating the electrophilicity of the π-bonding system. As a result,

after proton abstraction of the external aldimine, the carbanion intermediates (including the quinonoid species) are assumed to be highly unstable in Alr (Eliot & Kirsch, 2004). Yet, the unprotonated PLP feature has been proposed to aid in the control of reaction specificity in Alr by having intermediates with limited lifetime (Sun & Toney, 1999; Yow *et al.*, 2003; Toney, 2005; Rubinstein & Major, 2010; Griswold & Toney, 2011). Computational simulations and pre-steady state studies have demonstrated that Alr mutants with relatively stable intermediates were likely to catalyze transamination, instead of racemization (Sun & Toney, 1999; Yow *et al*, 2003; Rubinstein & Major, 2010).

At the outset of this study, it was not known if the protonation state of the PLP was important for the cystathionine β-lyase activity in Alr. Previously, an Arg219Glu mutation in *G. stearothermophilus* Alr resulted in protonated PLP, and consequently increased its transamination activity (Sun & Toney, 1999; Yow *et al*., 2003). To investigate the potential role of protonated PLP in increasing the cystathionine β-lyase activity in *E. coli* Alr, the Arg209 residue (that interacts with the pyridine nitrogen) was similarly mutated to Glu. The *alr* insert with the R209E mutation was cloned into the GFP-deleted over-expression vector, pCA24N (Section 2.4.9), to form pCA24N-*alr*(R209E). In the first instance, the *in vivo* enzyme activities of Alr(R209E) were assessed prior to *in vitro* studies.

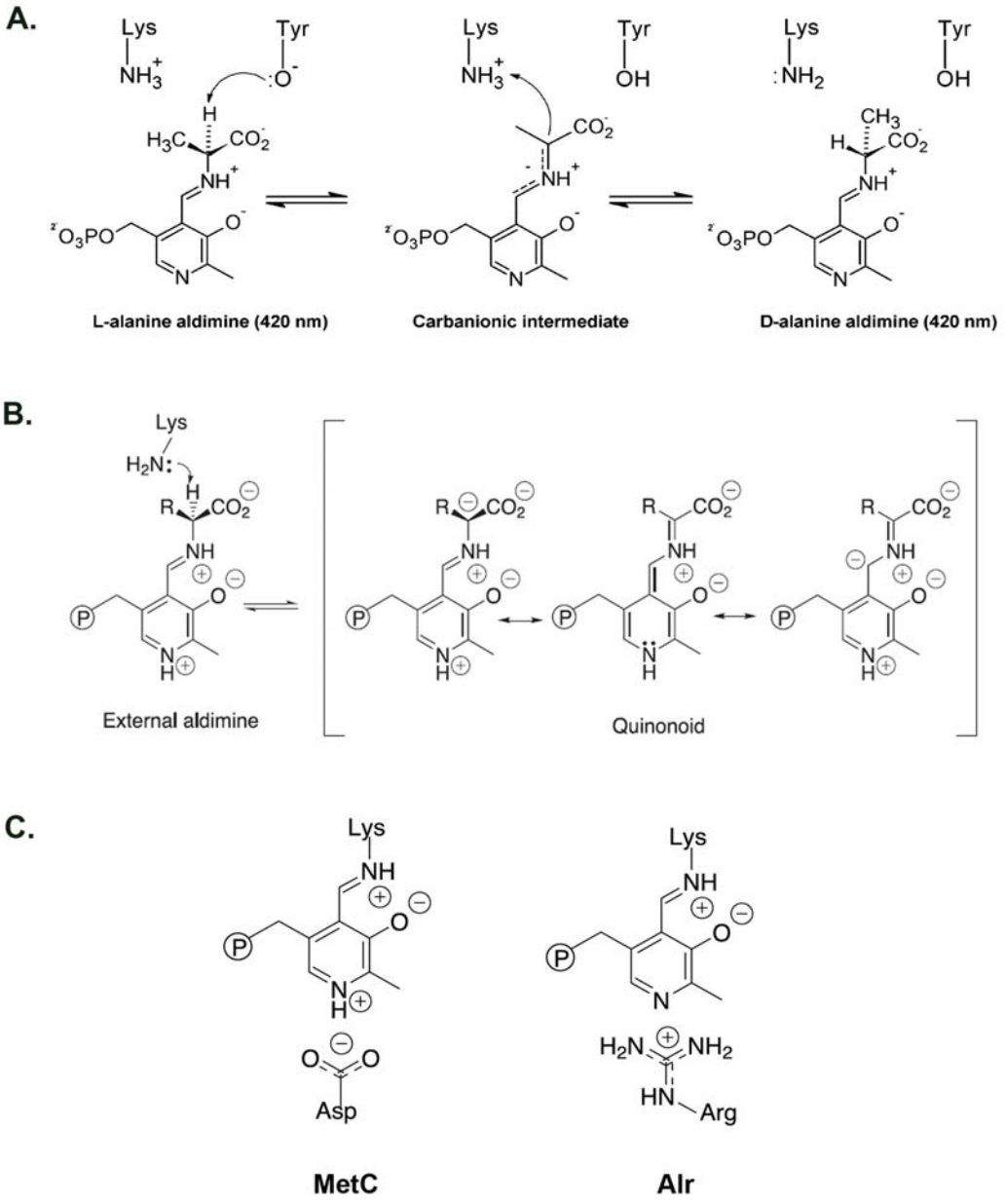


Figure 2.18: The alanine racemization mechanism of Alr.

(A) Proton abstraction and donation in alanine racemization are achieved by a pair of catalytic Lys and Tyr in the active site. (B) The formation of carbanion intermediates after the Cα of the external aldimine is deprotonated. Species shown in brackets are resonance structures of carbanion intermediates. (C) The protonation state of the pyridine nitrogen of PLP in MetC, and the lack thereof in Alr. Figure A was taken from Griswold and Toney (2011). Figures B and C were adapted from Eliot and Kirsch (2004).

2.2.5.1 THE *IN VIVO* ALANINE RACEMASE ACTIVITY OF ALR(R209E)

Previous studies have reported that protonated PLP in *G. stearothermophilus* Alr led to a ~1,000-fold decrease in its *in vitro* alanine racemase activity (Sun & Toney, 1999; Yow *et al*., 2003), but its *in vivo* effect has not been demonstrated. In this section, I will evaluate the *in vivo* alanine racemase activity of Alr(R209E).

E. coli MB2795 (Δ*alr* Δ*dadX*) is a D-alanine auxotroph (see Section I.3 in Appendix I), and its growth must be supplemented with D-alanine. This strain was transformed with pCA24N-NoIns, pCA24N-*alr*(GFP−) and pCA24N-*alr*(R209E) formed the negative control, positive control and test clone, respectively. As expected, the negative control did not grow on LB-chloramphenicol-IPTG agar (Plate 1 in Figure 2.19), even after 12 days of incubation at 28°C. In contrast, the positive control took 1 day to form full-sized colonies (Plate 3 in Figure 2.19). The growth of the Alr(R209E)-expressing clone, however, was worse than the positive control. Cells over-expressing Alr(R209E) grew only after 2–3 days of incubation (Plate 2 in Figure 2.19). Overall, the growth of the Alr(R209E)-expressing clone presented here is consistent with the previous, *in vitro* findings (Sun & Toney, 1999; Yow *et al*., 2003). The results also appeared to suggest that the R209E mutation had resulted in protonation of PLP, thereby reducing the racemization activity in Alr (Sun & Toney, 1999; Yow *et al*., 2003; Rubinstein & Major, 2010; Griswold & Toney, 2011).

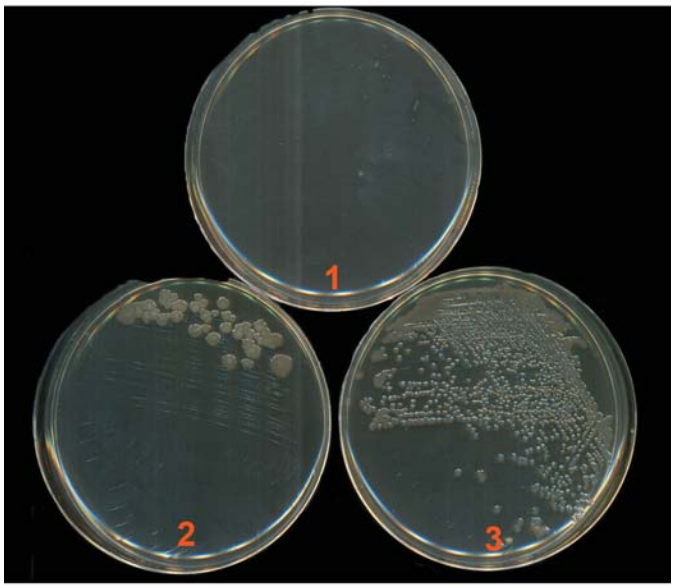


Figure 2.19: Growth complementation assays of *E. coli* MB2795 clones on LB-chloramphenicol-IPTG after 5 days of incubation at 28°C.

The agar plates were streaked with *E. coli* MB2795 expressing [1] the empty vector (pCA24N-NoIns); [2] pCA24N-*alr*(R209E) and [3] pCA24N-*alr*(GFP−).

2.2.5.2 THE *IN VIVO* CYSTATHIONINE β-LYASE ACTIVITY OF ALR(R209E)

Next, the *in vivo* cystathionine β-lyase activity of Alr(R209E) was assessed, to determine if the activity was increased by the R209E mutation. The plasmids (pCA24N-NoIns, pCA24N-*alr*(GFP−) and pCA24N-*alr*(R209E)] were introduced into *E. coli* Δ*metC* to form the negative control, parental clone and test clone, respectively. The strain was also transformed with pCA24N-*metC*(GFP−) to form the positive control. Similar to the results presented in Section 2.2.1, the MetC-expressing clone grew on M9/glucose-IPTG medium after 2 days of incubation at 28°C (Plate 2 of Figure 2.20A). The negative control did not grow, even up to 12 days of incubation (Plate 1 of Figure 2.20A). While the Alr-expressing clone grew as expected (4–5 days; Plate 4 of Figure 2.20A), the Alr(R209E)-expressing clone barely grew after 6 days of incubation (Plate 3 of Figure 2.20A). This implied that the R209E mutation did not increase the cystathionine β-lyase in Alr. Overall, the growth patterns of the Alr(R209E)-expressing clone indicated that the R209E mutation impeded both alanine racemase (Section 2.2.5.1) and cystathionine β-lyase activity in Alr.

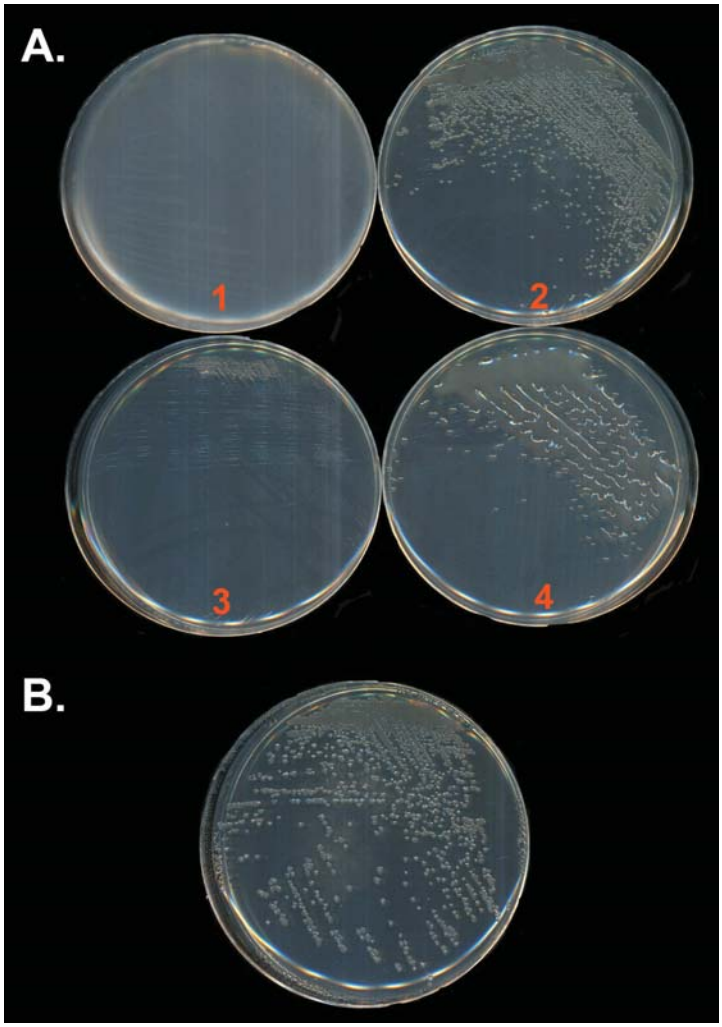


Figure 2.20: Growth complementation assays of *E. coli* Δ*metC* clones on M9/glucose-IPTG after 7 days of incubation at 28°C.

All agar plates were supplemented with kanamycin and chloramphenicol. (A) The agar plates were streaked with *E. coli* Δ*metC* expressing [1] the empty vector, pCA24N-NoIns; [2] pCA24N-*metC*(GFP–); [3] pCA24N-*alr*(R209E) and [4] pCA24N-*alr*(GFP–). (B) The M9/glucose agar supplemented with 100 μM L-methionine was streaked with *E. coli* Δ*metC* expressing pCA24N-*alr*(R209E).

2.2.5.3 THE *IN VITRO* ACTIVITY OF ALR(R209E)

Results from the growth complementation assays (Section 2.2.5.2) demonstrated that the *in vivo* cystathionine β-lyase activity of the Alr(R209E) mutant was worse than wildtype Alr. Hence, the possibility of detecting any cystathionine β-lyase activity in Alr(R209E) using *in vitro* assay was very low. This is confirmed further in this section by the lack of activity in the purified Alr(R209E).

The Alr(R209E) was purified from *E. coli* MB2795 (Δ*alr* Δ*dadX*), as previously described (Sections 2.2.4 and 2.4.5). The enzyme was largely soluble, and was purified to >95% purity,

as judged by SDS-PAGE (Figure 2.21). Consistent with the thick protein bands on the SDS-PAGE, the yield of purified Alr(R209E) was high (~31 mg/L of culture). However, the absorbance spectrum of Alr(R209E) revealed that no peak was observed at 420 nm (where PLP absorbs strongly) [Figure 2.22], thus indicating that the purified enzyme was not PLP-bound. In comparison, wildtype Alr [which had been purified simulatenously with Alr(R209E)] showed a peak at 420 nm (Figure 2.22) that corresponded to ~70% of PLP-bound enzyme.

Nevertheless, the purified Alr(R209E) was added to *in vitro* assays in an attempt to measure its cystathionine β-lyase activity. PLP was also added to the assays to reconstitute the apoenzymes. No activity was detected for Alr(R209E), with up to 10 mM L-cystathionine (results not shown). The lack of *in vitro* activity of Alr(R209E) was consistent with the low *in vivo* cystathionine β-lyase activity of the Alr(R209E)-expressing clone (Section 2.2.5.2). The overall implication was that the R209E mutation might have prevented strong anchoring of the PLP in the active site, thereby resulting in the loss of the cofactor from the enzyme upon purification. Due to the lack of PLP, it was likely that a majority of the Alr(R209E) enzymes were catalytically inactive.

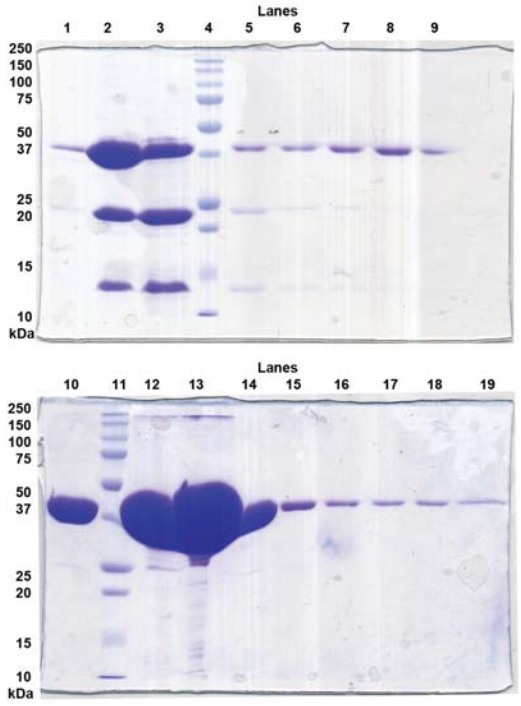


Figure 2.21: SDS-PAGE analyses of the purified Alr(R209E).

Each lane was loaded with a 6-μL aliquot from the corresponding fractions. Lane 1, Insoluble fraction. Lane 2, Soluble lysate. Lane 3, Proteins not bound to the Talon® metal affinity resin. Lanes 4 and 11, Bio-Rad Precision Plus protein All Blue standard. Lanes 5–6, Batch washes. Lane 7, Flow-through from the enzyme-bound resin in the gravity flow column. Lanes 8–9, Column washes with low imidazole concentration (3 mM). Lane 10, Column wash with high imidazole concentration (15 mM). Lanes 12–19, Elution fractions 1–8.

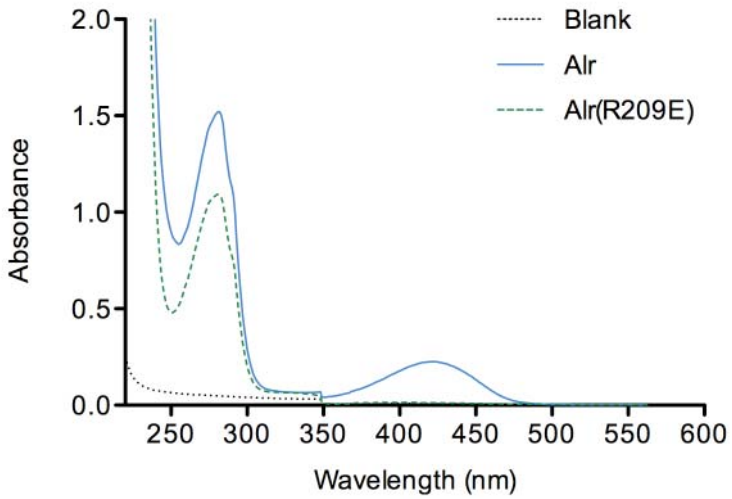


Figure 2.22: The absorbance spectra of purified Alr and Alr(R209E).

The peak at 280 nm represents the total protein absorbance, whereas the peak at 420 nm represents the absorbance of PLP in the sample.

2.2.6 Improving the cystathionine β-lyase activity of Alr via directed evolution

Since the *in vitro* cystathionine β-lyase activity was undetectable in wildtype Alr and the R209E variant, in this section I will attempt to improve this activity to a level that can be detected spectrophotometrically. This approach can demonstrate the existence of the weak cystathionine β-lyase activity in Alr (because enzyme activity can not be improved if there is no activity at all), and that this promiscuous activity is evolvable.

Random mutations were introduced into the *alr* gene using error-prone PCR (epPCR) (Cadwell & Joyce, 1992). Mutated *alr* genes were cloned into the pCA24N vector, and the ligated plasmids were subsequently introduced into *E. coli* Δ*metC*. The mutant clones with higher *in vivo* cystathionine β-lyase activity were selected based on their ability to grow faster on the M9/glucose medium than the parental clone with unmutated Alr.

2.2.6.1 An Alr library generated by error-prone PCR

A library consisting of 3.4×10^7 clones was constructed (see Section 2.4.10.1 for more details). A PCR screen of twenty unselected clones confirmed that the plasmid-encoded *alr* was present in each clone. These *alr* gene inserts were sequenced to obtain a general view on the mutational spectrum of the library (Table 2.2). A total of (20 sequences × 1,077 bp) = 21,540 bp were analyzed. Among these, 16 point mutations and 0 indel were found. The number of mutations found in each gene ranged from 0–4. In Table 2.2, it was apparent that

the mutational spectrum of the *alr* library biased towards G→N and C→N mutations. The number of G→N and C→N mutations was approximately twice as many as the number of A→N and T→N mutations. Overall, the average mutation rate was low (0.8 mutation per gene). This was advantageous for identifying beneficial mutations needed to improve low enzyme activity (Tracewell & Arnold, 2009). As most mutations are either deleterious or neutral, a high mutation rate would be more likely to yield false positives.

Table 2.2: The mutational spectrum of the error-prone *alr* library.

Types of mutation	Frequency	Proportion of total
Transitions		
A→G, T→C	1	6.3%
G→A, C→T	8	50.0%
Transversions		
A→T, T→A	3	18.8%
A→C, T→G	1	6.3%
G→C, C→G	2	12.5%
G→T, C→A	1	6.3%
Summary of bias		
Transitions/Transversions	1.3	N/A
AT→GC/GC→AT	0.2	N/A
A→N, T→N	5	31.5%
G→N, C→N	11	68.8%
Mutation per kb	0.7	N/A
Mutation per *alr* gene (1,077 bp)	0.8	N/A

N/A = not applicable.

PEDEL-AA was used to estimate the mutational diversity on an amino acid level of the *alr* library based on the DNA mutational spectrum (Firth & Patrick, 2008). The key characteristics of the library are summarized in Table 2.3. While it was previously calculated that the DNA mutation rate was 0.8 per gene, PEDEL-AA predicted an even lower amino acid substitution rate per protein (0.6 per variant). Therefore, it was estimated that up to ~62% of the library consisted of unmutated Alr sequences. However, the Alr library was still considerably large (~1.3×10^7 mutated variants), after the unmutated variants had been

subtracted from the total library size. Despite the low amino acid substitution rate per variant, the overall large library size rendered this Alr library suitable for subsequent experiment (Section 2.2.6.2).

Table 2.3: Summary of the key characteristics of the Alr library based on its DNA mutational spectrum.

Property	Estimate
Total library size	3.4×10^7
Number of variants with no indels or stop codons	3.3×10^7
Mean number of amino acid substitutions per variant	0.6
[#]Unmutated (wildtype) sequences (% of library)	61.8%
[#]Number of distinct full-length proteins in the library	2.4×10^6

2.2.6.2 SELECTION OF ALR VARIANTS WITH IMPROVED *IN VIVO* METC ACTIVITY

Having deemed the Alr library suitable for selection experiments, the library variants were selected using a more stringent growth complementation assay. Previously, all growth complementation assays were performed on M9/glucose medium supplemented with 50 μM IPTG. When the IPTG concentration was reduced to 5 μM, the duration required for the parental clone [*E. coli* Δ*metC* harbouring pCA24N-*alr*(GFP−)] to form colonies was extended to 12–13 days, instead of 3–4 days. Therefore, Alr variants with improved cystathionine β-lyase activity should form colonies faster than the parental clone.

E. coli Δ*metC* was transformed with the pooled plasmids of the Alr library. A total of 1.0×10^8 clones were then plated on the selective medium and incubated at 28°C. The number of clones plated represented a ~3-fold over-sampling of the initial Alr library (which consisted of 3.4×10^7 clones). This also corresponded to a sampling of 95.3% of the clones from the initial library [calculated by the library diversity and completeness programme, GLUE (Patrick *et al*., 2003)].

As expected, no growth was observed for the negative control (*E. coli* Δ*metC* expressing pCA24N-NoIns) throughout the selection experiment. For the Alr library, approximately 3,400 colonies had formed on the selective medium after only 2 days of incubation. In comparison to the parental clone (12-day growers), these Alr variants (2-day growers) represented a ~6-fold improvement in growth. Given that a total of 1.0×10^8 clones were initially spread on the agar, the number of surviving Alr variants represented a survival rate of 0.003%. The growth rate of these Alr variants was also comparable with that of the

positive control (the MetC-expressing clone), which also took 2 days to form colonies. The growth of all *E. coli* Δ*metC* clones on the selective medium (M9/glucose supplemented with 5 μM IPTG) is tabulated in Table 2.4. Since the number of surviving clones was large, 48 colonies were randomly picked for sequencing, in order to identify the mutations in their plasmid-encoded *alr*.

DNA sequences were obtained for 34 out of 48 clones. The sequencing chromatograms of the remaining 14 clones had overlapping peaks, which suggests the presence of mixed templates (*i.e.*, mixed clones). Among these 34 clones, 32 clones harboured mutated *alr*, which could be classified into eight mutated variants (Table 2.4). The number of amino acid mutations ranged from 1–3 substitutions per clone. In particular, every clone harboured the Y274F substitution. The Alr-3 variant with a single Y274F mutation dominated the pool with a frequency of (25/32) = 78%. Other clones possessed 1–2 substitutions, in addition to Y274F. The prevalence of the Y274F mutation in all 32 clones suggested that it was likely to be a beneficial, rather than a neutral, mutation. Furthermore, I later showed in Section 2.2.6.3 that the 2-day growth phenotype of these 32 clones was reproducible.

No DNA mutation was found in the plasmid-encoded *alr* in two clones, Alr-1 and Alr-2 (results not shown). The rapid growth of Alr-1 and Alr-2 was likely to be caused by regulatory mutations, such as those in the T5-*lacO* promoter of the pCA24N vector (Patrick & Matsumura, 2008). I tested this possibility by recloning the unmutated *alr* insert of Alr-1 and Alr-2 into fresh pCA24N vector (see Section 2.4.10.3 for details). Fresh *E. coli* Δ*metC* was then used to retransform the newly ligated plasmids. Cells over-expressing the newly recloned, unmutated *alr* insert were unable to form colonies after 2 days of incubation at 28°C (results not shown); instead, they grew only after 12 days of incubation, which was reminiscent of the growth of the parental Alr variant. This demonstrates that regulatory mutations were present in the original plasmid backbone of the Alr-1 and Alr-2 variants.

Table 2.4: Summary of the selected Alr variants and their amino acid mutations.

Variant	Frequency	Days to form colonies	Mutation(s)
Alr	[Template]	12–13	None
Alr-1*	1	2	None
Alr-2*	1	2	None
Alr-3	25	2	Y274F
Alr-11	1	2	Y274F, A335T
Alr-14	1	2	R266S, Y274F
Alr-19	1	2	M29I, Y274F
Alr-22	1	2	L241M, R263H, Y274F
Alr-24	1	2	T97S, P252R, Y274F
Alr-28	1	2	I242V, Y274F
Alr-41	1	2	V150I, Y274F
MetC	[Positive control]	2	None
NoIns	[Negative control]	No growth	None

*The growth phenotype of these clones was not reproducible upon recloning and retransformation tests.

A total of 10 amino acid substitutions were identified from the eight Alr variants described in Table 2.4. In seven of the mutated residues, the chemical property of the side chains remained similar. Examples include the T97S mutation, where both threonine and serine have a polar, uncharged side chain; as well as the V150I mutation, where both valine and isoleucine have a hydrophobic side chain. The three mutations with altered chemical property of the side chain were P252R (from hydrophobic to positively charged), R266S (from positively charged to polar and uncharged) and A335T (from hydrophobic to polar and uncharged).

The location of the mutated residues was mapped to the tertiary and quaternary structures of *E. coli* Alr (Figure 2.23). Of the 10 mutated residues, seven were located in the β-strands that populate the C-terminal domain (residues 232–359; Figure 2.23A). The remaining three mutated residues were buried in the core of the $(\beta/\alpha)_8$-barrel of the N-terminal domain (residues 1–231; Figure 2.23A). Y274 is the only residue located within the vicinity of the active site, which is a cleft formed between the N-terminal and C-terminal domains from opposite monomers. Along with P252, Y274 was also likely to be on the dimerization surface. Other residues were located on the surface of the protein.

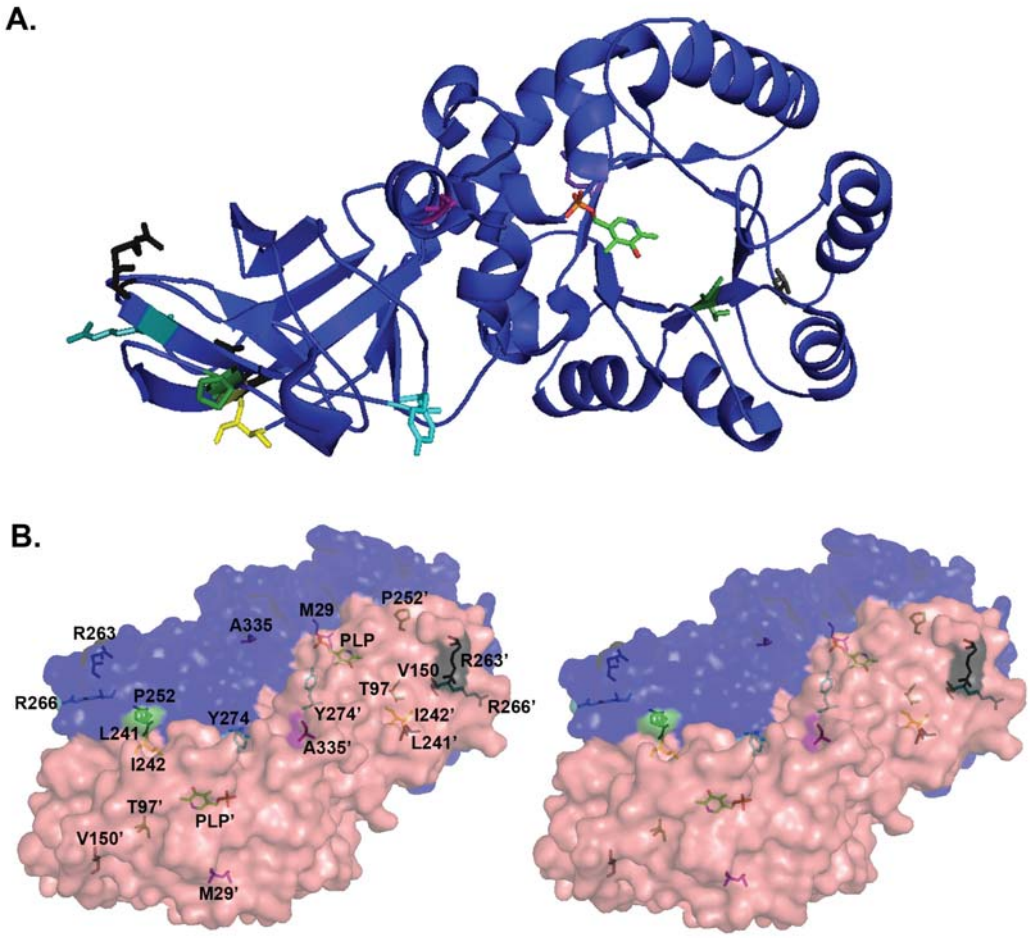


Figure 2.23: The position of the mutated residues in the Alr structure.

The mutated residues are mapped to the structure of *E. coli* Alr (PDB code: 2rjg). Each PLP molecule is coloured according to its elements, whereas each residue is coloured differently. (A) Ribbon diagram of an Alr monomer and the position of the mutated residues. Loops have been smoothed for clarity. (B) Surface diagram of an Alr dimer. The residues of both monomers are shown, and the residue numbering is given in the left panel.

2.2.6.3 CONFIRMATION OF THE IMPROVED GROWTH PHENOTYPE OF ALR MUTANTS

Retransformation and restreaking tests (see Section 2.4.10.3 for details) were used to confirm the growth phenotypes of the eight mutated clones (Alr-3, Alr-11, Alr-14, Alr-19, Alr-22, Alr-24, Alr-28 and Alr-41) that had been selected previously (Section 2.2.6.2). These tests can demonstrate that chromosomal mutations are not responsible for the rapid growth of the mutant clones.

Plasmids were isolated from these clones, and reintroduced into fresh *E. coli* Δ*metC* cells. The retransformed clones were then grown on the same selective medium, M9/glucose supplemented with 5 μM IPTG. To further identify the most active Alr

variant(s) among the eight clones, the clones were also streaked on M9/glucose agar without IPTG.

All eight retransformed clones formed colonies on the IPTG-supplemented M9/glucose agar after 2 days of incubation. These phenotypes were consistent with the 2-day growth phenotypes observed in the selection experiment (Section 2.2.6.2). By the third day, confluent growth was observed for all Alr variants on IPTG-supplemented M9/glucose agar, except for Alr-41 (Table 2.5; Figure 2.24). In particular, the Alr-41 variant did not manage to grow on the M9/glucose agar without IPTG. The growth of other clones varied greatly on the IPTG-less agar (Table 2.5; Figure 2.24). Therefore, the order of the *in vivo* cystathionine β-lyase activity in these Alr variants was established as follows: Alr-3 > Alr-11, Alr-14, Alr-19, Alr-24 > Alr-22, Alr-28 > Alr-41.

Overall, these results confirmed that all eight Alr mutants could rescue the *E. coli* MetC auxotroph faster than the parental Alr clone. The implication was that these Alr variants expressed higher *in vivo* cystathionine β-lyase activity.

Table 2.5: Summary of the growth of retransformed Alr mutant clones on selective media after 3 days of incubation.

Variant	Growth on M9/glucose agar + 5 μM IPTG	Growth on M9/glucose agar alone
Alr-3	++++	++++
Alr-11	++++	+++
Alr-14	++++	+++
Alr-19	++++	+++
Alr-22	++++	++
Alr-24	++++	+++
Alr-28	++++	++
Alr-41	++	–

Key for scoring growth: (–) No growth; (+) Little growth. Colony size was <0.1 mm; (++) Considerable growth. Colony size was ~0.1 mm; (+++) Good growth. Colony size was 0.1–0.5 mm; (++++) Confluent growth. Colony size was >0.5 mm.

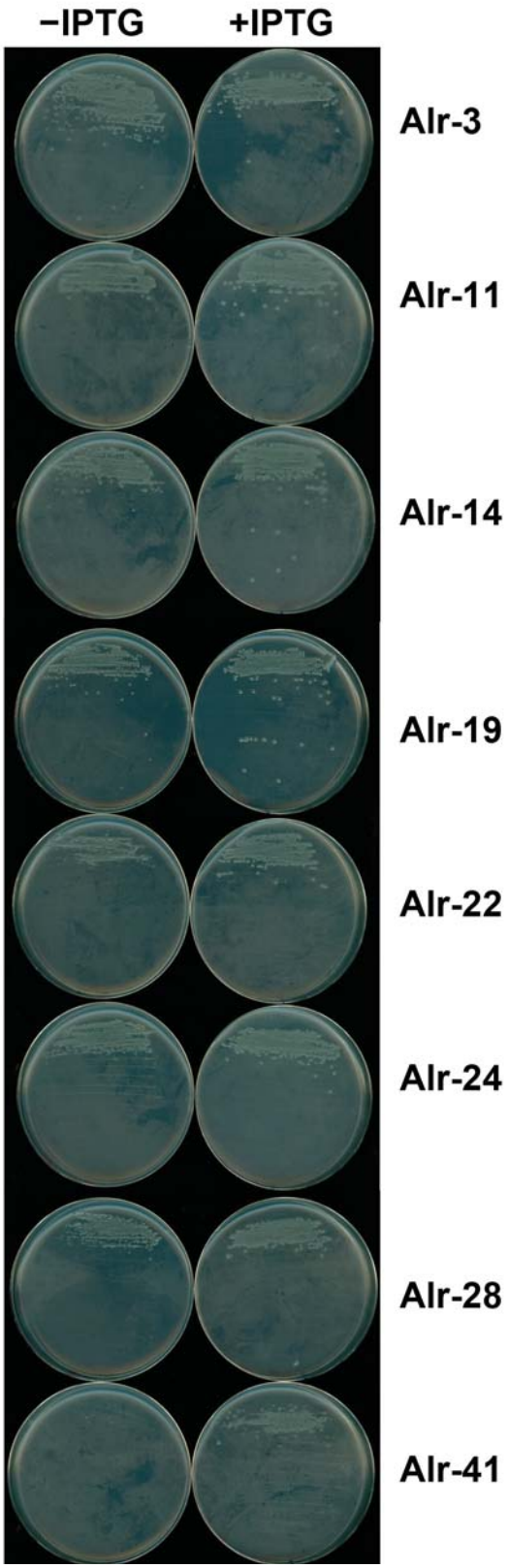


Figure 2.24: The growth of retransformed Alr mutant clones on M9/glucose agar with and without 5 μM IPTG, after 3 days of incubation at 28°C.

2.2.6.4 THE PARENTAL ACTIVITY OF THE ALR MUTANTS

In this section, I will test the *in vivo* alanine racemase activity of the eight Alr mutants. Previously, Alr variants with improved promiscuous activity have shown lower native activity (Seebeck & Hilvert, 2003; Yow *et al.*, 2003). As a preliminary test to probe the degree of activity trade-off, I expressed each of the eight Alr variants in the D-Ala auxotroph, *E. coli* MB2795 (Δ*alr* Δ*dadX*). The cells were grown on the selective medium, LB-chloramphenicol agar without D-Ala. No growth difference was noted for all eight Alr variants, regardless of the presence of IPTG (Table 2.6; Figure 2.25). Similar to the Alr-expressing clone, all clones expressing the mutated Alr variants formed full-sized colonies after 2 days of incubation at 28°C. The results suggested that *in vivo* activity trade-off was minimal in these mutated Alr variants.

Table 2.6: Summary of the growth of *E. coli* MB2795 that expressed mutated Alr variants, on selective media after 2 days of incubation.

Variant	Growth on LB agar + 50 μM IPTG	Growth on LB agar alone
Alr-3	++++	++++
Alr-11	++++	++++
Alr-14	++++	++++
Alr-19	++++	++++
Alr-22	++++	++++
Alr-24	++++	++++
Alr-28	++++	++++
Alr-41	++++	++++

Key for scoring growth: (–) No growth; (+) Little growth. Colony size was <0.1 mm; (++) Considerable growth. Colony size was ~0.1 mm; (+++) Good growth. Colony size was 0.1–0.5 mm; (++++) Confluent growth. Colony size was >0.5 mm.

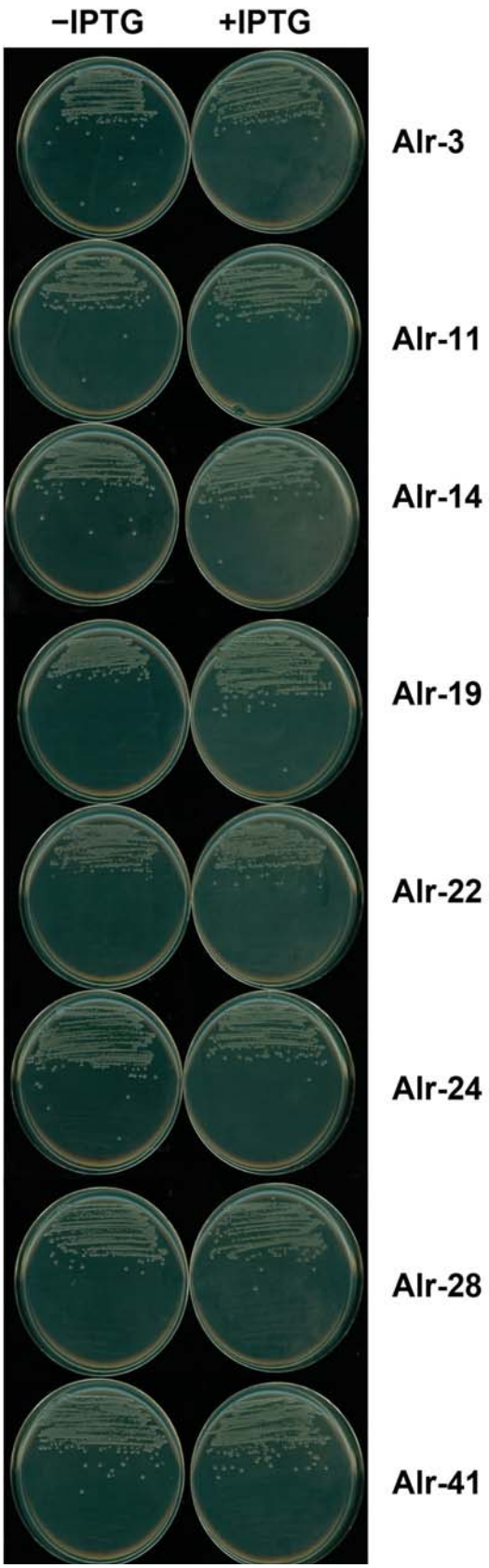


Figure 2.25: The growth of *E. coli* MB2795 that expressed mutated Alr variants, on LB medium with and without 50 μM IPTG, after 2 days of incubation at 28°C.

2.2.6.5 KINETIC CHARACTERIZATION OF ALR-3

After confirming the *in vivo* activities of the eight Alr mutants (Sections 2.2.6.3 and 2.2.6.4), I will test the *in vitro* activities in this section. Based on the growth complementation results (Section 2.2.6.3), the Alr variant with the highest *in vivo* cystathionine β-lyase activity was Alr-3, which was followed by Alr-11, Alr-14, Alr-19, Alr-24. Thus, Alr-3 and Alr-19 were initially selected for protein purification and kinetic characterizations.

The Alr-3 and Alr-19 variants were over-expressed in *E. coli* Δ*metC* strain. These enzymes were $(His)_6$-tagged purified to >95% purity, as judged by SDS-PAGE (Figure 2.26). Similar to wildtype Alr, both Alr-3 and Alr-19 were highly soluble despite having 1–2 amino acid substitution (Y274 for Alr-3; M29I and Y274F for Alr-19). The enzyme yields for Alr-3 and Alr-19 were 49 mg/L of culture and 68 mg/L of culture, respectively. The absorbance spectra of both enzyme variants revealed that ~82% of the Alr-3 enzymes, and 75% of the Alr-19 enzymes, were bound to PLP.

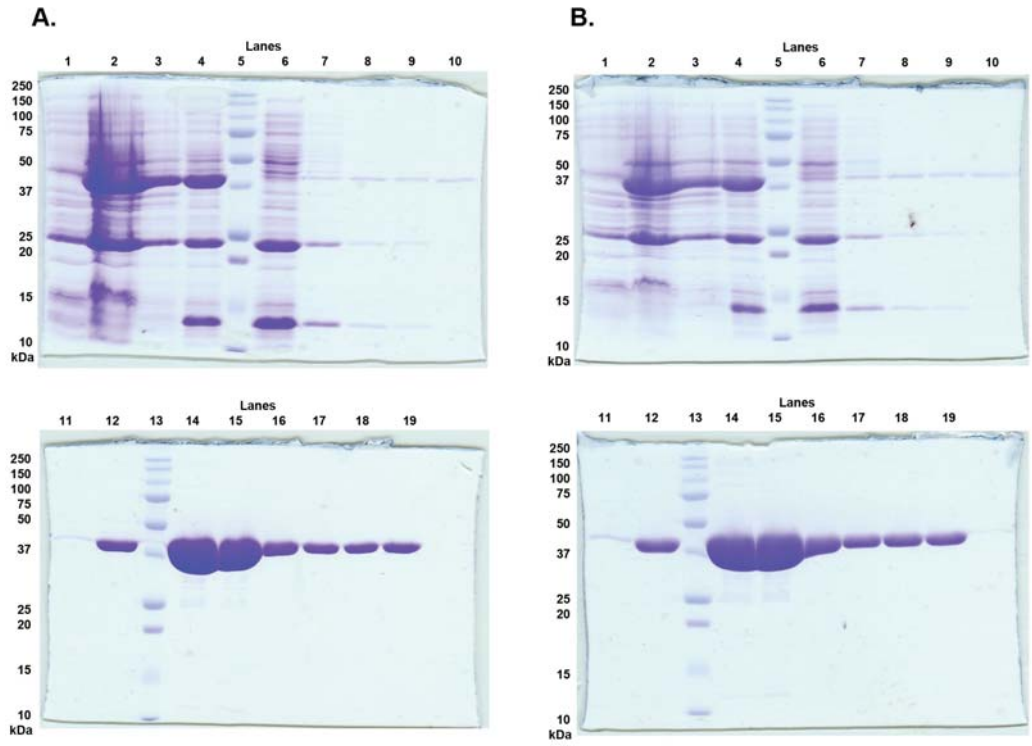


Figure 2.26: SDS-PAGE analyses of the purified (A) Alr-3 and (B) Alr-19.

Each lane was loaded with a 6-μL aliquot from the corresponding fractions. Lane 1, Crude cell lysate before IPTG induction. Lane 2, Crude cell lysate after 6 hours of protein over-expression. Lane 3, Insoluble fraction. Lane 4, Soluble lysate. Lanes 5 and 13, Bio-Rad Precision Plus protein All Blue standard. Lane 6, Proteins not bound to the Talon® metal affinity resin. Lanes 7–8, Batch washes. Lane 9, Flow-through from the enzyme-bound resin in the gravity flow column. Lanes 10–11, Column washes with low imidazole concentration (3 mM). Lane 12, Column wash with high imidazole concentration (15 mM). Lanes 14–19, Elution fractions 1–6.

Since Alr-3 displayed higher *in vivo* cystathionine β-lyase activity than Alr-19 (Section 2.2.6.3), the *in vitro* activities of Alr-3 were assayed in the first instance. Purified Alr-3 showed detectable cystathionine β-lyase activity, but the overall activity was low (<0.025 turnover/s at 10 mM L-cystathionine for all six repeats, which were assayed on different days; Figure 2.27). The overall catalytic efficiency (k_{cat}/K_M) was 0.85 $s^{-1}.M^{-1}$ (Table 2.7), which was ~2.7 × 10^6-fold lower than that of MetC. The low k_{cat}/K_M value of Alr-3 was largely due to poor k_{cat} (~10 turnover/h), rather than K_M. In particular, the K_M of Alr-3 for L-cystathionine was only 85-fold higher than MetC (Table 2.7). In contrast, the k_{cat} of Alr-3 was ~32,000-fold lower than MetC. Although the *E. coli* Δ*metC* clone that expressed Alr-3 was able to grow at the same rate as MetC-expressing clone (*i.e.*, 2 days to form colonies on M9/glucose agar), the low *in vitro* activity of Alr-3 did not seem to reflect its *in vivo* activity.

As evident in Figure 2.27, the error bars from each pooled data points were rather large. Previously, I had considered a possibility that the data points might be better fitted with a linear regression, instead of a hyperbolic regression. Linear regression analyses have been used previously to determine the k_{cat}/K_M of low promiscuous activities in enzymes, which could not be saturated with substrates during *in vitro* assays (Patrick & Matsumura, 2008; McLoughlin & Copley, 2008). When the data points of Figure 2.27 were fitted with a linear regression, the estimated k_{cat}/K_M of Alr-3 was 0.24 M $s^{-1}.M^{-1}$, with a R^2 value of 0.9283 (not shown). The k_{cat}/K_M value produced by linear regression was rather close to the k_{cat}/K_M produced by hyperbolic regression (~3.5-fold difference). However, the R^2 value of the linear fitting was slightly lower than that of the hyperbolic fitting (R^2 = 0.9636, as shown in Figure 2.27). Therefore, it was deemed that the hyperbolic regression fitted better than the linear fitting, and the final kinetic parameters were derived from hyperbolic regression.

Since the *in vitro* cystathionine β-lyase activity of Alr-3 was barely detectable, I did not proceed to assay the activities of Alr-19. Given that the *in vivo* cystathionine β-lyase activity of Alr-19 was lower than Alr-3 (Section 2.2.6.3), the *in vitro* activity of Alr-19 was expected to be lower, and hence, more difficult to be detected than Alr-3.

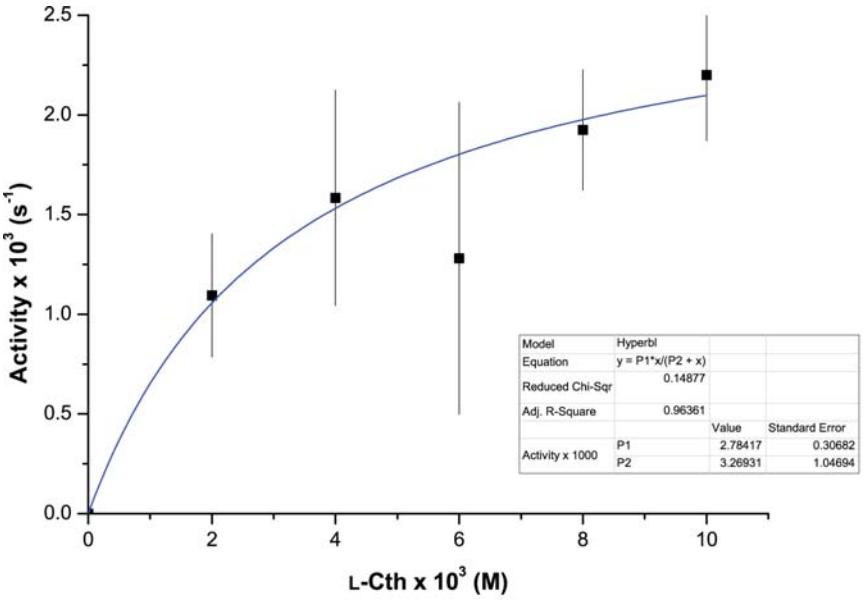


Figure 2.27: The cystathionine β-lyase activity of Alr-3.

The Michaelis-Menten plot was constructed using the mean ± S.E.M. from six data sets. L-Cth, L-cystathionine.

In addition, the native activity of Alr-3 and wildtype Alr was assayed spectrophotometrically to determine the degree of activity trade-off. The working principles of the racemase assay will be described in Section 3.2.3 (Chapter 3), and further details can be found in Section 2.4.7.1. Briefly, each racemase assay was coupled with NADH, which was oxidized by a coupled enzyme. The substrate for the coupled enzyme originated from the racemization of alanine. NADH oxidation was measured as a decrease in A_{340} over time.

In this study, the k_{cat} and K_M of the alanine racemization activity of Alr (see Table 2.7) were reasonably close to those of published data — Wu *et al*. (2008) has reported a k_{cat} of 53 s^{-1} and a K_M of 1.0 mM for *E. coli* Alr. In comparison to Alr, a subtle decrease in k_{cat} (~1.4-fold) and a subtle increase in K_M (~1.7-fold) were observed for Alr-3, thereby resulting in an overall ~2.5-fold decrease in k_{cat}/K_M (Table 2.7). The kinetic parameters of the alanine racemization activity of Alr-3 appeared to suggest that its Y274F mutation had minimal *in vivo* (Section 2.2.6.4) and *in vitro* activity trade-off.

Table 2.7: Kinetic parameters of MetC, Alr-3 and Alr.

Enzyme	Cystathionine → Homocysteine + Pyruvate + NH_3			L-Ala → D-Ala		
	k_{cat} (s^{-1})	K_M (M)	k_{cat}/K_M ($s^{-1}.M^{-1}$)	k_{cat} (s^{-1})	K_M (M)	k_{cat}/K_M ($s^{-1}.M^{-1}$)
MetC	89.5 ± 13.9	$(3.9 \pm 0.7) \times 10^{-5}$	2.3×10^{6}	Presented in Chapter 3		
Alr-3	[#]$(2.8 \pm 0.3) \times 10^{-3}$	[#]$(3.3 \pm 1.0) \times 10^{-3}$	0.85	39.6 ± 4.7	$(1.5 \pm 0.6) \times 10^{-3}$	2.6×10^{4}
Alr	[†]n.d.	n.d.	<0.85	55.3 ± 3.7	$(8.7 \pm 1.5) \times 10^{-4}$	6.4×10^{4}

Each value represented the mean ± S.E.M. (n = 3), and each value was the mean of duplicate assays.

[#]The values represented the mean ± S.E.M. of six repeats.

[†]n.d., not determined.

2.3 DISCUSSION

2.3.1 PROMISCUITY IN NON-HOMOLOGOUS ENZYMES

In this chapter, I demonstrated that the promiscuous activity of Alr was responsible for rescuing a methionine auxotroph, *E. coli* Δ*met*C. *In vitro* characterization of Alr-3 (an evolved Alr variant) revealed that the enzyme was able to catalyze the cystathionine β-lyase activity, the native function of MetC. Unlike other promiscuous activities of Alr such as transamination, aldol cleavage and α,β-elimination (Wang & Walsh, 1978; Wang & Walsh, 1981, Thornberry *et al*., 1991; Kurokawa *et al*., 1998; Seebeck & Hilvert, 2003; Yow *et al*., 2003), the weak cystathionine β-lyase activity was physiologically relevant in cells lacking MetC. This finding further emphasizes that promiscuous, pre-existing enzymes such as Alr can be recruited to catalyze functions in different metabolic pathways (Patrick & Matsumura, 2008; Kim *et al*., 2010).

Enzymes capable of catalyzing similar chemical transformations are usually descendants from a generalist ancestral enzyme, and hence, are sequence or structure homologues (Khersonsky & Tawfik, 2010). However, recent findings revealed that enzyme promiscuity could also exist in non-homologous enzymes (James & Tawfik, 2001; McLoughlin & Copley, 2008; Patrick & Matsumura, 2008; Kim *et al*., 2010). Results from this chapter further extend this line of reasoning. The lack of evolutionary relationship between Alr and MetC was evident from their minimal sequence and structural homologies. Pairwise alignment of protein sequences of *E. coli* Alr and MetC showed 9% sequence identity, whereas pairwise structural alignment of the Alr (input chain: 1sftA) and MetC (input chain: 1cl1A) using the DBAli algorithm (Martí-Renom *et al*., 2001) yielded a RMSD of 3.91Å over 54 residues. The overall structural alignment was interpreted as a lack of significant homology (P-value = 3.3; P-value >5.2 represents significant homology according to the algorithm).

Although the tertiary and quaternary structures of Alr and MetC are non-homologous, I further compared the active sites of both enzymes to identify any overlapping regions. Using a ligand-based structural alignment approach (Heifets & Lilien, 2010), the PLP molecules were aligned with a RMSD of 2.67Å over 15 atoms. As shown in Figure 2.28C, the internal aldimines formed between the PLP and the catalytic Lys residues were pointed at opposite directions in the aligned active sites. Within the vicinity of the PLP molecules, the residues from both enzymes could not be superimposed. Despite sharing the same PLP cofactor, the active site structure of Alr was spatially different from that of MetC. It seems reasonable to argue that the promiscuous cystathionine β-lyase activity of Alr could not have been predicted *a priori* based on sequence and structural comparisons alone.

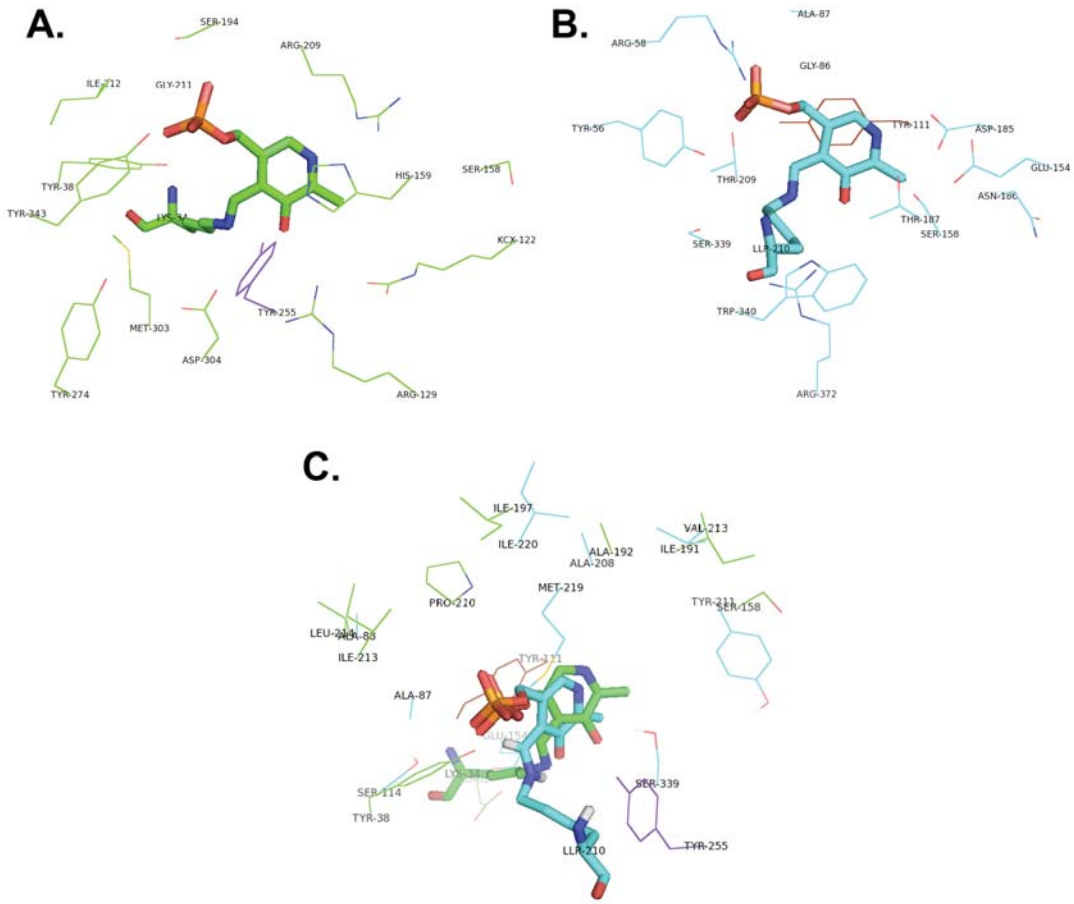


Figure 2.28: Poor superimposition of the active site of *E. coli* Alr and MetC.

The PLP molecules are illustrated in sticks, whereas the residues making up the active site are shown as lines. The catalytic Tyr is coloured purple in the active site of Alr (Tyr255'), and brown in the active site of MetC (Tyr111), respectively. (A) The active site of *E. coli* Alr. (B) The active site of *E. coli* MetC. (C) Structural alignment of the active site of Alr and MetC. Residues shown in the diagram are within 3.5Å of the internal aldimine. The alignment was performed using LigAlign, which aligns protein structures based on the ligand in the active site. The LigAlign software was interfaced to MacPymol to allow visualization of the alignment.

The different active site structure of Alr also indicated that the mechanism of the cystathionine β-lyase activity is likely to be different from the mechanism occuring in the active site of MetC. For instance, the pyridine ring of the PLP molecule in MetC was fully protonated by Asp185. To mimic the protonated state of PLP in MetC, a R209E mutation was introduced to protonate the pyridine ring of Alr. Instead of increasing the cystathionine β-lyase activity in Alr, growth complementation results indicated that the *in vivo* cystathionine β-lyase activity of Alr(R209E) was worse than Alr (Section 2.2.5.2). The reduced *in vivo* activity might have been a result of the Alr(R209E) favouring the transamination activity. As demonstrated previously (Sun & Toney, 1999; Yow *et al.*, 2003; Rubinstein & Major, 2010), a protonated PLP in Alr stabilized the quinonoid intermediate, which was otherwise highly unstable during alanine racemization. The stabilized quinonoid intermediate has been proposed to be critical for opening up possibilities for side activity such as alanine and α-ketoglutarate transamination to produce pyruvate and glutamate (Toney, 2005; Rubinstein & Major, 2010).

2.3.2 Directed evolution of Alr

Promiscuous activities are barely detectable *in vitro* because they are often orders of magnitude lower than native activities (Copley, 2012). A simple approach to circumvent the low activity level is to improve the activity by directed evolution. The underlying principle of a directed evolution experiment is "you get what you select for", and more importantly, it is unlikely to evolve an activity that is not already there (You & Arnold, 1996; Peisajovich & Tawfik, 2007). For instance, a previous study has shown that a 25-fold improvement of a promiscuous activity was sufficient to render the activity detectable using conventional spectrophotometric methods (Patrick & Matsumura, 2008).

The best-evolved Alr variant, Alr-3, was also the dominant variant from the selected pool. Cells expressing Alr with a single Y274F mutation grew most rapidly on the selective medium, thereby implying that its *in vivo* cystathionine β-lyase activity was the highest among the Alr mutants (Section 2.2.6.3). The Y274F mutation was not only found in Alr-3, but also in the rest of the selected variants that carried 1–2 additional amino acid substitutions. The prevalence of the Y274F mutation suggested that Y274′ (primed number denotes a residue from the other monomer) is an important determinant for the cystathionine β-lyase activity in Alr. Y274′ is located in the inner layer of the substrate entranceway of Alr, and is part of the active site (Figure 2.29). The residue is not known for taking part in the catalytic cycle of Alr, but it has been shown to interact with a stable aromatic adduct of D-cycloserine, a suicide inhibitor (Fenn *et al.*, 2003; Wu *et al.*, 2008). The aromatic adduct was a derivative of 3-hydroxyisoxazole pyridoxamine 5′-phosphate (PMP), of which its hydroxyimine formed a hydrogen bond with Y274′ in *E. coli* Alr (Figure 2.29A). This hydrogen bond hinted that Y274′ might interact with the substrate or product in the

active site. However, this hydrogen bond was likely to be abolished by a Y274F mutation, due to the lack of hydroxyl group in the side chain of F274′ (Figure 2.29B).

Therefore, the lack of a hydrogen bond between F274′ and the substrate or product was the likely factor in improving the overall cystathionine β-lyase activity in Alr. It is not possible to determine if the catalytic improvement was due to changed substrate specificity or higher product turnover because no individual kinetic parameters (K_M and k_{cat}) were obtained for the cystathionine β-lyase activity of Alr (Table 2.7 In Section 2.2.6.5). Hypothetically, if F274′ was responsible for altering the substrate preference of Alr-3, the K_M for alanine (the substrate for the native activity) would be drastically increased. However, the K_M for alanine shown by the Y274F variant was reduced by a mere 1.7-fold in comparison to wildtype Alr. In fact, the Y274F mutation did not seem to be too detrimental to the native activity of Alr, as the k_{cat}/K_M of the Y274F variant was only 2.5-fold lower than that of Alr.

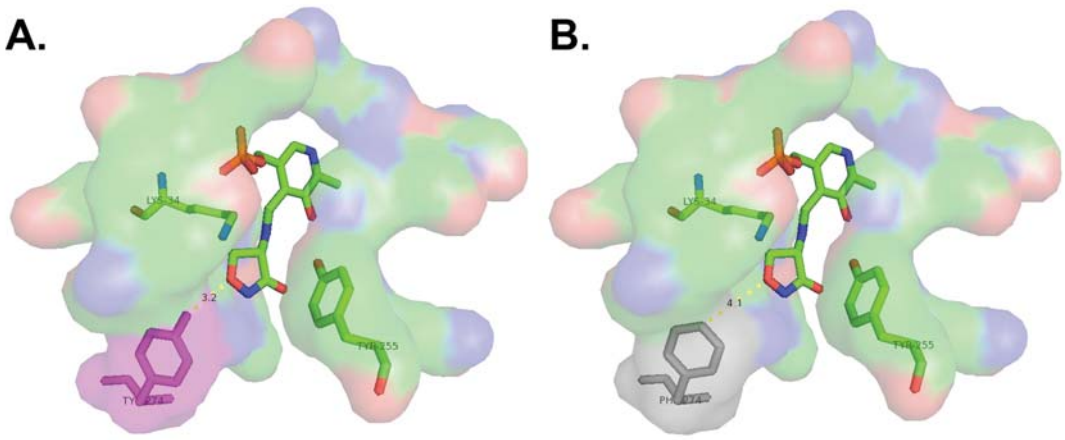


Figure 2.29: The surface diagram of the active site of *E. coli* Alr.

The 3-hydroxyisoxazole PMP and the catalytic pair (Lys34 and Tyr255′) are shown as sticks. The *E. coli* Alr bound to D-cycloserine (PDB code: 2rjh) was used as the template for modelling. (A) The Y274′ residue is highlighted in magenta. The distance between Y274′ and the hydroxyimine of the 3-hydroxyisoxazole PMP is 3.2Å. (B) The modelled mutated residue, F274′, is highlighted in grey. The distance between F274′ and the hydroxyimine of the 3-hydroxyisoxazole PMP is 4.1Å.

Instead, I speculate that the inability of F274′ to form hydrogen bond is likely to facilitate higher product turnover in the active site. Clausen and his colleagues (1996) have proposed that homocysteine and iminopropionate (the products of cystathionine β-lyase activity) were released in a two-step process. The details of the catalytic mechanism of MetC will be discussed in Section 3.3.2 (Chapter 3). Briefly, homocysteine is released after the protonation of Sγ of L-cystathionine by K210 in the MetC's active site (Figure 3.15 in Section

3.3.2, of Chapter 3). The resulting intermediate (a PLP derivative of aminoacrylate) is reprotonated to release iminopropionate. Iminopropionate is then hydrolyzed to pyruvate and ammonia outside the active site. The same authors also proposed that either the Cβ-S bond cleavage (of homocysteine), or the second transaldimination step (to release iminopropionate) is rate limiting based on unpublished results. Therefore, it was possible that product release and the internal aldimine regeneration were faster in the Y274F variant due to the loss of hydrogen bond in anchoring the products. This speculation is also consistent with the *in vitro* cystathionine β-lyase activity curves of Alr performed under near neutral pH conditions (Figure 2.17). Following the apparent burst phase, the plateau of each activity curve possibly reflects the rate-limiting step caused by iminopropionate release.

Although the Y274F variant was deemed the best-evolved variant based on growth complementation results, its *in vitro* cystathionine β-lyase activity was barely detectable. The k_{cat}/K_M of the Y274F variant (0.85 $s^{-1}.M^{-1}$) was slightly higher than that of the phosphoribosylanthranilate isomerase (TrpF) activity (k_{cat}/K_M ~0.3 $s^{-1}.M^{-1}$) reported for an evolved glutamine phosphoribosylpyrophosphate amidotransferase, PurF (Patrick & Matsumura, 2008). Such inefficient enzyme activity was also reported in another study, where the k_{cat}/K_M of the 3-phosphohydroxypyruvate phosphatase activity of an unevolved YeaB (predicted NUDIX hydrolase) was estimated to be >0.024 $s^{-1}.M^{-1}$ (Kim *et al.*, 2010). Clearly, promiscuous activities are often low.

Yet, inefficient enzyme activities can be physiologically relevant in a "life-or-death" situation. In this study, Alr-3 (carrying the Y274F mutation) managed to complement the methionine auxotrophy of *E. coli* Δ*metC* after 2 days. Since MetC also rescued the strain after 2 days, one could easily assume that the *in vitro* cystathionine β-lyase activity of Alr-3 was comparable to that of MetC. However, the superior *in vivo* activity of Alr-3 was not reflected in the *in vitro* activity, which was very low. This disconnect between the *in vitro* promiscuous activity and organismal fitness is an emerging theme in the literature. Previously, PurF was able to rescue *E. coli* Δ*trpF* at a rate similar to wildtype TrpF, but the *in vitro* kinetic activity was very low (see previous paragraph for the k_{cat}/K_M of the TrpF activity of PurF). In another study, the *N*-acetylglutamylphosphate reductase (ArgC) activity of glutamyl phosphate reductase (ProA) showed a k_{cat}/K_M of 0.4 $M^{-1}.s^{-1}$, which was approximately half of the cystathionine β-lyase activity of Alr-3. However, over-production of ProA did not rescue *E. coli* Δ*argC*, unless ProA carried a structural and a regulatory mutations (McLoughlin & Copley, 2008; Copley, 2012). Taken together, this suggests that there is no direct relationship between *in vitro* activity and *in vivo* fitness. A possible explanation for this phenomenon is related to the *in vivo* expression of the promiscuous protein. For instance, in this study, and that of Patrick and Matsumura (2008), a strong T5-*lacO* promoter was responsible for driving the expression of the pCA24N-encoded promiscuous proteins. Furthermore, the copy number of pCA24N was likely to be 300–400

per cell (Qiagen, 2011). Therefore, the amount of the promiscuous protein in a cell was likely to be very high, although the precise amount has not been investigated. Alternatively, regulatory mutations, such as that of ProA (McLoughlin & Copley, 2008), can also increase the amount of promiscuous protein in a cell. This high amount of protein could compensate for its low promiscuous activity in the course of effecting physiologically relevant activities (*i.e.*, rescuing conditionally auxotrophic strains). However, *in vitro* studies often showed that the promiscuous activities were low, because these *in vitro* assays measured the promiscuous activity itself, without taking the enzyme's *in vivo* expression into account. To what degree does the *in vivo* expression of the promiscuous proteins effect a selectable phenotype will require further investigation.

2.3.3 PROMISCUITY IS DRIVEN BY CHEMISTRY

The existence of promiscuous activities catalyzed by non-homologous enzymes poses important evolutionary implications. In particular, results from this study and those from Patrick *et al*. (2008) suggested that promiscuous activities from non-homologous scaffolds are evolvable under selection. Current knowledge of promiscuous activities was largely derived from biochemical characterization of enzyme superfamilies, of which their members share similar structural folds and/or sequence homologies but may catalyze different overall activities (Babbitt & Gerlt, 1997; Gerlt & Babbitt, 2001; Glasner *et al.*, 2006). The promiscuous activities found in extant members of enzyme superfamilies are thus believed to be catalytic vestiges from the ancestral enzymes. However, this cannot be said for Alr and MetC (as well as PurF and TrpF), because the enzymes did not have a common ancestral enzyme. Promiscuity appears to be inherent in enzymes, and remains difficult to predict based on protein sequence and structures.

Overall, results from this chapter support the notion that cofactor-mediated promiscuity (Dimitrov & Vassilev, 2009; Babtie *et al.*, 2010) can represent a plausible starting point in the course of evolving new biochemical functions. Of all the cofactors, PLP is known to be the most catalytically versatile (Eliot & Kirsch, 2004). It will not be surprising if there are more physiologically relevant, promiscuous activities yet to be discovered in other PLP-dependent enzymes. This theme will be explored further in Chapter 3, in which I searched for promiscuous alanine racemase activity in the *E. coli* proteome.

2.4 MATERIALS AND METHODS

All reagents were purchased from Sigma unless specifically noted. Common molecular biology materials, techniques and primer sequences are described in Appendix I. Specific materials and methods used in this chapter are described in the following subsections.

Lysozyme (20 mg/mL) was dissolved in 10 mM Tris-HCl (pH 8.0). One mg of DNase I (2000 U/mg; Roche) was dissolved in 1 mL Buffer A [20 mM Tris-HCl (pH 7.5), 1 mM $MgCl_2$, 50% (v/v) glycerol]. L-Cystathionine was dissolved in 0.04 N HCl and prepared as a 25 mM stock. PLP stock (100 mM) was prepared by dissolving the powder in 1 N HCl. NADH (Merck) was dissolved in 10 mM NaOH, and it was always prepared as 5 mM stocks. DTNB was prepared as a 10 mM stock, by dissolving the powder in 100 mM potassium phosphate buffer (pH 7.0). The NH_4HCO_3 buffer was dissolved in water without adjusting its pH (~8.1 at 20°C).

2.4.1 CONSTRUCTION OF AN EMPTY ASKA VECTOR (pCA24N-NoIns)

The backbone of the empty ASKA vector, pCA24N-NoIns, was derived from pCA24N-tS1-corr (Section I.4 in Appendix I). The 5.2-kb vector backbone was amplified using a pair of outward-facing primers, pCA24N-NoIns.for and pCA24N-NoIns.rev, which contained phosphorylated 5′ termini (Section I.6 in Appendix I; Figure 2.30). The forward primer was designed to bind to the 5′ start of the GFP region, and the reverse primer binds to the 3′ end of the $(His)_6$ tag. The 20-μL PCR reaction contained 1× iProof™ HF master mix (BioRad; also contained 0.4 U of DNA polymerase), 0.5 μM of each primer and 10 ng of the template (pCA24N-tS1-corr). The thermocycling conditions were 98°C for 30 s, 35 cycles of 98°C for 10 s, 67°C for 30 s, 72°C for 3 min, and 72°C for 5 min.

After purifying the PCR products using commercial spin columns (see Section I.8 in Appendix I for details), residual plasmid template was eliminated by *Dpn*I-digestion for 2 h. The 4-base cutter *Dpn*I only cleaves methylated restriction site (*i.e.*, DNA that has been originated from a *dam*$^+$ strain). Without cleaning up the digestion, the phosphorylated ends of the PCR products were ligated to form a circular plasmid, pCA24N-NoIns (Figure 2.30). The plasmid was then used to transform *E. coli* DH5α-E. After an overnight incubation, the plasmid was re-purified from a single transformed colony. The absence of a gene insert in pCA24N-NoIns was confirmed by sequencing, using a forward primer that anneals upstream of the 6×His region (pCA24N.for; Section 1.6 in Appendix I) and a reverse primer that anneals downstream of the GFP region (pCA24N.rev2; Section 1.6 in Appendix I). The molecular size of pCA24N-NoIns was also confirmed by a *Not*I-digestion. In addition, *E. coli* colonies carrying pCA24N-NoIns also appeared as fluorescing green when viewed using a UV transilluminator, indicating that the orientation of the GFP had not been altered during the cloning process.

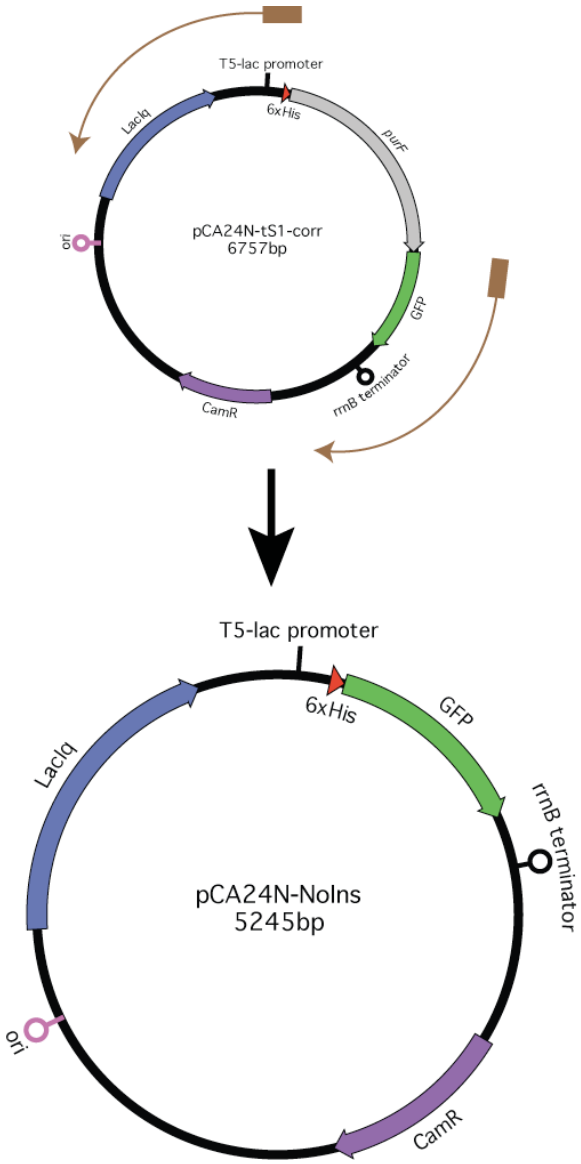


Figure 2.30: Schematic diagram of pCA24N-NoIns construction.

The 5.2-kb vector backbone was amplified from pCA24N-tS1-corr using a pair of 5′-phophorylated, outward-facing primers (brown rectangles). Brown arrows denoted the direction of PCR amplification. The phosphorylated ends of the linear PCR product were ligated to form a circular plasmid, pCA24N-NoIns.

2.4.2 Construction of pCA24N-*alr*(GFP−)

Similar to other plasmids from the ASKA collection, the *alr* gene of pCA24N-*alr* had been fused to a GFP. To remove the GFP, the *alr* gene was isolated and cloned into a pCA24N vector that lacks the GFP gene.

The oligonucleotides, pCA24N.for and Alr.rev, were used to amplify the *alr* gene insert from pCA24N-*alr*. As explained earlier (Section 2.4.1), pCA24N.for anneals upstream of the

6×His region. The reverse primer, Alr.rev, anneals to the 3′ of the *alr* gene. A stop codon and an *Sfi*I recognition site were also incorporated at the 3′ end of Alr.rev. The PCR reaction was set up with 1× Thermopol buffer, 1 U Vent$_R$® DNA polymerase (both from New England Biolabs), 0.2 mM dNTPs, 0.5 μM of each primer and ~20 ng of pCA24N-*alr*. The thermocycling conditions were 98°C for 30 s, 35 cycles of 98°C (30 s), 58°C (30 s), 75°C (90 s), and finally 75°C for 1 min. The amplified product was purified using commercial spin column.

The plasmid pCA24N-*trpD*(GFP−) was used to prepare the vector backbone since the GFP had been eliminated previously (see Section I.4 in Appendix I). The *alr* insert and pCA24N-*trpD*(GFP−) were *Sfi*I-digested to generate compatible sticky ends. The 4.5-kb linearized vector and 1.1-kb insert were then ligated to form pCA24N-*alr*(GFP−), and the plasmid was electroporated into *E. coli* Δ*metC*. Transformed colonies were PCR-screened for the presence of the plasmid using vector-specific primers, pCA24N.for and pCA24N.rev2. The plasmid purified from a single clone was also checked with *Sfi*I digestion. Sequencing of the *alr* insert (using pCA24N.for and pCA24N.rev2) confirmed that no mutation had been introduced into *alr* during the cloning process.

2.4.3 Construction of pCA24N-*metC*(GFP−)

To remove the GFP tag fused to the pCA24N-*metC* (from the ASKA collection), the pCA24N-*trpD*(GFP−) was used as the host vector (see Section 2.4.2 for details).

The *metC* gene insert was amplified from pCA24N-*metC* using the oligonucleotides, pCA24N.for and MetC(GFP−).rev. The reverse primer, MetC(GFP−).rev, anneals to the 3′ of the *metC* gene. The primer also has a stop codon and an *Sfi*I recognition site incorporated at the 3′ end of the sequence. The PCR reaction was set up with 1× GoTaq buffer (Promega), 1 U Phusion® DNA polymerase (Finnzymes), 0.2 mM dNTPs, 0.5 μM of each primer and ~20 ng of pCA24N-*metC*. The thermocycling conditions were 95°C for 1 min, 30 cycles of 95°C (30 s), 53°C (30 s), 72°C (60 s), and finally 72°C for 2 min. The amplified product was purified using commercial spin column.

The purified *metC* insert *Dpn*I-digested for 2.5 h to eliminate residual plasmid template that had been carried over from the PCR reaction. *Sfi*I was then added to the reaction to produce compatible sticky ends, and the reaction was further incubated for 3 h. The digested 1.2-kb fragment was gel-purified, and ligated into a *Sfi*I-digested vector to form pCA24N-*metC*(GFP−). *E. coli* MB2795 was transformed with the ligated plasmid and transformed cells were grown on LB-chloramphenicol supplemented with 0.5 mM D-Ala. A single colony that appeared the next day was inoculated in 3 mL LB-chloramphenicol with 0.5 mM D-Ala. The culture was grown overnight at 37°C. A 1-μL aliquot of the saturated culture was used for PCR screening with pCA24N-specific primers (pCA24N.for and

pCA24N.rev2) were used to detect the presence of the plasmid-encoded *metC*. The amplified *metC* insert was purified and sequenced to confirm that no mutation had been introduced during the cloning process. The remaining culture was used to make freezer stock, and stored at −80°C.

2.4.4 Growth complementation assays

2.4.4.1 Complementation of MetC deletion

A saturated culture of each *E. coli* Δ*metC* clone was prepared by inoculating 0.5 mL LB-chloramphenicol-kanamycin with the strain's glycerol stock. The culture was grown overnight, and pelleted at 12,000 ×g for 4 min on a microcentrifuge. After discarding the supernatant, the cell pellet was resuspended with 1 mL of 1× M9 medium, and re-pelleted. This pellet-washing step was repeated twice. The final cell resuspension was diluted to an OD_{600} ~0.1 using 1× M9 medium. A 2.5-μL aliquot was streaked or spread on the selective M9 agar plates. Cell count was obtained by plating appropriate dilutions of the culture on (non-selective) LB-chloramphenicol-kanamycin agar plates. The M9 agar plates were incubated at 28°C for at least 1 week, whereas the LB agar plates were incubated overnight at 37°C. All agar plates were placed in plastic containers lined with moistened paper towels.

The M9 agar plates were made up of 1× M9 medium, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$ (Merck), 0.4% (w/v) glucose, 34 μg/mL chloramphenicol and 30 μg/mL kanamycin. A final concentration of 50 μM IPTG (GBT GoldBio) was added if protein over-expression was desired. 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).

2.4.4.2 Complementation of Alr (and DadX) deletion

The glycerol stock of each *E. coli* MB2795 (Δ*alr* Δ*dadX*) clone was used to inoculate 0.5 mL LB-chloramphenicol supplemented with 0.5 mM D-Ala. The culture was grown overnight to saturation, and harvested by centrifuging at 12,000 ×g for 4 min. The supernatant was discarded, and the cell pellet was resuspended with 1 mL of LB. The cell resuspension was re-pelleted. This pellet-washing step was repeated twice. LB was used to resuspend the final cell pellet to an OD_{600} ~0.1. A 2.5-μL aliquot was streaked or spread on the selective LB-chloramphenicol agar plates supplemented with 50 μM IPTG to allow protein over-expression. Appropriate dilutions of the cell resuspension were plated on non-selective LB-chloramphenicol plates supplemented with 0.5 mM D-Ala to obtain cell counts. The selective agar plates were incubated at 28°C for at least 1 week, whereas the non-selective agar plates were incubated overnight at 37°C. All agar plates were incubated in plastic containers lined with moistened paper towels.

2.4.5 Enzyme over-expression, purification & dialysis

Unless otherwise specified, all enzyme variants described in this chapter were over-expressed, purified and dialyzed according to the following protocol. Metal affinity chromatography was used to purify all enzyme variants, as the N-terminus of all enzyme variants in this thesis were fused to a $(His)_6$ tag.

A 4-mL overnight culture was used to inoculate 250 mL LB supplemented with appropriate antibiotics. For the MB2795 strain, D-Ala (0.5 mM, final concentration) was also added to the media. The culture was incubated at 37°C until its OD_{600} reached ~0.5. IPTG (0.5 mM, final concentration) was added to the culture to induce enzyme over-expression. The culture was then incubated at 28°C for another 5–16 h. At the end of incubation, the cells were pelleted at 5,700 rpm and 4°C for 20 min. After discarding the supernatant, the pellet was stored overnight at −80°C.

A 4-mL aliquot of Column Buffer [50 mM potassium phosphate buffer (pH 8.0), 300 mM NaCl, 3 mM imidazole, 1 mM β-mercaptoethanol, 10% (v/v) glycerol, 10 μM PLP] was used to resuspend every 1 g of the thawed cell pellet. Next, the cell suspension was incubated with lysozyme (0.5 mg/mL, final concentration), DNase I (4 U, final amount) and 50 μL protease inhibitor cocktail (Sigma) on ice for 30–45 min. The cells were lysed by 10 cycles of sonication (amplitude 45%; Ultrasonic processor S-4000, Misonix) on ice. The lysate was clarified by centrifugation at 13,000 rpm for 35 min, at 4°C. The resulting supernatant was filtered through a 0.22-μm filter unit to remove possible aggregates. Filtered supernatant was incubated with 1 mL of Talon® metal affinity resin (bed volume, 0.5 mL; Clontech) at 4°C for 2 h with agitation, to allow the soluble $(His)_6$-tagged enzyme to bind to the resin. Prior to the incubation, the resin had been pre-washed twice with 8 mL of Column Buffer. At the end of the incubation, the resin was pelleted (800 ×g, 5 min, 4°C), and washed twice with 5 mL of Column Buffer. Then, the resin was suspended using 1 mL of Column Buffer, and transferred to a gravity flow column (Bio-Rad). The resin in the column was washed twice with 3 mL of Column Buffer and lastly, once with 3 mL of Column Buffer with a higher imidazole concentration (15 mM). The purified enzyme 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]. Eluted fractions usually appeared yellow due to the presence of PLP bound to enzymes.

Each enzyme was dialyzed overnight against Dialysis Buffer [50 mM potassium phosphate (pH 8.0), 150 mM NaCl, 10% (v/v) glycerol) at 4°C before being quantified (Section 2.4.5) and used for activity assays (Sections 2.4.7 and 2.4.8). Enzyme dialysis was performed in SpectraPor2 cellulose tubing (MWCO 12–14 kDa; SpectrumLabs) according to the manufacturer's instruction. Recovered enzyme was usually concentrated using a centrifugal filtration concentrator (Vivaspin 6; GE Healthcare) to 1–2 mL. The enzyme was finally passed through a 0.22-μm filter unit to remove any possible aggregates, and the

filtrate was aliquoted into microcentrifuge tubes. These enzyme aliquots were stored at either 4°C or −80°C. All enzymes were used within 2 weeks of being purified.

2.4.6 Enzyme quantification

Either a 10-fold or 20-fold dilution of each enzyme sample (prepared in a total volume of 100 μL) was scanned across the 220–600 nm wavelengths in a Cary 100 UV-Vis spectrophotometer (Agilent). The peaks at A_{280} and A_{420} corresponded to the amount of total enzyme and the amount of PLP-bound enzyme, respectively.

The background-corrected A_{280} of the enzyme sample was used to calculate the total enzyme concentration. The molar extinction coefficient of each enzyme variant was determined as previously described (Pace *et al.*, 1995). The ε_{280} used for each Trp residue was 5,500 $M^{-1}.cm^{-1}$, Tyr 1,490 $M^{-1}.cm^{-1}$ and Cystine 125 $M^{-1}.cm^{-1}$.

The background-corrected A_{420} of the enzyme sample was used to calculate the concentration of PLP-bound enzymes. The molar extinction coefficient of PLP used was 8,450 $M^{-1}.cm^{-1}$ (Inagaki *et al.*, 1986). A comparison of the concentrations of PLP-bound enzymes versus the total enzymes would reveal the fraction of catalytically active enzymes in the sample.

2.4.7 Steady state kinetic assays

All enzyme activity assays were performed using a thermostatically controlled Cary 100 UV-Vis spectrophotometer (Agilent). All reactions were measured in quartz cuvettes with 1-cm path length. All initial rates (V_0) were corrected for the background rate of spontaneous reaction (NADH oxidation for L-Ala racemization; DTNB decomposition for cystathionine β-elimination) in the absence of enzyme. Non-linear regression analyses of all V_0s and Michaelis-Menten curve fittings were performed using the Origin software (OriginLab). Unless specifically noted, all Michaelis-Menten plots were constructed with 7–9 data points of varying substrate concentrations.

2.4.7.1 Alanine racemization

A continuous coupled assay was set up for the racemization of L-Ala to D-Ala based on two previously reported assay systems (Esaki & Walsh, 1986; Patrick *et al.*, 2002). A 1-mL reaction mixture contained 100 mM *N*-cyclohexyl-2-aminoethanesulfonic acid (CHES, pH 9.0), various L-Ala concentrations (Merck), 0.2 mM NADH, 1 U D-amino acid oxidase and 120 U lactate dehydrogenase. Each mixture was pre-incubated at 37°C for about 2 min, and a purified Alr (8.4 nM, final concentration) aliquot was added to initiate the reaction. NADH was oxidized to NAD^+ during the reaction, and this could be monitored by a

decrease in A_{340}. The molar extinction coefficient of NADH (6,220 $M^{-1}.cm^{-1}$) was used to calculate the amount of NAD^+ formation. The substrate (L-Ala) concentrations used for all racemization assays typically ranged from (0.3–10) × K_M. In the absence of enzyme, the rate of spontaneous NADH oxidation measured throughout this study was ≤0.48 μM per min (which was approximately ≤2.0% of the maximal rate of Alr-catalyzed NAD^+ formation).

2.4.7.2 CYSTATHIONINE β-ELIMINATION

The continuous assay for cystathionine β-elimination was modified from the protocol described by Uren (1987). Every reaction mixture contained 50 mM Tris-HCl (pH 8.8), a fixed amount of purified enzyme and 0.4 mM DTNB. The final concentration of MetC used was 11 nM in 1 mL assays, whereas the concentration of Alr variants ranged from 0.8–7 μM in 0.5 mL assays. The reaction mixture was pre-incubated at 37°C for about 2 min, and L-cystathionine was added to initiate the reaction. The enzyme-catalyzed reaction produced homocysteine, of which its free thiol group would stoichiometrically react with DTNB to form 2-nitro-5-thiobenzoate dianion (TNB^{2-}). The formation of TNB^{2-} was monitored as an increase of A_{412}, and was calculated using its molar extinction coefficient of 14,150 $M^{-1}.cm^{-1}$ (Riddles *et al.*, 1979). For MetC assays, the substrate (L-cystathionine) concentrations typically ranged from (0.25–3) × K_M. Meanwhile, assays added with the Alr variants usually contained 0.5–10 mM L-cystathionine, which corresponded to (0.6–3) × K_M. Due to the low solubility of L-cystathionine (~5.6 mg/mL in 0.04 N HCl, which corresponded to a concentration of 25 mM), it was not possible to use >10 mM. To measure the cystathionine β-lyase activity of Alr-3, all reactions were performed in 0.5 mL assays, and the buffer was 100 mM Tris-HCl (pH 8.8). In the absence of enzyme, the rate of spontaneous DTNB decomposition oxidation measured throughout this study was ≤0.35 μM per min (which was approximately ≤0.6% of the maximal rate of MetC-catalyzed TNB^{2-} formation).

2.4.8 ENDPOINT ASSAYS AND ELECTROSPRAY MASS SPECTROMETRY

Endpoint activity assays and mass spectrometry were used to detect the reaction products of cystathionine β-elimination. A typical reaction mixture (total volume, 200 μL) contained 50 mM NH_4HCO_3 as the assay buffer, 5 mM L-cystathionine and a purified aliquot of MetC or Alr. A positive control containing 0.8 μM MetC, and a negative control without any enzyme, were prepared. The Alr was assayed at a concentration of 5.1 μM, and PLP (50 μM, final concentration) was optionally added. All reactions were foil-wrapped, and incubated at 37°C for 16 h in thin-walled, low-binding 0.5 mL tubes (Axygen).

At the end of incubation, 60 μL of Talon® metal affinity resin (bed volume, 30 μL) was added to each reaction. The resin was pre-washed twice with 200 μL assay buffer (50 mM NH_4HCO_3) prior use. The assays were further incubated at room temperature (~20°C) for 1 h to allow the $(His)_6$-tagged enzyme to bind to the resin. To remove the enzyme/resin

from the reactions, the tubes were pelleted at 12,000 ×g for 2 min. The supernatants lacking the enzyme/resin were transferred to new low-binding 0.5 mL tubes. The components in each supernatant fraction were then concentrated by evaporating all NH_4HCO_3 buffer in a vacuum concentrator (Concentrator 5301, Eppendorf), until no solution was observed at the bottom of the tube. The tubes were finally sent to the Norris Lab (Institute of Molecular Biosciences, Massey University, Palmerston North) for mass spectrometry analyses.

Briefly, the components in each tube were dissolved in 50 μL of 50% acetonitrile/water/0.1% formic acid. The tubes were incubated in a sonication bath (Elma S15H) for 3 min to assist component solubilization. Insoluble particulates were removed by pelleting at 16,300 ×g for 15 min. A 5-μL aliquot of the supernatant was injected into an Agilent 6520 Q-TOF mass spectrometer, using an Agilent 1200 LC system with a bypass run of 50% acetonitrile/water/0.1% formic acid at 0.1 mL/min. The samples were analyzed in positive mode and the collected spectra were examined using Mass Hunter Qualitative Analysis software (Rev B.03.01). Mass spectrometry conditions were set at: VCap = 3800 V, Fragmentor Voltage = 125 V, Skimmer Voltage = 65 V, OCT RF Voltage = 750 V, Gas Temperature = 325°C, Drying Gas = 5 L/min, Nebulizer = 40 psig, Acquisition Rate = 3 spectra/s, TOF spectra mass range = 100–4,500 m/z.

2.4.9 Construction of Alr(R209E)

A single codon mutation (R209E) was introduced to Alr by overlap extension PCR (Ho *et al.*, 1989). A pair of complementary primers, Alr-R290E.for and Alr-R209E.rev (Section I.6 in Appendix I), was designed to mutate the CGC codon to GAA. The mutagenic codon was in the middle of the 22-bp primer sequence, and was flanked by 10 and 9 nucleotides homologous to the upstream and downstream regions, respectively.

The plasmid carrying the *alr* gene [pCA24N-*alr*(GFP−)] was used as the DNA template for two separate PCR amplifications. First amplification involved using the vector-specific forward primer (pCA24N.for) and Alr-R209E.rev to produce a 718-bp DNA fragment. Second amplification involved using Alr-R209E.rev and the vector-specific reverse primer (pCA24N.rev2) to produce a 571-bp DNA fragment. Both PCR reactions were carried out in 1× Thermopol buffer, 2 U $Vent_R^®$ DNA polymerase, 0.2 mM dNTPs, 0.5 μM of each primer and ~20 ng of plasmid template. The thermocycling protocol for both reactions was 95°C for 1.5 min, 35 cycles of 95°C (30 s), 52°C (30 s), 72°C (1 min), and finally, 72°C for 30 s. Both amplified fragments were gel-purified.

The pair of mutagenic primers was used to produce two fragments with overlapping 3′ ends. The overlapping ends were annealed in a second PCR reaction. After the 3′ ends of both fragments were annealed, each strand could act as a primer for the 3′ extension of the complementary strand. Therefore, equal amount of purified overlapping fragments (10 ng

of each 718-bp and 571-bp fragments) was added to a PCR mix containing 1× Thermopol buffer, 2 U Vent$_R$® DNA polymerase, 0.2 mM dNTPs and 0.5 μM pCA24N.for and 0.5 μM pCA24N.rev2. The pCA24N-specific primers were used to amplify the entire 1.3-kb DNA product that had been reassembled from the two overlapping fragments. The PCR cycling conditions were as follows: 1 cycle of 95°C for 1 min, 35 cycles of 95°C (30 s), 54°C (20 s), 72°C (70 s), and finally 1 cycle of 72°C for 30 s.

The resulting 1.3-kb product (containing a codon mutation R209E) was digested with *Sfi*I. The fragment was then ligated to a *Sfi*I-digested 4.5-kb pCA24N(GFP−) vector (Section 2.4.2). *E. coli* MB2795 transformed with the ligated product was grown on LB-chloramphenicol supplemented with 0.5 mM D-Ala. The transformed colonies were PCR-screened using pCA24N-specific primers, and a 1.3-kb amplified product was sequenced. Nucleotide sequence of the fragment validated that the Arg209 codon had been changed to Glu209.

2.4.10 Random mutagenesis of alr

2.4.10.1 Construction of randomized Alr library

The GeneMorph II Random Mutagenesis kit (Stratagene) was used to introduce random mutations into the *alr* gene during PCR. An epPCR reaction mixture (total volume, 100 μL) contained 1× Stratagene Mutazyme buffer, 2.5 U Mutazyme DNA polymerase, 0.8 mM dNTPs, 0.5 μM of pCA24N.for primer, 0.5 μM of Alr.rev primer and 567 ng of plasmid template [pCA24N-*alr*(GFP−)]. That amount of plasmid template corresponded to 100 ng of initial target DNA (*alr*), which was recommended (by the manufacturer) to achieve a mutation rate of 1–4 amino acid mutations per gene. The thermocycling conditions were 1 cycle of 95°C for 2 min, 35 cycles of 95°C (30 s), 58°C (30 s), 72°C (70 s), and 1 final cycle of 72°C for 2 min.

The yield of the PCR product was estimated to be 1 μg by running a 1-μL aliquot of the product on an agarose gel. Given that the initial target template was 100 ng, the final yield reflected a 10-fold amplification of the initial target template and an overall duplication number of 3.3 ($2^{\text{duplication}}$ = amplification fold). The PCR efficiency (*i.e.*, the probability that any particular sequence is duplicated in a given PCR cycle) was therefore 0.068 (Patrick *et al.*, 2003).

Next, the 1.2-kb PCR product was purified using spin columns. Since a significant amount of plasmid template was co-purified with the PCR product, the sample was treated with *Dpn*I so that the plasmid template would be fragmented. *Sfi*I was also added to the reaction to generate compatible 3′ sticky ends with those of the pCA24N-tS1(no-ext) vector (Section I.4 in Appendix I), which also had been *Sfi*I-digested. The digested 1.1-kb fragment

(containing the randomized *alr* gene) and the linearized 4.5-kb vector backbone were gel-purified using the QIAEXII Gel Extraction kit (Qiagen). The purified *alr* insert (365 ng) was ligated to the vector (500 ng) at a 3 : 1 molar ratio. A "vector only" control was also set up using the same ligation components but without any gene insert. The ligation reaction was incubated for 16 h before being heat-inactivated. Then, the ligated products were cleaned up using the MinElute columns (Qiagen). A 1-μL aliquot of the ligated product was used to transform a fresh 50-μL aliquot of *E. coli* Δ*metC*. A total of 14 electroporations were performed for the *alr* library, and the time constant for each electroporation was >5.10 ms. The Δ*metC* cells were also transformed with the "vector only" control and pUC19 (a standard plasmid control used to determine transformation efficiency), respectively. Different dilutions of the recovered cells were plated on LB-kanamycin-chloramphenicol plates, and allowed to grow at 37°C for 15 h.

Based on the colony counts, the transformation efficiency of the Δ*metC* strain was found to be 2.5×10^9 cfu/μg of pUC19. After subtracting from the "vector only" background (~0.0014%), the total number of transformed clones for the *alr* library was 3.4×10^7 cfu. To check for the presence of the randomly mutated *alr* gene cloned in the pCA24N plasmid, 20 colonies were randomly picked from the library plates. The colonies were grown in 0.2 mL LB-kanamycin-chloramphenicol for 8 h. One μL of each saturated culture was used as the template for PCR screen. Vector-specific primers, pCA24N.for and pCA24N.rev2, were used. PCR results from all 20 clones showed the presence of the *alr* insert (carried in the plasmid), as indicated by a 1.3-kb fragment. After purifying these PCR products using commercial spin columns, they were sequenced in order to obtain the number and type of mutations introduced during epPCR.

Meanwhile, a spreader glass rod pre-wetted with LB-kanamycin-chloramphenicol was used to scrape the remaining colonies on the library agar plates. The ~45 mL cell suspension was concentrated by pelleting the cells at 3,500 ×g for 30 min, at 4°C. The LB supernatant was discarded, and the cell pellet was resuspended in 8 mL LB-kanamycin-chloramphenicol supplemented with 12.5% (v/v) glycerol. The OD_{600} of the final cell suspension was 512, which corresponded to ~1.3×10^{11} cfu/mL [assuming that OD_{600} ~1.0 is roughly equivalent to 2.5×10^8 cfu/mL (Sambrook & Russell, 2001)]. The cell suspension was evenly divided into 8 aliquots, and finally stored at −80°C.

2.4.10.2 Selection for *ALR* variants with higher lyase activity

A frozen aliquot of the *E. coli* Δ*metC* library harbouring the randomly mutated *alr* gene (pCA24N-*epAlr*) was thawed. Ten μL of the library was used to inoculate 50 mL LB-kanamycin-chloramphenicol, and grown at 37°C until OD_{600} reached ~0.8. The cells were harvested by centrifuging at 2,200 ×g for 16 min, at 4°C. After discarding the LB supernatant, the cell pellet was resuspended twice in 50 mL of 1× M9, and re-pelleted

similarly after each resuspension. The final pellet was resuspended in 50 mL of 1× M9. One mL of the cell suspension (~1.04 × 10^8 cfu; calculated based cells grown on non-selective medium) was plated on a selective medium (24.5 cm × 24.5 cm), M9-kanamycin-chloramphenicol supplemented with 0.4% (w/v) glucose and 5 μM IPTG.

For side-by-side comparison, the same *E. coli* strain carrying the parental *alr* gene [pCA24N-*alr*(GFP−)] was included in the selection experiment. In addition, the same *E. coli* cells harbouring the empty ASKA vector (pCA24N-NoIns) were used as a negative control, whereas those harbouring the pCA24N-*metC*(GFP−) [Section 3.4.2] was used as a positive control. Overnight cultures of these 3 clones were subjected to the same routine of 1× M9 washes as previously described. Cell densities were standardized by adjusting their OD_{600} to ~0.1, and a further 100-fold dilution was performed for each strain. A 2-μL aliquot of each diluted cell suspension was plated on separate M9-kanamycin-chloramphenicol agar plates supplemented with 0.4% (w/v) glucose and 5 μM IPTG. All M9 agar plates, including those plated with the *alr* library, were incubated at 28°C for at least 3 days. To enumerate the number of cells plated on M9 agar plates, appropriate dilutions were made for each strain, and plated on non-selective LB-kanamycin-chloramphenicol. The LB agar plates were grown overnight at 37°C.

After 2.5 days of incubation, the *alr* library plate was divided into 4 equal sections. Twelve colonies were randomly picked from each section, and inoculated in 0.2 mL LB-kanamycin-chloramphenicol. A total of 48 cultures were incubated overnight at 37°C. The next day, 1 μL from each saturated culture was used as the template for PCR screen. The remaining cultures were stored at −80°C after adding each with 50 μL of 50% (v/v) glycerol. The use of the vector-specific primers (pCA24N.for and pCA24N.rev2) in the PCR screen would ensure amplification of only pCA24N-encoded *alr* genes. The resulting PCR products were purified and sequenced to reveal the number and type of mutations in found in the *alr* gene that had allowed rapid colony formation on the selective medium.

2.4.10.3 Retransformation/Recloning and restreaking tests

The growth phenotypes of each *alr* variant from the epPCR library section (Section 2.4.10.2) were confirmed by retransformation/recloning and restreaking tests.

Plasmid was purified from all the mutated variants using commercial spin columns. A new batch of electrocompetent *E. coli* Δ*metC* cells was transformed with the purified plasmids, respectively. For each plasmid variant, one retransformed clone was picked and grown to saturation in 2 mL LB-kanamycin-chloramphenicol at 37°C. The cultures were PCR-screened using pCA24N-specific primers (pCA24N.for and pCA24N.rev2) to confirm the presence of the plasmid before being made as freezer stocks.

For (two) variants in which no mutation was been detected in the *alr* gene, the plasmid was similarly purified from the original clones (which had been maintained as freezer stocks). The plasmids were digested overnight with *Sfi*I, and the excised *alr* gene was gel-purified. The *alr* insert was ligated to the *Sfi*I-treated pCA24N vector in a molar ratio of 3 : 1. After heat-inactivating the ligation reaction, it was diluted 5 fold using sterile dH_2O. A 2-μL aliquot of each ligated product was used to transform freshly prepared electrocompetent *E. coli* Δ*metC* cells. One retransformed clone was selected and grown overnight in 2 mL LB-kanamycin-chloramphenicol at 37°C. The cultures were PCR-screened using pCA24N-specific primers (pCA24N.for and pCA24N.rev2) to confirm the presence of the plasmid before being made as freezer stocks.

For the restreaking experiments, 1 mL LB-kanamycin-chloramphenicol was inoculated with the freezer stock of either a retransformed or a recloned *alr* variant. After an overnight incubation at 37°C, the culture was pelleted at 12,000 ×g for 4 min in a benchtop microcentrifuge. The cell pellet was resuspended thrice in 1 mL of 1× M9 medium, and re-pelleted similarly after each resuspension. The final cell pellet was resuspended in 1 mL of 1× M9. A 20-fold dilution was performed for the cell suspension, and a 2.5-μL aliquot of the diluted suspension was streaked on two sets of agar, M9-kanamycin-chloramphenicol-glucose agar plates with and without 5 μM IPTG. All M9 agar plates were incubated at 28°C, in sealed container lined with moistened paper towels. Colony formation was monitored for at least 1 week.

CHAPTER 3

Reciprocal promiscuity in two non-homologous enzymes

Acknowledgement: Natasha le Roux assisted with the retransformation of ASKA clones described in Section 3.4.1.

3.1 INTRODUCTION

The catalytic diversity of PLP-dependent enzymes is evident from the number of enzyme classes they belong to: PLP-dependent enzymes are found in five out of six enzyme classes, according to the Enzyme Commission classification (Christen & Mehta, 2001). In addition to their native functions, many PLP-dependent enzymes are known to catalyze side reactions *in vitro* (Eliot & Kirsch, 2004). However, the *in vivo* relevance of these side activities is largely unexplored. Given that the PLP-dependent enzymes occupy a central role in carbon and nitrogen metabolism (by catalyzing a variety of reactions involving amino acids and their derivatives), the ability of these PLP-dependent promiscuous activities to alter the cell's metabolic flux, via changes in their activity or expression levels, is not clear.

In Chapter 2, the promiscuous activity of alanine racemase (Alr) was shown to rescue an *E. coli* strain with its cystathionine β-lyase (MetC) gene deleted. Both enzymes are PLP-dependent although they are not homologous. The results had two implications: (1) promiscuous activities mediated by PLP can be physiologically valuable under certain selective conditions; and (2) the native function of a PLP enzyme can be a promiscuous function of another non-homologous PLP enzyme. Enzyme promiscuity in non-homologous enzymes has gained attention in recent years because these enzymes did not diverge from a common ancestor, and thus, their promiscuous activities did not originate from a multifunctional ancestral protein (Jensen, 1976; also refer to Chapter 2 for details of this reasoning). Instead, their promiscuous activities have been described as a form of molecular contingency (Patrick & Matsumura, 2008). To date, only a handful of groups have explored protein promiscuity in non structural homologues (James & Tawfik, 2001; McLoughlin & Copley, 2008; Patrick & Matsumura, 2008; Kim *et al.*, 2010).

In this regard, I considered the PLP-dependent enzyme family a good system to explore promiscuous activities in evolutionarily unrelated enzymes. Seven structural scaffolds are adopted by modern PLP-dependent enzymes (Percudani & Peracchi, 2009), and enzymes belonging to each fold group are likely to represent an independent evolutionary lineage (Christen & Mehta, 2001). The very first ancestral PLP-dependent enzymes were probably multifunctional due to the innate catalytic potential of the cofactor. In the absence of an apoenzyme, the electrophilic PLP is believed to catalyze many reactions — which are normally catalyzed by PLP-dependent holoenzymes — very slowly (Eliot & Kirsch, 2004; Vacca *et al.*, 2008). This argument is based on the fact that almost all PLP-dependent enzymes (except glycogen phosphorylase) share the initial mechanistic steps during catalysis (*i.e.*, the formation of internal and external aldimine in the active site). However, as each lineage evolved, the reaction and substrate specificities of the enzymes narrowed. The enzymes became more specific for a certain type of catalysis and/or substrate due to the specialization of the protein scaffolds. In other words, the enzyme scaffolds may dictate the substrate choice as well as the type of catalysis possible for modern PLP-dependent enzymes.

In the course of studying the promiscuous activity of Alr in Chapter 2, I acquired *E. coli* MB2795 from Prof. Kurt Krause (University of Otago). As the two homologous alanine racemase genes (*alr* and *dadX*) of this strain have been deleted, the cells are not able to racemize alanine isomers (Strych *et al.*, 2001). The absence of D-Ala in this strain leads to the formation of a defective cell wall, which in turn effects cell death. Due to the potential for PLP-mediated promiscuity, I began searching for promiscuous alanine racemase activities that could complement the growth of *E. coli* MB2795. In this chapter, I describe the serendipitous discovery of a promiscuous alanine racemase activity shown by *E. coli* MetC (whose native activity is the promiscuous activity of Alr, as demonstrated in Chapter 2). The reciprocal activity profile shown by Alr and MetC is a novel finding, especially in light of their lack of structural homology. Overall, the findings from Chapter 2 and 3 support the notion that promiscuous activities can be found in non-homologous enzymes. Sequence or structural homology alone is not the only parameter to look for when predicting promiscuous activities.

3.2 RESULTS

3.2.1 SUPPRESSOR CLONES THAT RESCUE *E. COLI* MB2795 (Δ*alr* Δ*dadX*)

First, a multicopy suppression screen was set up in an attempt to search for promiscuous alanine racemase activity in the *E. coli* proteome. The *E. coli* MB2795 strain (Δ*alr* Δ*dadX*) is auxotrophic for D-Ala. The strain will not grow unless D-Ala is exogenously supplied. Therefore, the pooled plasmids of the ASKA library (Patrick *et al.*, 2007) were used to transform *E. coli* MB2795 to select for multicopy suppressors that could complement the growth of the strain. Transformed clones were selected on an IPTG-supplemented LB-chloramphenicol agar plate (see Section 3.4.1 for details). A total of 36,000 cells were initially spread on the selective agar. This amount of cells corresponded to ~7-fold over-sampling of the ASKA library, which consisted of 5,272 clones. This over-sampling ensured that ~99% of all ASKA plasmids were present [as calculated by the library diversity and completeness programme, GLUE (Patrick *et al.*, 2003)]. In the absence of exogenous D-Ala, ten clones survived and formed colonies on the agar plate after 1–4 days of incubation. No additional colonies were formed even after 2 weeks of incubation.

These ten clones were screened by PCR, in order to determine the identity of their plasmid-encoded ORF. Figure 3.1 shows the plasmid-encoded ORF amplified from each of the ten clones. The sizes of the ORFs varied between 1,000–1,650 bp, suggesting that more than one type of suppressor clone had been isolated. These ORFs were sequenced to reveal their identity. A total of five ORFs were identified from the ten clones: *alr*, *dadX*, *cycA*, *metC* and *wcaL* (Table 3.1).

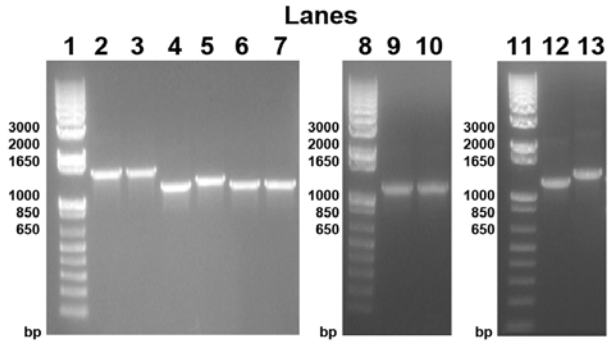


Figure 3.1: PCR product amplified from the ten suppressor clones using vector-specific primers.
Lanes 1, 8 and 11, Invitrogen 1 kb plus DNA ladder. Lanes 2–7, clones 1–6. Lanes 9–10, clones 7–8. Lanes 11–12, 9–10.

To confirm the rescuing activity of these ORFs, their individual ASKA plasmids were isolated and used to retransform fresh *E. coli* MB2795. The clones were grown on separate agar plates to verify that their growth phenotype was not a result of cross-feeding. Out of the five suppressor clones, only the WcaL-expressing clone did not grow after this retransformation test, indicating that this clone was likely to be a false positive. The remaining four clones formed colonies after 1–3 days of incubation, and this was consistent with the growth observed during the initial suppression test. In particular, the CycA-expressing clone grew after only 1 day of incubation. Other clones expressing Alr, DadX or MetC grew after 2–3 days of incubation. Table 3.1 summarizes the growth complementation results of the suppressor clones.

Table 3.1: List of over-expressed proteins identified from the suppressor clones.

Over-expressed protein	Frequency	Days to form colonies	Size of colony	Retransformed clone grew?
CycA (Serine/alanine/glycine APC transporter)	2	1	2 mm	Yes
Alr (Alanine racemase 1)	4	2–3	3–4 mm	Yes
DadX (Alanine racemase 2)	2	2	3–4 mm	Yes
MetC (Cystathionine β-lyase)	1	2	3–4 mm	Yes
WcaL (Predicted glycosyl transferase involved in colanic acid biosynthesis)	1	4	1 mm	No

As expected, over-expression of either Alr or DadX rescued the Δ*alr* Δ*dadX* strain, *E. coli* MB2795. The growth of the strain was also complemented by over-expression of CycA, a serine/alanine/glycine transporter (Keseler *et al*., 2009), which might have allowed the influx of residual D-Ala from LB. As explained in Chapter 2, MetC is neither a sequence nor structural homologue of Alr (or DadX). MetC is not known to racemize amino acids, nor to recognize alanine as a substrate. The native function of MetC is part of the methionine biosynthetic pathway, which has no direct relationship with alanine metabolism in cells. Therefore, the ability of MetC to rescue the D-Ala auxotroph implied that MetC might catalyze the formation of D-Ala.

3.2.2 Growth complementation tests

3.2.2.1 MetC(GFP−) complements *E. coli* (Δ*alr* Δ*dadX*)

To eliminate the possibility that the rescuing activity of MetC was due to the MetC-GFP fusion protein encoded by pCA24N-*metC*, the growth complementation test was repeated using a GFP-less MetC. *E. coli* Δ*alr* Δ*dadX* cells harbouring the pCA24N-*alr*(GFP−) [Section 2.4.2], pCA24N-*metC*(GFP−) [Section 2.4.3] and pCA24N-NoIns [Section 2.4.1] were used as the positive control clone, test clone and negative control clone, respectively. Saturated cultures of all three clones were plated on the selective agar (LB-chloramphenicol supplemented with 50 μM IPTG). The incubation temperature was increased to 37°C (from 30°C in the previous suppression test, described in Section 3.2.1) to provide optimal conditions for the growth of *E. coli*.

As expected, confluent growth was observed for the positive control after 1 day of incubation. The MetC-expressing clone also formed colonies at a similar rate. No mutation had been found in pCA24N-*metC*(GFP−) [Section 2.4.3], thus confirming that MetC, and not mutations in the plasmid, was responsible for the rescuing activity. No colonies were observed for the negative control clone, up to 1 week of incubation. Figure 3.2 shows an example of the growth results of the *E. coli* MB2795 clones on the selective agar.

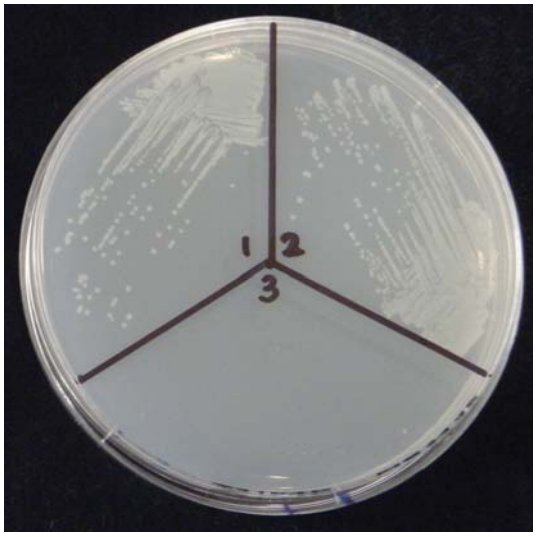


Figure 3.2: The growth of *E. coli* MB2795 clones on IPTG-supplemented LB-chloramphenicol agar after 1 day of incubation at 37°C.

For comparative purpose, different clones were streaked on the same agar plate. Different clones were normally plated or streaked on separate agar plates. Each section of the agar plate corresponded to the growth of (or the lack thereof) *E. coli* MB2795 clone harbouring the following plasmids (clockwise, from left): (1) pCA24N-*alr*(GFP−); (2) pCA24N-*metC*(GFP−); (3) pCA24N-NoIns.

3.2.2.2 Growth complementation of *E. coli* (Δ*alr* Δ*dadX* Δ*metC*)

I showed in Chapter 2 that cystathionine β-lyase activity was a promiscuous function of *E. coli* Alr. In this study, the ability of MetC to rescue the Δ*alr* Δ*dadX* strain (Section 3.2.2.1) suggested that MetC could catalyze alanine racemization. Thus, the overall implication is that MetC and Alr can catalyze each other's reaction, in addition to their native activity. However, it was not known whether both enzymes are bifunctional *in vivo*, *i.e.*, both the native and promiscuous activities are required for growth. Enzyme bifunctionality is important for addressing issues such as activity trade-offs and phenotypic robustness in the course of evolving new function (Copley, 2012). In this section, the bifunctionality of MetC and Alr are tested using *in vivo* assays.

To select for both alanine racemase and cystathionine β-lyase activities at the same time, a triple knockout strain, *E. coli* Δ*alr* Δ*dadX* Δ*metC*, was constructed. The chromosomal *metC* gene of *E. coli* MB2795 (Δ*alr* Δ*dadX*) was replaced with a kanamycin resistance cassette (refer to Section 3.4.2 for more details). This strain was transformed with pCA24N-NoIns to form the negative control. Cells that had been transformed with pCA24N-*alr*(GFP−) and pCA24N-*metC*(GFP−) formed the Alr-expressing and MetC-expressing clones, respectively. The growth of all three clones was tested on M9/glucose agar supplemented with 50 μM IPTG. The minimal medium selected for MetC activity, while the absence of D-Ala in the medium selected for Alr activity.

As expected, no growth was observed for the negative control, up to 4 weeks of incubation. Two phenotypes were observed for the Alr-expressing and MetC-expressing clones: the faster-growing "pseudo-revertants" and the slower-growing colonies. These "pseudo-revertants" were individual colonies that usually grew faster than other colonies that carry the same plasmid (as evident from the appearance of big colonies in the primary inocula in Figure 3.3). For both Alr-expressing and MetC-expressing clones, two to ten pseudo-revertant colonies appeared after 2–3 days of incubation, while the slower-growing colonies took 12–14 days to grow into pinprick sizes (Figure 3.3). These two phenotypes were still observed in independent repeats of the growth complementation assay. The presence of the pCA24N-encoded *alr* or *metC* insert in these pseudo-revertants was confirmed by PCR screens, thereby confirming that these pseudo-revertants were not cross contaminants. The rapid growth of these pseudo-revertants was likely due to the presence of chromosomal or regulatory mutations that increased both alanine racemase and cystathionine β-lyase activities in the cells.

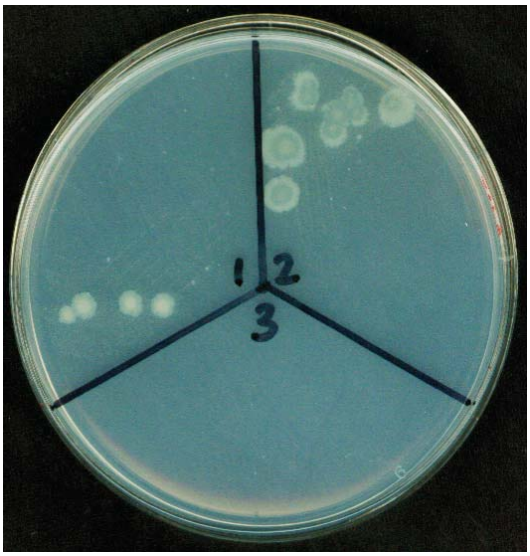


Figure 3.3: The growth of *E. coli* Δ*alr* Δ*dadX* Δ*metC* clones on IPTG-supplemented M9/glucose agar, after 14 days of incubation at 37°C.

For comparative purpose, different clones were streaked on the same agar plate. Different clones were normally streaked on separate agar plates. Each section of the agar plate corresponded to the growth of (or the lack thereof) *E. coli* Δ*alr* Δ*dadX* Δ*metC* clone harbouring the following plasmids (clockwise, from left): (1) pCA24N-*alr*(GFP–); (2) pCA24N-*metC*(GFP–); (3) pCA24N-NoIns.

The slower-growing colonies of the Alr-expressing and MetC-expressing clones appeared to be rescued by the plasmid-encoded *alr* or *metC*, instead of chromosomal or regulatory mutations. In independent repeats of the growth complementation assays, the formation rate of these pinprick-sized colonies was consistent (12–14 days of incubation).

Furthermore, these colonies grew along the streaked lines (Figure 3.3), rather than clustering around the faster-growing pseudo-revertants. This implied that the pinprick-sized colonies were not satellite colonies, which usually formed during a cross feeding phenomenon. Despite the appearance of the pseudo-revertants, the consistent growth of the pinprick-sized colonies provided a hint that Alr and MetC can be bifunctional *in vivo*.

3.2.3 Alanine racemase activity of MetC

In vivo results shown in Sections 3.2.1 and 3.2.2 suggested that MetC could catalyze alanine racemization as its promiscuous function. In this section, the alanine racemase activity of MetC was demonstrated using *in vitro* assays.

The $(His)_6$-tagged MetC was expressed and purified from *E. coli* Δ*metC*, as described in Sections 2.2.4.1 and 2.4.5 (Chapter 2). Figure 2.12A (in Section 2.2.4.1) shows an example of a SDS-PAGE analysis of various MetC fractions from the purification process. As shown in the same figure, the purity of the eluted MetC fractions reached >95%. The eluted fractions were further dialyzed into a buffer without β-mercaptoethanol, and the dialyzed enzyme fractions were used for *in vitro* assays.

The alanine racemase activity was measured using a coupled assay modified from previous protocols (Esaki & Walsh, 1986; Patrick *et al.*, 2002). Two coupling enzymes, D-amino acid oxidase (DAAO) and lactate dehydrogenase (LDH), were added to every reaction to convert D-Ala (the product of the alanine racemization reaction) to pyruvate and finally, lactate. During the conversion of pyruvate to lactate, LDH also oxidizes NADH to NAD^+. Since NADH absorbs strongly at A_{340} but NAD^+ does not, the rate of NAD^+ formation could be spectrophotometrically measured as a decrease of A_{340}. A schematic diagram detailing the coupled assay system is illustrated in Figure 3.4.

Unlike the promiscuous activity of Alr (Section 2.2.4.1), the alanine racemase activity of MetC was detectable *in vitro*. This racemase activity was confirmed by the lack of activity in assay controls without one of the coupling enzymes, DAAO. In the absence of DAAO, no decrease in A_{340} was detected (after the background rate of NADH oxidation was corrected; Figure 3.5). DAAO only recognizes D-amino acids (such as D-Ala) as substrate. Hence, the omission of DAAO from the assay should result in no substrate being produced for LDH (Figure 3.4). Furthermore, the possibility that MetC might be acting as a deaminase (*i.e.*, converting L-Ala into pyruvate) was eliminated, because in that case, a positive signal (*i.e.*, a decrease in A_{340}) would still be generated in the absence of DAAO. Instead, the signal was only detected when DAAO was present in the reaction, thereby confirming that D-Ala was produced from L-Ala via the activity of MetC.

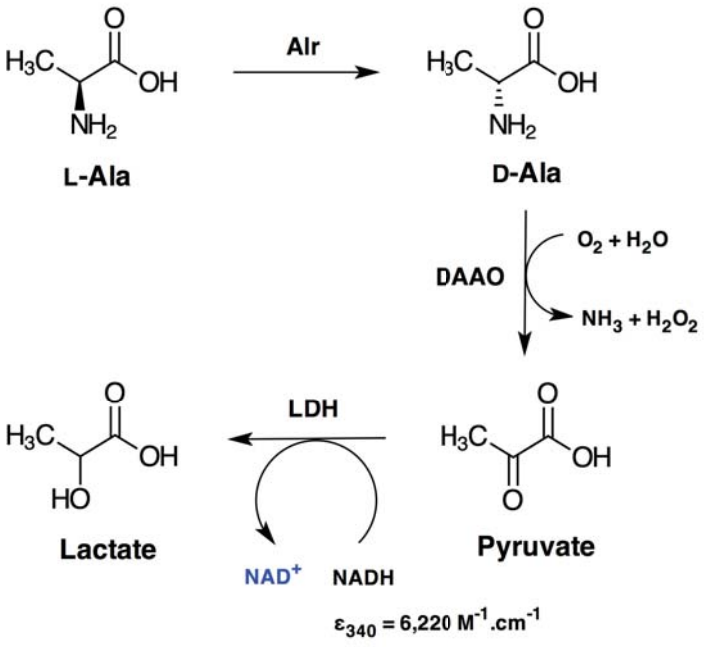


Figure 3.4: The *in vitro* coupled assay used to measure the alanine racemization activity.

The racemization of L-Ala produces D-Ala, which is then used as a substrate for DAAO. In the presence of O_2 and H_2O, D-Ala is converted to pyruvate. LDH further converts pyruvate to lactate via the oxidation of NADH. The amount of NADH can be measured spectrophotometrically at 340 nm using a molar extinction coefficient of 6,220 $M^{-1}.cm^{-1}$. Therefore, the formation of NAD^+ is measured as a decrease of absorbance at 340 nm. DAAO, D-amino acid oxidase; LDH, lactate dehydrogenase; NAD^+/NADH, oxidized/reduced form of β-nicotinamide adenine dinucleotide.

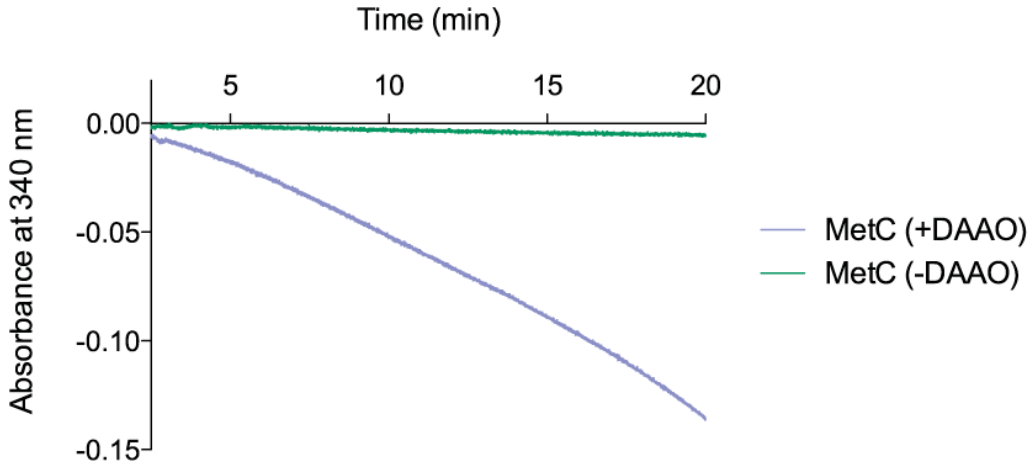


Figure 3.5: The alanine racemase assay of MetC performed in the presence and absence of DAAO.

The alanine racemase activity of MetC was measured in CHES buffer (pH 9.0) at 37°C. Each reaction contained 58 nM MetC, 5 mM L-Ala, 0.2 mM NADH, 100 U lactate dehydrogenase. The MetC(+DAAO) assay also contained 1 U of D-amino acid oxidase (DAAO).

The kinetic parameters of MetC for alanine racemization were derived from the mean of three repeats, with each repeat value representing the mean of duplicate assays. The K_M for L-Ala was found to be 51.0 ± 4.3 mM, whereas the k_{cat} was 3.3 ± 0.6 s^{-1}. Based on the catalytic efficiencies (k_{cat}/K_M), the alanine racemase activity of MetC was ~35,000-fold lower than the native activity of MetC (Table 3.2; data were also derived from Table 2.8 from Section 2.2.6.5, of Chapter 2). Although over-expression of MetC rescued *E. coli* MB2795 at a comparable rate with that of Alr (Section 3.2.2.1; both clones took 1 day to form colonies on the selective medium), the *in vitro* alanine racemase activity of MetC was still nearly 3 orders of magnitude lower than the native activity of Alr (Table 3.2). The lack of correlation between the *in vitro* promiscuous activity level and the *in vivo* fitness has also been previously noted and discussed (Section 2.3.2 of Chapter 2).

Table 3.2: Kinetic parameters of the alanine racemization of MetC.

Enzyme	Cystathionine → Homocysteine + Pyruvate + NH_3			L-Ala → D-Ala		
	k_{cat} (s^{-1})	K_M (M)	k_{cat}/K_M ($s^{-1}.M^{-1}$)	k_{cat} (s^{-1})	K_M (M)	k_{cat}/K_M ($s^{-1}.M^{-1}$)
MetC	89.5 ± 13.9	(3.9 ± 0.7) × 10^{-5}	2.3 × 10^6	3.3 ± 0.6	(5.1 ± 0.4) × 10^{-2}	65
Alr	[†]n.d.	n.d.	<0.85	55.3 ± 3.7	(8.7 ± 1.5) × 10^{-4}	6.4 × 10^4

The kinetic parameters of the alanine racemase activity of MetC are highlighted in blue.
[†]n.d., not determined.

3.2.4 Directed evolution of MetC (round 1)

Random mutagenesis and genetic selection can be used to improve very low promiscuous activities to detectable levels in the laboratory. These *in vitro* functional improvements have been demonstrated previously (Chapter 2; Patrick and Matsumura (2008)]. However, at the outset of this study, attempts to improve the alanine racemase activity in MetC were expected to be challenging, because the selection window for obtaining MetC variants with higher racemase activity was narrow — the starting racemase activity of MetC was only ~10^3 fold lower than Alr (Table 3.2 in Section 3.2.3). In this section, I probed the evolvability of the alanine racemase activity of MetC using two rounds of low-mutation-rate mutagenesis (1.0–3.0 DNA substitutions per *metC* gene) and genetic selection experiments.

3.2.4.1 First-generation library

EpPCR was used to introduce random mutations into the plasmid-encoded *metC* gene. For the first-generation library, the mutated *metC* gene was cloned into a pCA24N(*KanR*) vector,

which was a derivative of pCA24N with a kanamycin resistance cassette, instead of a chloramphenicol resistance cassette (courtesy of Dr. Monica Gerth). The change in resistance marker ensured that false positives arising from ASKA contaminants were avoided.

First, *E. coli* DH5α-E was used as a non-selective host to propagate the ligated plasmids. The reason for using the DH5α-E strain was its considerably higher transformation efficiency ($\geq 10^8$ cfu/μg of pUC19) than that of the D-Ala auxotroph, *E. coli* MB2975 ($<10^8$ cfu/μg of pUC19). In a preliminary study, I transformed both the DH5α-E and MB2795 strains with the same amount of ligated products. The number of transformed DH5α-E clones was ~2.2-fold higher than the number of MB2795 clones (results not shown). Therefore, the ligated *metC* products were initially electroporated into DH5α-E to maximize the library size. Circular plasmids purified from the transformed clones were then introduced into the MB2795 strain.

A library consisting of 1.5×10^5 distinct variants was constructed in *E. coli* DH5α-E (see Section 3.4.7.1 for more details). All 14 clones randomly chosen from the library were found to possess the plasmid-encoded *metC*. To estimate the overall mutation rate of the library, the *metC* genes from the 14 clones were sequenced. A total of (14 sequences × 1,185 bp) = 16,590 bp were analyzed. There were 42 point mutations and no indels in total, and each *metC* gene carried 1–5 mutations. As summarized in Table 3.3, the average DNA mutation rate was 3.0 substitutions per *metC* gene.

While the previous error-prone *alr* library was biased towards G→N and C→N mutations (Table 2.3 in Section 2.2.6.1), such bias was not observed for this *metC* library. In fact, the 42 mutations found in the *metC* library were almost evenly divided into A→N and T→N substitutions (47.6%), as well as G→N and C→N substitutions (52.4%). Yet, a slight bias towards DNA transitions (Ts; 59.5%), rather than transversions (Tv; 40.5%), was noted.

Table 3.4 summarized the key characteristics of the *metC* library, which had been analyzed by the programme that calculates the mutational diversity of an epPCR library on an amino acid level, PEDEL-AA (Firth & Patrick, 2008). On average, there were 2.0 amino acid substitutions per MetC variant. Less than 30% of the total library consisted of unmutated MetC variants, and thus, the estimated number of mutated MetC variants was $\geq 1.1 \times 10^5$ cfu. Overall, up to 7.4×10^4 cfu were estimated to carry distinct, full-length MetC variants.

Table 3.3: The mutational spectrum of the first-generation *metC* library.

Types of mutation	Frequency	Proportion of total
Transitions		
A→G, T→C	11	26.2%
G→A, C→T	14	33.3%
Transversions		
A→T, T→A	5	11.9%
A→C, T→G	4	9.5%
G→C, C→G	6	14.3%
G→T, C→A	2	4.8%
Summary of bias		
Transitions/Transversions	1.5	N/A
AT→GC/GC→AT	0.9	N/A
A→N, T→N	20	47.6%
G→N, C→N	22	52.4%
Mutation per kb	2.5	N/A
Mutation per *metC* gene (1,185 bp)	3.0	N/A

N/A = not applicable.

Table 3.4: Summary of the key characteristics of the first-generation *metC* library based on its DNA mutational spectrum.

Property	Estimate
Total library size	1.5×10^5
Number of variants with no indels or stop codons	1.4×10^5
Mean number of amino acid substitutions per variant	2.0
Unmutated (wildtype) sequences (% of library)	27.3%
Number of distinct full-length proteins in the library	7.4×10^4

3.2.4.2 Selection of first-generation MetC variants

The *metC* library (Section 3.2.4.1) was further introduced into the D-Ala auxotroph, *E. coli* MB2795, to select for MetC variants with higher alanine racemase activity on LB agar supplemented with 5 μM IPTG (see Section 3.4.7.2 for details). The 10-fold reduction in the IPTG concentration extended the duration taken for the MetC-expressing clone to form colonies at 28°C. Colonies of the MetC-expressing clone were usually observed after 1–2 days of incubation on LB agar supplemented with 50 μM IPTG (Section 3.2.2.1). However, reducing the IPTG concentration to 5 μM led to colony formation in 5–7 days. The reduced IPTG concentration on the selective medium did not affect the rate of colony formation for the positive control clone and the negative control clone — the Alr-expressing clone still formed colonies after 1 day of incubation, whereas no colonies were observed for the clone expressing the empty pCA24N vector.

In the absence of D-Ala, mutant clones that grew faster than the MetC-expressing clone were selected as "winners". A total of 1.0×10^6 cfu were plated on the selective agar, as calculated from colony counts on non-selective medium. The number of plated clones represented a ~6.6× coverage of the initial library, which consisted of 1.5×10^5 variants. The programme that calculates the library diversity and completeness, GLUE (Patrick *et al.*, 2003), suggested that 99.9% of the variants from the initial library were likely to be sampled during the selection experiment.

Pinprick-sized colonies were observed on the library plate after 1 day of incubation. By 42 h, the colony sizes on the library plate reached 0.5–1.5 mm. An estimated 37,000 colonies were present, which corresponded to a survival rate of 3.7%. The growth of these mutant clones was nearly as good as that of the positive control (the Alr-expressing clone), which took 1 day to form colonies.

3.2.4.3 Reselection of first-generation MetC variants

The survival rate of the first-generation *metC* library on the IPTG-supplemented LB agar was high (3.7%), even though the IPTG concentration had been lowered to 5 μM (from the 50 μM normally used for growth complementation assays). This implied that the selection system was not stringent enough to select for clear winners. The surviving MetC clones also showed a variety of colony sizes (0.5–1.5 mm) on the selective agar, which suggested that certain colonies grew faster than the rest. Therefore, the fastest-growing MetC clones were identified by another round of selection on LB-kanamycin agar alone. By omitting IPTG from the agar, it was hoped that only variants with high *in vivo* alanine racemase activities would be selected.

For the reselection test, I chose to focus on a subset of the 37,000 MetC clones that survived on the initial selective medium (Section 3.2.4.2), as it would be impractical to

screen all of them. Forty-eight colonies were randomly picked from the surviving populations, and restreaked on two sets of agar plates, LB with and without 5 μM IPTG. Consistent with the growth phenotypes described in the initial selection experiment (Section 3.2.4.2), all 48 clones grew on the IPTG-supplemented agar after 1 day of incubation (Figure 3.6). However, growth variations were observed for the clones in the absence of IPTG (Table 3.5; Figure 3.6). A general scale was used to rank the growth of each clone (Table 3.5), and clones that showed better growth (highlighted with red labels in Figure 3.6) were selected for further characterization.

Table 3.5: The growth phenotypes of the first-generation MetC clones after 1 day of incubation.

Clone(s)	Growth on LB-kanamycin agar + 5 μM IPTG	Growth on LB-kanamycin agar alone
A2, A4, A5, A6, A11, B1, B5, B6, B7, B9, B11, C2, C3, C7, C9, C11, C12, D1, D2, D4, D5, D6, D8, D9, D10, D11, D12	++	++
A1, A7, A9, A10, A12, B2, B3, B4, B8, B10, B12, C1, C4, C5, C6, C8, C10, D3, D7	++	+
A3, A8	++	−
Template (MetC)	+	−
Negative control (NoIns)	−	−

Key for scoring growth: (−) No growth; (+) Little growth. Isolated colonies of <0.5 mm were observed; (++) Considerable growth. Isolated colonies of ~0.5 mm were observed.

A total of 27 clones were selected based on their better growth phenotype in the absence of IPTG (Table 3.5; Figure 3.6). Their plasmid-encoded *metC* genes were PCR-amplified and sequenced to identify their mutations. DNA sequences were obtained for 24 clones. Out of 24 clones, only clone A11 did not harbour any mutation in the *metC* gene. Instead, sequencing analysis of the *lacI*^{q} gene and T5-*lacO* promoter region of the plasmid backbone revealed that a g→a DNA substitution was found in the *lacO2* region. Since the *lacO2* region is a binding site for LacI^q, the g→a mutation was likely to disrupt the LacI^q/*lacO* interaction, and led to leaky transcription. It seems likely that this leaky transcription had resulted in the increased levels of MetC in the cells, and further effected the improved growth of clone A11 in the absence of IPTG (Figure 3.6).

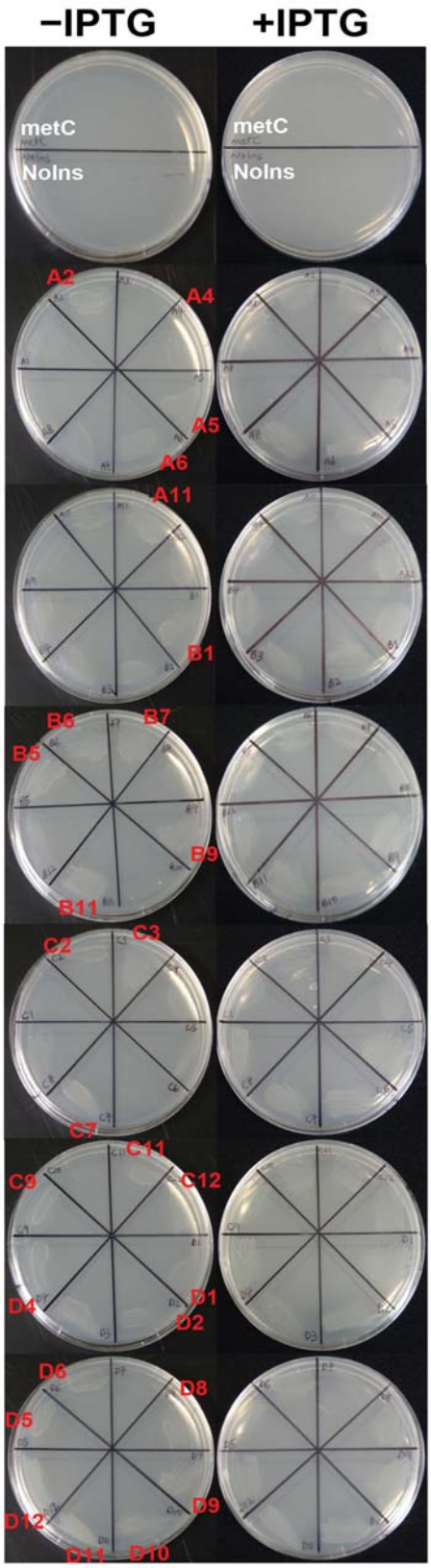


Figure 3.6: The growth of selected first-generation MetC mutants on LB-kanamycin agar plates with and without 5 μM IPTG, after 1 day of incubation at 28°C.

A general scale was used to rank the growth of each clone. Clones selected as "winners" based on their slightly better growth phenotype on LB agar without IPTG are labelled in red.

The remaining 23 clones harboured 1–2 amino acid substitutions per clone. Notably, residue P113 of all 23 clones was mutated into either glutamine, alanine, threonine or serine (Table 3.6). The pool was dominated by the P113S mutation, which appeared in 17 out of 23 clones. This was followed by the P113A mutation, which appeared in 4 clones. The remaining 2 clones carried the P113Q and P113T mutations, respectively. In total, only 8 residues were mutated in the 23 clones. Overall, the 23 clones represented 12 distinct variants (Table 3.6), whose growth phenotypes were essentially indistinguishable in the absence of IPTG (Figure 3.6).

Table 3.6: Summary of the selected MetC variants from the reselection experiment.

Variant	Frequency	Days to form colonies	Mutation(s)
MetC	[Template]	5–7	None
D11	10	1–2	P113S
A6	2	1–2	P113S; 1 silent
B7	2	1–2	P113A
D12	1	1–2	P113A, D247E
D2	1	1–2	L34M, P113A; 1 silent
C12	1	1–2	P113T
A4	1	1–2	P113Q; 1 silent
B9	1	1–2	P113S, A237V
A5	1	1–2	P113S, N321S
C2	1	1–2	P113S, A325V
D5	1	1–2	P113S, F331T
B11	1	1–2	P113S, A246V; 1 silent
A11*	1	1–2	None
Alr	[Positive control]	1	None
NoIns	[Negative control]	No growth	None

*A regulatory mutation was found in clone A11. See text for details.

The 8 mutated residues were mapped to the tertiary and quaternary structures of *E. coli* MetC (Figure 3.7). A literature search confirmed that these mutated residues are not known to possess catalytic or substrate-binding roles. Only one residue (L34) is in the N-terminal domain, which is important for tetramerization (Clausen *et al.*, 1996). Residue L34

is located on the monomer-monomer interface, which is formed around a crystallographic 2-fold axis (see Figure 3.7B for details). Four residues (P113, A237, A246, D247) are in the middle domain, where the active site and PLP are housed, but none of these residues are in the active site. The residue closest to the active site is P113, which was mutated in 23 out of 24 winners. Similar to L34, residues A237 and D247 are also located on the monomer-monomer interface. In particular, D247 is also located on the dimer-dimer interface, which is formed around a non-crystallographic diad axis (see Figure 3.7B for details). This puts D247 in a tetramer interface, where charged residues in this patch are known to interact with their symmetry mate from the opposite monomer (Clausen *et al*., 1996). The remaining three residues (N321, A325, F331) are in the C-terminal domain, and are physically located at the exposed surface area of the enzyme (Figure 3.7).

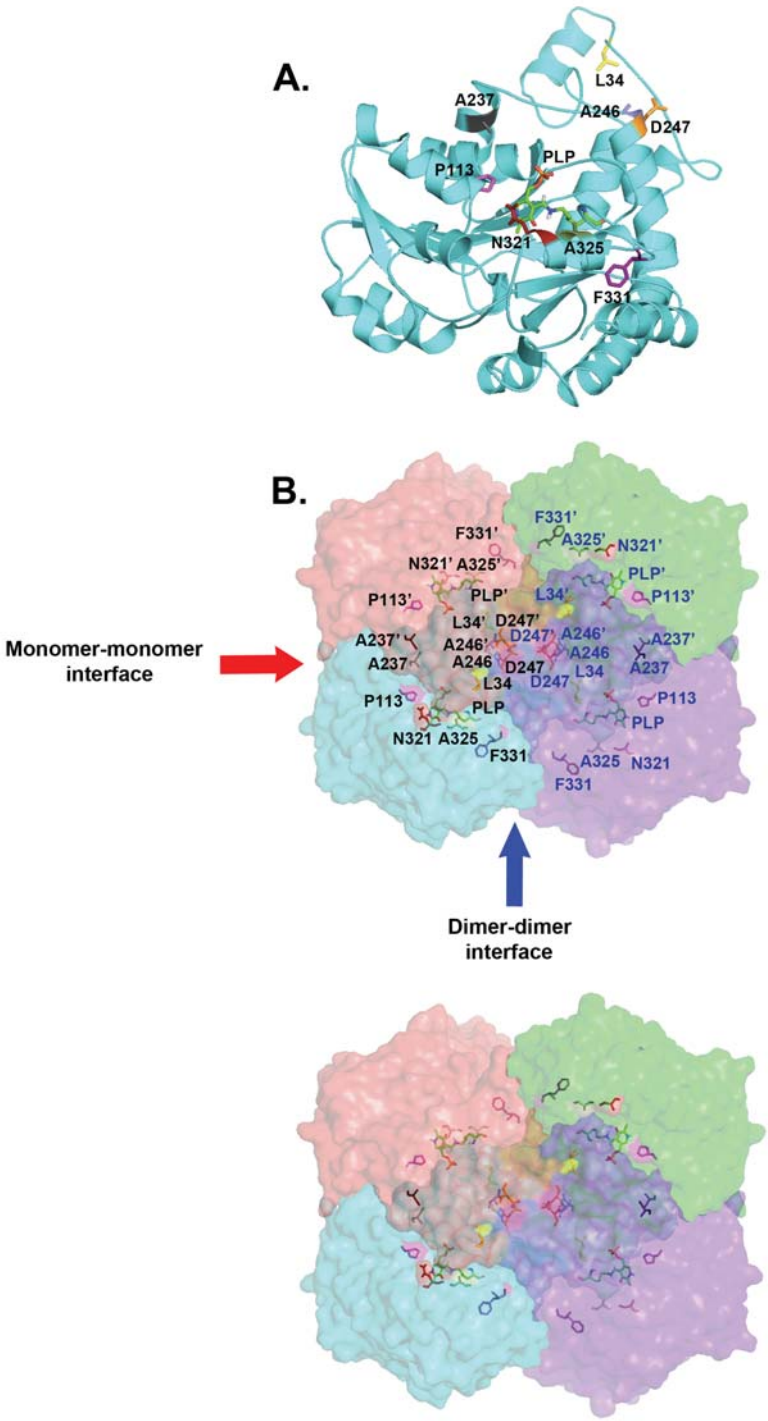


Figure 3.7: Mutated residues from the first-generation MetC winner clones.

The mutated residues from the first-generation MetC winner clones were mapped to the structure of *E. coli* MetC (PDB code: 1cl1). The biological unit of MetC is a tetramer, but the asymmetric unit in the PDB file is a dimer. The tetramer was recreated using the Sympexp command in MacPymol. Each PLP molecule is coloured according to its elements, while each residue is coloured differently. (A) Ribbon diagram of a MetC monomer and the position of the mutated residues. Loops have been smoothed for clarity. (B) Surface diagram of the quaternary structure of MetC. The mutated residues of each dimer are labelled in black and blue, respectively. The axis pointed by the red arrow is the monomer-monomer interface, whereas the axis pointed by the blue arrow is the dimer-dimer interface.

3.2.5 DIRECTED EVOLUTION OF METC (ROUND 2)

Results from the first round of directed evolution revealed that P113 is potentially an important residue for the promiscuous alanine racemase activity of MetC. The majority of the winner clones harboured one of the following mutations: P113Q, P113A, P113T or P113S. Although all four P113 mutants and the positive control clone (Alr-expressing cells) took the same duration to form colonies on the selective medium (~1 day), I realized that the magnitude of *in vivo* improvements does not strictly correlate with *in vitro* improvements of a promiscuous activity (see previous discussion in Section 2.3.2 of Chapter 2). For this reason, a second round of directed evolution was attempted to investigate if the *in vivo* and *in vitro* alanine racemase activity of the P113 mutants could be further improved.

3.2.5.1 SECOND-GENERATION LIBRARIES

All four P113 mutants (P113Q, P113A, P113T and P113S) were used as the templates for a second round of evolution. This would test whether the evolutionary trajectory of the various P113 mutants would differ in the course of evolving the alanine racemase activity. Random mutations were introduced into the *metC* inserts (from the P113 mutants) using epPCR. The inserts were then introduced into the pBAD vector (Invitrogen), which is a low-copy number plasmid [its pBR322 origin of replication maintains 15–20 plasmid copies in a cell; (Qiagen, 2011)]. In addition, the PBAD promoter allows titratable, arabinose-inducible protein expression (Guzman *et al.*, 1995). Such dose-dependent expression is in contrast with the all-or-nothing expression of the *lac* operon in the pCA24N vector of the first-generation winners. In principle, the use of a tightly regulated, low-copy number vector allows further repression of the *in vivo* enzyme expression level. When the enzyme expression level is low, only variants with alanine racemase activity higher than the parental P113 clones would be selected.

Similar to the first-generation library, the second-generation libraries were initially constructed in *E. coli* DH5α-E. The number of distinct variants in the P113Q, P113A, P113T and P113S libraries were 1.3×10^6 cfu, 9.1×10^5 cfu, 6.6×10^5 cfu and 2.2×10^6 cfu, respectively. The mutation rate in each library was determined by sequencing 11–14 clones. In general, all second-generation libraries showed lower mutation rates (1.0–2.4 mutations per *metC* gene) than the first-generation library (3.0 mutations per *metC* gene). Table 3.7 summarizes the type of mutations present in all four libraries, and their overall mutation rates. The variations in the mutational biases of the libraries emphasized the highly stochastic nature of epPCR, as these libraries were constructed simulatenously. For instance, the transition/transversion (Ts/Tv) ratios of the second-generation libraries ranged from 0.8–2.0 (Table 3.7), while Stratagene (the manufacturer of the GeneMorph II kit) has expected a Ts/Tv ratio of 0.9 for every epPCR library constructed using the components from the kit.

Table 3.7: The mutational spectrum of the second-generation *metC* libraries.

Types of mutation	Frequency (% proportion of total)			
	MetC(P113Q)	**MetC(P113A)**	**MetC(P113T)**	**MetC(P113S)**
Number of clones sequenced	11	12	12	14
Transitions				
A→G, T→C	3 (11.1%)	0 (0%)	4 (26.7%)	5 (29.4%)
G→A, C→T	13 (48.1%)	6 (40%)	5 (33.3%)	5 (29.4%)
Transversions				
A→T, T→A	3 (11.1%)	1 (6.7%)	0 (0%)	1 (5.9%)
A→C, T→G	2 (7.4%)	0 (0%)	3 (20.0%)	0 (0%)
G→C, C→G	2 (11.1%)	2 (13.3%)	2 (13.3%)	0 (0%)
G→T, C→A	3 (7.4%)	5 (33.3%)	1 (6.7%)	4 (23.5%)
Indels				
Insertion	1 (3.7%)	0 (0%)	0 (0%)	0 (0%)
Deletion	0 (0%)	1 (6.7%)	0 (0%)	2 (11.8%)
Summary of bias				
Transitions/Transversions	2.0	0.8	1.5	2.0
AT→GC/GC→AT	0.2	0	1.2	0.6
A→N, T→N	8 (29.6%)	1 (6.7%)	7 (46.7%)	6 (35.3%)
G→N, C→N	18 (66.7%)	13 (86.7%)	8 (53.3%)	9 (52.9%)
Mutation per kb	1.9	1.0	1.0	0.9
Mutation per *metC* gene (1,227 bp)	2.4	1.2	1.3	1.0

PEDEL-AA (Firth & Patrick, 2008) was used to estimate the mutational diversity of each second-generation library on an amino acid level. The key characteristics of each library are tabulated in Table 3.8. As expected from the low DNA substitution rate per sequence, all libraries showed even lower average number of amino acid substitutions per variant. Nevertheless, all the second-generation libraries were predicted to contain $>10^5$ distinct variants, which was greater than the number screened in the first-generation library (Table 3.4 in Section 3.2.4.1).

Table 3.8: Summary of the key characteristics of the second-generation *metC* libraries based on their DNA mutational spectrum.

Property	Estimate			
	MetC(P113Q)	MetC(P113A)	MetC(P113T)	MetC(P113S)
Vector background	3.1%	18.2%	10.7%	9.4%
Total library size after subtracting vector background	1.3×10^6	9.1×10^5	6.6×10^5	2.2×10^6
Number of variants with no indels or stop codons	1.1×10^6	8.0×10^5	6.4×10^5	1.8×10^6
Mean number of amino acid substitutions per variant	1.6	0.8	0.9	0.7
Unmutated (wildtype) sequences (% of library)	36.5%	49.2%	51.8%	50.9%
Number of distinct full-length proteins in the library	4.3×10^5	1.7×10^5	1.5×10^5	2.9×10^5

3.2.5.2 SELECTION OF SECOND-GENERATION METC VARIANTS

Plasmids from each of the second-generation libraries were pooled, and introduced into the selective host, *E. coli* MB2795. The four libraries were spread separately on LB-carbenicillin supplemented with different arabinose concentrations but without D-Ala (Table 3.9). The arabinose concentration chosen for each second-generation library represented the strongest selection stringency that could be applied. For every library, preliminary experiments were carried out to test the minimum arabinose concentration that permitted the growth of the parental clone (results not shown). When the arabinose concentrations were higher than the chosen concentration, the parental clones grew rapidly and thus, there was no room for selecting variants that outgrow the parental clones. In contrast, the parental clones did not grow if the arabinose concentrations were lower than the chosen concentration.

The arabinose concentration chosen for the P113Q, P113A, P113T and P113S libraries was 0.8 mM, 0.6 mM, 0.4 mM and 0.4 mM, respectively. Under these arabinose concentrations, the parental clones took 3–5 days to form full-sized colonies (*i.e.*, colonies >0.5 mm in diameter). During the selection experiment, thousands of full-sized colonies were observed for each library after only 1 day of incubation. Given that each library was over-sampled by 9- to 18-fold, the survival rate for these libraries was <0.02% (Table 3.9). The low survival rates of the second-generation libraries indicated that the selection stringencies were considerably higher than the first-generation library, whose survival rate was 3.7% (Section 3.2.4.2).

Table 3.9: Results summary of the selection of second-generation *metC* libraries on 0.4–0.8 mM arabinose.

Summary	Second-generation libraries			
	MetC(P113Q)	MetC(P113A)	MetC(P113T)	MetC(P113S)
Arabinose concentration	0.8 mM	0.6 mM	0.4 mM	0.4 mM
Coverage of the initial library	17.7×	8.5×	12×	9.8×
Survival rate	0.005%	0.075%	0.014%	0.019%
Growth improvements shown by winner clones	~3-fold	~3-fold	~4-fold	~4-fold

3.2.5.3 RESELECTION OF SECOND-GENERATION METC VARIANTS

Although the survival rates of the second-generation libraries were low, the large number of colonies in each library rendered further characterizations impractical within the time frame of this study. These large numbers of winners were likely due to the over-sampling of the libraries. For instance, the size of the P113Q library was 1.3×10^6, but a total of 2.3×10^7 cfu were plated on the selective agar, and thus ensuring that the initial library was covered by almost 18-fold. In other words, there would be 18 clones of the same winner in the pool of surviving colonies.

To further increase the stringency of selection, the arabinose concentration was reduced to 0.1 mM for all four second-generation libraries. A smaller number of cells were plated, so that the final coverage of each library was less than 6 fold (Table 3.10). For each library, full-sized colonies still appeared after 1 day of incubation, although the arabinose concentration had been reduced to a concentration at which all parental clones barely grew at all (results not shown). Approximately 180 cfu was observed for the P113Q library, 700 cfu for the P113T library, and 3,000 cfu for each of the P113A and P113S libraries. Despite the lower number of colonies appearing in each library, the survival rates did not change significantly (compare Table 3.9 and Table 3.10).

Table 3.10: Results summary of the reselection of second-generation *metC* libraries on 0.1 mM arabinose.

Summary	Second-generation libraries			
	MetC(P113Q)	MetC(P113A)	MetC(P113T)	MetC(P113S)
Arabinose concentration	0.1 mM	0.1 mM	0.1 mM	0.1 mM
Coverage of the initial library	2.9×	5.4×	3.9×	2.9×
Survival rate	0.005%	0.059%	0.027%	0.049%

The growth of the winner clones at such low arabinose concentration (0.1 mM) prompted me to investigate whether these clones would also grow in the absence of arabinose. Twenty-four colonies were randomly picked from each library and re-streaked on LB-carbenicillin agar alone. In the absence of arabinose (and D-Ala), all 24 clones from each library did indeed grow into full-sized colonies after 1 day of incubation (Figure 3.8). To test if the LB-carbenicillin agar plates were contaminated with D-Ala or arabinose (and hence, permitting the growth of all 96 winner clones), various parental and negative control clones were also streaked on the same batch of agar plates. However, no growth was observed for all the parental and negative control clones, even up to 1 week of incubation (Figure 3.9). Therefore, this confirmed that the growth of the winner clones (Figure 3.8) was not effected by arabinose or D-Ala contamination in the agar.

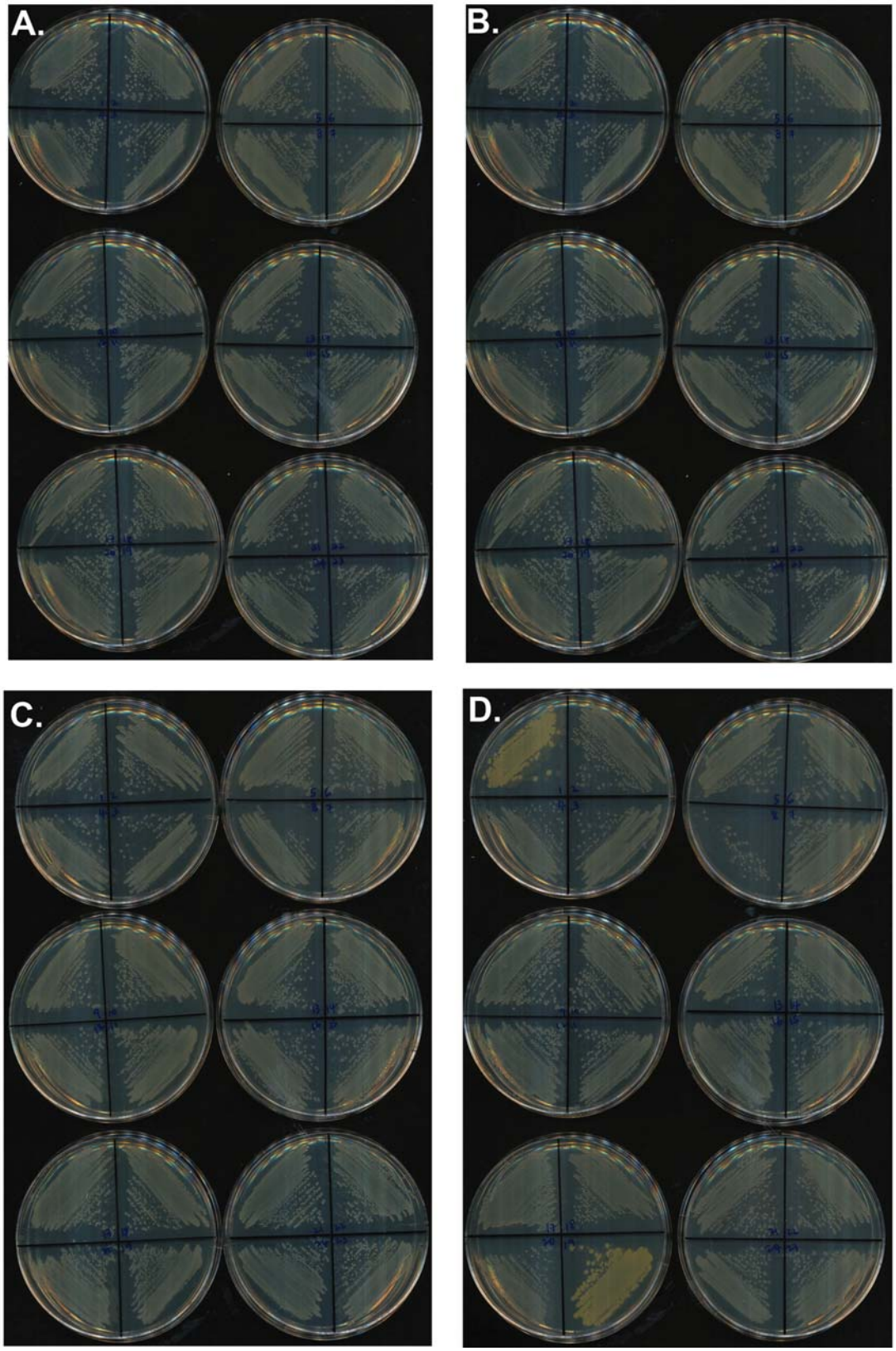


Figure 3.8: The growth of the selected second-generation MetC mutants in the absence of arabinose and D-Ala, after 1 day of incubation at 28°C.

The 24 winner clones were picked from the (A) P113Q library. (B) P113A library. (C) P113T library. (D) P113S library.

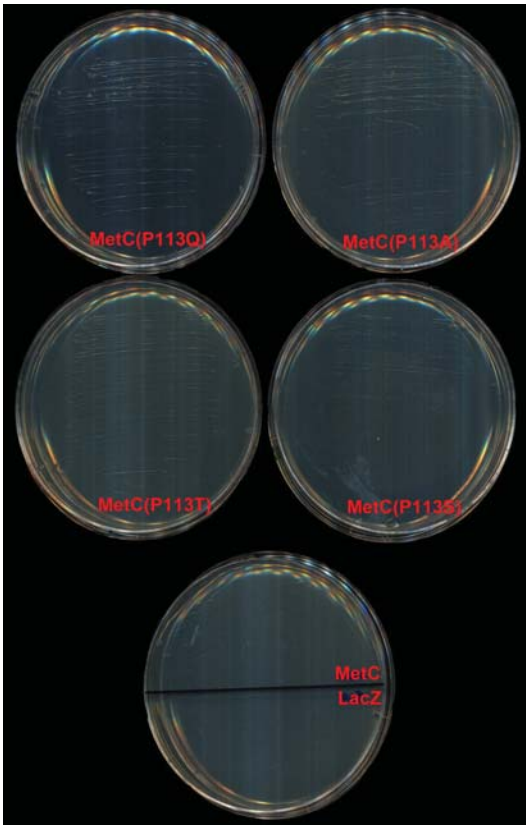


Figure 3.9: The lack of growth of various parental and negative control clones in the absence of arabinose, even after 1 week of incubation.

After PCR-screening the 24 clones from each library, less than half was tested positive for a pBAD-encoded *metC* gene. A total of eight positive clones were identified for the P113Q library, three positive clones for the P113A library, five positive clones for the P113T library and ten positive clones for the P113S library. The plasmid-encoded *metC* genes amplified from these 26 clones were sequenced.

Table 3.11 summarizes the mutations (or the lack thereof) in the pBAD-encoded *metC* gene of the sequenced clones. Sequencing analysis revealed that seven out of eight variants from the P113Q library, and one variant from the P113A library, consisted of heterogeneous populations based on the presence of mixed DNA templates. The remaining variant from the P113Q library, all five variants from the P113T library as well as four variants from the P113S library, did not harbour additional mutation apart from their P113 mutation. For the two variants from the P113A library, an A113P substitution had reverted their MetC to the wildtype enzyme. Overall, no new mutations were found in the *metC* genes of the winner clones from the P113Q, P113A and P113T libraries, even though the winner clones grew in the absence of arabinose.

On the contrary, additional mutations were found in the P113S-2, P113S-14 and P113-16 variants from the P113S library. The three variants carried two new amino acid substitutions, in addition to their parental P113S mutation. Furthermore, four out of the ten sequenced clones from the P113S library were the P113S-14 variant, which suggested that mutation at I251 (in addition to P113) might be important for improving the alanine racemase activity of MetC. On the whole, only three mutated variants were selected from all the second-generation libraries, and all three variants were from the P113S library.

Table 3.11: Mutations found in the second-generation winners.

Variant	Frequency	Day to form colonies in the absence of arabinose	Mutation(s)
The P113Q library			
MetC(P113Q)	[Template]	No growth up to 7 days	P113Q
P113Q-12	1	1	P113Q
The P113A library			
MetC(P113A)	[Template]	No growth up to 7 days	P113A
P113A-9	2	1	A113P; 2 silent
The P113T library			
MetC(P113T)	[Template]	No growth up to 7 days	P113T
P113T-20	5	1	P113T
The P113S library			
MetC(P113S)	[Template]	No growth up to 7 days	P113S
P113S-2	1	1	P113S, P156L, A227T
P113S-14*	4	1	P113S, I251M
P113S-16	1	1	P113S, K120N, P352A, 1 silent
P113S-18	4	1	P113S

*An R→W amino acid substitution was also found in the second residue of the linker sequence, which is located between the N-terminal $(His)_6$-tag and the *metC* gene.

For the variants whose genotype was identical to their parental P113 clone, it was hypothesized that their growth in the absence of arabinose was due to the presence of chromosomal mutations or regulatory mutations on the plasmid. To test this possibility, recloning, retransformation and restreaking tests were carried out. The pBAD-encoded *metC* insert from these clones were isolated and re-ligated with freshly linearized pBAD

vector. After transforming the ligated product into a new batch of host (*E. coli* MB2795), the clones were restreaked on the same selective medium (LB-carbenicillin) and incubated under identical conditions. However, no growth was observed for any of the retransformed clones, for up to 19 days of incubation (results not shown). The results were in contrast to the 1-day growth phenotype exhibited by these clones during the initial selection experiment (Figure 3.8). Hence, results from these recloning, retransformation and restreaking tests suggested that the growth of the "unmutated" clones in the absence of arabinose was caused by either chromosomal or regulatory mutations.

Retransformation and restreaking tests were also used to validate the growth phenotypes of the mutated variants, namely P113S-2, P113S-14 and P113S-16. Plasmids were extracted from these clones, and used to transform a fresh batch of competent *E. coli* MB2795. Retransformed clones were restreaked on the LB-carbenicillin agar lacking arabinose. After 1 day of incubation, full-sized colonies were observed for all three retransformed clones (Figure 3.10). This observation was consistent with the 1-day growth phenotype shown by these mutated clones during the initial selection experiment (Figure 3.8D). Thus, these mutated variants were deemed the winner clones of the entire second round of directed evolution.

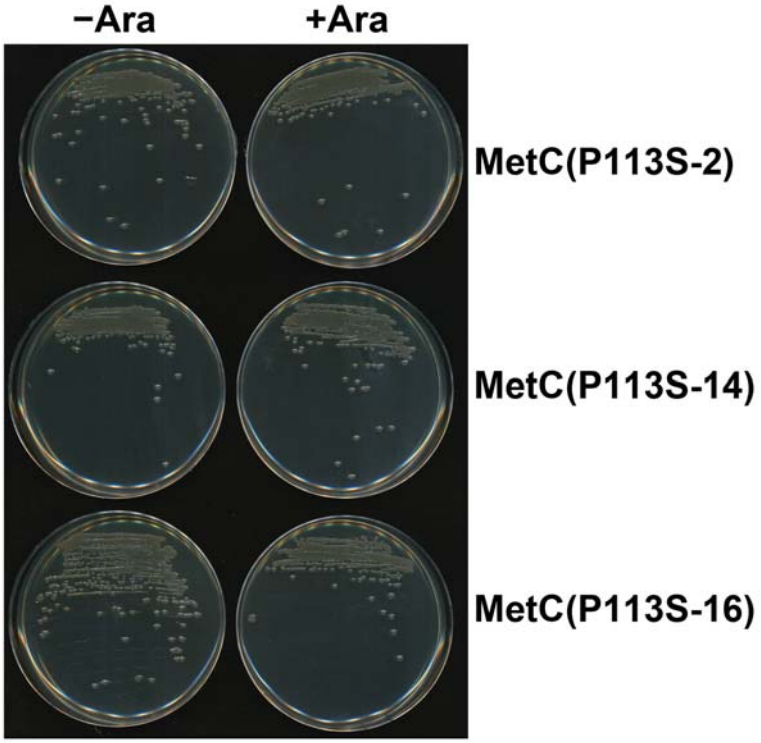


Figure 3.10: The growth of the retransformed clones of P113S-2, P113S-14 and P113S-16, on LB-carbenicillin after 1 day of incubation at 28°C.

The agar plates on the left did not contain any arabinose, whereas the agar plates on the right were supplemented with 0.1 mM arabinose.

The mutations found in the P113S-2, P113S-14 and P113S-16 variants were mapped to the tertiary and quaternary structures of *E. coli* MetC (Figure 3.11). A total of five new mutated residues were identified from these second-generation winners. Similar to the mutated residues found in the first-generation winners, little is known about the mutated residues of the P113S-2, P113S-14 and P113S-16 variants. Furthermore, the second-generation mutated residues are different from those of the first-generation. Most of the second-generation mutated residues are located in the middle-domain of MetC, but the first-generation mutated residues are distributed almost evenly between the middle and C-terminal domains. Specifically, four of the five mutated residues (K120, P156, A227 and I251) from the P113S-2, P113S-14 and P113S-16 variants are located in the middle-domain of MetC, while only 1 mutated residue (P352) is found in the C-terminal domain.

Yet, none of these five mutated residues is located near the active site — a trend that was also noted for the first-generation mutations. Three residues (K120, A227 and P352) were located on the exposed enzyme surface, with K120 close to the monomer-monomer interface (Figure 3.11). On the other hand, residue I251 is located near the tetramerization core of MetC. Along with mutations at A246 and D247 (found in the first-generation winner clones; Figure 3.7 in Section 3.2.4.3), mutation at I251 might therefore affect the oligomerization state of the protein.

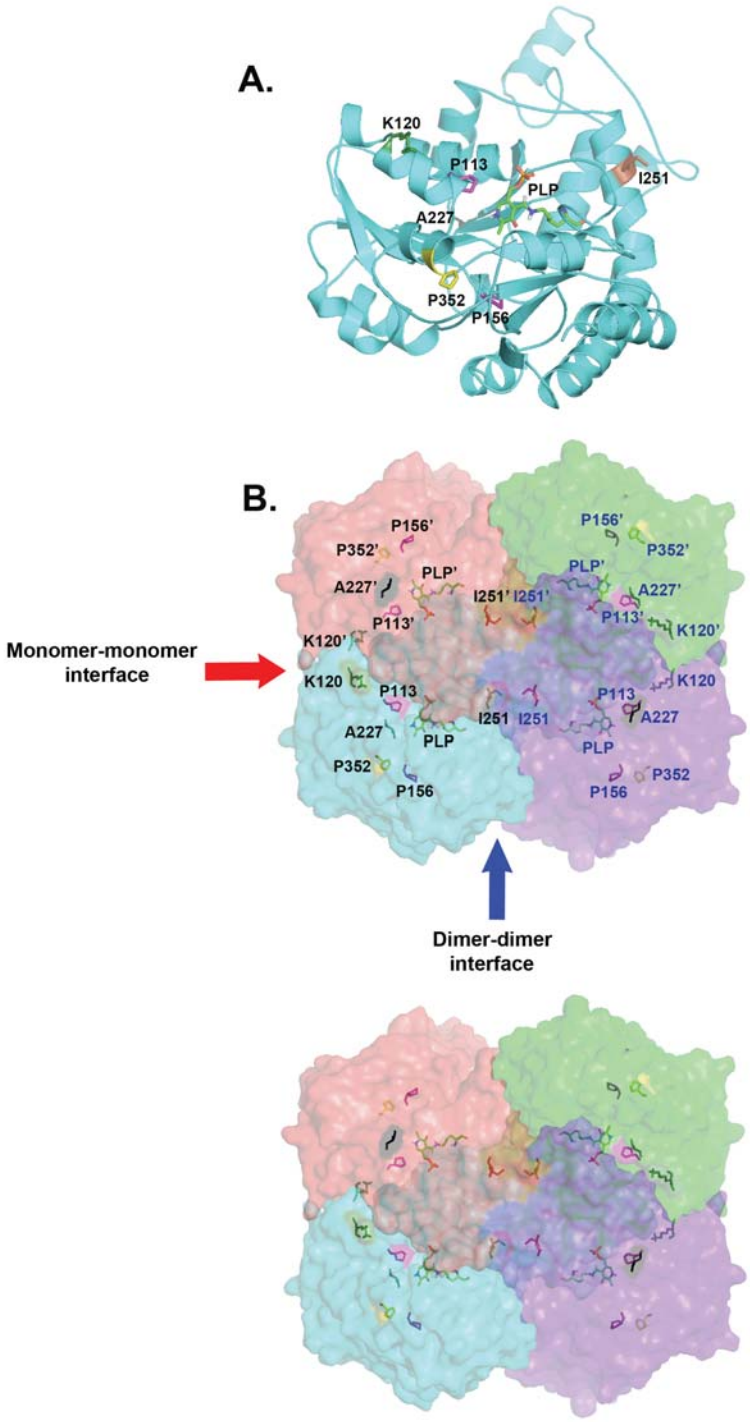


Figure 3.11: Mutated residues from the second-generation MetC winner clones.

The mutated residues from the second-generation MetC winner clones were mapped to the structure of *E. coli* MetC (PDB code: 1cl1). The biological unit of MetC is a tetramer, but the asymmetric unit in the PDB file is a dimer. The tetramer was recreated using the Sympexp command in MacPymol. Each PLP molecule is coloured according to its elements, while each residue is coloured differently. (A) Ribbon diagram of a MetC monomer and the position of the mutated residues. Loops have been smoothed for clarity. (B) Surface diagram of the quaternary structure of MetC. The mutated residues of each dimer are labelled in black and blue, respectively. The axis pointed by the red arrow is the monomer-monomer interface, whereas the axis pointed by the blue arrow is the dimer-dimer interface.

3.2.6 THE *IN VIVO* NATIVE ACTIVITY OF THE SELECTED METC VARIANTS

Two rounds of directed evolution yielded MetC variants that could efficiently rescue the D-Ala auxotroph, *E. coli* MB2795. The implication was that the *in vivo* alanine racemase activity of these MetC variants was higher after each round of selection. However, it was not known if the *in vivo* alanine racemase activity was improved at the expense of the enzyme's native activity. Hence, all four first-generation MetC variants (P113Q, P113A, P113T and P113S) and three second-generation MetC variants (P113S-2, P113S-14 and P113S-16) were expressed in *E. coli* Δ*metC* to investigate their *in vivo* cystathionine β-lyase activity. The clones were streaked on two sets of M9/glucose agar: one set with 50 μM arabinose and the other set without arabinose. Clone expressing the wildtype MetC served as the positive control, whereas a LacZ-expressing clone served as the negative control.

Table 3.12 summarizes the growth differences shown by various clones after 3 days of incubation at 28°C. Generally, the growth of the first-generation MetC variants was worse than the positive control on both sets of agar plates (Figure 3.12). In particular, the P113Q and P113T clones showed the poorest growth among all clones. This was followed by the P113A variant, whose growth was average in comparison to the confluent growth of the positive control clone. The growth of the P113S variant was reasonably good (although not as comparable to the growth of the positive control), and was considered the best among all first-generation MetC variants. These results suggested that the P113 mutations, which conferred improvements to the *in vivo* alanine racemase activity of MetC, were likely to be detrimental to the cystathionine β-lyase activity of MetC.

However, clones expressing the second-generation MetC variants grew better than the clones expressing first-generation MetC variants (Table 3.12). Even in the absence of arabinose, the growth of the second-generation variants was comparable to that of the positive control (Figure 3.12). This implied that the mutations of the second-generation variants seemed to be able to confer *in vivo* improvements to both alanine racemase (Figure 3.10) and cystathionine β-lyase activities (Figure 3.12). This observation was rather unexpected because the native activity of MetC had not been selected throughout the two rounds of directed evolution.

One possible explanation for the confluent growth shown by the second-generation variants was that there were regulatory mutations, in addition to the structural mutations detected previously (Table 3.11). It was realized at a later time that previous retransformation and restreaking tests used to confirm the growth phenotype of the second-generation winners (Figure 3.10) could not rule out the presence of regulatory mutations, because the *metC* inserts of the second-generation winners had not been recloned into fresh pBAD vectors. *In vitro* characterization of the activities of these MetC variants appeared to confirm this possibility (Section 3.2.7).

Table 3.12: Growth summary of *E. coli* Δ*metC* harbouring different MetC variants on selective medium after 3 days of incubation.

Variant	Growth on M9/glucose agar + 50 μM arabinose	Growth on M9/glucose agar alone
LacZ	−	−
MetC	++++	++++
MetC(P113Q)	+	+/−
MetC(P113A)	++	++
MetC(P113T)	+	+/−
MetC(P113S)	+++	++
MetC(P113S-2)	++++	++++
MetC(P113S-14)	++++	++++
MetC(P113S-16)	++++	++++

Key for scoring growth: (−) No growth; (+/−) Poor growth. Colony size was <0.1 mm; (+) Little growth. Colony sizes ranged from 0.1–0.5 mm; (++) Average growth. Colony sizes ranged from 0.5–1.0 mm; (+++) Good growth. Colony sizes ranged from 0.5–1.5 mm; (++++) Confluent growth. Colonies ranged from 1.0–2.0 mm.

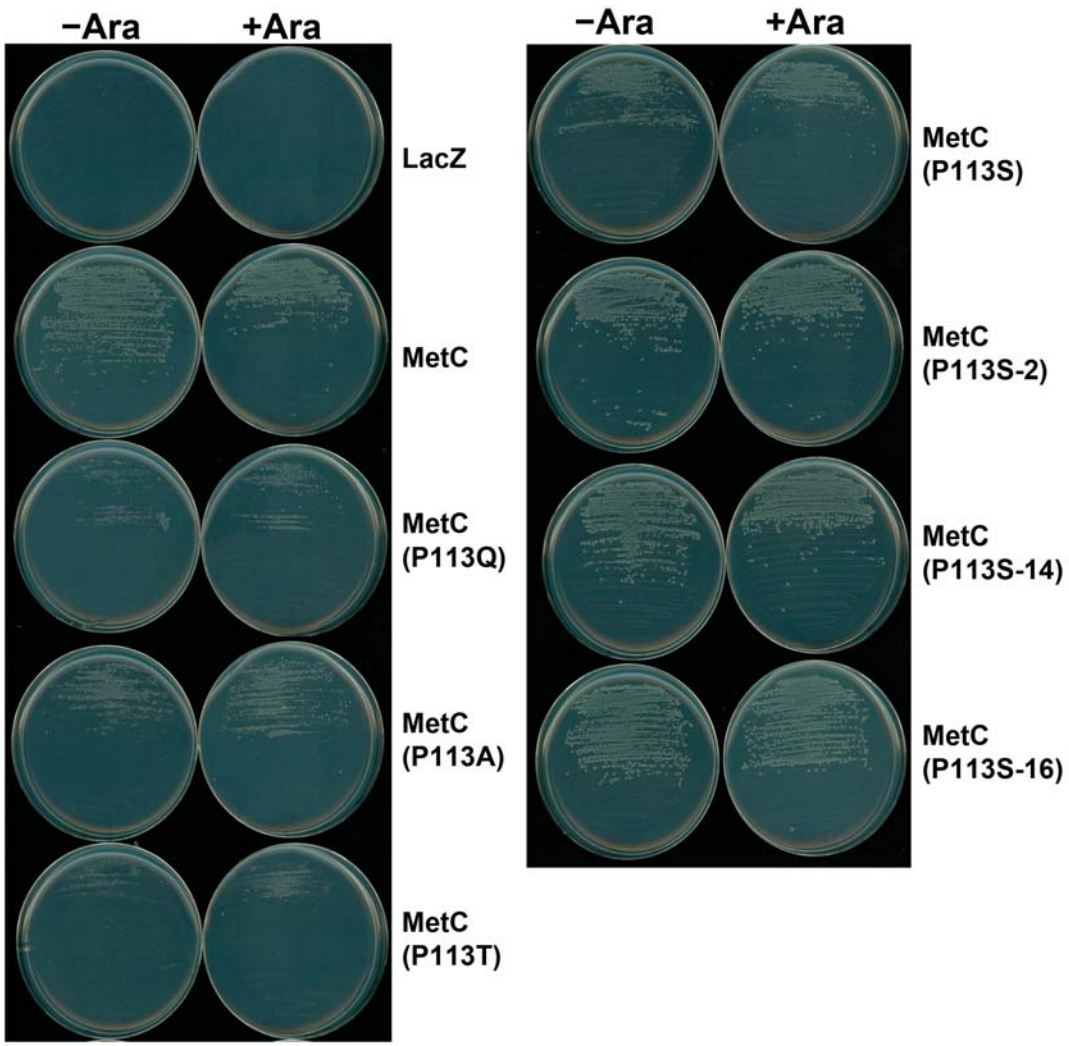
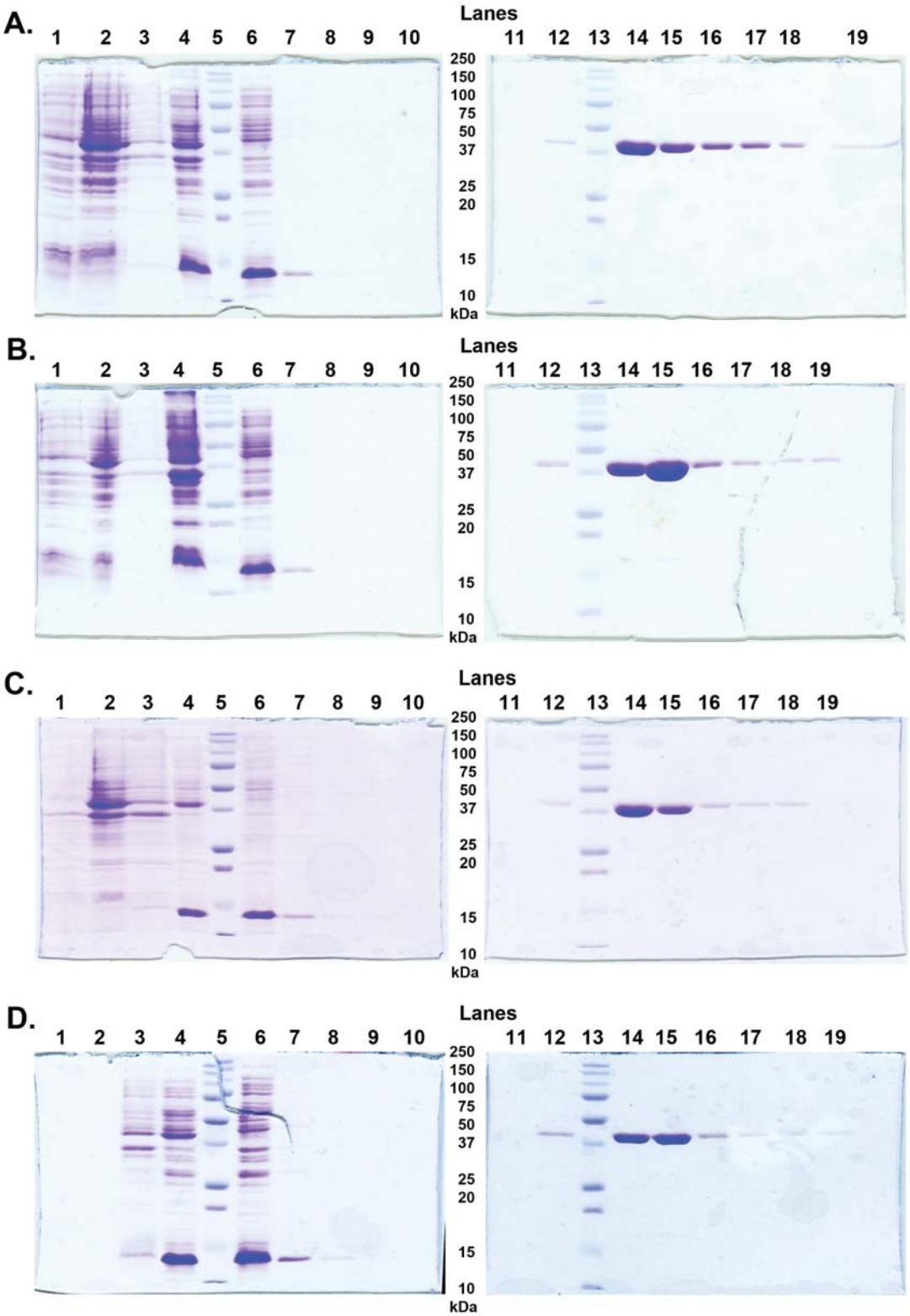


Figure 3.12: The growth of *E. coli* Δ*metC* clones harbouring different MetC variants on M9/glucose agar with and without 50 μM arabinose, after 3 days of incubation at 28°C.

3.2.7 Enzyme activities of MetC mutants

For *in vitro* characterizations, all seven MetC mutants (P113Q, P113A, P113T, P113S, P113S-2, P113S-14 and P113S-16) were expressed in *E. coli* MB2795, and purified to >95% purity, as previously described (Section 2.4.5 in Chapter 2). All enzyme variants appeared largely soluble, although their yields were lower than wildtype MetCs, as judged by SDS-PAGE (compare Figure 3.13 *versus* Figure 2.12A). This was likely to be caused by the lower expression, and copy number, of pBAD plasmids in the cell. The final enzyme yields for all MetC mutants were <10 mg/L of culture. The yield of the P113A variant was the highest (9.3 mg/L of culture), while the yield of the P113S-14 variant was the lowest (2.9 mg/L of culture). Overall, the proportion of PLP-bound enzymes in the MetC mutants (30% to 50%) was generally lower than that of wildtype MetC (>70%; Section 2.2.4.1 in Chapter 2).


A.
Lanes
1 2 3 4 5 6 7 8 9 10
11 12 13 14 15 16 17 18 19
250
150
100
75
50
37
25
20
15
10
kDa
B.
Lanes
1 2 3 4 5 6 7 8 9 10
11 12 13 14 15 16 17 18 19
250
150
100
75
50
37
25
20
15
10
kDa
C.
Lanes
1 2 3 4 5 6 7 8 9 10
11 12 13 14 15 16 17 18 19
250
150
100
75
50
37
25
20
15
10
kDa
D.
Lanes
1 2 3 4 5 6 7 8 9 10
11 12 13 14 15 16 17 18 19
250
150
100
75
50
37
25
20
15
10
kDa

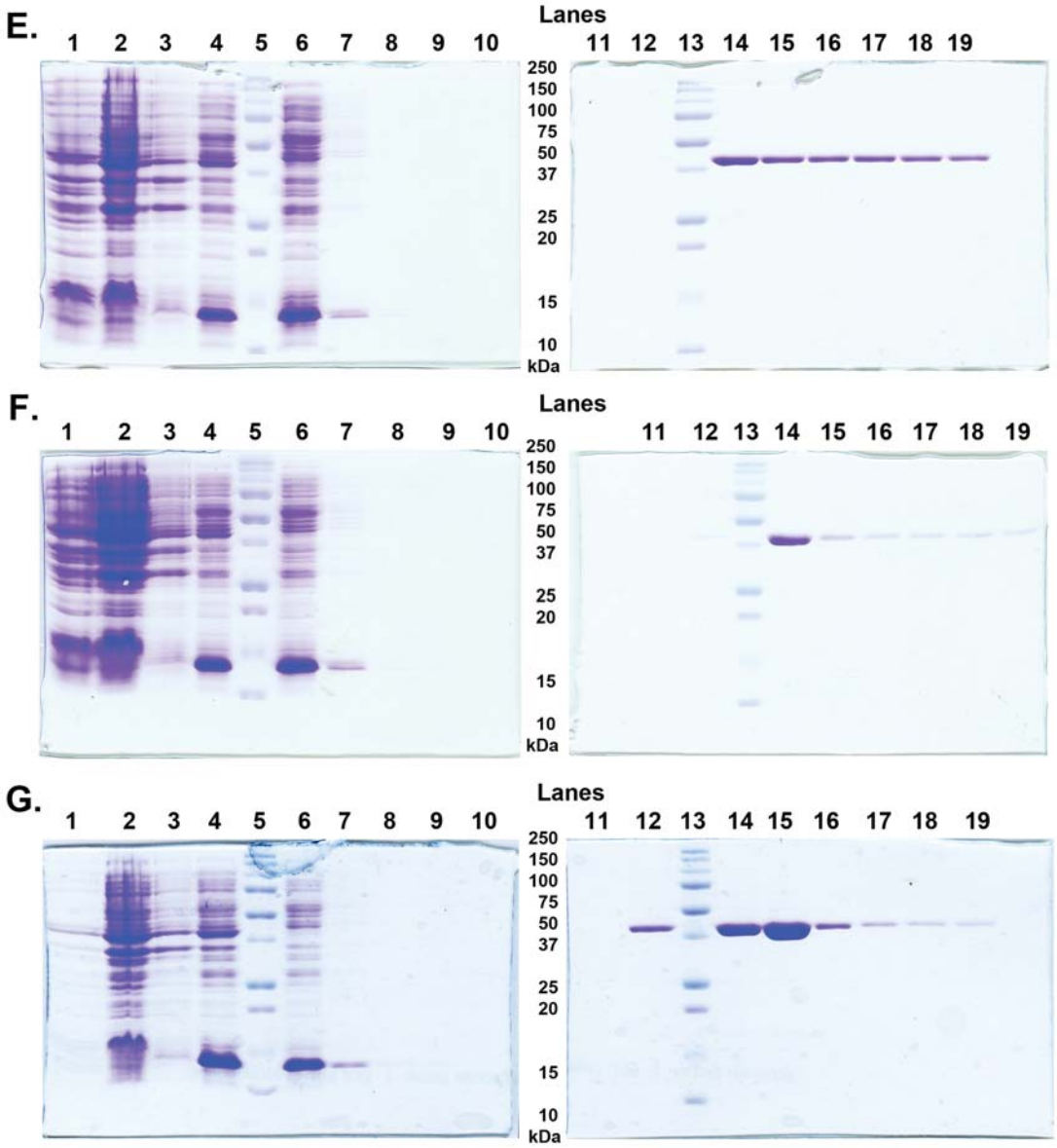


Figure 3.13: SDS-PAGE analyses of the purified MetC mutants.

(A) Purification of MetC(P113Q). (B) Purification of MetC(P113A). (C) Purification of MetC(P113T). (D) Purification of MetC(P113S). (E) Purification of MetC(P113S-2). (F) Purification of MetC(P113S-14). (G) Purification of MetC(P113S-16). Each lane [except lanes 1 and 2 of panel (D)] was loaded with a 6-µL aliquot from the corresponding fractions. Lane 1, Crude cell lysate before arabinose induction. Lane 2, Crude cell lysate after 6 hours of protein over-expression. Lane 3, Insoluble fraction. Lane 4, Soluble lysate. Lanes 5 and 13, Bio-Rad Precision Plus protein All Blue standard. Lane 6, Proteins not bound to the Talon® metal affinity resin. Lanes 7–8, Batch washes. Lane 9, Flow-through from the enzyme-bound resin in the gravity flow column. Lanes 10–11, Column washes with low imidazole concentration (3 mM). Lane 12, Column wash with high imidazole concentration (15 mM). Lanes 14–19, Elution fractions 1–6. The lanes 1 and 2 of panel (D) were empty due to accidental loss of samples.

After dialyzing each MetC variant into a buffer without β-mercaptoethanol, the dialyzed enzyme fractions were used for the *in vitro* assays. Both the alanine racemase and the cystathionine β-lyase activities of each MetC variant were assayed spectrophotometrically as previously described (Section 2.4.7 in Chapter 2; and Section 3.4.6). The kinetic parameters for all seven MetC variants are summarized in Table 3.13. The kinetic parameters were established from the mean of three repeats, with each repeat value representing the mean of duplicate assays.

Table 3.13: Kinetic parameters of MetC and selected variants from the first- and second-generation libraries.

Enzyme	Cystathionine → Homocysteine + Pyruvate + NH_3			L-Ala → D-Ala		
	k_{cat} (s^{-1})	K_M (M)	k_{cat}/K_M ($s^{-1}.M^{-1}$)	k_{cat} (s^{-1})	K_M (M)	k_{cat}/K_M ($s^{-1}.M^{-1}$)
Native enzymes						
MetC	89.5 ± 13.9	$(3.9 \pm 0.7) \times 10^{-5}$	2.3×10^{6}	3.3 ± 0.6	$(5.1 \pm 0.4) \times 10^{-2}$	65
Alr	†n.d.	n.d.	<0.85	55.3 ± 3.7	$(8.7 \pm 1.5) \times 10^{-4}$	6.4×10^{4}
First-generation mutants						
P113Q	95.6 ± 7.5	$(1.2 \pm 0.2) \times 10^{-4}$	8.0×10^{5}	3.6 ± 0.6	$(3.3 \pm 1.0) \times 10^{-2}$	1.1×10^{2}
P113A	81.1 ± 10.8	$(3.1 \pm 0.7) \times 10^{-4}$	2.6×10^{5}	4.6 ± 0.5	$(2.3 \pm 0.5) \times 10^{-2}$	2.0×10^{2}
P113T	30.4 ± 2.0	$(2.1 \pm 0.2) \times 10^{-4}$	1.4×10^{5}	5.8 ± 0.5	$(3.9 \pm 0.4) \times 10^{-2}$	1.5×10^{2}
P113S	94.6 ± 3.2	$(1.2 \pm 0.2) \times 10^{-4}$	7.9×10^{5}	21.6 ± 0.4	$(5.8 \pm 0.3) \times 10^{-2}$	3.7×10^{2}
Second-generation mutants						
P113S-2	76.0 ± 15.7	$(2.0 \pm 0.4) \times 10^{-4}$	3.8×10^{5}	11.9 ± 0.1	$(7.1 \pm 1.0) \times 10^{-2}$	1.7×10^{2}
P113S-14	103.2 ± 16.2	$(1.9 \pm 0.6) \times 10^{-4}$	5.4×10^{5}	20.5 ± 2.1	$(5.6 \pm 1.4) \times 10^{-2}$	3.7×10^{2}
P113S-16	76.0 ± 4	$(1.4 \pm 0.4) \times 10^{-4}$	5.4×10^{5}	16.3 ± 3.4	$(5.2 \pm 0.7) \times 10^{-2}$	3.1×10^{2}

Each value represented the mean ± S.E.M. (n = 3), and each value was obtained as the mean of duplicate assays.
†n.d., not determined.

The catalytic efficiency (k_{cat}/K_M) for alanine racemization by each of the first-generation MetC mutants ranged from 110 to 370 $s^{-1}.M^{-1}$, which corresponded to an increase of 1.7-fold to 5.7-fold over wildtype MetC (65 $s^{-1}.M^{-1}$). In comparison with wildtype MetC, the K_M for L-Ala was slightly decreased and the k_{cat} values were largely unchanged for the P113Q, P113A and P113T variants. On the other hand, the kinetic changes in the P113S variant were opposite to the other P113 mutants. A modest 6.5-fold increase was observed for the k_{cat} of P113S, while its K_M was similar to the wildtype MetC. The P113S mutant was also the variant with the highest alanine racemase activity among the first-generation mutants, with

a 5.7-fold improvement over wildtype MetC based on their catalytic efficiencies. Overall, the alanine racemase activity of the P113S variant was only 170-fold lower than wildtype Alr.

All three second-generation MetC variants were expected to display higher *in vitro* alanine racemase activity than the P113S variant, since the latter was their parental variant. However, this trend was not observed. The catalytic efficiency, K_M and k_{cat} values of the P113S-14 and P113S-16 variants were comparable to those of the P113S variant. The P113S-2 variant even showed a 2-fold lower k_{cat} and a 1.2-fold higher K_M than the P113S variant. Although these results appeared rather counter-intuitive, they appeared to support the hypothesis that additional regulatory mutations might be present in their plasmid backbone (see Section 3.2.6 for the discussion). These regulatory mutations might increase the expression, or even allow constitutive expression, of MetC in the cell. As a result, the second-generation MetC variants would rescue *E. coli* MB2795 (Figure 3.10) and *E. coli* Δ*metC* (Figure 3.12 in Section 3.2.6) better than the P113S variant. This elevated, or constitutive, expression of MetC could be apparent on the SDS-PAGE analyses, where the crude cell lyates at t = 0 (Lane 1 of Figure 3.13E, F and G) would show a heavy band that corresponded to the size of MetC. However, this was not observed. Further investigations are needed to confirm the presence of such regulatory mutations in the second-generation MetC variants, and to determine whether these regulatory mutations are solely responsible for the fitness-enhancing effects in growth complementation assays.

In comparison to wildtype MetC, the cystathionine β-lyase activity of the first-generation MetC variants was decreased by 3- to 16-fold (Table 3.13). The P113T variant showed the lowest cystathionine β-lyase activity due to a k_{cat} value that was 3-fold lower than that of wildtype MetC. The P113T variant was also the only one with a lowered k_{cat}, while the k_{cat} of the P113Q, P113A and P113S variants remained similar in comparison with the wildtype MetC. On the whole, the K_M of all the first-generation variants was increased by 3- to 5-fold. For the second-generation variants, their cystathionine β-lyase activity was lower than their parental P113S variant, albeit by a mere 1.4- to 2-fold (Table 3.13). Among these variants, the P113S-2 variant showed the lowest cystathionine β-lyase activity, which was approximately 6-fold lower than wildtype MetC.

Activity trade-offs of the MetC variants are illustrated in Figure 3.14, where both of their cystathionine β-lyase and alanine racemase activities were plotted. All P113 mutations appeared to decrease the enzyme's cystathionine β-lyase activity by ≥3-fold, while only improving the alanine racemase activity by ≥2-fold. Overall, the alanine racemase activity of these MetC variants appeared to be improved at the expense of their native activity.

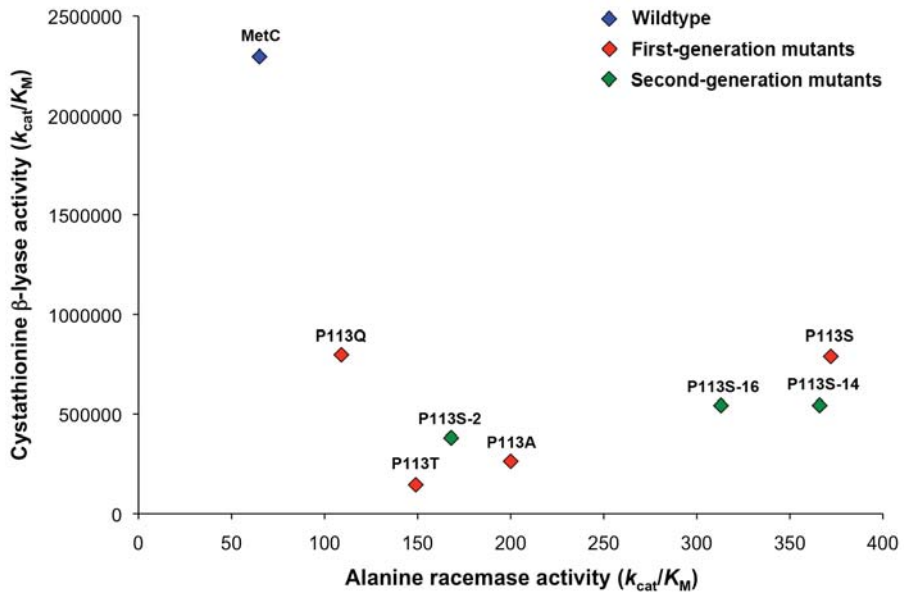


Figure 3.14: Activity trade-off in MetC variants.

3.3 DISCUSSION

3.3.1 ALANINE RACEMASE ACTIVITY OF METC

At the outset of this work, I expected that the *E. coli* MB2795 strain could be rescued by promiscuous alanine racemase activity in the cells. I speculated that the *in vivo* alanine racemase activity could potentially stem from the following enzymes:

(A) **General racemases or epimerases.** All racemases and epimerases, including *E. coli* Alr, interconvert L- and D-isomers in their active sites by having proton donors and acceptors arranged in opposite planes. The similar arrangements of proton donors and acceptors in the active sites would provide a favourable environment for alanine racemization. Besides, the chances that alanine will be recognized as a substrate in other racemases or epimerases are high because alanine is a relatively small substrate. Among the 4 racemases and 23 epimerases in *E. coli* (Keseler *et al.*, 2009), many isomerize amino acids and other small substrates. For instance, glutamate racemase isomerizes glutamate (Yoshimura *et al.*, 1993; Lundqvist *et al.*, 2007). It was appealing to hypothesize that the active site of glutamate racemase could also accommodate alanine at physiological levels since alanine is an even smaller amino acid than glutamate.

(B) **$(\beta/\alpha)_8$-barrel enzymes.** *E. coli* Alr possesses a $(\beta/\alpha)_8$-barrel (also known as TIM-barrel) fold, and promiscuous activities arising from TIM-barrel enzymes have been well-

documented. For instance, TIM-barrel enzymes found in enzyme superfamilies such as the amidohydrolases (Roodveldt & Tawfik, 2005; Afriat *et al.*, 2006) and the enolases (Palmer *et al.*, 1999; Thoden *et al.*, 2004) are known to catalyze secondary functions. Specific examples include HisA [*N*′-(5′-phosphoribosyl)formimino-5-aminoimidazole-4-carboxamide ribonucleotide isomerase], which is able to catalyze the TrpF (phosphoribosylanthranilate isomerase) reaction, although both enzymes belong to different biosynthetic pathways (Jürgens *et al.*, 2000; Henn-Sax *et al.*, 2002). As the TIM-barrel fold is a common fold in prokaryotes (Wolf *et al.*, 1999), I speculated that there could be TIM-barrel enzymes in *E. coli* that might catalyze alanine racemization.

(C) **Other PLP-dependent enzymes.** Promiscuous activities are rather common in PLP-dependent enzymes due to their shared mechanistic chemistry. It was possible that other PLP-dependent enzymes in *E. coli*, in addition to Alr, might be able to catalyze alanine racemization. One of the probable candidates was diaminopimelate decarboxylase (PDB code: 1KNW, Unpublished work by Levdikov *et al.*), as this also possesses a TIM-barrel domain. However, following the results from Chapter 2, I also expected that promiscuous alanine racemase activity could arise in other non-homologous, PLP-dependent enzymes.

Indeed, results from the Δ*alr* Δ*dadX* suppression screen revealed that the last speculation was right — one of the suppressors was found to express MetC, a PLP-dependent enzyme that is non-homologous to Alr (Section 3.2.1). Even so, the discovery that MetC could rescue *E. coli* MB2795 was unexpected. Recall that, in Chapter 2, Alr was found to catalyze MetC's activity *in vivo* and *in vitro*. In a reciprocal manner, the *in vivo* and *in vitro* alanine racemase activity of MetC were discovered and characterized in this chapter. Overall, the discovery of this reciprocal activity profiles shown by MetC and Alr was far from the initial predictions.

The *in vitro* alanine racemase activity of MetC was initially expected to be low, as promiscuous activities stemming from non-homologous scaffolds are generally low [Alr from Chapter 2; YeaB from Kim *et al.* (2010); PurF from Patrick & Matsumura (2008)]. However, the k_{cat}/K_M of MetC for alanine racemization (65 $s^{-1}.M^{-1}$) was at least 80-fold higher than other promiscuous activities previously reported (0.85 $s^{-1}.M^{-1}$ for an evolved Alr variant; ~0.3 $s^{-1}.M^{-1}$ for an evolved PurF variant; >0.024 $s^{-1}.M^{-1}$ for YeaB). In particular, the k_{cat} for L-Ala turnover by MetC (~3 s^{-1}) was less than 4-fold lower than the average "global" k_{cat} of 10 s^{-1}, which was derived from several thousand enzymes (Bar-Even *et al.*, 2011). Based on the k_{cat} value, alanine racemization may occur in MetC at a rate of approximately 200 turnovers per minute (or 12,000 turnovers per hour) — a rate that was clearly sufficient to sustain the growth of the D-Ala auxotroph, *E. coli* MB2795, in the absence of D-Ala.

Another group has also recently discovered that upregulation of MetC in an *E. coli* BL21(DE3)/Δ*alr* Δ*dadX* strain led to the formation of prototrophic revertants in the absence of D-Ala (Kang *et al.*, 2011). Point mutations in the *metJ* repressor were ascribed to the increased MetC expression level in the revertants. Their results were analogous to the findings reported here, where the growth of *E. coli* MB2795 was effected by over-expression of the plasmid-encoded MetC. The overall implication is that MetC expression in *E. coli* could be up-regulated to physiologically relevant levels by regulatory mutations. These regulatory mutations can increase enzyme expression to compensate for the low promiscuous activity (see previous discussion in Section 2.3.2 of Chapter 2).

3.3.2 *IN VITRO* EVOLUTION OF METC

Despite the structural differences shown by MetC and Alr (see previous discussion in Section 2.3.1 of Chapter 2), I attempted to improve the alanine racemase activity on the MetC scaffold. After one round of directed evolution, mutations were observed for P113 in almost all selected MetC variants. The prevalence of P113 mutations implied that the residue was critical for the fitness improvements seen in MetC clones selected in the absence of D-Ala (Table 3.5 in Section 3.2.4.3). To my knowledge, the role of P113 has not been studied in any MetC enzyme.

Residue P113 is located in the vicinity of the active site of MetC (Figure 3.16), especially close to Y111. Two roles have been proposed for Y111 (Clausen *et al.*, 1996; Clausen *et al.*, 1997b): proton abstraction from the α-amino group of cystathionine (the substrate) during the transaldimination step (Panel II of Figure 3.15), and re-protonation of the PLP derivative of aminoacrylate after the elimination of homocysteine (Panel IV of Figure 3.15). The re-protonated aminoacrylate forms iminopropionate, which would diffuse from the active site (and spontaneously hydrolyze into pyruvate and ammonia). As shown in Figures 3.16, the short distance (3.6 Å) between the phenolate ion of Y111 and the α-amino group of a MetC substrate analogue [aminoethoxyvinylglycine (AVG), which is also an inhibitor] may facilitate proton transfer. While P113 is not directly involved in the catalytic mechanism of MetC, the kinetic data suggested that mutations at P113 might mediate catalytic changes in an alternative manner. It was rather unlikely that the P113 mutations altered the hydrophobicity or steric effects in the active site, because the side chains of all four P113 mutations are not of the same length, nor belong to the same chemical group. For example, the side chains of S113 and A113 are shorter in length than those of T113 and Q113. Also, S113, T113 and Q113 carry polar side chains, whereas A113 harbours a hydrophobic side chain. There was also lack of evidence that the P113 mutations were changing the substrate preference of MetC from cystathionine to alanine (Table 3.13 in Section 3.2.7). The K_M values of the P113 mutants for alanine were generally unchanged in comparison to wildtype MetC.

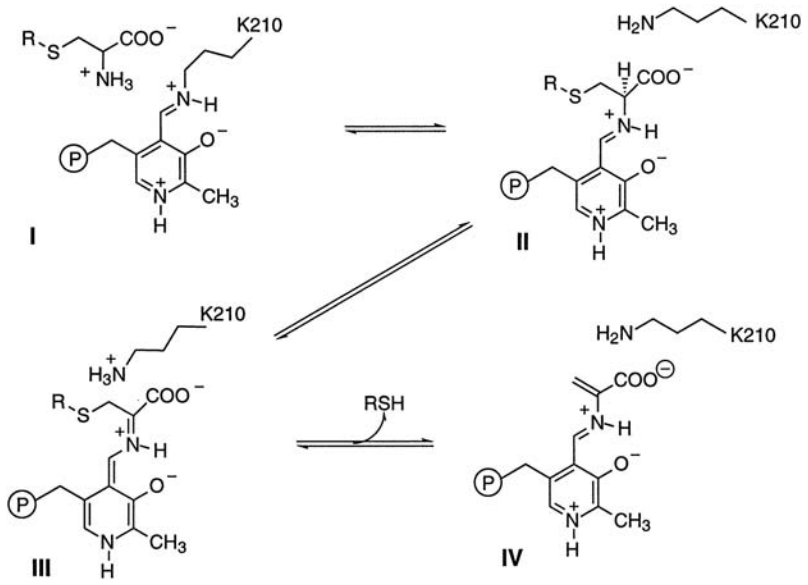


Figure 3.15: The proposed cystathionine β-elimination mechanism of MetC.

Transaldimination occurs after cystathionine (Panel I) binds to the PLP, and results in the formation of an external aldimine (Panel II). The proton of cystathionine is abstracted to yield a resonance stabilized ketimine quinonoid intermediate (Panel III). Homocysteine (abbreviated as RSH) is then eliminated (as a result of protonation of Sγ by K210), and a PLP derivative of aminoacrylate is formed (Panel IV). Re-protonation of the aminoacrylate by Y111 yields iminopropionate, and leads to regeneration of the enzyme-bound PLP. The iminopropionate is hydrolyzed to pyruvate and ammonia after it is released from the active site. Figure taken from Clausen *et al*. (1997a).

The racemization activity in Alr requires a catalytic pair (K34 and Y255') at opposite planes for proton transfers (Figure 2.28 in Section 2.3.1). A similar arrangement was also found in MetC — K210 and Y111 are arranged in such a way that alanine racemization appears spatially and chemically possible (Figures 3.16B and D). Based on the arrangement of K210 and Y111 in MetC, it was initially speculated that the P113 mutations might have caused a subtle rearrangement of Y111 in the active site. Since both P113 and Y111 are located at the start of helix 6, mutations at P113 might bring Y111 closer to the substrate (and PLP), thereby facilitating proton transfers. Consequently, the more efficient proton transfers would hypothetically lead to higher racemization activity, as evident from the 7-fold improvement in k_{cat} for the P113S mutant (Table 3.13 in Section 3.2.7). However, the unaffected k_{cat} for alanine racemization in the P113Q, P113A and P113T mutants do not appear to support this speculation. Furthermore, the closer distance between Y111 and the substrate might also affect the turnover of cystathionine in MetC, since the cystathionine β-lyase activity also requires the protonation and deprotonation steps (Figure 3.15). Yet, the k_{cat} for cystathionine turnover was largely unaffected for the P113 mutants (Table 3.13 in Section 3.27).

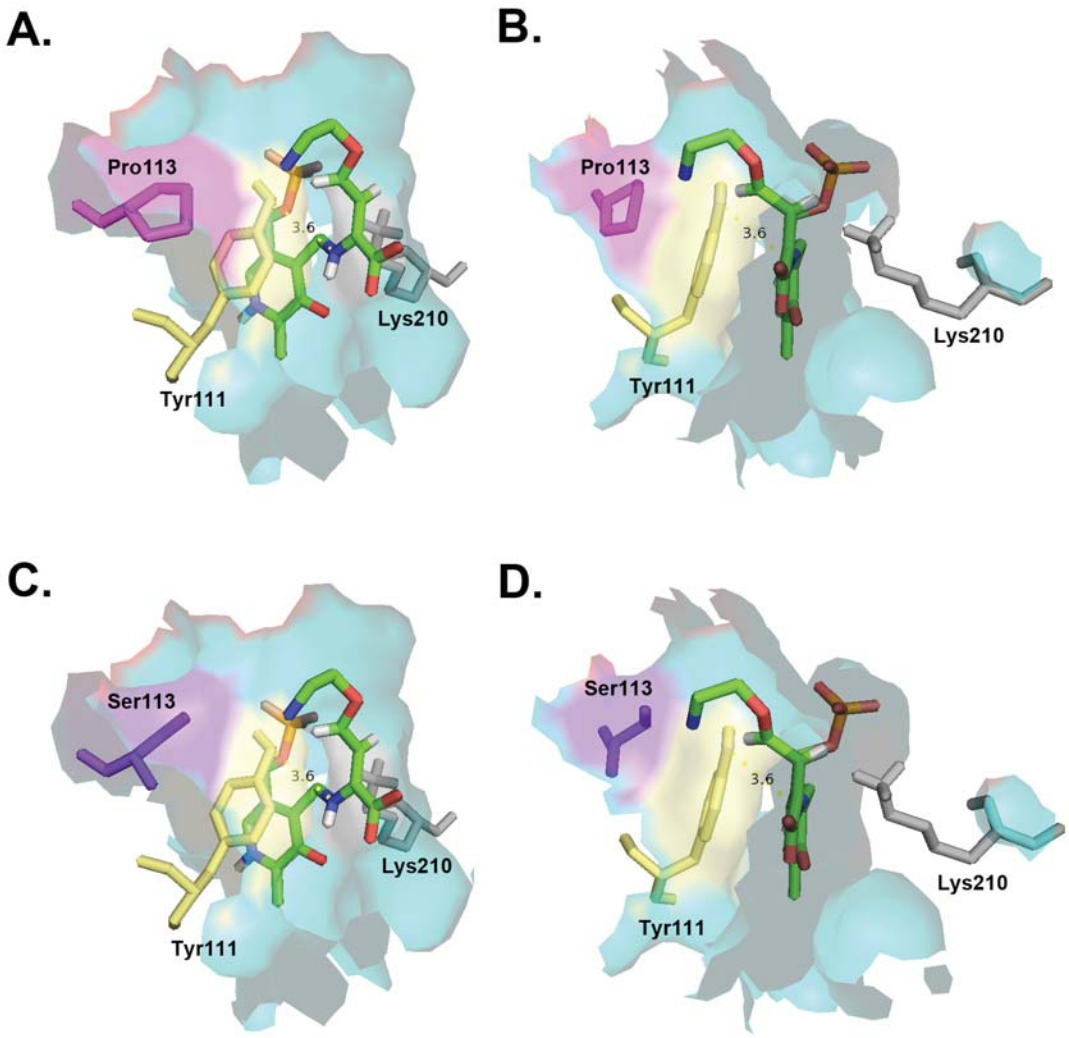


Figure 3.16: Surface diagrams of the active site of *E. coli* MetC.

The PLP-AVG complex, P113 and the catalytic pair (K210 and Y111) are shown as sticks. The *E. coli* MetC bound to AVG (PDB code: 1cl2) was used as the template for modelling. The distance between the phenolate ion of the Y111 and the α-amino group of the AVG is 3.6 Å. (A) & (C) The active site viewed from the substrate entranceway. Residue P113 is highlighted in magenta, whereas S113 is modelled as purple. (B) & (D) Side view of the active site.

Mutations that improve the promiscuous activity often compromise the native activity of the enzyme (Jürgens *et al.*, 2000; Seebeck & Hilvert, 2003), because such mutations tend to disrupt the optimal catalytic or binding environments for the native substrate. Indeed, moderate activity trade-offs were observed for the first-generation MetC mutants (Figure 3.14 in Section 3.2.7). However, the moderate reductions in the cystathionine β-lyase activity (Table 3.13 in Section 3.2.7) may be pivotal for the overall improvements of the alanine racemase activity, because such reductions may divert the activity flux in MetC towards the alanine racemase activity (Copley, 2012). Activity flux, or catalytic preference, can be

defined as the ratio of activity A to activity B. Figure 3.17 illustrates the catalytic preference of each MetC variant as a ratio of its promiscuous (alanine racemase) activity to its native (cystathionine β-lyase) activity. Wildtype MetC shows a low catalytic preference towards the alanine racemase activity, as its cystathionine β-lyase activity is four orders of magnitude higher than the racemase activity (Table 3.13 in Section 3.2.7). In comparison to wildtype MetC, the MetC mutants showed significantly higher preference towards the alanine racemase activity due to moderate decreases in the cystathionine β-lyase activity and modest increases in the alanine racemase activity. Among all the mutants, the P113T variant showed the highest catalytic preference towards the alanine racemase activity — its catalytic preference of 0.001 is 38-fold higher than the wildtype MetC. The high preference for alanine racemase activity in the P113T variant is attributed to its poor k_{cat}/K_M for its native activity (Table 3.13 in Section 3.2.7). On the other hand, the variants with the highest alanine racemase activity among all mutants, P113S and P113S-14, display a slight difference in their catalytic preferences. Due to a lower cystathionine β-lyase activity, the catalytic preference of P113S-14 variant was 1.5-fold higher than P113S (Figure 3.17). Taken together, mutations at the residue P113 can either increase the promiscuous alanine racemase activity, or decrease the native cystathionine β-lyase activity on the MetC scaffold. Based on the catalytic efficiencies, the P113S mutation seems to increase the promiscuous activity, while the P113T was responsible for the decrease in MetC's native activity.

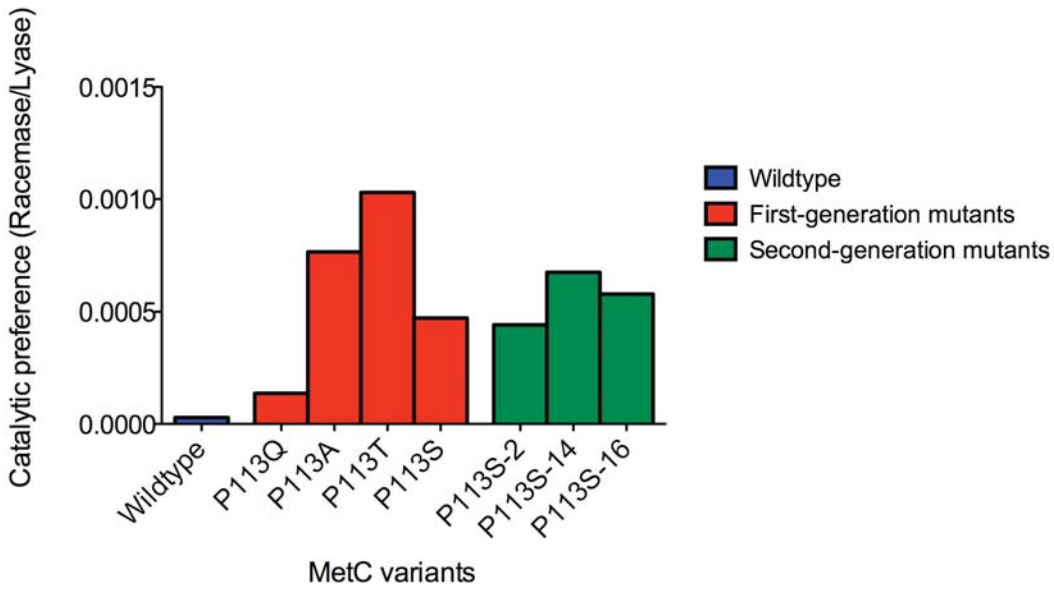


Figure 3.17: Catalytic preferences of MetC variants

For each MetC variant, the catalytic preference was derived from the ratio of k_{cat}/K_M of the alanine racemase activity to k_{cat}/K_M of the cystathionine β-lyase activity.

To elucidate the roles of P113S and P113T mutations in altering the catalytic preferences on the MetC scaffold, the Squire group (University of Auckland) is attempting to solve the structures of both mutant enzymes. Crystal structures of the native and the inhibitor-bound P113S and P113T mutants will be solved. The inhibitor of the enzyme, (*R*)-1-aminoethylphosphonic acid [L-Ala-P], is a synthetic L-Ala analogue that forms an external aldimine with alanine racemases (Stamper *et al.*, 1998). As L-Ala-P is neither racemized nor hydrolyzed, the L-Ala-P-bound enzymes will be trapped in a non-reactive intermediate state. Although L-Ala-P is not known to inhibit MetC, crystals of the P113S variant saturated with L-Ala-P, have recently been obtained (Christopher Squire, personal communication). It is hoped that the native and the L-Ala-P-complexed structures of the P113 mutants will be informative, especially regarding the effects of P113 mutations and the residues required for the racemization activity in MetC. The mechanistic analysis of the alanine racemase activity in MetC may also help to resolve a conundrum about the role of Y111 as a proton carrier. While Clausen *et al*. (1996) initially proposed that Y111 is involved in proton transfers for the cystathionine β-lyase activity, another group has recently argued that Y111 has no role in proton transfers, based on their relatively unchanged activity observed in the Y111F variant (Lodha & Aitken, 2011).

3.3.3 RECIPROCAL PROMISCUITY IN NON-HOMOLOGOUS SCAFFOLDS

The overall findings from Chapter 2 and 3 substantiate the hypothesis that promiscuous activities can arise in non-homologous enzymes. For *E. coli* MetC and Alr, the PLP cofactor and other catalytic residues in their active sites provide environments permissible for secondary activities. The abilities of MetC and Alr to catalyze each other's reaction are particularly noteworthy, because to my knowledge, such reciprocal promiscuity has not been noted elsewhere.

From the literature, cross-reactivities, or overlapping activities among homologous enzymes is an occurrence that appears most similar to reciprocal promiscuity shown by MetC and Alr. Several groups have previously reported the overlapping activity profiles in enzyme homologues, such as the paraoxonase and dihydroorotase activities in *P. diminuta* phosphotriesterase (PTE) and *E. coli* dihydroorotase (DHO). The DHO enzyme catalyzed both its native dihydroorotase activity and paraoxonase activity, but the PTE enzyme only catalyzed its "native" paraoxonase activity (Roodveldt & Tawfik, 2005). Similarly, Jürgens *et al*. (2000) observed that an artificially evolved HisA of *Thermotoga maritima* shows low TrpF activity (46 $s^{-1}.M^{-1}$), while maintaining its native activity (4.6×10^4 $s^{-1}.M^{-1}$). However, no HisA activity was detected on the TrpF scaffold.

Yet, the reciprocal promiscuity shown by MetC and Alr is different from these examples of enzyme cross-reactivities in three aspects. First, Met and Alr are not homologues, while the enzymes showing cross-reactivities are structural homologues

(Roodvelt & Tawfik, 2005; Jürgens *et al*., 2000). Second, the native and promiscuous activities of MetC are the promiscuous and native activities of Alr, respectively. In contrast, only one enzyme from either the PTE/DHO or HisA/TrpF combination is able to catalyze two activities. In particular, the paraoxonase activity of PTE is not even considered as the enzyme's native activity, because the substrates for the activity are synthetic organophosphates, such as paraoxon (Roodvelt & Tawfik, 2005). Lastly, and perhaps most importantly, the promiscuous activities of both MetC and Alr are relevant for *E. coli* fitness, *i.e*., both enzymes rescued *E. coli* when either protein was absent. The ability of PTE to replace the function of DHO in the cell, and vice versa, has not been demonstrated. While the *T. maritima* HisA was previously shown to rescue an *E. coli* Δ*trpF* strain (Jürgens *et al*., 2000), no suppressor genes were able to rescue the *E. coli* Δ*hisA* strain (Patrick *et al*., 2007).

For both MetC and Alr, mutations at a single residue were sufficient to improve their promiscuous activities. Mutations at the P113, such as P113S, on the MetC scaffold were particularly insightful. Based on the activities of the P113S variant, the high turnover rates for both cystathionine (~90 s^{-1}) and alanine (~20 s^{-1}) indicated that the MetC scaffold is capable of supporting both activities, since there is only a 4-fold difference between the two turnover rates. In contrast, the MetC scaffold does not appear to bind both cystathionine and alanine equally well. The difference between the K_M values of both substrates is 1,500-fold, and the P113S variant bound alanine as poorly as the wildtype MetC. On the whole, it seems that the barrier to switching the activity in the MetC scaffold lies with the constrained substrate specificity.

However, the kinetic changes of the P113S mutant appear to be consistent with the current arguments for the evolution of PLP-dependent enzymes. Due to the catalytic potential of PLP, the ancestral PLP-dependent enzymes were likely to be multifunctional after PLP was incorporated into the rudimentary active sites. Over time, the scaffolds evolved to optimize substrate binding and possibly, reaction specificity (Christen & Mehta, 2001; Eliot & Kirsch, 2004). As a result, the substrate and reaction specificities of modern PLP-dependent enzymes are largely controlled by their protein scaffolds. Therefore, mutations on the protein scaffolds are likely to disrupt the optimal binding of the native substrate, but not the alternative substrate (as observed for the P113 mutants; see kinetic data in Table 3.13 of Section 3.2.7). In some cases such as the P113S mutant, the spatial changes in the binding of native substrate may remove the physical barrier for the catalysis of side reactions. This leads to an increase of product turnover, such as the significant increase of k_{cat} in the P113S variant.

3.3.4 Increased protein expression as a route to fitness?

Overall, weak promiscuous activities, such as the cystathionine β-lyase activity of Alr (Chapter 2) and the alanine racemase activity of MetC (Chapter 3), could be physiologically

relevant when their expression is increased. This can be inchieved by over-expression from plasmids, regulatory mutations or gene duplications (or higher-order amplifications). The results from Chapters 2 and 3 implied that elevated protein expression is likely to be an important determinant in the evolution of new functions, and possibly, new phenotypes. This possibility will be investigated further in Chapter 4, in which over-expressed proteins from *E. coli* were screened for their ability to mediate the emergence of new phenotypes.

3.4 MATERIALS AND METHODS

Materials and methods specific to this chapter are described in the following sections as well as Section 2.4 (of Chapter 2). Common molecular biology materials and techniques, as well as the primer sequences, can be found in Appendix I.

3.4.1 IDENTIFICATION OF SUPPRESSOR GENES THAT COMPLEMENT ALANINE RACEMASE

An aliquot of *E. coli* MB2795 (Section I.3 in Appendix I) was transformed with 3.2 ng of pooled ASKA plasmids (Patrick *et al.*, 2007). The cells were recovered in 0.5 mL SOC supplemented with 0.5 mM D-Ala at 37°C for 1 h. The recovered cells were pelleted at 10,000 ×g for 3 min in a benchtop microcentrifuge. After discarding the supernatant, the cell pellet was washed thrice with 1× M9 medium to eliminate residual D-Ala from the growth medium. During each wash, the cell pellet was resuspended in 0.5 mL of 1× M9 medium, and centrifuged at 10,000 ×g for 3 min. The cell pellet was finally resuspended in 0.5 mL of 1× M9. A 50-μL aliquot of the cell suspension was plated on the selective agar (24.5 cm × 24.5 cm), LB-chloramphenicol supplemented with 50 μM IPTG. The selective agar plate was incubated in a sealed container lined with moistened paper towels, at 30°C for 7 days. Colony formation was observed throughout the incubation period. Appropriate dilutions of the cell suspension were also spread on non-selective medium, LB-chloramphenicol supplemented with 0.5 mM D-Ala, to enumerate the number of cells plated on the non-selective medium. Non-selective agar plates were incubated overnight at 37°C.

Colonies that formed on the selective agar plate (LB-chloramphenicol-IPTG) were inoculated in 0.5 mL LB-chloramphenicol supplemented with 0.5 mM D-Ala. The cultures were incubated overnight at 37°C. One μL from each saturated culture was used as the template for a PCR screen. The universal pCA24N-specific primers, pCA24N.for and pCA24N.rev (Section I.6 in Appendix I), were added to the PCR reaction to amplify each pCA24N-encoded ORF. The resulting PCR products were analyzed by gel electrophoresis, and purified using commercial spin columns. The purified PCR products were sent for sequencing to reveal the identity of the ORFs.

Once the identity of the ORFs was revealed, the corresponding plasmid from the individual ASKA clone was isolated for retransformation test. The retransformation tests were carried out to confirm the rescuing phenotype displayed by cells over-expressing the pCA24N-encoded ORFs. The same host strain, *E. coli* MB2795, was transformed with the purified plasmid. The cells were recovered and washed with 1× M9 medium as previously described. A total of 250–1,000 transformed cells (calculated by the number of colonies grown on non-selective medium in parallel) were spread on the selective medium, LB-

chloramphenicol supplemented with 50 μM IPTG. The agar plates were incubated at 30°C for at least 1 week.

3.4.2 Construction of *E. coli* (Δ*alr* Δ*dadX* Δ*metC*)

The protocol for in-frame deletion of the genomic *metC* gene in *E. coli* was modified from Datsenko and Wanner (2000). *E. coli* MB2795 was used as the target host since both the *alr* and *dadX* genes have been deleted from this strain (Strych *et al.*, 2001). Using a homologous recombination approach mediated by λ Red recombination, the *metC* gene was replaced with a kanamycin resistance (Kan^R) cassette (Figure 3.18).

3.4.2.1 Amplification of the Kan^R cassette

The 1.4-kb Kan^R cassette was derived from the *E. coli* Δ*metC* strain in the Keio collection (Baba *et al.*, 2006). A 1-μL aliquot of a saturated culture of *E. coli* Δ*metC* was added to a PCR reaction containing 1× Thermopol buffer, 1 U $Vent_R^®$ DNA polymerase (both from New England Biolabs), 0.2 mM dNTPs and 0.5 μM of each forward and reverse primer (metC-upstr.for and metC-dwnstr.rev; Section I.6 in Appendix I). The thermocycling conditions were 95°C for 6 min, 35 cycles of 95°C (30 s), 67°C (30 s), 75°C (45 s), and finally 75°C for 2 min. The amplified Kan^R product was purified using a commercial spin column.

3.4.2.2 Introduction of the Kan^R cassette into MB2795

Approximately 30 ng of the helper plasmid, pKD46 (Section I.4 in Appendix I), was used to transform *E. coli* MB2795. The electroporated cells were recovered in 0.5 mL LB supplemented with 0.5 mM D-Ala at 28°C for 90 min. The recovered cells were then spread on LB agar plates supplemented with 0.5 mM D-Ala and 10 μg/mL ampicillin (to select for clones transformed with pKD46). All agar plates were incubated overnight at 28°C. The next day, a colony was inoculated in 1 mL LB supplemented with 0.5 mM D-Ala and 10 μg/mL ampicillin, and grown overnight at 28°C. The saturated culture was made as a freezer stock, and was stored at −80°C.

This freezer stock (MB2795 cells harbouring the pKD46) was used to inoculate 0.1 mL LB supplemented with 0.5 mM D-Ala and 10 μg/mL ampicillin. The culture was inoculated overnight at 30°C. The next day, 2 mL LB supplemented with 0.5 mM D-Ala and 10 μg/mL ampicillin was added to the overnight culture, and the cells were incubated at 30°C for another 4 h. L-Arabinose (final concentration, 5 mM) was added to the culture to induce the expression of the recombinases from pKD46. Induction was allowed to proceed at 37°C for 1.5 h. At the end of the induction, the culture was evenly divided into two aliquots, which were pelleted at 17,000 ×g for 1 min. The cell pellet in each aliquot was resuspended twice with 1 mL of pre-chilled 10% (v/v) glycerol, and pelleted similarly after each resuspension.

The final cell pellet was resuspended using 50 μL of the pre-chilled 10% (v/v) glycerol. The cells were transformed with ~100 ng of the purified Kan^R cassette (Section 3.4.3.1), and the cells were recovered in 0.5 mL SOC supplemented with D-Ala (0.5 mM) and L-Arabinose (1 mM) at 37°C for 3.5 h. All the recovered cells were spread on LB agar plates supplemented with 0.5 mM D-Ala and 30 μg/mL kanamycin, which were then incubated overnight at 37°C.

After the incubation, 16 colonies were selected for further screening. The colonies were grown to saturation in 0.15 mL LB-kanamycin supplemented with 0.5 mM D-Ala, at 37°C for 6.5 h. A 2-μL aliquot from each saturated culture was streaked on two different agar plates: LB-kanamycin-D-Ala alone and LB-kanamycin-D-Ala plus 50 μg/mL ampicillin. All agar plates were incubated overnight at 42°C to promote the loss of the helper plasmid, pKD46, from the cells. At the end of the incubation, 15 clones (out of 16 screened) grew only on LB-kanamycin-D-Ala and none on LB-kanamycin-D-Ala-ampicillin, indicating that pKD46 had been lost from these 15 clones.

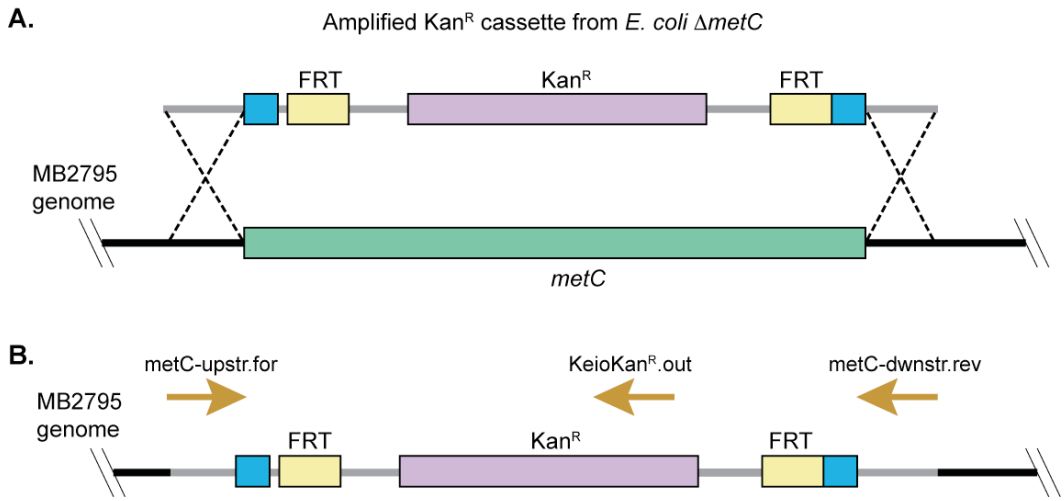


Figure 3.18: Schematic diagram of gene replacement mediated by λ Red recombination.

(A) The 1.4-kb Kan^R cassette flanked with FRT sites at both ends was amplified from *E. coli* Δ*metC* from the Keio collection using the metC-upstr.for and metC-dwnstr.rev primers (illustrated as brown arrows). The 54-nucleotide 5′ extension of the amplified product is homologous to the 5′ end of the genomic *metC* gene. Similarly, the 47-nucleotide 3′ extension of the product is homologous to the 3′ end of the genomic *metC* gene. The linear product was introduced to the host strain (MB2795), which expressed the λ Red recombinases from the helper plasmid (pKD46). The homologous ends of the linear product facilitated recombination (as shown by the double crossovers), resulting in the genomic *metC* gene being replaced by the linear product (Kan^R cassette). (B) In addition to a new selectable phenotype (kanamycin resistance), MB2795 cells with their *metC* gene replaced by the Kan^R cassette could be PCR-verified using the metC-upstr.for and metC-dwnstr.rev primers. In another similar PCR screen, the use of the metC-upstr.for and KeioKan^R.out primers (brown arrows) yielded a 750-bp product, confirming the presence of the Kan^R cassette in the *metC* locus.

3.4.2.3 VERIFICATION OF THE RECOMBINED KAN^R CASSETTE IN MB2795

Among those 15 clones that grew on LB-kanamycin-D-Ala agar plates, Clone 1 was grown overnight in 2 mL LB-kanamycin-D-Ala. A 300-μL aliquot of the saturated culture was then made as freezer stock, and kept at −80°C.

The genomic DNA of the remaining culture was extracted using the Wizard® Genomic DNA Purification Kit (Promega). The purified genomic DNA of Clone 1 was used as the template for PCR screens. The PCR mixture contained 1× Green GoTaq® Reaction buffer, 1 U i-Taq™ DNA polymerase, 0.2 mM dNTPs and 0.5 μM of each forward and reverse primer (metC-upstr.for and metC-dwnstr.rev). The thermocycling conditions were 95°C for 6 min, 40 cycles of 95°C (60 s), 51°C (60 s), 72°C (45 s), and finally 72°C for 2 min. The size of the amplified PCR product was ~1.4 kb according to agarose gel electrophoresis. After purifying the PCR product using commercial spin columns, the PCR product was sent for sequencing using three primers (metC-upstr.for, metC-dwnstr.rev and KeioKan^R.out). Sequencing results verified the presence of Kan^R cassette flanked by FRT sites at both ends.

3.4.3 GROWTH COMPLEMENTATION ASSAYS

All growth complementation assays were performed as previously described in Section 2.4.4 (of Chapter 2). Selective and non-selective agar plates were supplemented with appropriate antibiotics depending on the antibiotic resistance marker present in the plasmid and the *E. coli* genome. Specifically, cells carrying the pBAD plasmid were always grown in 50 μg/mL carbenicillin. *E. coli* Δ*alr* Δ*dadX* Δ*metC* was always grown in 30 μg/mL kanamycin. Thiamine (final concentration, 5 μg/mL) was added to the M9/glucose agar used for the growth complementation assays with *E. coli* Δ*alr* Δ*dadX* Δ*metC* clones.

3.4.4 ENZYME OVER-EXPRESSION, PURIFICATION & DIALYSIS

All enzyme variants in this study were purified according to the protocols described in Section 2.4.5 (of Chapter 2).

3.4.5 ENZYME QUANTIFICATION

All enzymes were quantified as previously described in Section 2.4.6 (of Chapter 2).

3.4.6 STEADY STATE KINETIC ASSAYS

Generally, all *in vitro* enzymatic activities were assayed as outlined in Section 2.4.7 (of Chapter 2).

The alanine racemization activity for all MetC variants was measured as described in Section 2.4.7.1 (of Chapter 2). The final concentration of each MetC variant ranged from 29–158 nM. The L-Ala concentrations typically ranged from ≤0.5 to ≥2 × K_M.

The cystathionine β-elimination activity for all evolved MetC variants was assayed as described in Section 2.4.7.2 (of Chapter 2). The final concentration of the evolved MetCs ranged from 6.5–20.2 nM. The L-cystathionine concentrations typically ranged from 0.1 to ≥1.6 × K_M.

3.4.7 Random mutagenesis of MetC (round 1)

3.4.7.1 Construction of first-generation MetC library

The GeneMorph II Random Mutagenesis kit (Stratagene) was used to introduce random mutations into the plasmid-encoded *metC* gene. The components in a 100-μL epPCR reaction were 1× Stratagene Mutazyme buffer, 2.5 U Mutazyme DNA polymerase, 0.8 mM dNTPs, 0.5 μM of pCA24N.for primer, 0.5 μM of MetC(GFP−).rev primer and 1.325 μg of plasmid template [pCA24N-*metC*(GFP−)]. The amount of plasmid template corresponded to 300 ng of initial target DNA (*metC*), which was recommended to achieve a mutation rate of 4.5–9 mutations per kb (Stratagene). The thermocycling conditions were 1 cycle of 95°C for 1.5 min, 30 cycles of 95°C (30 s), 54°C (30 s), 72°C (90 s), and a final cycle of 72°C for 1 min. The yield of the 1.3-kb PCR product was estimated to be 2 μg. Given that the amount of the initial DNA target was 300 ng, the final yield reflected a 6.67× amplification (and 2.74 duplications) of the initial DNA target. The PCR efficiency was calculated as 0.065 (Firth & Patrick, 2005).

After the PCR product was purified using spin columns, it was restricted with *Dpn*I for 2 h to digest any plasmid template that had been carried over from the purification step. Next, *Sfi*I was added to the reaction to produce 3′ sticky ends compatible with those of the *Sfi*I-linearized pCA24N(*KanR*)-tS1 vector (Section I.4 in Appendix I). After verifying the size of the digested *metC* gene insert (1.2 kb) and the linearized vector (5.0 kb) on an agarose gel, the DNA bands containing the gene insert and vector were excised and purified using the QIAEXII Gel Extraction kit (Qiagen).

The purified *metC* insert was ligated to the linearized pCA24N(*KanR*) vector at a 3 : 1 molar ratio. A total of 290 ng of *metC* insert and 400 ng of vector were added to the ligation reaction. A "vector only" control was also set up using the same components but without the *metC* insert. After 16 hours of incubation, the reactions were heat-inactivated and cleaned up using MinElute columns (Qiagen). Due to the low transformation efficiency of *E. coli* MB2795 (typically <10^8 cfu/μg of pUC19), the ligated products were first electroporated

into a fresh batch of *E. coli* DH5α-E (whose transformation efficiency was routinely higher than that of *E. coli* MB2795).

A 50-μL aliquot of *E. coli* DH5α-E was transformed with 1 μL of the ligated product by electroporation. A total of 22 electroporations were carried out, and the time constant for each electroporation was ≥4.50 ms. Cells were also transformed with the "vector only" control and pUC19 (to determine transformation efficiency of the cells). Each aliquot of electroporated cells was recovered in 0.5 mL of SOC at 37°C for 1 h. The recovered cells were plated on LB-kanamycin agar plates (24.5 cm × 24.5 cm), which were then incubated at 37°C for 16 h.

Based on the colony counts, the transformation efficiency of the DH5α-E cells was 1.6×10^8 cfu/μg of pUC19. After subtracting the "vector only" background (0.36%), the effective MetC library size was 1.5×10^5 cfu. Sixteen colonies were randomly picked from the library and PCR-screened using the vector-specific primers, pCA24N.for and MetC(GFP−).rev. The PCR products were purified and sent for sequencing. All sequences were used for estimating the mutation rate and the overall type of mutations present in the library.

The remaining colonies were washed off the LB-kanamycin plates, and pooled in ~45 mL of LB-kanamycin. The cell suspension was concentrated by centrifuging at 2,500 ×g for 20 min, at 4°C. After discarding the supernatant, the cell pellet was resuspended in 7 mL LB-kanamycin. One mL of the cell suspension was stored at −80°C for posterity, whereas the remaining cell suspension was evenly divided into 10 aliquots for plasmid extraction [using the QIAprep Spin Miniprep kit (Qiagen)]. The plasmids from each aliquot were eluted in 30 μL of Elution Buffer. All plasmid aliquots were pooled to form a 300 μL stock, at a concentration of 340 ng/μL. The pooled plasmids were used to transform fresh *E. coli* MB2795 cells.

For electroporation, a 20-fold dilution of the pooled plasmid was prepared. A 1-μL aliquot of the plasmid (~17 ng) was used to electroporate 50 μL of *E. coli* MB2795. A total of 10 electroporations were carried out and the time constant for every electroporation exceeded 5.20 ms. An aliquot of the cells was transformed with pUC19 to estimate transformation efficiency. The electroporated cells were recovered in 0.5 mL SOC-D-Ala at 37°C for 1 h. Recovered cells were grown on LB-kanamycin agar plates supplemented with 0.5 mM D-Ala at 37°C for 16 h.

The transformation efficiency of *E. coli* MB2795 was 6.2×10^7 cfu/μg of pUC19. The total number of transformed clones was 3.0×10^7. The size of the library should theoretically ensure that 100% of the initial library (1.5×10^5 cfu) was sampled. Using a spreader glass rod, all the colonies were washed off the library agar plates and pooled in 20 mL LB-kanamycin-D-Ala. The cell suspension was further concentrated by pelleting at 2,500 ×g for

20 min, at 4°C. The cell pellet was finally resuspended in 5 mL LB-kanamycin-D-Ala supplemented with 12.5% (v/v) glycerol. The OD_{600} of the cell suspension was 350, which corresponded to ~8.8×10^{10} cfu/mL. The cell suspension was evenly divided into 25 aliquots, which were stored at −80°C.

3.4.7.2 Selection for MetC variants with higher racemase activity

A 10-μL aliquot of the thawed *E. coli* M2795 cells harbouring the MetC epPCR library was used to inoculate 50 mL LB-kanamycin supplemented with 0.5 mM D-Ala. The culture was incubated at 37°C until an OD_{600} of 0.4 was reached. After diluting the culture 25-fold using LB, a 350-μL aliquot of the diluted culture was spread on the selective agar plate (24.5 cm × 24.5 cm), LB-kanamycin supplemented with 5 μM IPTG. Dilutions of the cultures were also spread on the non-selective agar plates, LB-kanamycin supplemented with 0.5 mM D-Ala. The selective agar plate was incubated at 28°C, whereas the non-selective agar plates were incubated overnight at 37°C.

It was estimated that a total of 1.0×10^6 cfu was plated on the selective medium based on the colony counts of the non-selective medium. After 42 h of incubation, the library plate was divided into 4 equal sections. Twelve colonies were randomly picked from each section, and grown to saturation in 0.2 mL LB-kanamycin-D-Ala. The cultures were stored at −80°C for restreaking tests (Section 3.4.7.3).

3.4.7.3 Restreaking tests

The 48 clones isolated from the selection plate (Section 3.4.7.2) were individually restreaked on new selective plate to confirm the growth phenotype of the clones.

The freezer stock of the 48 clones was used to inoculate 0.2 mL LB-kanamycin supplemented with 0.5 mM D-Ala. After an overnight incubation at 37°C, the saturated cultures were pelleted at 12,000 ×g for 3 min. Cell pellets were resuspended and diluted to OD_{600} ~0.1 using LB. A 3-μL aliquot of each cell suspension was streaked on two sets of agar, LB-kanamycin with and without 5 μM IPTG. All agar plates were incubated at 28°C for at least 1 week.

To sequence the plasmid-encoded *metC* gene of the selected clones, a 1-μL aliquot of the saturated culture from the selected clones was used as the template for PCR screen. Vector-specific primers, pCA24N.for and pCA24N.rev2, were used to amplify the plasmid-encoded *metC* gene. The resulting 1.4-kb PCR product was purified using spin columns, and sent for DNA sequencing. The following primers were used for sequencing the *metC* gene: pCA24N.for (N-terminus), pCA24N.rev2 (C-terminus) and metC877.rev (middle region). When no mutation was found in the *metC* gene, the Rev+270 and Univ-70 primers were used to sequence the $lacI^q$ repressor of the vector. The pUCrev+120 primer was used to

sequence the promoter region and the $(His)_6$ tag located upstream of the *metC* gene. All primer sequences are listed in Section I.6 of Appendix I.

3.4.8 RANDOM MUTAGENESIS OF METC (ROUND 2)

3.4.8.1 SUBCLONING *metC* INSERTS INTO pBAD VECTOR

The wildtype *metC* gene and the four mutated variants selected from the first-generation library were subcloned into pBAD/*Myc-His-B* (Invitrogen) to test the expression of MetC under the PBAD promoter of the *araBAD* operon. The plan was to clone the *metC* gene inserts into the *Nco*I and *Xba*I sites of the pBAD plasmid. However, since there are multiple *Nco*I recognition sites in the *metC* gene, *Pci*I was used to produce 3′ ends, which are compatible with those produced by *Nco*I on the 5′ end of the vector backbone.

All pCA24N-encoded *metC* gene inserts (wildtype, P113Q, P113A, P113T and P113S) were PCR-amplified using the primers, His-metC.for and His-metC.rev. The amplified product also included a $(His)_6$ tag fused to the N-terminus of the *metC* insert via a 15-amino acid linker sequence (MSHHHHHHTDPALRA), which was produced after the MRG tripeptide was removed from the original 17-amino acid linker (MRGSHHHHHHTDPALRA). The first methionine codon was reintroduced into the linker sequence via the *Pci*I recognition site (ACATGT) in the His-metC.for sequence. For the His-metC.rev sequence, an *Xba*I recognition site was engineered on the C-terminus of the primer.

The *E. coli* clones harbouring the pCA24N-encoded *metC* variants were grown to saturation. A 1-μL aliquot of each saturated culture was added to a PCR reaction containing the 1× Phusion HF buffer, 1 U Phusion® High-Fidelity DNA polymerase (Finnzymes), 0.8 mM dNTPs, 0.5 μM of His-metC.for, 0.5 μM of His-metC.rev. The thermocycling conditions were 1 cycle of 98°C for 1 min, 35 cycles of 98°C (15 s), 59°C (20 s), 72°C (30 s), and a final cycle of 72°C for 5 min. The 1.2-kb PCR products were cleaned up using commercial spin columns.

Next, the purified *metC* gene inserts were digested with *Pci*I and *Xba*I for 7 h. The pBAD vector was treated with *Nco*I and *Xba*I for the same duration to generate sticky ends compatible with those of the *metC* gene inserts. After verifying the size of the DNA bands on an agarose gel, the 1.2-kb gene inserts and the 4.3-kb vector were gel-purified using MinElute columns. Each *metC* variant was then ligated to the linearized pBAD vector in a molar ratio of 3 : 1. The MinElute columns were used to clean up the ligation reactions. Two μL of the ligated products were used electroporate a 50-μL aliquot of *E.coli* MB2795. Recovered cells were grown on LB agar plates supplemented with 50 μg/mL carbenicillin and 0.5 mM D-Ala. Transformed colonies were PCR-screened for the presence of pBAD-

encoded *metC* using vector-specific primers, pBAD-60 and pBAD.rev. The PCR products were purified and sequenced to confirm that no extra mutation had been introduced during the subcloning process.

3.4.8.2 Construction of second-generation MetC libraries

Similar to the first-generation library, random mutations were introduced to the *metC* gene using the GeneMorph II Random Mutagenesis kit. The pCA24N(*KanR*)-encoded *metC* variants (P113Q, P113A, P113T and P113S) selected from the first-generation library were used as DNA templates for constructing four separate second-generation libraries.

A 50-μL epPCR reaction was set up for each library. An epPCR reaction consisted of 1× Stratagene Mutazyme buffer, 2.5 U Mutazyme DNA polymerase, 0.8 mM dNTPs, 0.5 μM of His-metC.for primer, 0.5 μM of His-metC.rev primer and 1.5 μg of plasmid template. This amount of plasmid template corresponded to 300 ng of initial target DNA (*metC*). The thermocycling protocols were 1 cycle of 95°C for 1.5 min, 30 cycles of 95°C (30 s), 55°C (30 s), 72°C (75 s), and a final cycle of 72°C for 1 min. The yield of the 1.2-kb PCR product was estimated to be 1 μg for all four reactions. Given that 300 ng of the initial DNA template was used, the final yield reflected a 3.33× amplification (and 1.74 duplications) of the initial DNA target. The PCR efficiency was calculated as 0.041 (Firth & Patrick, 2005).

After cleaning up the PCR reactions using spin columns, the PCR product was digested with *Dpn*I for 2 h to eliminate the DNA template that had been carried over from the purification step. To generate the sticky ends for cloning into the pBAD vector, *Pci*I and *Xba*I were added to the reaction, which was further incubated for another 6 h. At the same time, the pBAD/*Myc-His-B* vector was digested with *Nco*I and *Xba*I to linearize the vector. The QIAEXII Gel Extraction kit was used to purify the digested 1.2-kb *metC* insert and the 4.3-kb linearized vector. Approximately 280 ng of the purified insert was ligated to 340 ng of the linearized pBAD vector at a molar ratio of 3 : 1. A ligation control was also set up using the same ligation components but without the *metC* insert. After 16 h of incubation, all ligation reactions were heat-inactivated, and cleaned up with MinElute columns.

Similar to the first-generation library, the ligated products were first electroporated into *E. coli* DH5α-E. A 1-μL aliquot of the ligated product was used to electroporate a 50-μL aliquot of *E. coli* DH5α-E. A total of 11 electroporations were performed for each of the four libraries. The time constants for all the electroporations exceeded 5.0 ms. After recovering the electroporated cells in 0.5 mL of SOC at 37°C for 1 h, the cells were spread and incubated overnight on LB-carbenicillin library plates (24.5 cm × 24.5 cm). Based on the number of colonies grown on the agar plates, the library size could be estimated.

Twelve to seventeen colonies were randomly selected from each library and PCR-screened using the vector-specific primers, pBAD-60 and pBAD.rev. The presence of a 1.3-

kb PCR product indicated the presence of the pBAD-encoded *metC* gene. The PCR products were purified and sent for sequencing to allow determination of the mutation rate and bias of the libraries.

The remaining colonies on each library plate were separately pooled in 50 mL LB-carbenicillin. The cell suspensions were pelleted at 2,300 ×g for 20 mins, at 4°C. Each cell pellet was resuspended thoroughly in 10 mL LB-carbenicillin. A 400-μL aliquot from each library suspension was stored at −80°C for posterity. For each library, the remaining suspension was evenly divided to 10 aliquots for plasmid extraction (using QIAprep Spin Miniprep kit). After eluting the plasmids from the spin columns, a 300-μL plasmid stock was obtained for each library. The plasmid concentration for the MetC(P113Q), MetC(P113A), MetC(P113T) and MetC(P113S) were 43 ng/μL, 77 ng/μL, 77 ng/μL and 102 ng/μL, respectively.

Next, the pooled plasmids were introduced to *E. coli* MB2795 for function selection. A 50-μL aliquot of fresh *E. coli* MB2795 was transformed with 1 μL of the pooled plasmids. A total of 14 electroporations were performed for the MetC(P113Q) library, 13 electroporations for MetC(P113A) library, 10 electroporations for the MetC(P113T) library and 12 electroporations for the MetC(P113S) library. The time constant for all electroporations were >5.30 ms. All electroporated cells were recovered in 0.5 mL SOC-D-Ala at 37°C for 1 h. Recovered cells were spread on LB-carbenicillin library plates (24.5 cm × 24.5 cm) supplemented with 0.5 mM D-Ala, and grown at 37°C. After 16 h of incubation, colony counts were obtained for every library plate.

The next day, all colonies were scrapped off each library plate and pooled into 40 mL using LB-carbenicillin-D-Ala. The cell suspension of each library was concentrated by centrifuging at 2,500 ×g for 30 min, at 4°C. After decanting the supernatant, each cell pellet was resuspended in 5 mL LB-carbenicillin-D-Ala supplemented with 12.5% (v/v) glycerol. The OD_{600} readings of the cell suspensions ranged from 190–300, which corresponded to approximately 4.8×10^{10} to 7.5×10^{10} cfu/mL. Each cell suspension was divided into 10 aliquots, and stored at −80°C.

3.4.8.3 Selection for MetC variants with higher racemase activity

For every second-generation library [MetC(P113Q), MetC(P113A), MetC(P113T) and MetC(P113S)], a 10-μL aliquot of the thawed *E. coli* M2795 harbouring the MetC epPCR library was used to inoculate 50 mL LB-carbenicillin supplemented with 0.5 mM D-Ala. The culture was grown to OD_{600} of 0.4, at 37°C. The culture was diluted 25-fold using LB, followed by spreading an appropriate volume of the diluted culture (~0.8–2.0 mL) on the selective agar plate (24.5 cm × 24.5 cm), LB-carbenicillin supplemented with arabinose to induce MetC expression. The arabinose concentrations used in the selection processes ranged from 0–0.8 mM (see Section 3.2.5.2 and 3.2.5.3 for details). Dilutions of the cultures

were also spread on the non-selective agar plates, LB-carbenicillin supplemented with 0.5 mM D-Ala. The selective agar plate was incubated at 28°C, whereas the non-selective agar plates were incubated overnight at 37°C.

At the end of selection, each library plate was divided into 4 equal sections. Six colonies were randomly picked from each section, and used to inoculate 0.2 mL LB-carbenicillin-D-Ala. The cultures were grown to saturation, and stored at −80°C for posterity. A 1-μL aliquot of the saturated cultures was used for PCR screen to amplify the plasmid encoded *metC* gene. The vector-specific primers, pBAD-60 and pBAD.rev, were used. The resulting PCR products were then purified and sent for sequencing.

3.4.8.4 RETRANSFORMATION/RECLONING AND RESTREAKING TESTS

The plasmids from selected clones were extracted using commercial spin columns. The purified plasmids were used to transform a new batch of electrocompetent *E. coli* MB2795. The retransformed colonies were grown to saturation in 2 mL LB-carbenicillin-D-Ala at 37°C. The saturated cultures were always PCR-screened to confirm the presence of the plasmid-encoded *metC* before being made as freezer stocks.

For several clones with genotypes similar to their first-generation parental clone (*i.e.*, no extra mutation was introduced to the *metC* gene), their plasmid-encoded *metC* insert was amplified using the His-metC.for and His-metC.rev primers. The amplified insert was digested with *Pci*I and *Xba*I to generate 3′ ends compatible with those of pBAD vector, which were produced by *Nco*I/*Xba*I digestion. The restricted insert and vector was ligated, and the ligated plasmid was electroporated into fresh *E. coli* MB2795. One retransformed colony was grown to saturation in 1 mL LB-carbenicillin-D-Ala at 37°C. The saturated culture was screened for the presence of the plasmid using PCR before being used for restreaking test.

For each restreaking test, the freezer stock of the selected clones was used to inoculate 1 mL LB-carbenicillin supplemented with 0.5 mM D-Ala. After an overnight incubation at 37°C, the saturated culture was pelleted at 12,000 ×g for 4 min. The cell pellet was washed twice with 1 mL LB, with a similar pelleting step (12,000 ×g for 4 min) after each wash. The cell pellet was finally resuspended in 1 mL LB. The cell resuspension was further diluted 200-fold dilution using LB. A 2.5-μL aliquot of the cell dilution was streaked on two sets of agar, LB-carbenicillin with and without 0.1 mM arabinose. All agar plates were incubated at 28°C for at least 1 week.

CHAPTER 4

Evolution of new phenotypes mediated by protein over-expression

Sections 4.1, 4.2 and 4.4 contain published materials from:

> **Soo, V. W. C., Hanson-Manful, P. & Patrick, W. M.** (2011). Artificial gene amplification reveals an abundance of promiscuous resistance determinants in *Escherichia coli*. *Proc Natl Acad Sci USA,* **108**:1484-1489.

Acknowledgement: Ilana Gerber performed half of the PCR screens, and purified all of the PCR products, described in Section 4.4.4.

4.1 INTRODUCTION

In Chapters 2 and 3, I used multicopy suppression experiments to identify promiscuous proteins (Alr and MetC) that could rescue conditionally auxotrophic strains of *E. coli*. This approach is a tractable — and therefore popular — way of demonstrating the connection between over-expression mutations, protein promiscuity and bacterial adaptation (Berg *et al.*, 1988; Miller & Raines, 2004; Patrick *et al.*, 2007; Christ & Chin, 2008; Morett *et al.*, 2008; Patrick & Matsumura, 2008; Kim *et al.*, 2010).

Two general conclusions can be drawn from these multicopy suppression experiments: (i) some fraction of proteins in modern proteomes are multifunctional and/or display broad substrate specificities; and (ii) their secondary functions can maintain metabolite fluxes *in vivo*, particularly when these fluxes are disrupted by deletion of metabolic enzymes (D'Ari & Casadesús, 1998; Kim *et al.*, 2007). Utilization of a few novel nutrients (*e.g.* xylitol and amides) in bacteria has also been attributed to induced expression of proteins with secondary activities (Betz *et al.*, 1974; Mortlock, 1984). However, at the outset of this study, the ability of proteome-wide promiscuity to drive an organism's adaptation to new environments remained unknown.

Therefore, in this chapter, I attempted to survey the entire *E. coli* proteome for its latent ability to confer genuinely new phenotypes (rather than to re-create old ones). The aim was to gauge the extent to which a simple adaptive response — over-expression of a pre-existing protein — could impart new phenotypes on *E. coli*, when the cells were exposed to novel nutrients. In contrast to Chapters 2 and 3 where single examples of promiscuous enzymes were examined, the global search for promiscuous proteins in this chapter was designed to offer insights into the evolutionary potential of a contemporary proteome.

4.1.1 THE GENETICS OF ADAPTIVE EVOLUTION

All species must adapt to survive in changing environments. At the molecular level, the presence of novel nutrients and toxins can drive the evolution of proteins that recognize and/or metabolize them (Hegeman & Rosenberg, 1970; Mortlock, 1984). Duplicated copies of pre-existing genes provide the primary genetic source for functional innovation (Lewis, 1951; Ohno, 1970). However, the relationship between duplication and the emergence of new biochemical functions is complicated because gene duplications are generally expected to be either selectively neutral (Lynch & Conery, 2003) or deleterious (Wagner, 2005).

Thus, numerous hypotheses have been proposed to explain the fates of duplicated genes within populations (Roth *et al.*, 2007; Conant & Wolfe, 2008; Innan & Kondrashov, 2010). A particularly appealing model identifies a mechanism by which selection can act continuously to favour both an increase in gene dosage, and divergence of one copy from the parental gene. This model, termed adaptive radiation (Francino, 2005) or the Innovation,

Amplification and Divergence (IAD) model (Bergthorsson *et al.*, 2007) is rooted in the notion that many existing proteins display broad specificity and/or secondary activities, in addition to the function that they evolved to carry out (Khersonsky & Tawfik, 2010). In primordial times, this 'promiscuity' would have allowed many metabolic functions to be carried out by a minimal number of multi-tasking proteins (Jensen, 1976). In more specialized, modern day proteins, promiscuity is likely to be a product of contingency: active sites typically contain a variety of reactive groups (proton donors and acceptors, metal ions, *etc.*) that could (by chance) play roles in secondary reactions (O'Brien & Herschlag, 1999). These secondary activities are assumed to provide the necessary reservoir of novel and selectable functions.

In the IAD model, selection is imposed on a weak, promiscuous activity. In order to increase fitness, duplications (and higher-order amplifications) of the promiscuous gene are selected and maintained in the population. In turn, this increase in copy number improves the likelihood of point mutations that enhance the promiscuous activity, by increasing the number of mutational targets while allowing at least one copy to retain the parental activity (Bergthorsson *et al.*, 2007).

Two central tenets of the IAD model are: (i) that gene amplification events are common; and (ii) that promiscuous activities are widespread, as well as being biochemically and mechanistically diverse. Experiments with unselected bacterial populations have shown the first point to be accurate (Anderson & Roth, 1981), with cells that bear duplications at any given locus reaching steady-state frequencies of ~10^{-3} (Reams *et al.*, 2010). In this chapter, I aimed to catalogue the proteome-wide promiscuity that could allow the evolution of new phenotypes, thereby addressing the second tenet and connecting protein biochemistry to the genetics of the IAD model.

4.1.2 The evolutionary potential of a modern proteome

In the present study, the expression levels of *E. coli* proteins were increased artificially in an attempt to mimic the effects of the amplified gene copies, as described in the IAD model (Bergthorsson *et al.*, 2007). After transforming the ASKA library into a common, laboratory *E. coli* strain, the plasmid-encoded protein in each clone was over-expressed. This *E. coli* pool with over-expressed proteins was then introduced into ~1,000 unique environments, where each contained a specific nutrient that needed to be either metabolized or tolerated if cells were to thrive. I demonstrated that *E. coli* could adapt to ~35% of these environments just by over-expressing one of its native proteins. Overall, the results imply that protein over-expression is a feasible biochemical route to cellular adaptation in novel environments.

4.2 RESULTS

4.2.1 LIBRARY-ON-LIBRARY SCREEN (PM 1 TO PM 10)

To search for promiscuous proteins that can impart new phenotypes on *E. coli*, two tools from functional genomics were utilized: the ASKA library and Biolog Phenotype Microarray (PM) plates (Bochner *et al.*, 2001; Bochner, 2009). The PM plates are 96-well assays that were designed to test metabolic capabilities (Bochner *et al.*, 2001), such as cellular utilization of carbon (PM 1 and PM 2), nitrogen and peptides (PM 3, PM 6 to PM 8), phosphorus and sulfur (PM 4), organic supplements (PM 5), as well as cell sensitivity to osmolytes and pH (PM 9, PM 10). However, in this study, I considered each well of the PM plates as an evolutionary scenario, where cells in both the test and control populations must either grow in the presence of a foreign nutrient, or face death. Every well contains a dehydrated base medium and a specific nutrient source. The base medium provides all the ingredients essential for cellular growth, except the assayed metabolic source. For instance, there is a different carbon source in every well of PM 1. A tetrazolium dye was also added to each well to monitor cell growth. When the cells grow, electrons generated from cellular respiration reduce the tetrazolium dye, which in turn leads to a colour change from colourless to purple in the well (Bochner, 2009).

First, the pooled plasmids of the ASKA library were used to transform *E. coli* strain DH5α-E. IPTG was added to this ASKA pool to induce over-expression (Figure 4.1A). The cells of the ASKA pool (each over-expressing a single *E. coli* ORF) were then used to inoculate every well of PM 1 to PM 10. The inoculum size was approximately 50,000 cells in each well to ensure that the entire ASKA pool of 5,272 clones was represented (an average of 9–10 copies of each ORF). Another set of PM plates was inoculated with *E. coli* DH5α-E containing the empty ASKA vector, pCA24N-NoIns (see Section 2.4.1). Both sets of inoculated PM plates were incubated in parallel at 37°C for 1 week (Figure 4.1B). The growth patterns between the ASKA and control clones were compared daily according to the tetrazolium colour intensities that had developed in the wells (Figure 4.1C). Two independent replicates of the entire screen were performed, and only wells that had been inoculated with the ASKA pool, and had showed darker shades of purple both times, were considered for subsequent steps (see Section 4.2.2).

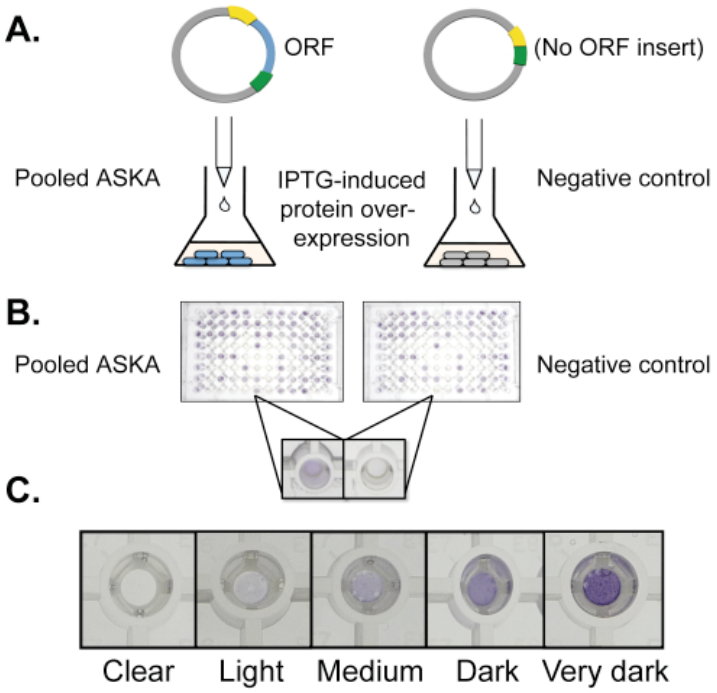


Figure 4.1: The library-on-library screen.

(A) *E. coli* cells harbouring the pooled plasmids of the ASKA ORF collection (blue) and the negative control clone (grey) were grown to mid-log phase in parallel, and protein over-expression was induced by the addition of IPTG (50 μM). (B) Both cultures were used to inoculate every well of two sets of PM plates. A well was scored positive when the ASKA pool out-grew the negative control, as shown by more rapid tetrazolium colour development (Inset). (C) Examples of different tetrazolium colour intensities after incubation.

In two independent screens, the ASKA pool out-grew the negative control in 380 of the 960 conditions screened (40%). The increased growth rates in the ASKA pool were noted in every type of environment: carbon (14% of wells); nitrogen and peptides (51%); phosphorus and sulfur (61%); biosynthetic supplements (90%); osmolytes and pH stresses (7%). The improved growth phenotypes of the ASKA pool (relative to the negative control) in the presence of various nutrient sources demonstrated the cells' flexibility in metabolizing additional nutrients in defined, minimal environments. In comparison to the nutrient sources, the ASKA pool did poorly in conditions with different osmolarities or pH, suggesting that *E. coli* cells have limited metabolic capacity in tolerating changes in salt content and acidity of their environments. A more detailed analysis of each type of environment will be further described in Sections 4.2.3 to 4.2.7.

4.2.2 Enrichment of the fittest clone(s)

In wells where the ASKA pool reproducibly out-grew the negative control, serial enrichment was employed to isolate the individual clones (out of 5,272 ASKA clones) that

were responsible for the overall improved growth phenotypes. The basis for the enrichment approach was that the ASKA pool in these selective environments would be dominated by populations of faster-growing clones, which harboured at least one type of ASKA ORF. As soon as the first sign of growth was detected, a small fraction of the ASKA pool was passaged into the same well of a new plate. Each passage represented a 50-fold dilution of the ASKA population in the well. This passage was repeated once in an identical fashion, prior to plating the clones on non-selective agar plates (Figure 4.2A). Colonies were screened using PCR with pCA24N-specific primers, and the amplified products were sequenced to determine the ASKA ORFs that were responsible for the enhanced phenotypes observed during the initial screens. Eight clones were screened from every enriched population. I considered an ASKA ORF as a positive hit only if it was isolated at least twice from the PCR screen (Figure 4.2B). By chance alone, it was highly unlikely that any two ORFs would be identical ($P \approx 0.005$; refer to Section 4.4.4.1 for more details), unless a particular ORF had been positively enriched during the serial transfer.

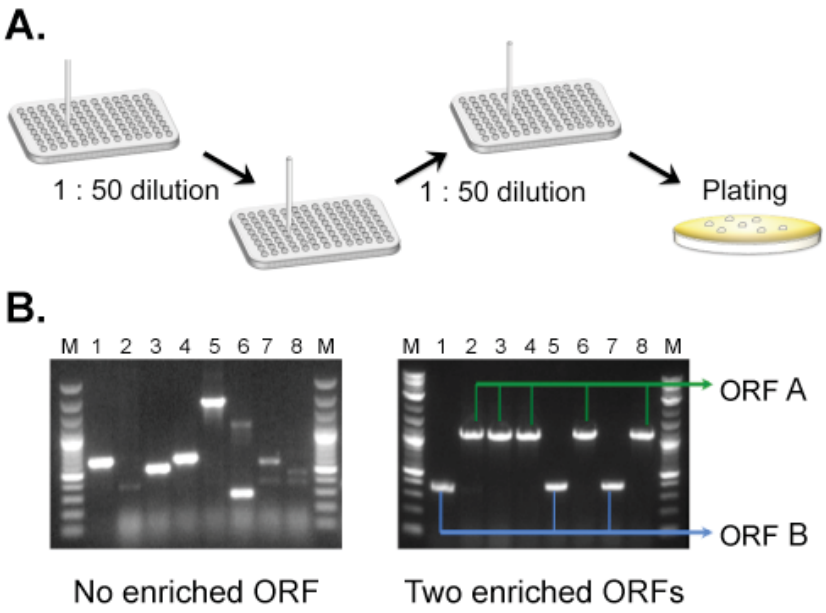


Figure 4.2: Isolation of ASKA-encoded ORFs via a serial enrichment approach.

(A) To enrich the populations of fitter clones, fresh aliquots of the ASKA pool (50,000 cells) were inoculated into a new set of PM plates. As soon as growth was detected (as indicated by light purple in the well), a 2-μL aliquot was transferred to the corresponding well of a new plate. This transfer was repeated once more before an aliquot of the enriched culture was plated on non-selective medium. (B) Eight random colonies that formed on the non-selective medium were screened using PCR with plasmid-specific primers. An ORF was considered as a positive hit only if it was isolated at least twice from a population (*i.e.*, the amplified ASKA ORF showed up twice in the agarose gel). In the left gel image, no ORF was considered as a positive hit since no two PCR products were of identical sizes (Lanes 1 to 8). On the other hand, two enriched ORFs were apparent in the right gel image. PCR products from Lanes 1, 5 and 7 represented one ORF, whereas PCR products from Lanes 2 to 4, 6 and 8 represented another ORF. Lanes M, DNA ladders from Fermentas (left gel image) and New England Biolabs (right gel image).

Using the serial enrichment approach, a total of 402 ASKA-encoded ORFs that were responsible for enhancing growth in 332 PM wells were isolated (Table 4.1 to Table 4.6). For a majority of the environments, only one ASKA ORF was identified. At maximum, three different ASKA ORFs were isolated from a single environment. The identities of the ASKA ORFs are grouped according to the types of environments from which they were isolated, and will be described in Section 4.2.3 to Section 4.2.7. The recurrence of several ASKA ORFs enriched in many different environments will also be discussed in subsequent sections. The EcoCyc (Keseler *et al.*, 2009) and UniProt (The UniProt Consortium, 2010) databases were used to infer the functions of all isolated ASKA-encoded ORFs.

4.2.3 Carbon sources

4.2.3.1 ASKA-encoded ORFs isolated from PM 1 & PM 2

The growth of the ASKA pool was compared to the negative control in a total of 192 wells, which contain a rich base medium and a unique carbon source in each well. According to Zhou *et al*. (2003), the base medium of PM 1 and PM 2 consists of 2.0 g/L of tryptone, 1.0 g/L of yeast extract and 1.0 g/L of NaCl. This medium is a 5-fold dilution of LB-Lennox (Sambrook & Russell, 2001).

The ASKA pool displayed an enhanced growth phenotype in 27 of these environments. A total of 22 ASKA-encoded ORFs were isolated from 20 carbon-containing wells (Table 4.1). Hence, the hit rate (*i.e.*, the percentage of environments where at least one ASKA-encoded ORF was enriched) for carbon sources was (20/192) = 10%. There were no wells in which the ASKA pool grew more poorly than the negative control.

Two ORFs were enriched from multiple wells containing different carbon sources: *ycbS* (6 times) and *agaX* (6 times). AgaX (a domain of an uncharacterized carbohydrate transport system) and YcbS (an uncharacterized outer membrane protein) appeared to be involved in the transport of a variety of sugar derivatives (Table 4.1). Another recurring ORF was *bglA*, which was isolated from two different wells (Table 4.1). *E. coli* BglA is a β-glucosidase, which will be discussed further in Section 4.2.3.2. The remaining eight ORFs were isolated once (Table 4.1).

Along with AgaX and YcbS, other ORFs that code for membrane proteins were *lacY* (a lactose permease) and *yedZ* (an uncharacterized membrane protein), which were simultaneously enriched from the same carbon source (D-Mannitol). In two cases (α-D-lactose and caproic acid), the isolated ORFs (*ybcD* and *ymfL*) code for uncharacterized, prophage-derived proteins (Table 4.1). The relationship between the expression of prophage-related proteins and carbon utilization in bacteria has not been investigated,

although activation of prophage genes in *E. coli* has been shown to mediate non-specific cellular responses against environmental stress (Wang *et al.*, 2009) [also see Chapter 5].

Table 4.1: The complete list of over-expressed ORFs that conferred enhanced growth in 27 carbon-containing environments.

Carbon source	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
PM 1						
D-Mannitol	1	Dark	Medium	*lacY*	2	Lactose permease
				yedZ	2	Uncharacterized inner membrane protein
D-Ribose	1	Medium	Light	*ycbS*	8	Uncharacterized outer membrane β-barrel protein
Tween 20	4	Medium	Light	*agaX*	8	Uncharacterized domain from a phosphotransferase system
L-Rhamnose	1	Medium	Light	*agaX*	8	Uncharacterized domain from a phosphotransferase system
Acetic acid	1	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
Thymidine	1	Medium	Light	*agaX*	4	Uncharacterized domain from a phosphotransferase system
Tween 40	5	Light	Clear	*yjcF*	2	Uncharacterized protein
α-D-Lactose	2	Light	Clear	*ybcD*	3	Pseudogene from DLP12 prophage; predicted fragment of replication protein
Lactulose	3	Medium	Clear	*ydjO*	2	Uncharacterized protein
Uridine	1	Light	Clear	*agaX*	8	Uncharacterized domain from a phosphotransferase system
m-Tartaric acid	3	Medium	Light	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
β-methyl-D-glucoside	1	Light	Clear	*bglA*	4	6-phospho-β-glucosidase A
				ycbS	4	Uncharacterized outer membrane β-barrel protein
Glycyl-L-aspartic acid	3	Medium	Light	*ycbS*	2	Uncharacterized outer membrane β-barrel protein

Carbon source	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
Bromosuccinic acid	2	Light	Clear	None isolated	N/A	
Glycolic acid	3	Light	Clear	*ycbS*	5	Uncharacterized outer membrane β-barrel protein
Glyoxylic acid	2	Light	Clear	*agaX*	8	Uncharacterized domain from a phosphotransferase system
N-acetyl-β-D-mannosamine	2	Medium	Light	*agaX*	2	Uncharacterized domain from a phosphotransferase system
Mono-methyl-succinate	2	Light	Clear	None isolated	N/A	
D-Malic acid	1	Light	Clear	*wbbL_1*	2	Putative lipopolysaccharide biosynthesis glycosyltransferase
L-Lyxose	3	Medium	Clear	None isolated	N/A	
Glucuronamide	3	Medium	Light	*bglA*	6	6-phospho-β-glucosidase A
L-Galactonic acid-γ-lactone	2	Very dark	Dark	None isolated	N/A	
PM 2						
3-O-β-D-galactopyranosyl-D-arabinose	3	Light	Clear	*cpdA*	3	cAMP phosphodiesterase
Caproic acid	5	Light	Clear	*ymfL*	5	Uncharacterized transcriptional regulator from e14 prophage
β-Hydroxy Butyric acid	3	Light	Clear	None isolated	N/A	
5-Keto-D-Gluconic acid	3	Dark	Light	None isolated	N/A	
Glycine	4	Medium	Light	None isolated	N/A	

* Days of incubation before a change in growth phenotype (*i.e.,* tetrazolium colour intensity) was observed.

Key for scoring growth phenotype (*i.e.,* tetrazolium colour intensity):

Very dark Dark Medium Light ☐ Clear

4.2.3.2 GLUCURONAMIDE UTILIZATION IN *E. COLI*

Among the ASKA-encoded ORFs enriched from different carbon sources (Section 4.2.3.1), *bglA* is a particularly interesting case. BglA, a β-glucosidase, was thought to hydrolyze only phosphorylated β-glucosides (Prasad *et al.*, 1973; Hall *et al.*, 1983), such as 6-phospho-β-D-glucosyl-(1,4)-D-glucose (also known as cellobiose-6-phosphate) and arbutin-6-phosphate (Figure 4.3A). One product of these reactions is glucose-6-phosphate, which can be utilized as the carbon source for cellular metabolism. β-glucosides such as cellobiose and arbutin are phosphorylated during their uptake via a specific β-glucoside transporter, BglF (Kricker & Hall, 1984; Kricker & Hall, 1987). Point mutations and insertions of transposable IS elements at the regulatory regions activate the expression of BglF (and BglB, another β-glucosidase), which is otherwise cryptic under laboratory conditions (Prasad & Schaefler, 1974). In contrast, the expression of BglA is constitutive (Prasad *et al.*, 1973; Hall *et al.*, 1983).

The enrichment of the ASKA-encoded *bglA* from the well containing β-methyl-D-glucoside (Figure 4.3B; Table 4.1 in Section 4.2.3.1) also appeared to be reasonable. Unlike cellobiose and arbutin, phosphorylation of the aliphatic β-methyl-D-glucoside in *E. coli* is carried out by the glucose permease (PtsG), instead of the cryptic BglF (Schaefler, 1967; Prasad *et al.*, 1973). Hence, this phosphorylated form of β-methyl-D-glucoside is also recognized as a substrate for BglA. The product of hydrolysis of the phosphorylated β-methyl-D-glucoside is glucose.

However, *bglA* was also enriched in another well containing glucuronamide (Figure 4.3C; Table 4.1 in Section 4.2.3.1), implying that BglA may potentially recognize non-phosphorylated substrates. In contrast to β-glucosides, the C6 position of the glucuronamide is attached to an amide group, instead of a hydroxyl group. As a result, glucuronamide cannot be phosphorylated. Thus far, glucuronamide uptake and utilization in *E. coli* has not been described in the literature.

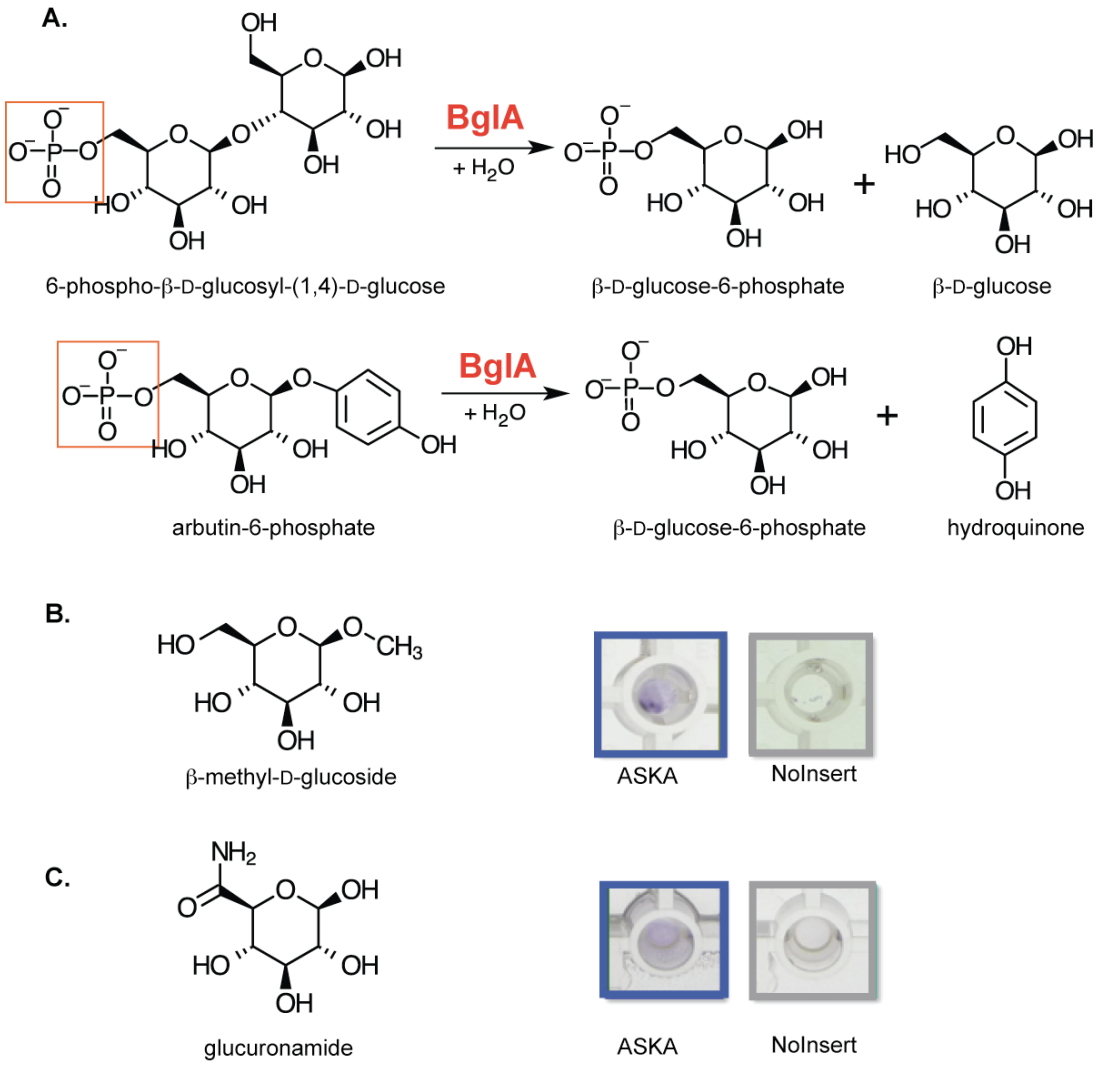


Figure 4.3: The structures of the previously reported, and alternative, substrates of BglA.

(A) Two reactions catalyzed by BglA. The enzyme was thought to recognize only phosphorylated glucosides. The phosphate groups are shown in red boxes. (B) The ASKA pool out-grew the negative control (NoInsert) in the well containing β-methyl-D-glucoside after 1 day of incubation. (C) Similarly, the ASKA pool also out-grew the negative control in the well containing glucuronamide after 3 days of incubation.

The growth phenotypes of *E. coli* in glucuronamide were re-tested, in order to confirm that: (i) *E. coli* could utilize glucuronamide as the sole carbon source; and (ii) cells over-expressing BglA could utilize glucuronamide better than the negative control. Wildtype *E. coli* strain K-12 (Section I.3 in Appendix I) was used for further testing, in order to eliminate the possibility that the growth phenotypes observed in the glucuronamide-containing well (during the initial screen) were specific to the DH5α-E strain. In the absence of pCA24N plasmid, *E. coli* K-12 was found to grow slowly in M9 media supplemented with 0.1% to 0.5% glucuronamide as the sole carbon source (Figure 4.4). Although the overall growth of *E. coli* K-12 was low (maximum OD_{600} <0.2) for all glucuronamide concentrations, the

growth rates were correlated with glucuronamide concentrations — the higher the glucuronamide concentration, the higher the *E. coli* growth rate. Growth saturation, as normally shown as a plateau in a growth curve, was not present after 24 h of incubation. In comparison, the K-12 strain grown in a good carbon source (glycerol) reached saturation (OD_{600} ~0.8) after 16 h of incubation.

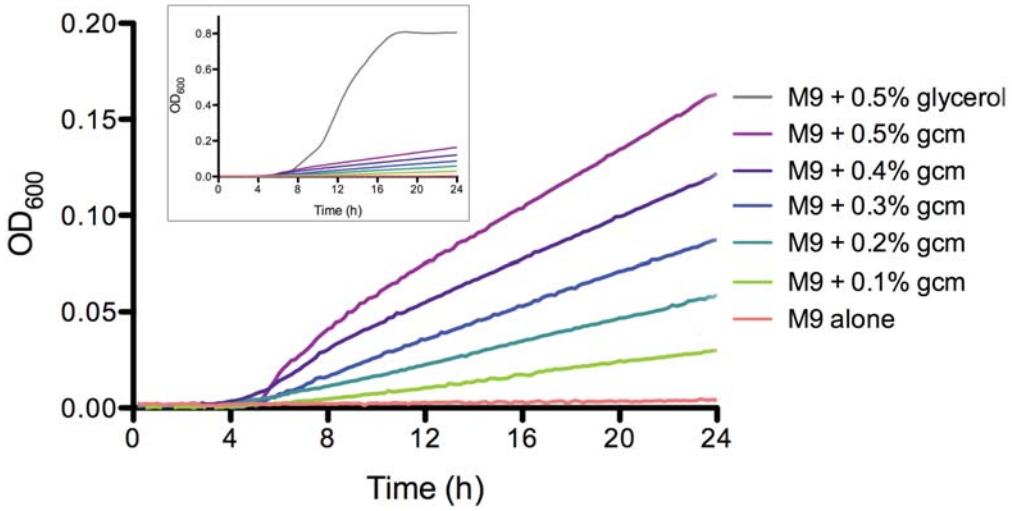


Figure 4.4: The growth of wildtype *E. coli* (strain K-12) in different glucuronamide concentrations.

E. coli K-12 was grown in M9 supplemented with various concentrations of glucuronamide, for 24 h. For comparative purposes, M9-glycerol was used as the positive control (inset), whereas the M9 medium alone served as the negative control. Each growth curve is a mean representative of eight replicates. Abbreviation: gcm, glucuronamide.

The growth curves shown in Figure 4.4 suggested that *E. coli* K-12 could utilize glucuronamide, albeit poorly. Next, the effect of over-expressing BglA on *E. coli* growth was examined. The pCA24N-encoded *bglA* was introduced into *E. coli* K-12, and its growth in M9-glucuronamide was compared against the negative control used during the initial screen. The growth of the *E. coli* clones was tested in two glucuronamide concentrations, 0.2% and 0.5%.

In both glucuronamide concentrations, cells over-expressing BglA appeared to grow faster than the negative control, especially after t = 11 h (Figures 4.5A and B). In the absence of IPTG, no differences were noted for the growth of the clone harbouring BglA and the negative control. The implication was that over-expressed levels of BglA could indeed impart improved growth phenotypes in *E. coli* by allowing better utilization of glucuronamide as the carbon source. The BglA activity towards glucuronamide is expected

to be low, as the plasmid-encoded *bglA* had to be over-expressed in order to elicit enhanced growth phenotypes (Figures 4.5A and B). Overall, the preliminary results presented in this section appear to support the initial enrichment of the ASKA-encoded *bglA* from the glucuronamide-containing well (Section 4.2.3.1).

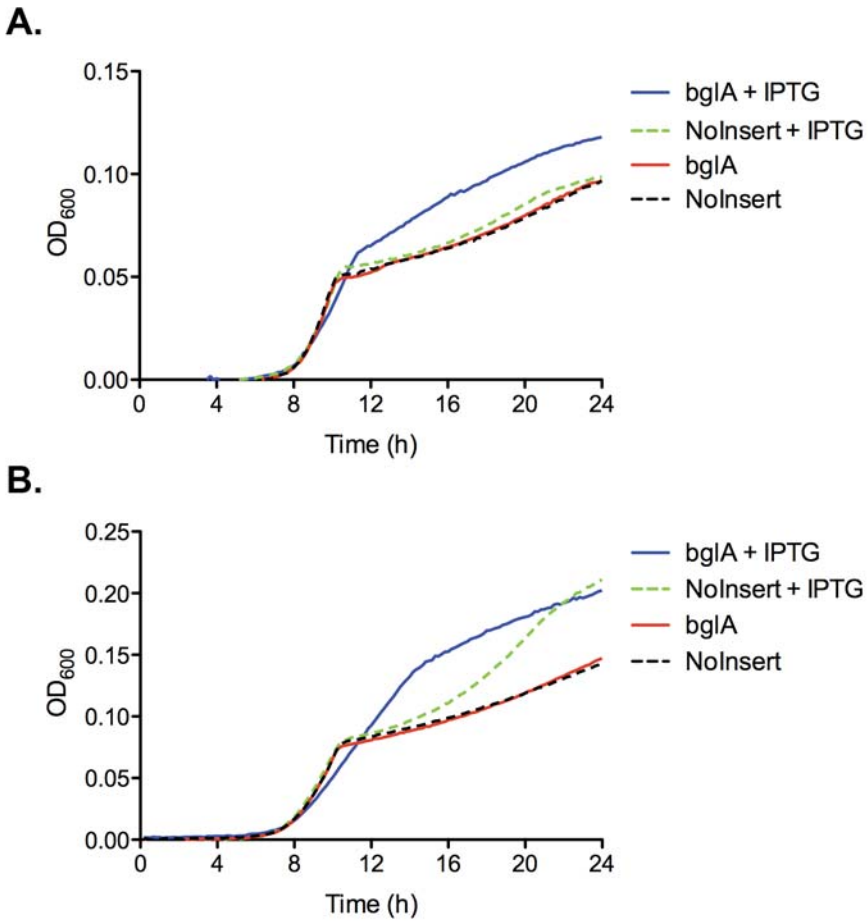


Figure 4.5: *E. coli* growth in M9 medium supplemented with glucuronamide as the sole carbon source.

(A) Growth of different strains in 0.2% glucuronamide. (B) Growth of different strains in 0.5% glucuronamide. Each growth curve is the mean of eight replicates.

4.2.4 NITROGEN SOURCES

For PM 3, PM 6, PM 7 and PM 8, every well contains a base medium and a unique nitrogen source. The medium consists of 100 mM NaCl, 30 mM triethanolamine-HCl (pH 7.1), 2 mM NaH_2PO_4, 250 μM Na_2SO_4, 50 μM $MgCl_2$ and 1 mM KCl (Zhou *et al.*, 2003). The carbon sources were 20 mM sodium succinate and 2 μM ferric citrate (see Section 4.4.3 for more details).

Of the 384 conditions screened, the ASKA pool out-grew the negative control in 194 conditions (with different nitrogen sources). A total of 206 ASKA-encoded ORFs that were responsible for the improved growth phenotypes in 169 conditions (Table 4.2) were isolated. The overall hit rate (*i.e.*, the percentage of environments where at least one ASKA-encoded ORF was enriched) for nitrogen sources was (169/384) = 44%.

In PM 3, the nitrogen source in every well is either an amino acid or its derivative. There were 31 wells, where the ASKA pool out-grew the negative control. At least one ASKA-encoded ORF was enriched from each of these wells. One particular ORF, *cfa* (which codes for cyclopropane fatty acyl phospholipid synthase), was over-represented and isolated from 30 of the 31 wells. The only well in which *cfa* was not isolated contained L-tyrosine, and the enriched ORFs were *glk* (a glucokinase) and *yhdZ* (an ATP-binding component of a putative amino acid transporter). Other ORFs identified from the remaining wells of PM 3 were *ylcG* (an uncharacterized prophage protein), *ybgA* (an uncharacterized protein), *grxA* (a glutaredoxin), *csiE* (a stationary phase-inducible protein) and *ymfL* (an uncharacterized prophage protein).

In contrast, no well from PM 6 to PM 8 yielded clones that were over-expressing Cfa. Instead, the over-represented ORFs were *ycbS* (isolated from 81 wells) and *ylcG* (33 wells). These two ORFs constituted 70% of the total ORFs isolated from PM 6 to PM 8, in which each well contained a particular peptide as the sole nitrogen source. Out of 288 peptide-containing wells, a total of 259 wells selected for clones over-expressing either YcbS or YlcG. Only nine ORFs were enriched once: *ibpA* (from the well containing Arg-Asp), *yjjN* (Gly-Leu), *mgsA* (Leu-Asp), *ubiX* (Lys-Arg), *pncA* (Lys-Glu), *yqcC* (Lys-Leu), *yciK* (Lys-Phe), *rfaS* (Lys-Tyr) and *yeaD* (Leu-Pro).

Overall, the prevalence of multiple-occurring ORFs in Table 4.2 suggested that a non-specific cellular response was elicited in *E. coli* cells upon nutrient starvation. These results were rather unexpected. The possible factors leading to this phenomenon, and its implications, will be discussed further in Section 4.3.3.1.

Table 4.2: The complete list of over-expressed ORFs that conferred enhanced growth in 194 nitrogen- or peptide-containing environments.

Nitrogen source	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
PM 3						
Ammonia	2	Light	Clear	*cfa*	5	Cyclopropane fatty acyl phospholipid synthase
Nitrite	3	Light	Clear	*cfa*	6	Cyclopropane fatty acyl phospholipid synthase
L-Alanine	3	Very dark	Medium	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
L-Arginine	3	Medium	Light	*ylcG*	4	Uncharacterized protein from DLP12 prophage
				cfa	4	Cyclopropane fatty acyl phospholipid synthase
L-Asparagine	2	Light	Clear	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
L-Aspartic acid	2	Light	Clear	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
L-Glutamine	2	Light	Clear	*cfa*	6	Cyclopropane fatty acyl phospholipid synthase
Glycine	2	Light	Clear	*cfa*	7	Cyclopropane fatty acyl phospholipid synthase
L-Histidine	3	Light	Clear	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
L-Phenylalanine	2	Light	Clear	*cfa*	7	Cyclopropane fatty acyl phospholipid synthase
L-Serine	2	Medium	Light	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
L-Threonine	2	Light	Clear	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
L-Tryptophan	2	Medium	Light	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
L-Tyrosine	5	Dark	Medium	*glk*	2	Glucokinase
				yhdZ	5	Putative amino acid transporter
D-Alanine	2	Medium	Light	*cfa*	4	Cyclopropane fatty acyl phospholipid synthase
				ybgA	2	Uncharacterized protein
				ylcG	2	Uncharacterized protein from DLP12 prophage

Nitrogen source	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
D-Serine	2	Medium	Light	*cfa*	7	Cyclopropane fatty acyl phospholipid synthase
L-Citrulline	3	Light	Clear	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
Glucuronamide	2	Medium	Light	*cfa*	7	Cyclopropane fatty acyl phospholipid synthase
N-Acetyl-D-glucosamine	1	Light	Clear	*cfa*	3	Cyclopropane fatty acyl phospholipid synthase
				grxA	2	Glutaredoxin
Adenine	2	Light	Clear	*cfa*	7	Cyclopropane fatty acyl phospholipid synthase
Cytidine	2	Dark	Light	*cfa*	4	Cyclopropane fatty acyl phospholipid synthase
				csiE	3	Stationary phase-inducible protein
Guanine	3	Medium	Light	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
Thymine	4	Light	Clear	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
Uracil	4	Light	Clear	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
Xanthine	3	Medium	Light	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
Xanthosine	2	Light	Clear	*cfa*	7	Cyclopropane fatty acyl phospholipid synthase
Uric acid	3	Medium	Light	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
δ-Amino-N-valeric acid	2	Light	Clear	*cfa*	7	Cyclopropane fatty acyl phospholipid synthase
Ala-His	2	Light	Clear	*cfa*	5	Cyclopropane fatty acyl phospholipid synthase
				ymfL	2	Uncharacterized transcriptional regulator from e14 prophage
Ala-Leu	3	Light	Clear	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
Gly-Glu	2	Light	Clear	*cfa*	7	Cyclopropane fatty acyl phospholipid synthase

Nitrogen source	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
PM 6						
Negative control	3	Light	Clear	*ylcG*	5	Uncharacterized protein from DLP12 prophage
Positive control: L-Glutamine	3	Very dark	Dark	*ylcG*	6	Uncharacterized protein from DLP12 prophage
Ala-His	2	Light	Clear	None isolated	N/A	
Ala-Leu	3	Light	Clear	*ybgA*	4	Uncharacterized protein
Ala-Ser	2	Light	Clear	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
Arg-Ala	2	Light	Clear	*glk*	3	Glucokinase
Arg-Arg	2	Light	Clear	*ylcG*	5	Uncharacterized protein from DLP12 prophage
Arg-Asp	2	Light	Clear	*ibpA*	5	Small heat shock protein
Arg-Gln	2	Light	Clear	*csiE*	4	Stationary phase-inducible protein
Arg-Glu	3	Dark	Medium	*csiE*	2	Stationary phase-inducible protein
				glk	2	Glucokinase
				ylcG	3	Uncharacterized protein from DLP12 prophage
Arg-Ile	2	Light	Clear	None isolated	N/A	
Arg-Leu	4	Light	Clear	*ligT*	2	2′-5′ RNA ligase
				yjfN	2	Uncharacterized protein
Arg-Met	2	Light	Clear	None isolated	N/A	
Arg-Phe	2	Light	Clear	*ylcG*	2	Uncharacterized protein from DLP12 prophage
				ymgC	2	Uncharacterized protein involved in biofilm formation
Arg-Ser	2	Light	Clear	*glk*	2	Glucokinase
Asn-Glu	2	Light	Clear	*ycbS*	4	Uncharacterized outer membrane β-barrel protein
Asn-Val	5	Light	Clear	*yjfN*	4	Uncharacterized protein
				ylcG	2	Uncharacterized protein from DLP12 prophage
Asp-Leu	3	Light	Clear	*ligT*	2	2′-5′ RNA ligase

Nitrogen source	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
Asp-Val	5	Light	Clear	*yjfN*	3	Uncharacterized protein
				ylcG	5	Uncharacterized protein from DLP12 prophage
Cys-Gly	2	Medium	Light	None isolated	N/A	
Gln-Gly	2	Dark	Medium	*csiE*	2	Stationary phase-inducible protein
Glu-Gly	2	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
				ylcG	2	Uncharacterized protein from DLP12 prophage
Glu-Trp	3	Dark	Medium	*csiE*	2	Stationary phase-inducible protein
Gly-Arg	2	Light	Clear	*yjfN*	2	Uncharacterized protein
				ylcG	6	Uncharacterized protein from DLP12 prophage
Gly-Cys	2	Light	Clear	None isolated	N/A	
Gly-His	2	Light	Clear	None isolated	N/A	
Gly-Leu	4	Light	Clear	*ligT*	2	2′-5′ RNA ligase
				yjjN	2	Uncharacterized oxidoreductase
Gly-Met	2	Light	Clear	None isolated	N/A	
Gly-Phe	2	Light	Clear	*ybgA*	3	Uncharacterized protein
				ylcG	2	Uncharacterized protein from DLP12 prophage
Gly-Thr	3	Medium	Light	*ylcG*	3	Uncharacterized protein from DLP12 prophage
Gly-Trp	2	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
Gly-Val	7	Light	Clear	*ilvB*	3	Acetohydroxybutan-oate synthase/ acetolactate synthase
				yjfN	2	Uncharacterized protein
				ylcG	2	Uncharacterized protein from DLP12 prophage
His-Asp	3	Very dark	Dark	*frvA*	2	Uncharacterized domain of a phosphotransferase system

Nitrogen source	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
				ycbS	2	Uncharacterized outer membrane β-barrel protein
				ymgC	3	Uncharacterized protein involved in biofilm formation
His-Gly	3	Dark	Medium	None isolated	N/A	
His-Leu	3	Light	Clear	*ylcG*	4	Uncharacterized protein from DLP12 prophage
Ile-Arg	3	Medium	Light	*yjfN*	2	Uncharacterized protein
Ile-Gln	3	Medium	Light	*ylcG*	5	Uncharacterized protein from DLP12 prophage
Ile-Gly	2	Light	Clear	*yjfN*	2	Uncharacterized protein
Ile-His	3	Light	Clear	None isolated	N/A	
Ile-Phe	2	Light	Clear	*ylcG*	3	Uncharacterized protein from DLP12 prophage
Ile-Ser	2	Light	Clear	*ylcG*	4	Uncharacterized protein from DLP12 prophage
Ile-Trp	3	Light	Clear	None isolated	N/A	
Ile-Tyr	2	Light	Clear	*ymfL*	2	Uncharacterized transcriptional regulator from e14 prophage
Ile-Val	4	Light	Clear	None isolated	N/A	
Leu-Ala	4	Light	Clear	*ybgA*	2	Uncharacterized protein
Leu-Arg	2	Light	Clear	*frvA*	2	Uncharacterized domain of a phosphotransferase system
Leu-Asp	2	Light	Clear	*mgsA*	4	Methylglyoxal synthase
				ylcG	2	Uncharacterized protein from DLP12 prophage
Leu-Glu	2	Light	Clear	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
Leu-Leu	5	Light	Clear	*csiE*	3	Stationary phase-inducible protein
				ylcG	4	Uncharacterized protein from DLP12 prophage
Leu-Met	2	Light	Clear	*ylcG*	4	Uncharacterized protein from DLP12 prophage

Nitrogen source	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
Leu-Phe	3	Light	Clear	None isolated	N/A	
PM 7						
Leu-Trp	2	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
Lys-Ala	3	Light	Clear	*frvA*	4	Uncharacterized domain of a phosphotransferase system
				ycbS	2	Uncharacterized outer membrane β-barrel protein
Lys-Arg	2	Light	Clear	*ubiX*	2	3-octaprenyl-4-hydroxybenzoate decarboxylase
Lys-Glu	3	Dark	Medium	*pncA*	2	Pyrazinamidase/ Nicotinamidase
				ycbS	6	Uncharacterized outer membrane β-barrel protein
Lys-Ile	3	Light	Clear	None isolated	N/A	
Lys-Leu	3	Light	Clear	*ligT*	2	2′-5′ RNA ligase
				yqcC	2	Uncharacterized protein
Lys-Phe	2	Light	Clear	*yciK*	2	Uncharacterized oxidoreductase
Lys-Pro	3	Dark	Medium	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
				ylcG	2	Uncharacterized protein from DLP12 prophage
Lys-Ser	2	Light	Clear	*ycbS*	8	Uncharacterized outer membrane β-barrel protein
Lys-Thr	3	Medium	Light	*ycbS*	7	Uncharacterized outer membrane β-barrel protein
Lys-Tyr	3	Medium	Light	*rfaS*	3	Lipopolysaccharide core biosynthesis protein
Met-Arg	3	Medium	Light	*frvA*	5	Uncharacterized domain of a phosphotransferase system
Met-Asp	3	Medium	Light	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
Met-Gln	2	Medium	Light	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
Met-Glu	2	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein

Nitrogen source	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
Met-Gly	2	Light	Clear	*ycbS*	4	Uncharacterized outer membrane β-barrel protein
Met-Ile	3	Light	Clear	None isolated	N/A	
Met-Leu	2	Light	Clear	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
Met-Lys	2	Light	Clear	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
Met-Met	2	Light	Clear	*ylcG*	4	Uncharacterized protein from DLP12 prophage
Met-Phe	2	Light	Clear	*ycbS*	6	Uncharacterized outer membrane β-barrel protein
Met-Pro	2	Medium	Light	*ycbS*	6	Uncharacterized outer membrane β-barrel protein
Met-Trp	2	Medium	Light	*ycbS*	5	Uncharacterized outer membrane β-barrel protein
Phe-Ala	4	Dark	Medium	*ylcG*	4	Uncharacterized protein from DLP12 prophage
Phe-Gly	2	Light	Clear	*ycbS*	7	Uncharacterized outer membrane β-barrel protein
Phe-Ile	3	Light	Clear	None isolated	N/A	
Phe-Pro	3	Medium	Light	*ycbS*	8	Uncharacterized outer membrane β-barrel protein
Phe-Ser	3	Medium	Light	*ycbS*	6	Uncharacterized outer membrane β-barrel protein
Phe-Trp	3	Medium	Light	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
Pro-Asp	2	Light	Clear	*higB*	5	Toxin of the HigA-HigB toxin-antitoxin system
Pro-Gln	3	Very dark	Medium	None isolated	N/A	
Pro-Hyp	4	Dark	Medium	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
Pro-Leu	3	Medium	Light	*higB*	2	Toxin of the HigA-HigB toxin-antitoxin system
Pro-Pro	3	Very dark	Medium	*csiE*	2	Stationary phase-inducible protein
Ser-Ala	2	Light	Clear	*ycbS*	3	Uncharacterized outer membrane β-barrel protein

Nitrogen source	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
Ser-Gly	2	Light	Clear	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
Ser-His	4	Very dark	Dark	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
Ser-Leu	3	Light	Clear	*ycbS*	6	Uncharacterized outer membrane β-barrel protein
Ser-Met	3	Very dark	Medium	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
Ser-Phe	4	Dark	Medium	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
Ser-Ser	3	Medium	Clear	*ycbS*	4	Uncharacterized outer membrane β-barrel protein
Ser-Tyr	4	Dark	Medium	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
Thr-Arg	3	Medium	Light	None isolated	N/A	
Thr-Glu	4	Dark	Medium	*ycbS*	4	Uncharacterized outer membrane β-barrel protein
Thr-Gly	2	Light	Clear	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
Thr-Leu	2	Light	Clear	*csiE*	2	Stationary phase-inducible protein
Thr-Met	2	Light	Clear	*ylcG*	3	Uncharacterized protein from DLP12 prophage
Trp-Ala	3	Dark	Medium	*ycbS*	7	Uncharacterized outer membrane β-barrel protein
Trp-Leu	3	Light	Clear	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
Trp-Phe	2	Light	Clear	None isolated	N/A	
Trp-Ser	2	Medium	Light	*ycbS*	8	Uncharacterized outer membrane β-barrel protein
Tyr-Gly	2	Light	Clear	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
Tyr-Leu	2	Light	Clear	*ymfL*	3	Uncharacterized transcriptional regulator from e14 prophage
Tyr-Lys	3	Medium	Light	*ycbS*	2	Uncharacterized outer membrane β-barrel protein

Nitrogen source	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
Tyr-Phe	3	Medium	Light	*ymfL*	2	Uncharacterized transcriptional regulator from e14 prophage
Val-Ile	3	Light	Clear	None isolated	N/A	
PM 8						
Negative control	3	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
Asp-Ala	3	Dark	Medium	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
Asp-Gly	2	Light	Clear	*ycbS*	5	Uncharacterized outer membrane β-barrel protein
Gly-Asp	2	Light	Clear	*ycbS*	6	Uncharacterized outer membrane β-barrel protein
Gly-Ile	3	Light	Clear	None isolated	N/A	
His-Ala	2	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
				yciW	2	Uncharacterized oxidoreductase
His-Glu	2	Light	Clear	*ybgA*	2	Uncharacterized protein
				ycbS	5	Uncharacterized outer membrane β-barrel protein
His-His	3	Light	Clear	*ycbS*	4	Uncharacterized outer membrane β-barrel protein
Ile-Asn	3	Light	Clear	*ylcG*	6	Uncharacterized protein from DLP12 prophage
Ile-Leu	2	Light	Clear	None isolated	N/A	
Leu-Asn	3	Light	Clear	*ycbS*	4	Uncharacterized outer membrane β-barrel protein
Leu-His	3	Light	Clear	*ycbS*	5	Uncharacterized outer membrane β-barrel protein
Leu-Pro	2	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
				yeaD	2	Putative mutarotase
				yjbI	2	Uncharacterized protein
Leu-Tyr	2	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein

Nitrogen source	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
Lys-Asp	2	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
				yjbI	2	Uncharacterized protein
Lys-Gly	2	Light	Clear	None isolated	N/A	
Lys-Met	3	Light	Clear	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
Met-Thr	2	Light	Clear	None isolated	N/A	
Met-Tyr	2	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
Phe-Asp	2	Light	Clear	*ycbS*	6	Uncharacterized outer membrane β-barrel protein
Phe-Met	2	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
Phe-Tyr	2	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
Pro-Ile	2	Light	Clear	*ylcG*	4	Uncharacterized protein from DLP12 prophage
Pro-Ser	2	Medium	Light	*ycbS*	5	Uncharacterized outer membrane β-barrel protein
Pro-Val	5	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
				ylcG	4	Uncharacterized protein from DLP12 prophage
Ser-Asn	3	Dark	Medium	*ycbS*	6	Uncharacterized outer membrane β-barrel protein
Ser-Asp	2	Light	Clear	*ycbS*	5	Uncharacterized outer membrane β-barrel protein
Ser-Gln	2	Medium	Light	*ycbS*	7	Uncharacterized outer membrane β-barrel protein
Ser-Glu	2	Medium	Light	*ycbS*	7	Uncharacterized outer membrane β-barrel protein
Thr-Asp	3	Medium	Light	*ycbS*	4	Uncharacterized outer membrane β-barrel protein
Thr-Phe	3	Light	Clear	*ylcG*	4	Uncharacterized protein from DLP12 prophage
Thr-Ser	4	Dark	Medium	*ycbS*	4	Uncharacterized outer membrane β-barrel protein

Nitrogen source	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
Tyr-Ile	3	Medium	Light	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
Val-Ala	4	Light	Clear	*ycbS*	4	Uncharacterized outer membrane β-barrel protein
				yciW	2	Uncharacterized oxidoreductase
Val-Gln	4	Light	Clear	*ilvB*	3	Acetohydroxybutan-oate synthase/ acetolactate synthase
Val-Glu	6	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
				ylcG	5	Uncharacterized protein from DLP12 prophage
Val-Lys	6	Light	Clear	*ilvB*	6	Acetohydroxybutan-oate synthase/ acetolactate synthase
				ylcG	2	Uncharacterized protein from DLP12 prophage
Val-Pro	6	Light	Clear	*ycbS*	6	Uncharacterized outer membrane β-barrel protein
β-Ala-Phe	5	Light	Clear	*ymfL*	3	Uncharacterized transcriptional regulator from e14 prophage
D-Ala-D-Ala	2	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
D-Ala-Gly	2	Light	Clear	*ycbS*	6	Uncharacterized outer membrane β-barrel protein
D-Leu-Gly	4	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
				ylcG	3	Uncharacterized protein from DLP12 prophage
γ-Glu-Gly	4	Light	Clear	None isolated	N/A	
Gly-D-Ala	2	Light	Clear	*ycbS*	4	Uncharacterized outer membrane β-barrel protein
Gly-D-Ser	2	Light	Clear	*ycbS*	6	Uncharacterized outer membrane β-barrel protein
Leu-β-Ala	5	Medium	Light	*ylcG*	2	Uncharacterized protein from DLP12 prophage
Phe-β-Ala	3	Light	Clear	*ylcG*	2	Uncharacterized protein from DLP12 prophage

Nitrogen source	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
Ala-Ala-Ala	2	Medium	Light	None isolated	N/A	
Gly-Gly-Gly	2	Light	Clear	*ycbS*	4	Uncharacterized outer membrane β-barrel protein
Gly-Gly-Ile	2	Light	Clear	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
Gly-Gly-Leu	2	Light	Clear	*ylcG*	2	Uncharacterized protein from DLP12 prophage
Gly-Gly-Phe	2	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
Leu-Gly-Gly	4	Light	Clear	*frvA*	2	Uncharacterized domain of a phosphotransferase system
Leu-Leu-Leu	5	Light	Clear	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
Phe-Gly-Gly	2	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
Tyr-Gly-Gly	3	Medium	Light	*ycbS*	4	Uncharacterized outer membrane β-barrel protein

* Days of incubation before a change in growth phenotype (*i.e.*, tetrazolium colour intensity) was observed. Key for scoring growth phenotype (*i.e.*, tetrazolium colour intensity):

Very dark Dark Medium Light Clear

4.2.5 PHOSPHORUS & SULFUR SOURCES

The ASKA pool and the negative control were screened in a total of 60 phosphorus sources and 36 sulfur sources (found in PM 4). The base medium used for the wells containing phosphorus sources comprises 100 mM NaCl, 30 mM triethanolamine-HCl (pH 7.1), 5 mM NH_4Cl, 250 μM Na_2SO_4, 50 μM $MgCl_2$ and 1 mM KCl (Zhou *et al.*, 2003). For the wells containing sulfur sources, the medium components are 100 mM NaCl, 30 mM triethanolamine-HCl (pH 7.1), 5 mM NH_4Cl, 2 mM NaH_2PO_4, 50 μM $MgCl_2$ and 1 mM KCl (Zhou *et al.*, 2003). In both types of environment, 20 mM sodium succinate and 2 μM ferric citrate were used as the carbon sources (see Section 4.4.3 for more details).

The ASKA pool showed improved growth in 37 (out of 60) phosphorus-containing environments. A total of 38 ASKA-encoded ORFs were enriched from 34 of the phosphorus-containing environments (Table 4.3). Hence, the overall hit rate (*i.e.*, the percentage of environments where at least one ASKA-encoded ORF was enriched) for phosphorus

sources was (34/60) = 57%. Similar to the nitrogen sources (Section 4.2.4), the list of enriched ORFs was dominated by *cfa* (14 times), *ylcG* (9 times) and *ycbS* (6 times). The only ORFs that were isolated once were *dgsA* (from the well with 2-deoxy-D-glucose-6-phosphate), *yhdZ* (guanosine-2′,3′-cyclic monophosphate), *ybgA* (cytidine-3′,5′-cyclic monophosphate) and *glk* (cytidine-2′-monophosphate).

Table 4.3: The complete list of over-expressed ORFs that conferred enhanced growth in 37 phosphorus-containing environments.

Phosphorus source	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
PM 4 (Phosphorus)						
Phosphate	2	Light	Clear	*ylcG*	6	Uncharacterized protein from DLP12 prophage
Pyrophosphate	2	Light	Clear	*ylcG*	2	Uncharacterized protein from DLP12 prophage
Trimetaphosphate	3	Light	Clear	*ylcG*	2	Uncharacterized protein from DLP12 prophage
Tripolyphosphate	2	Light	Clear	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
Adenosine-2′-monophosphate	2	Dark	Medium	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
Adenosine-2′,3′-cyclic monophosphate	3	Dark	Light	*cfa*	7	Cyclopropane fatty acyl phospholipid synthase
Thiophosphate	3	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
Dithiophosphate	2	Light	Clear	*cfa*	4	Cyclopropane fatty acyl phospholipid synthase
				ylcG	2	Uncharacterized protein from DLP12 prophage
β-Glycerol phosphate	2	Light	Clear	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
Carbamyl phosphate	2	Light	Clear	*cfa*	7	Cyclopropane fatty acyl phospholipid synthase
D-3-Phospho-glyceric acid	5	Light	Clear	*ylcG*	2	Uncharacterized protein from DLP12 prophage
Guanosine-2′-monophosphate	2	Light	Clear	*cfa*	5	Cyclopropane fatty acyl phospholipid synthase
Guanosine-3′-monophosphate	2	Light	Clear	*cfa*	2	Cyclopropane fatty acyl phospholipid synthase

Phosphorus source	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
Guanosine-5′-monophosphate	2	Light	Clear	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
Guanosine-2′,3′-cyclic monophosphate	5	Light	Clear	*yhdZ*	2	Putative amino acid transporter
Phosphoenol pyruvate	2	Light	Clear	*ylcG*	2	Uncharacterized protein from DLP12 prophage
D-Glucose-1-phosphate	2	Medium	Light	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
D-Glucose-6-phosphate	2	Medium	Light	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
2-Deoxy-D-glucose-6-phosphate	2	Light	Clear	*dgsA*	7	DgsA transcriptional repressor
Cytidine-2′-monophosphate	2	Light	Clear	*csiE*	3	Stationary phase-inducible protein
				glk	2	Glucokinase
Cytidine-5′-monophosphate	2	Light	Clear	*csiE*	6	Stationary phase-inducible protein
Cytidine-2′,3′-cyclic monophosphate	3	Light	Clear	*ylcG*	3	Uncharacterized protein from DLP12 prophage
Cytidine-3′,5′-cyclic monophosphate	2	Light	Clear	*ybgA*	5	Uncharacterized protein
				ylcG	2	Uncharacterized protein from DLP12 prophage
D-Mannose-1-phosphate	2	Light	Clear	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
D-Mannose-6-phosphate	3	Medium	Light	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
Cysteamine-S-phosphate	2	Light	Clear	*ylcG*	2	Uncharacterized protein from DLP12 prophage
Phospho-L-arginine	2	Light	Clear	*cfa*	7	Cyclopropane fatty acyl phospholipid synthase
O-Phospho-D-serine	5	Light	Clear	None isolated	N/A	
O-Phospho-L-serine	2	Light	Clear	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
Uridine-3′-monophosphate	2	Light	Clear	*csiE*	2	Stationary phase-inducible protein
Uridine-5′-monophosphate	3	Light	Clear	*csiE*	6	Stationary phase-inducible protein
Uridine-2′,3′-cyclic monophosphate	5	Dark	Medium	*ycbS*	8	Uncharacterized outer membrane β-barrel protein

Phosphorus source	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
Phosphocreatine	2	Light	Clear	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
O-Phosphoryl-ethanolamide	4	Light	Clear	None isolated	N/A	
Thymidine-3′-monophosphate	2	Light	Clear	None isolated	N/A	
Thymidine-5′-monophosphate	2	Light	Clear	*ycbS*	7	Uncharacterized outer membrane β-barrel protein
Inositol hexaphosphate	2	Light	Clear	*csiE*	2	Stationary phase-inducible protein
				ycbS	3	Uncharacterized outer membrane β-barrel protein

* Days of incubation before a change in growth phenotype (*i.e.*, tetrazolium colour intensity) was observed. Key for scoring growth phenotype (*i.e.*, tetrazolium colour intensity):

Very dark Dark Medium Light ☐ Clear

Of the 36 environments with different sulfur sources, the ASKA pool out-grew the negative control in 22 of them. A total of 24 enriched ASKA-encoded ORFs were isolated from 20 environments (Table 4.4). The overall hit rate (*i.e.*, the percentage of environments where at least one ASKA-encoded ORF was enriched) for sulfur sources was (20/36) = 56%. The ORFs that were isolated most frequently were *ylcG* (6 times) and *ymfL* (6 times), followed by *ycbS* (5 times) and *cfa* (4 times). Altogether, these ORFs represented >70% of the total ORFs enriched from sulfur-containing environments.

Table 4.4: The complete list of over-expressed ORFs that conferred enhanced growth in 22 sulfur-containing environments.

Sulfur source	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
PM 4 (Sulfur)						
Sulfate	2	Light	Clear	*csiE*	2	Stationary phase-inducible protein
				ycbS	2	Uncharacterized outer membrane β-barrel protein
Thiosulfate	2	Light	Clear	*ycbS*	8	Uncharacterized outer membrane β-barrel protein
Tetrathionate	2	Light	Clear	*ycbS*	4	Uncharacterized outer membrane β-barrel protein

Sulfur source	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
Thiophosphate	2	Light	Clear	*ylcG*	2	Uncharacterized protein from DLP12 prophage
				ymfL	5	Uncharacterized transcriptional regulator from e14 prophage
Dithiophosphate	2	Light	Clear	*ylcG*	2	Uncharacterized protein from DLP12 prophage
L-Cysteine	2	Medium	Clear	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
D-Cysteine	2	Light	Clear	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
L-Cysteine-glycine	2	Light	Clear	None isolated	N/A	
L-Cysteine sulfinic acid	2	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
Lanthionine	2	Light	Clear	*cfa*	7	Cyclopropane fatty acyl phospholipid synthase
Glutathione	4	Medium	Clear	*ylcG*	4	Uncharacterized protein from DLP12 prophage
				ymfL	4	Uncharacterized transcriptional regulator from e14 prophage
L-Methionine	3	Light	Clear	*cfa*	8	Cyclopropane fatty acyl phospholipid synthase
D-Methionine	4	Light	Clear	*ylcG*	2	Uncharacterized protein from DLP12 prophage
Glycyl-L-methionine	2	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
N-Acetyl-D,L-methionine	2	Light	Clear	*ylcG*	2	Uncharacterized protein from DLP12 prophage
L-Methionine sulfoxide	3	Light	Clear	*ymfL*	3	Uncharacterized transcriptional regulator from e14 prophage
L-Djenkolic acid	3	Medium	Light	*ylcG*	3	Uncharacterized protein from DLP12 prophage
Taurine	2	Light	Clear	*csiE*	2	Stationary phase-inducible protein
				ymfL	4	Uncharacterized transcriptional regulator from e14 prophage

Sulfur source	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
Hypotaurine	2	Light	Clear	*ymfL*	3	Uncharacterized transcriptional regulator from e14 prophage
Butane sulfonic acid	2	Light	Clear	*wbbL_1*	2	Putative lipopolysaccharide biosynthesis glycosyltransferase
2-Hydroxyethane sulfonic acid	2	Light	Clear	None isolated	N/A	
Methane sulfonic acid	2	Light	Clear	*ymfL*	3	Uncharacterized transcriptional regulator from e14 prophage

* Days of incubation before a change in growth phenotype (*i.e.*, tetrazolium colour intensity) was observed. Key for scoring growth phenotype (*i.e.*, tetrazolium colour intensity):

Very dark Dark Medium Light ☐ Clear

4.2.6 BIOSYNTHETIC NUTRIENTS

Instead of the nitrogen-free, phosphorus-free and sulfur-free versions of the minimal base medium used for PM 3, PM 4, PM 6, PM 7 and PM 8, every well of PM 5 contains the complete version of the minimal base medium as well as a biosynthetic nutrient. These nutrient supplements include various amino acids, vitamins and metabolic intermediates. The components of the base medium are 100 mM NaCl, 30 mM triethanolamine-HCl (pH 7.1), 5 mM NH_4Cl, 2 mM NaH_2PO_4, 250 μM Na_2SO_4, 50 μM $MgCl_2$ and 1 mM KCl. Similar to PM 3, PM 4, PM6, PM 7 and PM 8, the carbon sources were 20 mM sodium succinate and 2 μM ferric citrate (see Section 4.4.3 for more details).

Out of 96 conditions with different biosynthetic nutrients, the ASKA pool out-grew the negative control in 86 conditions. A total of 104 ASKA-encoded ORFs were enriched from 84 conditions. The overall hit rate (*i.e.*, the percentage of environments where at least one ASKA-encoded ORF was enriched) for biosynthetic nutrients was (84/96) = 88%. Similar to PM 1 and PM 2, there was no condition, in which the ASKA pool was out-grown by the negative control.

Two ORFs, *ylcG* and *ycbS*, dominated the pool of ORFs enriched from the 84 wells. YlcG-expressing clones were enriched 44 times, whereas YcbS-expressing clones were enriched 42 times. Other ORFs isolated multiple times were *glk* (9 times), *yciI* (3 times) and *yhdZ* (2 times). The only ORFs that were isolated once were *thiE* (from the well containing adenosine), *ilvN* (cytosine), *csiE* [(5)4-amino-imidazole-4(5)-carboxamide] and *yjfN* (pyridoxine).

Table 4.5: The complete list of over-expressed ORFs that conferred enhanced growth in 86 environments, which contained different biosynthetic supplements.

Biosynthetic nutrient	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
PM 5						
Negative control	3	Light	Clear	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
				ylcG	3	Uncharacterized protein from DLP12 prophage
Positive control (LB)	2	Very dark	Dark	*ycbS*	5	Uncharacterized outer membrane β-barrel protein
L-Alanine	3	Light	Clear	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
				ylcG	3	Uncharacterized protein from DLP12 prophage
L-Arginine	3	Light	Clear	*glk*	4	Glucokinase
L-Asparagine	3	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
				ylcG	2	Uncharacterized protein from DLP12 prophage
L-Aspartic acid	3	Light	Clear	*glk*	2	Glucokinase
L-Glutamic acid	3	Light	Clear	*ylcG*	3	Uncharacterized protein from DLP12 prophage
Adenosine-3′,5′-cyclic monophosphate	3	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
Adenine	3	Light	Clear	*glk*	8	Glucokinase
Adenosine	3	Light	Clear	*thiE*	6	Thiamine phosphate synthase
2′-Deoxyadenosine	3	Medium	Light	*ycbS*	4	Uncharacterized outer membrane β-barrel protein
L-Glutamine	3	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
Glycine	3	Light	Clear	*ylcG*	5	Uncharacterized protein from DLP12 prophage
L-Histidine	3	Light	Clear	*ylcG*	2	Uncharacterized protein from DLP12 prophage
L-Isoleucine	3	Light	Clear	*yciI*	3	Uncharacterized protein
L-Leucine	3	Light	Clear	*ylcG*	8	Uncharacterized protein from DLP12 prophage
L-Lysine	3	Light	Clear	*ylcG*	3	Uncharacterized protein from DLP12 prophage

Biosynthetic nutrient	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
L-Methionine	3	Light	Clear	*glk*	2	Glucokinase
				ylcG	2	Uncharacterized protein from DLP12 prophage
L-Phenylalanine	3	Light	Clear	*ylcG*	2	Uncharacterized protein from DLP12 prophage
Guanosine-3′,5′-cyclic monophosphate				*ycbS*	3	Uncharacterized outer membrane β-barrel protein
	3	Light	Clear	*ylcG*	2	Uncharacterized protein from DLP12 prophage
Guanine	3	Light	Clear	None isolated	N/A	
Guanosine	3	Light	Clear	*ylcG*	8	Uncharacterized protein from DLP12 prophage
2′-Deoxyguanosine	3	Light	Clear	*yhdZ*	4	Putative amino acid transporter
L-Serine	3	Light	Clear	*glk*	3	Glucokinase
L-Threonine	3	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
				ylcG	3	Uncharacterized protein from DLP12 prophage
L-Tryptophan	3	Light	Clear	*ylcG*	4	Uncharacterized protein from DLP12 prophage
L-Tyrosine	3	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
L-Isoleucine + L-Valine	3	Light	Clear	*glk*	4	Glucokinase
				ylcG	2	Uncharacterized protein from DLP12 prophage
trans-4-Hydroxy-L-proline	3	Light	Clear	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
(5)4-Amino-imidazole-4(5)-carboxamide	3	Light	Clear	*csiE*	4	Stationary phase-inducible protein
Hypoxanthine	3	Medium	Clear	*glk*	2	Glucokinase
				ycbS	2	Uncharacterized outer membrane β-barrel protein
Inosine	3	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
2′-Deoxyinosine	3	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein

Biosynthetic nutrient	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
L-Ornithine	3	Light	Clear	*ylcG*	4	Uncharacterized protein from DLP12 prophage
L-Citrulline	3	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
				ylcG	2	Uncharacterized protein from DLP12 prophage
L-Homoserine lactone	3	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
				ylcG	6	Uncharacterized protein from DLP12 prophage
D-Alanine	3	Light	Clear	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
				ylcG	2	Uncharacterized protein from DLP12 prophage
D-Aspartic acid	3	Light	Clear	*ylcG*	2	Uncharacterized protein from DLP12 prophage
D-Glutamic acid	3	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
D,L-ε-Diaminopimelic acid	3	Light	Clear	*ylcG*	3	Uncharacterized protein from DLP12 prophage
Cytosine	3	Light	Clear	*ilvN*	2	Acetohydroxybutanoate synthase/ acetolactate synthase
				ycbS	2	Uncharacterized outer membrane β-barrel protein
				ylcG	3	Uncharacterized protein from DLP12 prophage
Cytidine	3	Light	Clear	*yciI*	2	Uncharacterized protein
2′-Deoxycytidine	3	Light	Clear	*ylcG*	3	Uncharacterized protein from DLP12 prophage
Putrescine	3	Light	Clear	*ylcG*	3	Uncharacterized protein from DLP12 prophage
Spermidine	3	Light	Clear	*glk*	2	Glucokinase
Spermine	3	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
				ylcG	4	Uncharacterized protein from DLP12 prophage
Pyridoxine	4	Medium	Light	*ycbS*	3	Uncharacterized outer membrane β-barrel protein

Biosynthetic nutrient	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
				yjfN	2	Uncharacterized protein
Pyridoxal	3	Light	Clear	*ylcG*	4	Uncharacterized protein from DLP12 prophage
Pyridoxamine	4	Medium	Light	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
β-Alanine	3	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
D-Pantothenic acid	3	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
				ylcG	2	Uncharacterized protein from DLP12 prophage
Orotic acid	3	Light	Clear	*ycbS*	4	Uncharacterized outer membrane β-barrel protein
Uracil	3	Light	Clear	*ylcG*	2	Uncharacterized protein from DLP12 prophage
Uridine	3	Light	Clear	*ycbS*	4	Uncharacterized outer membrane β-barrel protein
2′-Deoxyuridine	3	Light	Clear	*yhdZ*	2	Putative amino acid transporter
Quinolinic acid	3	Light	Clear	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
				ylcG	4	Uncharacterized protein from DLP12 prophage
Nicotinic acid	3	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
Nicotinamide	3	Light	Clear	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
				ylcG	4	Uncharacterized protein from DLP12 prophage
β-Nicotinamide adenine dinucleotide	3	Medium	Light	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
δ-Amino-levulinic acid	3	Light	Clear	*ycbS*	5	Uncharacterized outer membrane β-barrel protein
Hematin	3	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
				ylcG	4	Uncharacterized protein from DLP12 prophage
Deferoxamine mesylate	3	Light	Clear	*ycbS*	4	Uncharacterized outer membrane β-barrel protein

Biosynthetic nutrient	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
Thymine	3	Light	Clear	*ylcG*	4	Uncharacterized protein from DLP12 prophage
Glutathione (reduced)	3	Light	Clear	*ylcG*	8	Uncharacterized protein from DLP12 prophage
Thymidine	3	Light	Clear	*ylcG*	6	Uncharacterized protein from DLP12 prophage
Oxaloacetic acid	3	Light	Clear	*ylcG*	2	Uncharacterized protein from DLP12 prophage
D-Biotin	3	Light	Clear	*ylcG*	2	Uncharacterized protein from DLP12 prophage
Cyano-cobalamine	3	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
p-Amino-benzoic acid	3	Light	Clear	*ylcG*	2	Uncharacterized protein from DLP12 prophage
Folic acid	3	Light	Clear	*ylcG*	3	Uncharacterized protein from DLP12 prophage
Riboflavin	3	Light	Clear	*ycbS*	4	Uncharacterized outer membrane β-barrel protein
Pyrrolo-quinoline quinone	3	Light	Clear	*ycbS*	4	Uncharacterized outer membrane β-barrel protein
Menadione	3	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
Myo-inositol	3	Light	Clear	None isolated	N/A	
Butyric acid	3	Light	Clear	*ylcG*	8	Uncharacterized protein from DLP12 prophage
D,L-α-Hydroxybutyric acid	3	Light	Clear	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
α-Ketobutyric acid	3	Light	Clear	*glk*	8	Glucokinase
Caprylic acid	3	Light	Clear	*ylcG*	2	Uncharacterized protein from DLP12 prophage
D,L-α-Lipoic acid (oxidized)	3	Light	Clear	*ycbS*	2	Uncharacterized outer membrane β-barrel protein
				ylcG	6	Uncharacterized protein from DLP12 prophage
D,L-Mevalonic acid	3	Light	Clear	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
D,L-Carnithine	3	Light	Clear	*ylcG*	2	Uncharacterized protein from DLP12 prophage

Biosynthetic nutrient	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
Choline	3	Light	Clear	*ycbS*	3	Uncharacterized outer membrane β-barrel protein
Tween 20	3	Light	Clear	*ycbS*	4	Uncharacterized outer membrane β-barrel protein
Tween 40	3	Light	Clear	*ylcG*	4	Uncharacterized protein from DLP12 prophage
Tween 60	3	Light	Clear	*yciI*	2	Uncharacterized protein
Tween 80	3	Light	Clear	*ylcG*	8	Uncharacterized protein from DLP12 prophage

* Days of incubation before a change in growth phenotype (*i.e.*, tetrazolium colour intensity) was observed. Key for scoring growth phenotype (*i.e.*, tetrazolium colour intensity):

Very dark Dark Medium Light ☐ Clear

4.2.7 Osmolytes & pH stresses

Similar to the rich base medium used in every well of PM 1 and 2, the base medium used for PM 9 and PM 10 consists of 2.0 g/L of tryptone, 1.0 g/L of yeast extract and 1.0 g/L of NaCl (Zhou *et al*., 2003).

As mentioned previously (Section 4.2.1), there were not many wells in PM 9 and PM 10, where the ASKA pool out-grew the negative control. Out of 192 conditions screened, there were only 14 cases where the ASKA pool grew better than the negative control. However, only eight ASKA-encoded ORFs were isolated from five of these conditions (Table 4.6). Therefore, the hit rate (*i.e.*, the percentage of environments where at least one ASKA-encoded ORF was enriched) for osmolyte and pH stresses was only (5/192) = 3%.

Among the ORFs isolated, *cpdA* (a cAMP phosphodiesterase) was isolated twice (from the wells with potassium chloride 5% and potassium chloride 6%). Two other ORFs were also enriched from each of these potassium chloride-containing wells: *bdm* (a biofilm-dependent modulation protein) and *acpP* (an acyl carrier protein). Although it was not surprising that such similar environments yielded a clone (CpdA) in common, it is less clear why they also yielded different clones (Bdm versus AcpP).

For the alkaline environments (pH 9.5 supplemented with glycine and L-lysine, respectively), the ASKA pool out-grew the negative control. However, no ASKA-encoded ORF was enriched from these environments. In contrast, three ORFs were enriched successfully from two acidic environments: *yhiS* and *yfjL* (pH 4.5 + L-alanine) and *frwB* (pH 4.5 + β-hydroxyglutamate).

Table 4.6: The complete list of over-expressed ORFs that conferred enhanced growth in 14 environments, which contained different osmolyte concentration and pH.

Osmolyte and pH	Days of inc.*	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product
PM 9						
NaCl 6% + Creatinine	2	Light	Clear	None isolated	N/A	
Potassium chloride 5%	1	Medium	Light	*bdm*	3	Biofilm-dependent modulation protein
				cpdA	5	cAMP phosphodiesterase
Potassium chloride 6%	5	Dark	Medium	*acpP*	5	Acyl carrier protein
				cpdA	3	cAMP phosphodiesterase
Sodium formate 3%	2	Medium	Light	*tesB*	2	Thioesterase II
Sodium formate 5%	2	Light	Clear	None isolated	N/A	
Urea 3%	2	Light	Clear	None isolated	N/A	
Sodium nitrite 100 mM	2	Medium	Light	None isolated	N/A	
PM 10						
pH 4.5 + L-Alanine	4	Medium	Light	*yhiS*	3	Uncharacterized protein
				yfjL	2	Uncharacterized protein from CP4-57 prophage
pH 4.5 + Phenylalanine	3	Light	Clear	None isolated	N/A	
pH 4.5 + L-Serine	2	Light	Clear	None isolated	N/A	
pH 4.5 + L-Valine	3	Light	Clear	None isolated	N/A	
pH 4.5 + β-hydroxyglutamate	2	Medium	Clear	*frwB*	7	Putative subunit of a phosphotransferase system
pH 9.5 + Glycine	1	Dark	Medium	None isolated	N/A	
pH 9.5 + L-Lysine	1	Light	Clear	None isolated	N/A	

* Days of incubation before a change in growth phenotype (*i.e.*, tetrazolium colour intensity) was observed.

Key for scoring growth phenotype (*i.e.*, tetrazolium colour intensity):

■ Very dark ■ Dark ■ Medium ■ Light □ Clear

4.2.8 CASES WHERE THE NEGATIVE CONTROL OUT-GREW THE ASKA POOL

In total, the negative control grew better than the ASKA pool in 18 environments (2% of 960 environments screened; Table 4.7). In particular, the negative control out-grew the ASKA pool in 9 wells with various osmolyte concentrations and pH values. In these 18 environments, protein over-expression might have incurred a high fitness cost, compared to the negative control. Although the exact cause of the elevated cost was not clear, the overall slower growth of the ASKA pool (relative to the negative control) implied that *E. coli* would require a different adaptive approach (*e.g.* mutations in a nutrient transporter) rather than up-regulating its protein expression in these environments. A further investigation of the cost of protein over-expression under different conditions was outside the scope of this study, and therefore was not undertaken. The emphasis of this study was on finding *E. coli* proteins that, when over-expressed, were able to impart improved growth phenotypes in new environments.

Table 4.7: The 18 environments where the negative control out-grew the ASKA pool.

Environment	Days of incubation*	ASKA growth	NoInsert growth
PM 3			
Thymidine	5	Clear	Light
Uridine	5	Clear	Light
PM 4			
O-Phospho-L-tyrosine	2	Clear	Light
PM 6			
His-Met	4	Light	Medium
His-Tyr	2	Clear	Light
Ile-Ala	3	Light	Medium
PM 7			
Negative control (No nitrogen source)	5	Clear	Light
Phe-Phe	5	Medium	Dark
Pro-Ala	2	Light	Medium
PM 9			
NaCl 1%	6	Dark	Very dark
NaCl 2%	6	Dark	Very dark

Environment	Days of incubation*	ASKA growth	NoInsert growth
NaCl 3%	6	Dark	Very dark
NaCl 6% + L-Carnitine	2	Clear	Light
Sodium sulfate 4%	2	Medium	Dark
Sodium sulfate 5%	2	Medium	Dark
Urea 2%	1	Medium	Dark
Sodium benzoate pH 5.2, 20 mM	1	Light	Medium
PM 10			
pH 9.5 + L-Leucine	2	Light	Medium

* Days of incubation before a change in growth phenotype (*i.e.*, tetrazolium colour intensity) was observed. Key for scoring growth phenotype (*i.e.*, tetrazolium colour intensity):

Very dark Dark Medium Light Clear

4.3 DISCUSSION

4.3.1 AN OVERVIEW OF THE RESULTS

The library-on-library screen described in this chapter was the first part of a comprehensive experiment to catalogue the evolutionary potential of the *E. coli* proteome. The emphasis was on finding *E. coli* proteins that are able to impart improved growth phenotypes in the presence of foreign compounds, when they are over-expressed. Although the screening methodology was initially conceived to discover enzymes with secondary activities, new phenotypes were also expected to be underpinned by gene regulatory effects and by metabolite influx via ambiguous transporter proteins. To date, this library-on-library screen is the most systematic screen conducted to probe the global phenotypic effects of protein over-expression in a variety of growth environments.

Over-expression from each ASKA plasmid is *lac*-controlled, and hence, is amenable to IPTG induction. Over-expression of the plasmid-encoded ORFs mirrored the effects of gene amplifications in the cells. The pCA24N backbone also encodes the *lacI*^{q} repressor, for tightly controlling the expression of each ASKA ORF. The ASKA expression vector, pCA24N, has the same modified pMB1 replication origin as pQE30 (Qiagen), which gives it a copy number of 300–400 per cell (Qiagen, 2011). Therefore, this library-on-library screen was specifically aimed at examples where protein over-expression was essential for increased fitness. However, a previous study showed that induction with a high concentration of IPTG (1 mM) severely inhibits the growth of 51% of ASKA strains (including almost all that over-express membrane proteins), while another 28% show

moderate growth inhibition (Kitagawa *et al.*, 2005). To avoid this source of bias, a significantly lower concentration of IPTG (50 μM) was used to induce protein expression. This is sufficient for derepression of the pCA24N T5-*lacO* promoter, but minimizes fitness differences associated with over-expression of different ORFs (Miller & Raines, 2004; Patrick *et al.*, 2007).

In 335 of the 380 environments where the growth of ASKA pool was better than the negative control (Section 4.2.1), the colour differences between the two populations were observed within 1–3 days of incubation. I found that visual inspection was adequate for differentiating the intensity of the tetrazolium dye in each well. In a preliminary screen, I had attempted to quantify the A_{595} (as recommended by Biolog Inc.) of the colour intensities developed in the wells of PM 11 using a microtitre plate reader. However, the absorbance readings were not strictly correlated with the overall colour intensity in the wells, especially after 2 days of incubation (results not shown). This was because most of the intensely-coloured cell clumps, which were formed after prolonged incubation (2–7 days), tend to sediment at the bottom of the wells. The presence of these cell clumps often interfered with the absorbance readings; manual resuspension of these cell clumps usually provided a more accurate representation of the colour intensity in each well. Furthermore, many wells inoculated with the ASKA pool were only a shade darker than the negative control (*e.g.* a "Light *versus* Clear" comparison, or a "Medium *versus* Light" comparison). In these cases, it was difficult to employ an absolute cut-off value for each colour shade, if the colour intensity of the PM wells was to be measured spectrophotometrically.

In many cases, the minimal colour differences between the ASKA pool and the negative control implied that the ASKA populations were only slightly fitter than the negative control. In order to isolate the ASKA clone that was responsible for each improved growth phenotype, the ASKA populations were serially enriched. Each serial transfer enabled a 50-fold enrichment, *i.e.*, the numbers of each surviving ASKA clone in a well could be amplified 50 times. Accordingly, the ASKA clone(s) that propagated faster would outnumber other slow-growing clones, and thus, the final ASKA population would be dominated by the fitter ASKA clone(s) after 2 rounds of serial enrichment (50-fold × 50-fold = 2,500-fold). Enrichment of the faster-growing clones in the ASKA population also manifested in growth improvements during each transfer. For example, the ASKA population from the well containing L-methionine as the sulfur source (Table 4.4 in Section 4.2.5) took 3 days to develop a noticeable purple colour. After the first round of enrichment, the duration for colour development was reduced to 2 days. Finally, the colour development took only 1 day after the second round of enrichment. Overall, this serial enrichment strategy yielded winner clones from 87% of the 380 environments (Section 4.2.2). There were 41 environments a single ORF was enriched to apparent homogeneity, *i.e.* eight of eight sequenced ORFs were identical.

In a minority of cases (~13%), serial transfer failed to enrich any individual ASKA clone to a frequency of two or more in the eight sequenced clones. One explanation is that there were many plasmid-encoded ORFs that all imparted fitness advantages (of similar magnitudes) in those conditions. It is also possible that chromosomal mutations were responsible for improved growth, although this is unlikely when the experimental design (two independent ASKA pools, both required to out-grow negative control populations) and the small size of each input population (~50,000 cells/well) are considered. Alternatively, further transfers may have been required to identify the fastest-growing clones in the ASKA pool. By imparting stringent criteria for identifying genes that imparted a growth advantage, the probability of isolating false positives was minimized.

4.3.2 Over-represented ORFs

The standout feature of the results for PM 3 to PM 8 (Section 4.2.4 to Section 4.2.6) was the dominance of three ASKA-encoded ORFs: *cfa*, *ycbS* and *ylcG*. This observation was unexpected, since the environments that yielded these multi-occurring ORFs were very different in their chemical composition. Each environment contained a unique nitrogen source (PM 3, PM 6, PM 7 and PM 8; Section 4.2.4), phosphorus source, sulfur source (both in PM 4; Section 4.2.5) or biosynthetic nutrient (PM 5; Section 4.2.6). In total, the *cfa*, *ycbS* and *ylcG* genes were isolated from 275 of the 576 wells in plates PM 3 to PM 8.

The literature contains little information about Cfa, YcbS and YlcG. Before speculating on their roles in enhancing the growth phenotypes in *E. coli* (Section 4.3.3.1), summaries of what is known about each are provided in Sections 4.3.2.1, 4.3.2.2 and 4.3.2.3.

4.3.2.1 Cfa

One of the three ORFs isolated multiple times was *cfa*, which codes for cyclopropane fatty acyl phospholipid synthase (Keseler *et al.*, 2009). The Cfa enzyme methylates an unsaturated fatty acyl chain to produce a cyclopropane derivative of the phospholipid fatty acid (PCFA), a component of the membrane lipid bilayer (Figure 4.6). In total, *cfa* was isolated from 47 environments in PM 3 and PM 4 (in which the nitrogen source is varied).

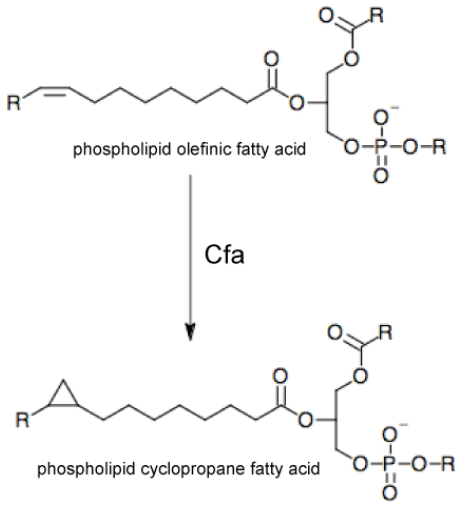


Figure 4.6: The reaction catalyzed by Cfa.

The unsaturated bond of the acyl chain of phospholipid olefinic fatty acid is methylated via a methyl group donated by *S*-adenosyl-L-methionine, hence forming a cyclopropane derivative of the fatty acid (PCFA).

The physiological role of the product of Cfa catalysis (PCFA), and hence Cfa itself, has not been elucidated, although PCFAs are normally found in cells entering stationary phase, or growing in adverse conditions (Grogan & Cronan, 1997). Cfa catalyzes the methylation of mature phospholipid fatty acids found in membrane bilayers, but not free fatty acids nor intermediates of phospholipid biosynthesis. Previously, *E. coli* mutants lacking the *cfa* gene were shown to be more susceptible to acid shock (Chang & Cronan, 1999). *E. coli* cells over-expressing Cfa did not appear to possess any growth defect (Grogan & Cronan, 1984).

4.3.2.2 YcbS

ASKA-encoded *ycbS* was the most over-represented ORF isolated from PM 3 to PM 8. The *ycbS* gene constituted 35% of all ASKA ORFs isolated, and was enriched from 136 environments in PM 3 to 8. The YcbS protein is fairly large (866 amino acids), and is part of the *ycbQRSTUVF* operon in *E. coli* (Keseler *et al.*, 2009). The operon codes for a set of putative chaperone-usher fimbrial proteins, which comprise a type of bacterial secretion system for transporting substrates across their cell membranes (Rêgo *et al.*, 2010). The protein structure predicted for YcbS (Figure 4.7A) showed that it is likely to be an outer membrane β-barrel protein. The closest homologue of YcbS is FimD, which shares 48% sequence identity. FimD, the usher protein for the fimbriae assembly complex (Figure 4.7B), forms a pore in the outer membrane to allow translocation of fimbrial subunits to the cell surface (Keseler *et al.*, 2009).

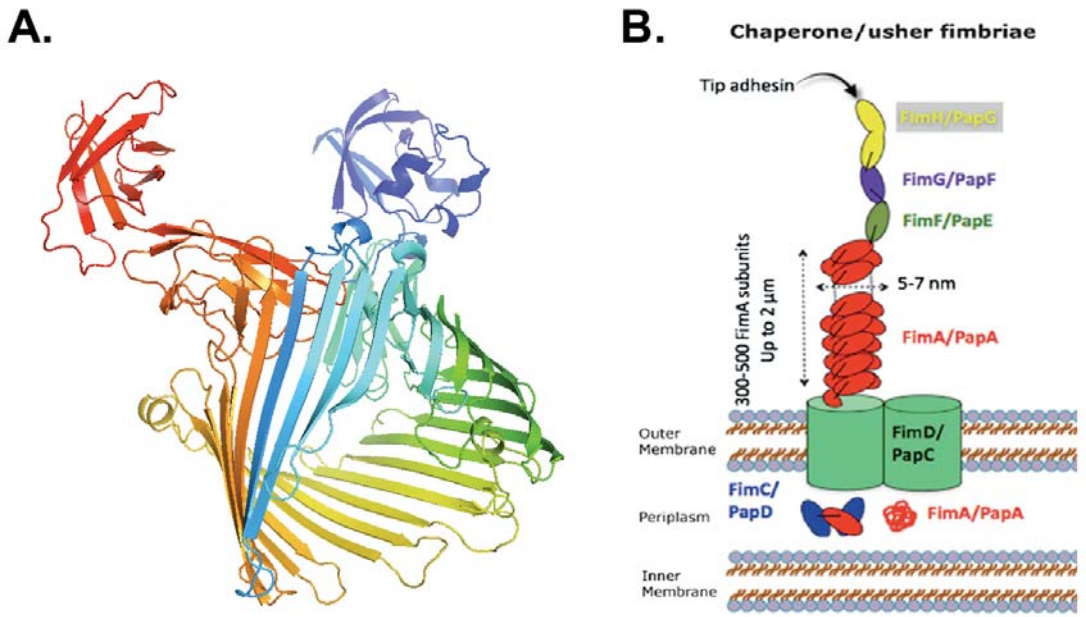


Figure 4.7: YcbS, and the Fim assembly protein complex in *E. coli*.

(A) A predicted protein structure for YcbS. The structure was modelled using the Phyre[2] server (Kelley & Sternberg, 2009), in October 2011. 773 residues of YcbS were modelled with 100% confidence based on the structure of the Fim outer membrane usher protein, FimD (PDB code: 3rfz; Chain B). (B) The Fim protein complex. FimD is an outer membrane usher protein that forms pore to allow the passage of the individual fimbrial subunits (FimA, FimF, FimG and FimH). The fimbrial subunits form the fimbrial rod on the surface of *E. coli*. Figure B was adapted from Korea *et al*. (2011).

Under standard culturing conditions (*i.e.*, LB as growth medium, 37°C incubation), Korea and her colleagues (2010) observed that the *ycb* operon in *E. coli* is cryptic. However, the insertion of a constitutive promoter in front of the *ycb* operon resulted in increased biofilm formation on abiotic surfaces (Korea *et al.*, 2010). The conclusion was that the activated *ycb* operon codes for functional fimbriae. In the same study, the short and rigid Ycb fimbriae were proposed to function as adhesins, thereby allowing the Ycb-expressing cells to adhere to polysterene surfaces 7-fold more than cells that did not express the operon.

4.3.2.3 YLCG

Unlike Cfa and YcbS, YlcG is a prophage-derived protein. ASKA-encoded *ylcG* constituted 25% of all the ASKA ORFs isolated from PM 3 to PM 8, and was isolated from 96 environments. In contrast to YcbS, YlcG is a very small protein (46 amino acids) with unknown function. Predicting the structure (Figure 4.8) and function of YlcG were not possible, in light of the low sequence and structural homologies with existing proteins. A previous study demonstrated that YlcG was mildly expressed in *E. coli* during stationary phase (Hemm *et al.*, 2008). The same study also predicted that YlcG is likely to contain a

single transmembrane helix. A lab member has attempted to express and purify the ASKA-coded YlcG but the protein was largely insoluble (Ilana Gerber, personal communication). Recently, *E. coli* cells harbouring the ASKA-encoded *ylcG* were shown to have a reduced glycogen content (Eydallin *et al.*, 2010), which led to a proposal that YlcG is involved in regulating carbon metabolism.

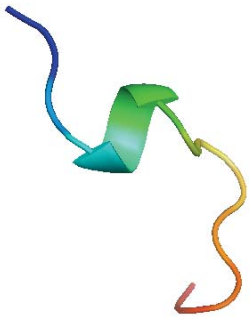


Figure 4.8: A predicted protein structure for YlcG.

The structure was modelled using the Phyre[2] server (Kelley & Sternberg, 2009), in October 2011. Only 14 residues of YlcG were modelled with 6.9% confidence based on the structure of a putative Blasticidin S deaminase from *Bacillus anthracis* (PDB code: 3b8f; Chain B).

4.3.3 THIAMINE AUXOTROPHY IN *E. COLI* DH5α-E

The over-representation of *cfa*, *ycbS* and *ylcG* in PM 3 to PM 8 hinted that there may be a drawback in the experimental design. The first clue to the drawback was offered by the overall growth patterns of the ASKA pool and the negative control in PM 5. This plate is normally used to test the nutrient requirements of auxotrophic strains. Three of the wells in PM 5 are supplemented with thiamine and its derivatives: well G6 contains thiamine and inosine; G7 contains thiamine; and G8 contains thiamine pyrophosphate. In duplicate screens, no difference was noted between the growth of the ASKA pool and the negative control in these three wells (Figure 4.9). In particular, the growth of both the ASKA and negative control populations was fairly rapid in all three thiamine-containing wells. The tetrazolium dye turned purple in these wells after only 1–2 days of incubation. In comparison, growth in other wells was undetectable until ≥3 days of incubation (see Table 4.5 in Section 4.2.6). These observations led me to postulate that the cells themselves — regardless of whether they were harbouring an ASKA plasmid or an empty vector — were starving from the lack of thiamine in their growth environment.

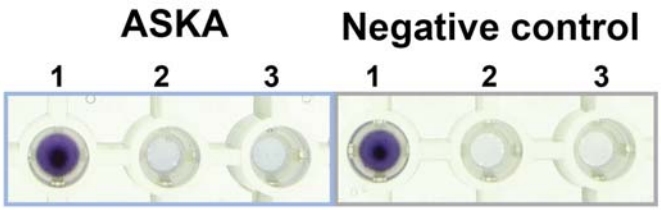


Figure 4.9: The growth of the ASKA pool and the negative control in wells containing thiamine and its derivatives, after 2 days of incubation.

Well 1 (G6) contained thiamine and inosine; Well 2 (G7) contained thiamine alone; Well 3 (G8) contained thiamine pyrophosphate. The growth patterns shown here were from one of the duplicate screens.

The host strain used for the entire library-on-library screen, and the serial enrichment experiment, was *E. coli* DH5α-E (Invitrogen), which has a *thi*−1 mutation (Section I.3 in Appendix I). Strains carrying this mutation do not require nutrient supplementation when grown in rich media, but their growth in minimal media must be supplemented with thiamine (Casali, 2003). However, at the outset of the screen, the formulations of the pre-made base medium used in PM 3 to PM 8 were not known, as these were proprietary information (Biolog Inc., 2011). Throughout this study, the DH5α-E strain was grown without thiamine supplementation.

The manufacturer of the PM plates, Biolog Inc., has recently revealed that the base medium of PM 3 to PM 8 is similar — but not identical — to the defined medium previously reported by Zhou *et al*. (2003). The complete version of this medium is 100 mM NaCl, 30 mM triethanolamine HCl (pH 7.1), 5 mM NH_4Cl, 2 mM NaH_2PO_4, 0.25 mM Na_2SO_4, 50 μM $MgCl_2$ and 1 mM KCl. As recommended by Biolog Inc., a mixed carbon source (20 mM sodium succinate and 2 μM ferric citrate) was always added to the base medium to support cellular growth. Overall, the growth medium used for PM 3 to PM 8 does not appear to contain any trace elements or vitamins, such as thiamine. Therefore, the absence of thiamine could be a major selective pressure for the repetitive enrichment of *cfa*, *ycbS* and *ylcG*.

The thiamine auxotrophy of the DH5α-E strain was not apparent when the screens and serial transfers were conducted. One would expect that the auxotrophy could be easily complemented by the over-expression of a thiamine transporter or enzymes involved in the thiamine biosynthesis. Yet, this was not observed. Only one thiamine-related ORF was isolated from the entire screen. The ORF for thiamine phosphate synthase [which catalyzes the penultimate step in thiamine biosynthesis (Keseler *et al*., 2009)], *thiE*, was enriched from a well containing adenosine (PM 5). Other ORFs that might encode for enzymes with low thiamine synthesis activity, such as YjbQ with promiscuous ThiE activity (Morett *et al*., 2008), were not isolated either. Therefore, the under-represented thiamine-related ORFs implied that the over-expression of thiamine transporter proteins and thiamine biosynthetic

enzymes are likely to be costly in *E. coli*, especially in the presence of a poor carbon source (*i.e.*, succinate) and the omission of crucial nutrient sources from the base medium (nitrogen in PM 3, PM 6, PM 7 and PM 8; phosphorus in PM 4; sulfur in PM 4).

The nature of the *thi*–1 mutation in the DH5α-E strain would be informative for predicting which ASKA-encoded ORF would complement the auxotrophy. However, no information is known about the *thi*–1 mutation (The Coli Genetic Stock Center). Since the DH5α-E strain in our lab had been commercially obtained from Invitrogen (Carlsbad, CA), it was difficult to trace the origins of the strain. According to the records provided by The Coli Genetic Stock Center, it is likely that the DH5α-E strain is a descendant of the MM294 strain (Meselson & Yuan, 1968). The latter had been used to construct DH1 and DH5 (Hanahan, 1983), which were presumably the ancestors for the DH5α-E strain. Notably, the thiamine auxotrophy (resulting from the *thi*–1 mutation) had been observed and reported for the MM294 strain (Meselson & Yuan, 1968). Similar to other bacterial gene operons, the expression of the thiamine operon (*thiCEFSGH*) in *E. coli* is tightly coordinated (Jurgenson *et al.*, 2009). Therefore, over-expression of a single *thi* ORF, such as *thiE*, is unlikely to derepress the expression of other chromosomal genes in the thiamine operon.

In hindsight, thiamine should have been added to every well of PM 3 to PM 8, to complement the auxotrophy of the DH5α-E strain. A concentration of 1 μg/mL thiamine (Meselson & Yuan, 1968) should be sufficient to support growth of *E. coli* strains with the *thi*–1 mutation. Similarly, thiamine should have been added to the Inoculation Fluid (IF) media (Biolog Inc.), which were used to resuspend the ASKA pool and the negative control prior to inoculation in the PM wells (see Section 4.4.3 for details). Although the formulations of the IF media are proprietary (Biolog Inc., 2011), it is likely that thiamine is absent from the media.

4.3.3.1 Speculation on the effects of Cfa, YcbS and YlcG

The repetitive enrichment of *cfa*, *ycbS* and *ylcG* from PM 3 to PM 8 implied that the over-expression of these proteins was a generic fitness-enhancing response in *E. coli* when thiamine was absent. In particular, clones over-expressing either Cfa, YcbS or YlcG were isolated from several environments where no growth was expected (and indeed, no growth was observed in these PM wells that had been inoculated with the negative control clone). For instance, the ASKA-encoded *ycbS* was enriched from a peptide-less well in PM 8 (termed the negative control well in Table 4.2 of Section 4.2.4); whereas *ylcG* was enriched from a peptide-less well in PM 6 (termed the negative control well in Table 4.2 of Section 4.2.4). Taken together, the over-expression of Cfa, YcbS or YlcG appeared to confer new phenotypes on *E. coli* in the absence of not only thiamine, but also major nutrient sources, such as nitrogen.

The information summarized in Sections 4.3.2.1 to 4.3.2.3 allowed me to speculate the roles of Cfa, YcbS and YlcG in improving the growth of *E. coli* in environments with minimal amounts of nutrients (*e.g.* the wells of PM 3 to PM 8). First, Cfa was likely to buffer *E. coli* from starvation of both thiamine and other nutrients, since the conditional formation of PCFAs, the products of Cfa catalysis (see Figure 4.6 in Section 4.3.2.1), has been proposed as a response to an exhausted metabolic state of the cells (Grogan & Cronan, 1997). On the other hand, increased levels of YcbS might promote the formation of fimbriae in *E. coli* by allowing more fimbrial subunits to traverse the outer membrane. Eventually, this might lead to increased cell adherence and biofilm formation in the PM wells. The ability of cells to withstand adverse conditions in the biofilms is well documented in the literature. For YlcG, its potential in regulating carbon metabolism (Eydallin *et al*., 2010) might translate as a fitness advantage, particularly when cells were struggling to adapt to a poor carbon source (*i.e.*, succinate) and were starving from a lack of thiamine. Overall, the (known) functions of Cfa, YcbS and YlcG suggest they can act as general stress responses against nutrient starvation. The likelihood of any of these three ORF-expressing clones being enriched in a particular well (of PM 3 to PM 8) was probably determined by sampling biases, such as the degree of cell resuspension prior to serial transfers.

Alternatively, over-expression of Cfa, YcbS or YlcG might improve the scavenging capabilities of the *E. coli* clones. Recall that about ~50,000 cells were competing for nutrients against one another in every well. Apart from the possibility of being able to grow faster than the neighbouring clones, a clone could also scavenge nutrients, such as thiamine, from dead neighbours in order to survive. The ability to prolong survival would seem reasonable in a life-or-death situation. However, confirmation of such scavenging capabilities would require different approaches. The library-on-library screens and the serial enrichment described in this chapter were designed to only differentiate fit clones (such as those over-expressing Cfa, YcbS and YlcG), from unfit clones.

4.3.4 ORFs ENRICHED FROM THIAMINE-CONTAINING ENVIRONMENTS

In spite of the issues around thiamine auxotrophy, the experiments described in this chapter still revealed many examples of apparent promiscuity. Enriched ORFs from PM 1, PM 2, PM 9 and PM 10 appeared to be more specific to individual environments. This was because the base medium used for PM 1, PM 2, PM 9 and PM 10 was derived from complex nutrients: tryptone and yeast extract [see Sections 4.2.3.1 and 4.2.7 for medium formulation; Zhou *et al*. (2003)]. In particular, thiamine could be derived from the yeast extract in the medium. Therefore, *E. coli* DH5α-E was not starved for thiamine in these plates.

In one case, *E. coli* cells over-expressing BglA appeared to grow more rapidly when glucuronamide was present as the sole carbon source (Table 4.1 in Section 4.2.3.1). Confirmation of the improved growth phenotype of the BglA-expressing clone was

subsequently demonstrated in Section 4.2.3.2. However, biphasic growth curves were observed for the BglA-expressing clones and the negative control grown using glucuronamide as the carbon source (Figure 4.5 in Section 4.2.3.2). These results suggested that trace amounts of a richer nutrient source were present. This was likely to be glycerol, which is a good carbon source for *E. coli*. Glycerol could have been carried over from the M9-glycerol media used to pre-condition the cells to the M9 medium, prior to growth in M9-glucuronamide (see Section 4.4.5 for details). However, the presence of mixed carbon sources in the M9-glucuronamide media was analogous to the growth environments where ASKA-encoded *bglA* was originally enriched. Recall that the base medium in every well of PM 1 consisted of tryptone and yeast extract. Carbon sources, such as glucose, are present in yeast extract and could contribute to basal levels of growth in PM wells. Thus, BglA appears to be a genuine example, where over-expression imparted a new growth phenotype in *E. coli*, *i.e.*, the ability to grow using glucuronamide as a carbon source. The implication was that BglA, a glucosidase known to hydrolyze phosphorylated glucosides only (Prasad *et al.*, 1973; Hall *et al.*, 1983), could also recognize glucuronamide as a substrate. As glucuronamide does not contain a glycosidic bond (Figure 4.3C), BglA cannot be acting as a glucosidase. Instead, it is tempting to speculate that BglA could possess glycosyltransferase activity, and that this is the key reaction for allowing growth on glucuronamide. Glycosidases, an enzyme family to which glucosidase belongs, are known to catalyze the formation of glycosidic linkages using monosaccharides and an excess of acceptor molecule (Hancock *et al.*, 2006). Glucuronamide could be used as an acceptor molecule in this case to form glycosides, which in turn could be metabolized. Further experiments will be required to confirm this tentative hypothesis.

In addition to BglA, AgaX also appeared to be an interesting case since *E. coli* that were over-expressing AgaX managed to out-grow the negative control in six different carbon sources (Table 4.1 in Section 4.2.3.1). Bioinformatic analysis of the *agaX* gene predicted that AgaX is a component of an uncharacterized phosphotransferase system (Keseler *et al.*, 2009), which is normally involved in the uptake and phosphorylation of a variety of substrates. The enrichment of the ASKA-encoded *agaX* from six different carbon-containing environments suggests that AgaX is likely to show broad substrate specificities. Thus, it is appealing to hypothesize that AgaX is indeed part of the uncharacterized phosphotransferase system, which may also show broad substrate specificities.

Another interesting example isolated from an environment with moderately high osmolyte concentration (3% sodium formate) was TesB (Table 4.6 in Section 4.2.7). TesB is a thioesterase that cleaves long-chain acyl-CoA into coenzyme A and a carboxylate (Keseler *et al.*, 2009). Nie and her colleagues (2008) recently reported that *E. coli* TesB also cleaves a variety of medium-chain (C6–C16) acyl-CoA substrates, which indicated that the enzyme does not seem to discriminate against the chain length of acyl-CoA. However, sodium formate (HCOONa) does not have a CoA, nor an aliphatic chain. One possible scenario that

could contribute to the enrichment of *tesB* was that *E. coli* could metabolize sodium formate into a compound that resembles a long/medium chain acyl-CoA. Therefore, TesB, which appears to show broad substrate specificities, may react with the compound derived from sodium formate.

4.3.5 PROTEIN OVER-EXPRESSION CAN MEDIATE THE EMERGENCE OF NEW PHENOTYPES

As shown in Chapter 2 and 3 (as well as several references therein), over-expression of pre-existing proteins in *E. coli* can rescue auxotrophic strains. In this chapter, I demonstrated that over-expression of pre-existing proteins can also drive the emergence of new phenotypes in *E. coli*. For instance, an alternative carbon source, glucuronamide, could be utilized in cells that over-expressed BglA. This suggests that *E. coli* with increased levels of BglA could also utilize glucuronamide as the carbon source in natural settings. Instead of IPTG induction, gene amplifications or regulatory mutations could provide a means to increase enzyme levels in the cells. Similarly, the over-expression of Cfa, YcbS and YlcG could be a means for *E. coli* to adapt to thiamine starvation (and possibly, other nutrient starvation), since cells over-expressing these proteins displayed enhanced growth phenotypes. Taken together, the protein over-expression mechanism represents a viable solution for *E. coli* in the course of adapting to nutritionally challenging environments. Another common scenario for adaptive evolution is to evolve resistance to antibiotics and toxins. This will be explored further in Chapter 5.

4.4 Materials and Methods

All reagents were from Sigma, unless otherwise indicated. Specific materials and methods are described in the following subsections, while general molecular biology materials and techniques are explained in Appendix I.

Sodium succinate was prepared as a 2 M stock in water, and ferric citrate was prepared as a 0.2 mM stock in water. D-Glucuronamide (Chemos GmbH) was prepared as a 10% (w/v) stock in water. All three reagents were filter-sterilized before use.

4.4.1 Construction of the ASKA library in DH5α-E

All 5,272 plasmids from the ASKA collection had been pooled previously (Patrick *et al.*, 2007). Approximately 10 ng of these pooled plasmids were used to transform an aliquot of *E. coli* DH5α-E. The transformed cells were grown on LB-chloramphenicol plates at 28°C. To enumerate the library size, dilutions of the transformed cells were plated on separate LB-chloramphenicol plates. After 20 h of incubation, a total of ~723,600 cfu were obtained. The vast over-sampling ensured that >99% of the ASKA clones were represented [calculated by the library diversity and completeness programme, GLUE (Patrick *et al.*, 2003)].

All clones were scraped off the agar plates, and pooled in 2.5 mL LB supplemented with chloramphenicol (34 μg/mL) and glycerol [12.5% (v/v)]. The pooled cell suspension was stored in aliquots at −80°C.

4.4.2 Construction of the control clone

A negative control was constructed, for comparison with the ASKA pool (Section 4.4.1). The same host strain, *E. coli* DH5α-E, was transformed with pCA24N-NoIns (Section 2.4.1) to create the negative control. The plasmid pCA24N-NoIns shares the same vector backbone as all the ASKA plasmids, but lacks a gene insert.

4.4.3 Library-on-library screen

Overnight cultures of the control clone and ASKA pool were prepared by inoculating 10 mL LB-chloramphenicol with 10 μL of the respective frozen stocks. A 1-mL aliquot of each overnight culture was used to inoculate 50 mL LB-chloramphenicol. Both cultures were grown at 37°C to mid-log phase (OD_{600} ~0.5) before adding 50 μM IPTG to induce protein over-expression. After induction for 2 h, cells were harvested by centrifugation (2,000 ×g, 15 min, 4°C). The cell pellets were resuspended to OD_{600} ~0.2 in sterile water, and 100 μL of each cell suspension was added to the Inoculation Fluid (IF) medium (Biolog Inc.) supplemented with 50 μM IPTG. Assuming that OD_{600} corresponded to 2.5×10^8 cells/mL

(Sambrook & Russell, 2001), the final cell density in the IF medium was ~5 × 10^5 cells/mL. The cell suspensions were starved by incubating at room temperature for 1 h. The starvation step was critical for two reasons: (i) to exhaust rich nutrient sources (carbon, nitrogen, phosphorus, sulfur, vitamins, *etc*.) carried over from the previous culture medium (LB), as well as any metabolite reserves in the cells; and (ii) to ensure brutal selection pressure for the cells to metabolize the exogenous compounds in each PM well.

After starvation, aliquots of each culture (100 μL) were used to inoculate the wells of PM 1 to PM 10 (Biolog Inc.). The PM plates were placed in airtight containers with moistened paper towels (to prevent evaporation), and incubated at 37°C for 7 days. Growth in each well, as indicated by purple colour development, was monitored visually and scored daily. Two independent replicates of the complete screen were performed.

According to Biolog Inc., the IF medium is a balanced salt solution with a gelling polysaccharide that keeps the microbial cells in suspension in the wells. As recommended, the IF-0 medium was used for PM 1 to PM 8, whereas the IF-10 medium was used for PM 9 and PM 10. All IF media contained 1× IF and 1× Biolog Redox dye mix.

Another recommendation by Biolog Inc. was the addition of a mixed carbon source to the minimal, defined base medium of PM 3 to PM 8. Sodium succinate (final concentration, 20 mM) and ferric citrate (final concentration, 2 μM) were added to the cell suspensions after the starvation step (and prior to PM well inoculation). Although *E. coli* does not normally utilize citrate as a carbon source (Blount *et al*., 2008), the low amount of citrate could potentially act as an anti-coagulant by chelating divalent ions, such as Ca^{2+}, present in the well of the PM plates.

4.4.4 Isolation of winner genes via serial enrichment

For conditions in which the ASKA pool had reproducibly out-grown the negative control, fresh aliquots of the pool were now used to inoculate the corresponding wells of new PM plates. At the first sign of purple colour development in a well (usually 1–5 days), a 2-μL aliquot of the culture from that well was mixed with 98 μL IF medium (supplemented with 50 μM IPTG) and transferred to the corresponding well of a new plate. For PM 3 to PM 8, the IF medium was also supplemented with 20 mM sodium succinate and 2 μM ferric citrate. After a second transfer (carried out in an identical fashion), cells from the final passage were spread on LB-chloramphenicol plates. The universal pCA24N-specific primers, pCA24N.for and pCA24N.rev (Patrick *et al*., 2007), were used to amplify the ASKA ORFs (see Section I.10) from at least eight of the resulting colonies. The thermocycling parameters were as follows: 95°C for 5 min, 30 cycles of 95°C for 30 s, 50°C for 30 s, 72°C for 2 min, and finally, 72°C for 5 min. PCR products were sequenced to reveal each ORF that was responsible for enhanced growth.

4.4.4.1 Probability of sampling the same ORF twice or more, by chance alone

Given that there were 5,272 ORF variants in the ASKA collection, the probability of sampling the same ORF twice or more, in eight random clones is small.

First, consider the probability that all eight clones are *unique* (*i.e.*, different from each other):

The first isolated clone is always *unique*, so $P = 1$.

The probability of the second clone being different from the first clone is:

$$P = \frac{5{,}271}{5{,}272}$$

Similarly, the probability of the third clone being different from both the first and second clones is:

$$P = \frac{5{,}270}{5{,}272}$$

Therefore, the overall probability of getting eight *unique* clones is:

$$P = 1 \times \frac{5{,}271}{5{,}272} \times \frac{5{,}270}{5{,}272} \times \frac{5{,}269}{5{,}272} \times \frac{5{,}268}{5{,}272} \times \frac{5{,}267}{5{,}272} \times \frac{5{,}266}{5{,}272} \times \frac{5{,}265}{5{,}272}$$
$$P = 0.9947$$

That is, there is a 99.47% chance that all eight clones, which are randomly sampled from the pool of 5,272 variants, carry different ASKA ORFs. Hence, the probability that all eight clones are not all different (*i.e.*, two or more are the same) is:

$$P = 1 - 0.9947$$
$$P = 0.0053$$

That is, by chance alone, the probability of isolating two clones harbouring identical ASKA-encoded ORFs was only 0.53%.

4.4.5 *E. coli* growth using glucuronamide as a carbon source

As described in Section 2.4.4.1, the 1× M9 broth was diluted from a 5× M9 stock. The broth was also supplemented with 0.5% (v/v) carbon source (either glycerol or glucuronamide) and 34 μg/mL chloramphenicol. If necessary, IPTG (50 μM, final concentration) was added to the M9 medium to induce protein over-expression.

The growth of *E. coli* cells in every test condition was measured in eight replicates, in microtitre plates. First, freezer stocks of both the test strain and the negative control were

used to inoculate 1.5 mL LB-chloramphenicol, respectively. The cells were grown to saturation (~6.5 h), and were pelleted at 10,000 ×g for 5 min in a benchtop microcentrifuge. The cell pellets were washed once with 1 mL of 1× M9 broth, and resuspended in 1 mL of 1× M9 broth. To pre-condition the cells to the M9 medium, 5 μL of each resuspended culture were used to inoculate 5 mL of M9-glycerol, and the cultures were grown at 37°C for 16 h. The overnight cultures were then diluted 10-fold, and 1 μL of each diluted culture was used to inoculate 200 μL M9 supplemented with either glycerol or glucuronamide as the sole carbon source. Cell growth was monitored at OD_{600} using a Synergy 2 plate reader (Biotek Instruments), which was programmed to agitate continuously at 37°C. The OD_{600} readings were obtained every 10 min, for 24 h.

CHAPTER 5

Toxin resistance mediated by protein over-expression

Sections 5.2, 5.3 and 5.4 contain published materials from:

> **Soo, V. W. C., Hanson-Manful, P. & Patrick, W. M.** (2011). Artificial gene amplification reveals an abundance of promiscuous resistance determinants in *Escherichia coli*. *Proc Natl Acad Sci USA*, **108**:1484-1489.

Acknowledgements: Ilana Gerber assisted with half of the PCR screens and purification of PCR products described in Section 5.4.1. Susan Morton, Laura Nigon and Paulina Hanson-Manful provided technical assistance in the pairwise fitness assays (Section 5.4.3). Paulina Hanson-Manful also tested the antibiotic susceptibilities of the ASKA clones in benzethonium chloride, oxytetracycline, puromycin, sisomicin and spiramycin (Section 5.4.4).

5.1 INTRODUCTION

The library-on-library screen described in Chapter 4 enabled a global search for promiscuous proteins that could mediate the evolution of novel phenotypes in *E. coli*, provided protein expression was up-regulated. This approach yielded many cases where over-expression of pre-existing proteins led to enhanced cellular growth in the presence of uncommon nutrients. However, the results were also dominated by a handful of proteins that appeared to mediate non-specific stress responses. In this chapter, I aimed to avoid some of the pitfalls encountered in Chapter 4. I utilized the same screening strategy to identify *E. coli* proteins that could impart resistance to the cells upon exposure to antibiotics and toxins. In contrast to common pathogenic bacteria, the laboratory strain of *E. coli* used for the screen has not been exposed to any clinical settings that may favour the evolution of resistance. Therefore, this library-on-library screen tested the ability of the *E. coli* proteome to mediate the *de novo* evolution of toxin resistance.

5.1.1 MOLECULAR ORIGINS OF ANTIBIOTIC RESISTANCE

Antibiotic resistance in bacterial populations is an outcome of successful adaptation to a toxic environment. From an evolutionary perspective, resistant clones are fitter than susceptible clones because they have gained an improved capacity for survival and reproduction in the presence of a toxin. This can be made possible by one or more of the following physiological alterations: (i) enzymatic inactivation or degradation of the antibiotic; (ii) reduction of intracellular antibiotic concentrations by drug transporters and pumps; (iii) mutating antibiotic targets in such a way that they are no longer inhibited by the antibiotic; and (iv) the appearance of compensatory mutations in response to costly resistance mechanisms. The genetics and biochemistry of these resistance mechanisms have been the subjects of extensive research for the past 70 years, and have been reviewed widely in the literature (Levy & Marshall, 2004; Courvalin, 2008; Sandegren & Andersson, 2009; Davies & Davies, 2010).

The need to understand the diverse nature of resistance, and its rapid spread, has spurred recent efforts to focus on the origins and evolution of resistance proteins. Protein structure and biochemical analyses have revealed that many modern resistance proteins are evolutionarily related to proteins that are not involved in drug resistance (Wright, 2007; Morar & Wright, 2010). For example, β-lactamases (the penicillin-hydrolyzing enzymes) are likely to have evolved from bacterial DD-peptidases, which are involved in peptidoglycan biosynthesis and maintenance (Kelly *et al.*, 1986; Urbach *et al.*, 2008). One of the aminoglycoside-inactivating enzymes, aminoglycoside kinase (APH), shares high structural similarities with Ser/Thr/Tyr protein kinases, and this observation led to the hypothesis that both types of kinases share a common ancestor (Hon *et al.*, 1997; Daigle *et al.*, 1999; Fong *et al.*, 2011). However, the origins of resistance proteins with broad substrate

specificities (also known clinically as "super" resistance proteins) are less clear. For instance, an aminoglycoside acetyltransferase (AAC) variant from clinical bacterial isolates was able to acetylate, and therefore reduce the activities of, both aminoglycosides and fluoroquinolones (Robicsek *et al.*, 2006). The aminoglycosides and fluoroquinolones belong to two separate compound classes with distinctive functional groups and modes of inhibition. In a separate example, *E. coli* and *Klebsiella pneumoniae* strains that caused an epidemic outbreak in 2010 are now known to have been expressing the NDM-1 metallo-β-lactamase, which shows activities for a broad range of β-lactams, including the last-line antibiotic for β-lactam-resistant infections, the carbapenems (King & Strynadka, 2011; Thomas *et al.*, 2011). Resistance proteins with secondary activities are not only found in clinical variants, but also in environmental isolates. The discovery of a bifunctional β-lactamase from a remote Alaskan soil sample suggested that resistance proteins have been evolving in the environment for a long time (Allen *et al.*, 2009). This notion was further supported by a study that unearthed several 30,000-year-old resistance genes from permafrost sediments (D'Costa *et al.*, 2011). The evolution of resistance in the environment is a natural response to the diversity of antibiotics produced by antibiotic-producing microorganisms. Overall, these examples imply that resistance proteins with secondary activities are not uncommon.

Wright (2007) proposed that it is highly likely that modern resistance proteins have originated from ancestral metabolic proteins, which had fortuitous activities towards and/or affinities for antibiotics. Over time, these promiscuous activities evolved to provide robust resistance mechanisms as a result of antibiotic exposures from the clinical (Rice, 2009) and environmental (Fajardo & Martínez, 2008; Martínez, 2008) reservoirs. This model is identical to the one discussed previously for the evolution of new biochemical functions from enzymes with secondary activities [Jensen (1976); Bergthorsson *et al.* (2007); see also Section 4.1.1 in Chapter 4]. Apart from allowing *E. coli* to grow better in environments containing new nutrients (Chapter 4), increases in gene dosage, or protein expression, could potentially confer new resistance phenotypes on bacteria. Therefore, in this chapter, I set out to catalogue all the possible resistance activities in an *E. coli* proteome using the library-on-library screen described previously (Chapter 4). This experiment would offer insights into all the resistance pathways (towards all the antibiotics) that a modern proteome can facilitate. I provide experimental evidence that the over-expression of pre-existing *E. coli* proteins can provide resistance to >80 antibiotics and toxins, through a variety of proposed mechanisms. The results suggest that the evolution of resistance is surprisingly likely, and that even the proteome of a non-pathogenic bacterium harbours substantial reservoirs of latent resistance proteins.

5.2 RESULTS

5.2.1 LIBRARY-ON-LIBRARY SCREEN (PM 11 TO PM 20)

The library-on-library screen described in Chapter 4 was used to uncover latent resistance proteins in the *E. coli* proteome. The conceptual and methodological details of the screen has been described in Sections 4.2.1 and 4.4.3, respectively. Briefly, *E. coli* cells were transformed with the pooled plasmids of the ASKA library, and then used to inoculate every well of PM 11 to PM 20 (Figure 4.1 in Section 4.2.1, Chapter 4). In total, these ten PM plates contained 237 toxins (including anti-bacterial agents), with each toxin present at four concentrations (Bochner *et al.*, 2001). Similar to PM 1 to PM 10 (Chapter 4), growth in each toxin-containing well was monitored by purple colour development, due to the presence of a tetrazolium indicator dye. The negative control (*E. coli* harbouring the empty ASKA vector, pCA24N-NoIns) was used to inoculate a second set of PM plates. By comparing the growth of the ASKA pool and the control in parallel, I was able to screen for examples in which the over-expression of ASKA-encoded ORFs conferred increased fitness (*i.e.*, an improved capacity for survival and reproduction) in the presence of a toxin.

5.2.1.1 CASES WHERE THE ASKA POOL OUT-GREW THE NEGATIVE CONTROL

In duplicate screens, the ASKA pool out-grew the negative control in the presence of 99 toxins (42% of those tested). In the majority of cases, these improved growth phenotypes manifested as small but reproducible changes in the intensity of tetrazolium colour development, within the first 48 h of incubation (Table 5.1). Specifically, the ASKA pool showed enhanced growth in 50 of the 78 anti-bacterial agents (64%) contained within the PM plates. The improved growth phenotypes was also noted for 49 of the other 159 toxins that were tested (31%). These included anti-fungal and anti-parasitic agents, inorganic salts and dyes (Table 5.1).

Table 5.1: The complete list of over-expressed genes isolated from 99 toxin-containing environments, and the proposed mechanism by which each gene product may act to impart resistance to the corresponding toxin.

Toxic compound	PM well*	Days of inc.†	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product	Proposed mechanism	Reference(s)
β-Lactams									
Amoxicillin	PM11, B4	1	Medium	Clear	None isolated	N/A			
Aztreonam	PM18, F11	1	Very dark	Dark	*yeaD*	2	Putative mutarotase	Catalytic promiscuity and/or substrate ambiguity	(Chittori *et al.*, 2007)
					ycgZ	3	Small protein induced by YcgF	Stress response	(Tschowri *et al.*, 2009)
					rbsR	2	Ribose transcriptional repressor	Regulatory effect	(Keseler *et al.*, 2009; The UniProt Consortium, 2010)
Carbenicillin	PM14, G6	1	Medium	Light	None isolated	N/A			
Cefazolin	PM11, E4	1	Very dark	Light	*bdm*	8	Biofilm-dependent modulation protein	Biofilm formation & regulation	(Keseler *et al.*, 2009; The UniProt Consortium, 2010)
Cefmetazole	PM15, A12	2	Medium	Clear	*ycgZ*	8	Small protein induced by YcgF	Stress response	(Tschowri *et al.*, 2009)
Cefoperazone	PM17, G10	1	Medium	Clear	*bcr*	8	Bcr multidrug transporter	Efflux pump/transporter	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Cefuroxime	PM13, D4	1	Light	Clear	*ycgZ*	7	Small protein induced by YcgF	Stress response	(Tschowri *et al.*, 2009)
Cephalothin	PM11, H4	1	Light	Clear	*bdm*	8	Biofilm-dependent modulation protein	Biofilm formation & regulation	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Cloxacillin	PM11, B7	1	Medium	Clear	None isolated	N/A			
Moxalactam	PM13, H8	1	Medium	Clear	*ycgZ*	8	Small protein induced by YcgF	Stress response	(Tschowri *et al.*, 2009)
Nafcillin	PM11, D11	1	Dark	Light	*bdm*	3	Biofilm-dependent modulation protein	Biofilm formation & regulation	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)

Toxic compound	PM well*	Days of inc.†	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product	Proposed mechanism	Reference(s)
Oxacillin	PM12, B3	1	Medium	Clear	None isolated	N/A			
Penicillin G	PM12, A4	1	Dark	Clear	*bfd*	2	Bacterioferritin-associated ferredoxin	Stress response	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Phenethicillin	PM19, F4	1	Light	Clear	*ydaC*	5	Uncharacterized protein from Rac prophage	Prophage gene	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Piperacillin	PM14, F8	1	Dark	Light	*yeeX*	2	Uncharacterized cytoplasmic protein	Unknown	
					ugpC	2	Glycerol-3-phosphate ABC transporter; ATP-binding subunit	Efflux pump/transporter	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Aminoglycosides									
Apramycin	PM20, A8	1	Light	Clear	*yejG*	8	Uncharacterized protein	Unknown	
Dihydrostreptomycin	PM19, G8	1	Medium	Light	*cpdA*	8	cAMP phosphodiesterase	Regulatory effect	(Imamura *et al.*, 1996)
Paromomycin	PM12, C4	1	Light	Clear	None isolated	N/A			
Sisomicin	PM12, D3	1	Dark	Medium	*yejG*	6	Uncharacterized protein	Unknown	
Tobramycin	PM12, F3	1	Medium	Clear	*yejG*	3	Uncharacterized protein	Unknown	
Anti-folates									
Sulfachloropyridazine	PM17, C6	1	Light	Clear	*bcr*	8	Bcr multidrug transporter	Efflux pump/transporter	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)

Toxic compound	PM well*	Days of inc.†	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product	Proposed mechanism	Reference(s)
Sulfadiazine	PM12, E6	2	Light	Clear	*nudB*	3	Dihydroneopterin triphosphate pyrophosphatase	Maintaining metabolic flux	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
					luxS	2	S-ribosylhomocysteine lyase	Biofilm formation & regulation	(Li *et al.*, 2007)
Sulfamethazine	PM12, D6	2	Very dark	Dark	*rbsR*	5	Ribose transcriptional repressor	Regulatory effect	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
					ycgZ	3	Small protein induced by YcgF	Stress response	(Tschowri *et al.*, 2009)
Sulfamonomethoxine	PM17, C10	3	Light	Clear	*rbsR*	4	Ribose transcriptional repressor	Regulatory effect	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Sulfathiazole	PM12, F7	4	Light	Clear	*phnC*	2	Alkylphosphonate ABC transporter, ATP-binding subunit	Efflux pump/transporter	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
					bcr	6	Bcr multidrug transporter	Efflux pump/transporter	(Vedantam *et al.*, 1998)
Sulfisoxazole	PM18, C5	2	Light	Clear	*bcr*	8	Bcr multidrug transporter	Efflux pump/transporter	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Trimethoprim	PM16, B12	1	Light	Clear	*folA*	8	Dihydrofolate reductase	Overexpression of the drug's target	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Quinolones									
Ciprofloxacin	PM20, D8	1	Medium	Light	*marA*	8	MarA transcriptional activator	Regulatory effect	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Enoxacin§	PM11, E8	1	Light	Clear	*soxS*	8	SoxS transcriptional activator	Regulatory effect	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
					cynT	8	Carbonic anhydrase	Catalytic promiscuity and/or substrate ambiguity	(Gould & Tawfik, 2005)

Toxic compound	PM well*	Days of inc.†	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product	Proposed mechanism	Reference(s)
Nalidixic acid	PM11, E12	1	Medium	Light	*galE*	3	UDP-glucose 4-epimerase	Envelope / capsule biosynthesis	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
					yfdO	2	Uncharacterized protein from CPS-53 (KpLE1) prophage	Prophage gene	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Ofloxacin	PM11, H12	1	Light	Clear	*marA*	8	MarA transcriptional activator	Regulatory effect	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Tetracyclines									
Demeclocycline	PM11, D7	2	Light	Clear	None isolated	N/A			
Minocycline	PM11, C10	3	Light	Clear	None isolated	N/A			
Oxytetracycline	PM20, F8	1	Light	Clear	*bcr*	8	Bcr multidrug transporter	Efflux pump / transporter	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Penimepicycline	PM12, B6	3	Very dark	Light	*ycgZ*	2	Small protein induced by YcgF	Stress response	(Tschowri *et al.*, 2009)
					marA	2	MarA transcriptional activator	Regulatory effect	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Rolitetracycline	PM13, D11	1	Light	Clear	*bcr*	8	Bcr multidrug transporter	Efflux pump / transporter	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Macrolide-Lincosamide-Streptogramin									
Erythromycin	PM11, F8	3	Very dark	Dark	*ydaC*	3	Uncharacterized protein from Rac prophage	Prophage gene	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
					ydfW	2	Uncharacterized protein from Qin prophage	Prophage gene	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)

Toxic compound	PM well*	Days of inc.†	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product	Proposed mechanism	Reference(s)
Josamycin	PM19, A2	1	Medium	Clear	*yhfK*	2	Uncharacterized inner membrane protein	Unknown	
					cpdA	6	cAMP phosphodiesterase	Regulatory effect	(Imamura *et al.*, 1996)
Lincomycin	PM11, A12	2	Dark	Medium	*cmr*	4	MdfA/Cmr multidrug transporter	Efflux pump/transporter	(Edgar & Bibi, 1997)
					rbsR	3	Ribose transcriptional repressor	Regulatory effect	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Oleandomycin	PM15, F8	1	Light	Clear	*ydfW*	8	Uncharacterized protein from Qin prophage	Prophage gene	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Spiramycin	PM12, H3	1	Dark	Clear	*tusE*	4	Sulfurtransferase	Catalytic promiscuity and/or substrate ambiguity	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
					ydfW	2	Uncharacterized protein from Qin prophage	Prophage gene	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Tylosin	PM13, H12	1	Light	Clear	*marA*	7	MarA transcriptional activator	Regulatory effect	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Glycopeptides									
Phleomycin	PM15, D1	1	Medium	Light	None isolated	N/A			
Vancomycin	PM12, C8	2	Light	Clear	*mzrA*	8	Modulator of EnvZ and OmpR	Stress response	(Gerken *et al.*, 2009)
Nitrofurans									
5-Nitro-2-furaldehyde semicarbazone	PM15, E5	2	Light	Clear	*mdtM*	8	MdtM multidrug transporter	Efflux pump/transporter	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)

Toxic compound	PM well*	Days of inc.†	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product	Proposed mechanism	Reference(s)
Furaltadone	PM14, A6	3	Light	Clear	*yagT*	5	Aldehyde oxidoreductase, Fe-S subunit	Catalytic promiscuity and/or substrate ambiguity	(Neumann *et al.*, 2009)
					mdtM	3	MdtM multidrug transporter	Efflux pump/transporter	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Nitrofurantoin	PM14, E5	1	Light	Clear	*tfaX*	8	Uncharacterized protein from DLP12 prophage	Prophage gene	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Rifamycins									
Rifampicin	PM12, H7	2	Medium	Clear	*feoC*	2	Transcriptional regulator of ferrous iron transport	Regulatory effect	(Cartron *et al.*, 2006)
					tusE	6	Sulfurtransferase	Catalytic promiscuity and/or substrate ambiguity	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Rifamycin SV	PM16, E12	3	Light	Clear	None isolated	N/A			
Steroids antibacterials									
Fusidic acid	PM15, C8	2	Medium	Light	*hcaR*	7	Hca operon transcriptional activator	Regulatory effect	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Other toxins									
2,2'-Dipyridyl	PM13, B5	1	Light	Clear	*yqjD*	3	Uncharacterized protein	Unknown	

Toxic compound	PM well*	Days of inc.†	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product	Proposed mechanism	Reference(s)
2,4-Diamino-6,7-diisopropyl-pteridine	PM12, E4	1	Medium	Clear	*folA*	8	Dihydrofolate reductase	Catalytic promiscuity and/or substrate ambiguity	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
3,4-Dimethoxybenzyl alcohol	PM15, F3	2	Medium	Light	*csiE*	2	Stationary phase-inducible protein	Stress response	(Marschall & Hengge-Aronis, 1995)
3,5-Dinitrobenzene	PM20, G6	1	Medium	Light	*gmr*	2	Cyclic-di-GMP phosphodiesterase	Stress response	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
6-Mercaptopurine	PM13, C4	2	Light	Clear	*yrfG*	7	Purine nucleotidase	Catalytic promiscuity and/or substrate ambiguity	(Kuznetsova *et al.*, 2006)
9-Aminoacridine	PM14, B3	2	Medium	Light	*ycbS*	6	Uncharacterized outer membrane β-barrel protein	Efflux pump/transporter	(Zhai & Saier, 2002)
Amitriptyline	PM20, A3	2	Light	Clear	*galE*	8	UDP-glucose 4-epimerase	Envelope/capsule biosynthesis	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Antimony (III) chloride	PM18, E9	1	Medium	Light	*ykiA*	2	Uncharacterized protein	Unknown	
					csiE	3	Stationary phase-inducible protein	Stress response	(Marschall & Hengge-Aronis, 1995)
Benzethonium chloride	PM12, E10	3	Medium	Clear	*yhiK*	2	Uncharacterized protein	Unknown	
					puuD	2	γ-glutamyl-γ-aminobutyrate hydrolase	Catalytic promiscuity and/or substrate ambiguity	(Kurihara *et al.*, 2006)
					ybcD	2	Pseudogene from DLP12 prophage; predicted fragment of replication protein	Prophage gene	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Blasticidin S	PM19, F7	2	Medium	Light	*csiE*	3	Stationary phase-inducible protein	Stress response	(Marschall & Hengge-Aronis, 1995)

Toxic compound	PM well*	Days of inc.†	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product	Proposed mechanism	Reference(s)
					yjbI	2	Uncharacterized protein	Unknown	
Boric acid	PM14, C3	3	Dark	Medium	None isolated	N/A			
Caffeine	PM17, H7	1	Light	Clear	*ydaC*	7	Uncharacterized protein from Rac prophage	Prophage gene	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Chelerythrine	PM14, G4	1	Light	Clear	*ydaC*	6	Uncharacterized protein from Rac prophage	Prophage gene	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
					ydfW	2	Uncharacterized protein from Qin prophage	Prophage gene	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Chlorpromazine	PM17, D10	2	Dark	Medium	*galE*	8	UDP-glucose 4-epimerase	Envelope/capsule biosynthesis	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Crystal violet	PM20, E3	2	Medium	Light	*mdtM*	8	MdtM multidrug transporter	Efflux pump/transporter	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
D-Serine	PM17, A3	1	Light	Clear	*tabA*	2	Toxin-antitoxin biofilm protein	Biofilm formation & regulation	(Kim *et al.*, 2009)
					yidF	5	Uncharacterized protein; predicted transcriptional regulator	Regulatory effect	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
D,L-Methionine hydroxamate	PM17, F7	3	Dark	Medium	*aroL*	8	Shikimate kinase II	Catalytic promiscuity and/or substrate ambiguity	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Dodecyltrimethyl ammonium bromide	PM12, H11	1	Light	Clear	*cmr*	8	MdfA/Cmr multidrug transporter	Efflux pump/transporter	(Edgar & Bibi, 1997)
FCCP	PM19, E3	2	Dark	Medium	*ycbS*	7	Uncharacterized outer membrane β-barrel protein	Efflux pump/transporter	(Zhai & Saier, 2002)

Toxic compound	PM well*	Days of inc.†	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product	Proposed mechanism	Reference(s)
Fusaric acid	PM14, B6	1	Dark	Medium	*yghW*	2	Uncharacterized protein	Unknown	
Guanazole	PM18, G7	1	Light	Clear	*hda*	8	Regulator of DnaA	Regulatory effect	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Guanidine hydrochloride	PM15, A6	1	Medium	Clear	*csiE*	7	Stationary phase-inducible protein	Stress response	(Marschall & Hengge-Aronis, 1995)
Harmane	PM19, B7	2	Light	Clear	*yjfK*	8	Uncharacterized protein	Unknown	
Hydroxyurea	PM15, H7	5	Medium	Light	*cpdA*	8	cAMP phosphodiesterase	Regulatory effect	(Imamura *et al.*, 1996)
Iodoacetate	PM14, D7	2	Light	Clear	*yedA*	8	Uncharacterized inner membrane transporter	Efflux pump/transporter	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
L-Aspartic-β-hydroxamate	PM12, G12	1	Medium	Clear	*cpdA*	3	cAMP phosphodiesterase	Regulatory effect	(Imamura *et al.*, 1996)
					iaaA	3	Isoaspartyl peptidase	Catalytic promiscuity and/or substrate ambiguity	(Hejazi *et al.*, 2002)
Lidocane	PM18, D10	1	Dark	Light	*wbbL_1*	2	Putative lipopolysaccharide biosynthesis glycosyltransferase	Envelope/capsule biosynthesis	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
					yfdO	2	Uncharacterized protein from CPS-53 (KpLE1) prophage	Prophage gene	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Manganese chloride	PM13, G7	2	Very dark	Medium	None isolated	N/A			
Methyl viologen	PM15, E9	4	Very dark	Dark	*ycbS*	2	Uncharacterized outer membrane β-barrel protein	Efflux pump/transporter	(Zhai & Saier, 2002)
Niaproof	PM17, E4	4	Medium	Light	*ymjB*	2	Uncharacterized protein	Unknown	

Toxic compound	PM well*	Days of inc.†	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product	Proposed mechanism	Reference(s)
					ygeN	4	Uncharacterized protein	Unknown	
Ornidazole	PM20, C10	2	Light	Clear	*ypjC*	8	Uncharacterized protein	Unknown	
Orphenadrine	PM20, B2	2	Dark	Light	*galE*	7	UDP-glucose 4-epimerase	Envelope / capsule biosynthesis	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Patulin	PM20, H1	1	Medium	Clear	*ysgA*	8	Putative dienelactone hydrolase	Catalytic promiscuity and / or substrate ambiguity	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Plumbagin	PM18, H11	1	Very dark	Dark	*ydaC*	8	Uncharacterized protein from Rac prophage	Prophage gene	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Potassium chromate	PM13, C10	4	Light	Clear	*yeiW*	8	Uncharacterized protein	Unknown	
Potassium tellurite	PM11, G11	1	Medium	Clear	*tehB*	4	Tellurite resistance protein	Catalytic promiscuity and / or substrate ambiguity	(Liu *et al.*, 2000)
Potassium tellurite	PM16, F3	1	Light	Clear	*tehB*	8	Tellurite resistance protein	Catalytic promiscuity and / or substrate ambiguity	(Liu *et al.*, 2000)
Pridinol	PM20, F10	1	Medium	Light	*galE*	2	UDP-glucose 4-epimerase	Envelope / capsule biosynthesis	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Proflavine	PM20, D3	2	Medium	Light	*marA*	8	MarA transcriptional activator	Regulatory effect	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Puromycin	PM15, F11	1	Light	Clear	*mdtM*	8	MdtM multidrug transporter	Efflux pump / transporter	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Sanguinarine	PM14, A11	1	Light	Clear	*ydaC*	3	Uncharacterized protein from Rac prophage	Prophage gene	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)

Toxic compound	PM well*	Days of inc.†	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product	Proposed mechanism	Reference(s)
					tmcA	2	$tRNA^{Met}$ cytidine acetyltransferase	Catalytic promiscuity and/or substrate ambiguity	(Chimnaronk *et al.*, 2009)
					ycbS	3	Uncharacterized outer membrane β-barrel protein	Efflux pump/transporter	(Zhai & Saier, 2002)
Sodium bromate	PM18, D7	2	Medium	Light	*elaD*	2	Deubiquitinase	Catalytic promiscuity and/or substrate ambiguity	(Catic *et al.*, 2007)
Sodium dichromate	PM14, D12	1	Light	Clear	*cmr*	2	MdfA/Cmr multidrug transporter	Efflux pump/transporter	(Edgar & Bibi, 1997)
					yhbT	6	Uncharacterized protein	Unknown	
Sodium *m*-arsenite	PM18, D1	1	Light	Clear	*hcaR*	7	Hca operon transcriptional activator	Regulatory effect	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Sodium metaborate	PM14, E11	2	Light	Clear	None isolated	N/A			
Sodium nitrite	PM14, G11	4	Medium	Light	*yjeH*	2	Uncharacterized inner membrane transporter	Efflux pump/transporter	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
					pheS	2	Phenylalanyl-tRNA synthetase α-chain	Unknown	
Thallium (I) acetate	PM13, F11	2	Light	Clear	None isolated	N/A			
Thioridazine	PM20, C1	2	Very dark	Dark	*galE*	2	UDP-glucose 4-epimerase	Envelope/capsule biosynthesis	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Tinidazole	PM18, F7	3	Light	Clear	*wrbA*	2	NAD(P)H:quinone oxidoreductase	Catalytic promiscuity and/or substrate ambiguity	(Patridge & Ferry, 2006)

Toxic compound	PM well*	Days of inc.†	ASKA growth	NoInsert growth	Enriched gene	Freq. (n/8 clones)	Product	Proposed mechanism	Reference(s)
					mdtM	6	MdtM multidrug transporter	Efflux pump/transporter	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)
Trifluoperazine	PM13, G12	2	Medium	Clear	*galE*	8	UDP-glucose 4-epimerase	Envelope/capsule biosynthesis	(Keseler *et al.*, 2009; The Uniprot Consortium, 2010)

* The toxin-containing PM well from which the corresponding gene(s) was isolated. † Days of incubation before a change in growth phenotype (*i.e.*, tetrazolium colour intensity) was observed.

§ Two ASKA ORFs were co-isolated from each of the screened colonies.

Key for scoring growth phenotype (*i.e.*, tetrazolium colour intensity): Very dark Dark Medium Light Clear

5.2.1.2 CASES WHERE THE NEGATIVE CONTROL OUT-GREW THE ASKA CLONE

In contrast to the improved growth phenotypes shown by the ASKA pool (Section 5.2.1.1), there were four toxin-containing environments (2% of the all conditions screened) where the negative control out-grew the ASKA pool (Table 5.2) in two independent screens. A possible reason was that one or more ASKA clones were sensitizing the whole population to these compounds, *e.g.* by metabolizing the compounds into more toxic products. Unlike the 99 cases described in Section 5.2.1.1, protein over-expression did not seem to be able to confer fitness to any member of the ASKA population. This was also noted previously (Section 4.2.8 in Chapter 4), where the ASKA pool was out-grown in 2% of the 960 environments of PM 1 to PM 10. Since the number of such cases was small, and the difference in growth rates were slight, the exact reasons leading to this phenomenon were not investigated further.

Table 5.2: The four toxin-containing environments in which the negative control out-grew the ASKA pool.

Compound	PM well*	Days of incubation†	ASKA growth	NoInsert growth
β-Lactam				
Cefoxitin	PM14, E4	3	Light	Medium
Anti-folate				
Sulfamethoxazole	PM12, G7	3	Clear	Light
Polymyxin				
Polymyxin B	PM12, B10	1	Dark	Very dark
Other toxins				
1-chloro-2,4-dinitrobenzene	PM16, D2	2	Light	Dark

* The toxin-containing PM well of which the difference of colour intensity was noted. † Days of incubation before a change in growth phenotype (*i.e.,* tetrazolium colour intensity) was observed.

Key for scoring growth phenotype (*i.e.,* tetrazolium colour intensity):
Very dark Dark Medium Light Clear

5.2.2 IDENTIFICATION OF NOVEL RESISTANCE GENES

To deconvolute the results of the library-on-library screen, the individual ASKA clones that were responsible for the observed improvements in growth rates were isolated. In each of the 99 conditions identified above, serial transfer was used to enrich for the fittest clone(s) within the ASKA pool (Figure 4.2 in Section 4.2.2, Chapter 4). After two transfers, cells from

the enriched population were plated under non-selective conditions. PCR and DNA sequencing revealed the plasmid-encoded genes that had imparted improved growth.

An ASKA ORF was scored as a hit if it occurred at least twice among the eight clones sequenced from a serially-passaged population. Given the large size of the ASKA pool (5,272 clones), it was highly unlikely ($P \approx 0.005$; refer to Section 4.4.4.1 for more details) that any two of the eight randomly-chosen colonies would contain the same ASKA plasmid, unless that clone had been positively selected during serial transfer. This approach yielded hits from 86 of the 99 conditions tested, including representatives of all the compound classes that were tested (Table 5.3). For example, the PM plates contained 22 β-lactam antibiotics, and ASKA clones that improved growth in 11 of them were isolated (Table 5.3, row 1). The serial enrichment strategy yielded a single hit from 60 of the 86 conditions. In 38 of these, the ASKA clone was enriched to apparent homogeneity (*i.e.*, eight of eight sequenced clones were identical). However, serial transfer gave two different hits in 23 conditions, and three hits in another three cases (the β-lactam, aztreonam; the antiseptic, benzethonium chloride; and the toxic alkaloid, sanguinarine). Overall, therefore, 115 positively-selected hits were identified from 86 toxin-containing environments (Table 5.1 in Section 5.2.1.1; Table 5.3).

Table 5.3: Summary of hits obtained from the library-on-library screen for PM 11 to PM20.

Class of toxic compound	Number of positives*	Number of ASKA hits†
Anti-bacterials		
β-lactams	11 (of 22)	14
Aminoglycosides	4 (of 14)	4
Anti-folates	7 (of 9)	10
Quinolones	4 (of 9)	6
Tetracyclines	3 (of 8)	4
Macrolide-Lincosamide-Streptogramin (MLS) antibiotics	6 (of 7)	10
Glycopeptides	1 (of 3)	1
Nitrofurans	3 (of 3)	4
Rifamycins	1 (of 2)	2
Steroid anti-bacterial	1 (of 1)	1
Other toxins	45 (of 159)	59
Total	86 (of 237)	115

* The number of toxic compounds for which screening and serial enrichment yielded at least one ASKA-encoded resistance gene. The total number of toxins that were screened in each class is listed in parentheses.

† Total number of ASKA ORFs that were isolated from PM wells containing toxins of each compound class.

In 13 cases, no individual ASKA clones were successfully enriched to a frequency of two or more in the eight sequenced clones (Table 5.1 in Section 5.2.1.1). The possible explanations for this phenomenon has been discussed previously (Section 4.3.1 in Chapter 4).

5.2.3 DIVERSE MECHANISMS OF RESISTANCE

The initial focus was on identifying proteins — and particularly enzymes — that possessed weak secondary activities towards antibiotics and toxins. However, it rapidly became apparent that this library-on-library screen had uncovered many different classes of protein that were contributing to resistance. Therefore, I used the EcoCyc (Keseler *et al.*, 2009) and UniProt (The UniProt Consortium, 2010) databases, in addition to the references cited in Table 5.1 (Section 5.2.1.1), to assign proposed modes of action for each of the 115 cases that had been positively selected and identified during the screen (Table 5.4).

Table 5.4: Proposed modes of action for 115 cases of resistance imparted by ASKA ORFs.

Proposed mechanism of resistance	Frequency
Efflux pump/transporter	22
Regulatory effect	19
Catalytic promiscuity and/or substrate ambiguity	15
Prophage gene	14
Stress response	13
Envelope/capsule synthesis	8
Biofilm formation and regulation	5
Over-expression of the drug's target	1
Maintenance of metabolic flux	1
Unknown	17
Total	115

Eighteen genes occurred multiple times, and were responsible for improving growth in a total of 72 toxin-containing environments (Table 5.5), suggesting broad and non-specific mechanisms. Consistent with previous results (Nishino & Yamaguchi, 2001; Breidenstein *et al.*, 2008; Duo *et al.*, 2008; Liu *et al.*, 2010), the expression of efflux pumps and transporters was common among this subset of non-specific resistance determinants. For example, expression of the Bcr multidrug transporter increased growth in the presence of three anti-folate antibiotics (sulfachloropyridazine, sulfathiazole and sulfisoxazole), two tetracyclines (oxytetracycline and rolitetracycline), and one β-lactam (cefoperazone). An outer membrane

β-barrel, YcbS (see Section 4.3.2.2 of Chapter 4 for a summary of its function), improved growth in the presence of four toxic compounds: a mutagen, 9-aminoacridine; an uncoupling agent, FCCP; a herbicide, methyl viologen; and an alkaloid, sanguinarine. Similarly, the MdtM and Cmr efflux pumps improved growth in the presence of five and three toxins respectively (Table 5.5). The selection of these membrane proteins confirmed that any bias associated with IPTG-induced over-expression had been minimized (refer to Section 4.3.1 for the potential fitness bias associated with IPTG induction).

I also discovered transcriptional regulators, stress response proteins and proteins from capsule biosynthesis that each improved fitness in the presence of multiple toxins (Table 5.5). The only readily predictable example was the transcriptional activator MarA, which was isolated from environments containing two quinolones (ciprofloxacin and ofloxacin), a tetracycline (penimepicycline), a macrolide (tylosin), and a mutagen (proflavine). MarA is known to mediate multidrug resistance by up-regulating the efflux system and down-regulating membrane permeability (Cohen *et al.*, 1993). In other cases, proteins that had not previously been associated with resistance were found, including the biofilm modulation protein, Bdm, the small stress response protein, YcgZ, and the global regulator of transcription, CpdA (cAMP phosphodiesterase). The likely pleiotropic effects of these in altering the expression of downstream resistance determinants can be rationalized. Another prominent hit was the UDP-glucose 4-epimerase (GalE) ASKA strain, which was isolated from seven PM wells. GalE is involved in a variety of galactose metabolic processes, including glycolysis and the biosynthesis of lipopolysaccharide and capsule components. The host strain, *E. coli* DH5α-E, is *gal*$^{-}$, although the exact mutation(s) responsible for this phenotype are not known. It is possible that the fitness effects seen when GalE was over-expressed were simply due to complementation of the *gal*$^{-}$ mutation. However, if this was the case, the GalE-expressing ASKA strain would have out-grown the negative control in many more than seven of the 237 environments that had been tested. Instead, GalE expression appears to mediate a suite of pleiotropic effects that are relevant for the evolution of resistance. Finally, other multiple-occurring ORFs hint at the functions of previously uncharacterized proteins (Table 5.5). For example, the YejG-expressing clone was isolated from PM wells containing apramycin, sisomicin and tobramycin, suggesting that it may possess broad enzymatic activity towards these aminoglycoside antibiotics.

Table 5.5: Genes that conferred increased resistance to more than one toxic compound.

Gene	Gene product	Freq.*	Toxic compounds	Proposed mechanism of resistance
galE	UDP-glucose 4-epimerase	7	Amitriptyline, Chlorpromazine, Nalidixic acid, Orphenadrine, Pridinol, Thioridazine, Trifluoperazine	Envelope/capsule biosynthesis
bcr	Bcr multidrug transporter	6	Cefoperazone, Oxytetracycline, Rolitetracycline, Sulfachloropyridazine, Sulfathiazole, Sulfisoxazole	Efflux/transport
ycgZ	Small protein induced by YcgF	6	Aztreonam, Cefmetazole, Cefuroxime, Moxalactam, Penimepicycline, Sulfamethazine	Stress response
ydaC	Uncharacterized protein from Rac prophage	6	Caffeine, Chelerythrine, Erythromycin, Phenethicillin, Plumbagin, Sanguinarine	Prophage gene
marA	MarA transcriptional activator	5	Ciprofloxacin, Ofloxacin, Penimepicycline, Proflavine, Tylosin	Regulatory effect
mdtM	MdtM multidrug transporter	5	5-Nitro-2-furaldehyde, Crystal violet, Furaltadone, Puromycin, Tinidazole	Efflux/transport
cpdA	cAMP phosphodiesterase	4	Dihydrostreptomycin, Hydroxyurea, Josamycin, L-Aspartic-b-hydroxamate	Regulatory effect
csiE	Stationary phase-inducible protein	4	3,4-Dimethoxybenzyl alcohol, Antimony (III) chloride, Blasticidin S, Guanidine hydrochloride,	Stress response
rbsR	Ribose transcriptional repressor	4	Aztreonam, Lincomycin, Sulfamethazine, Sulfamonomethoxine	Regulatory effect
ycbS	Uncharacterized outer membrane β-barrel protein	4	9-Aminoacridine, FCCP, Methyl viologen, Sanguinarine	Efflux/transport
ydfW	Uncharacterized protein from Qin prophage	4	Chelerythrine, Erythromycin, Oleandomycin, Spiramycin,	Prophage gene
bdm	Biofilm-dependent modulation protein	3	Cefazolin, Cephalothin, Nafcillin	Biofilm formation & regulation
cmr	MdfA/Cmr multidrug transporter	3	Dodecyltrimethyl ammonium bromide, Lincomycin, Sodium dichromate	Efflux/transport
yejG	Uncharacterized protein	3	Apramycin, Sisomicin, Tobramycin	Unknown
folA	Dihydrofolate reductase	2	2,4-Diamino-6,7-diisopropyl-pteridine, Trimethoprim	Catalytic promiscuity and/or substrate ambiguity; over-expression of drug target
hcaR	Hca operon transcriptional activator	2	Fusidic acid, Sodium *m*-arsenite	Regulatory effect
tusE	Sulfurtransferase	2	Rifampicin, Spiramycin	Catalytic promiscuity and/or substrate ambiguity
yfdO	Uncharacterized protein from CPS-53 [KpLE1] prophage	2	Lidocane, Nalidixic acid	Prophage gene

* The number of PM wells from which the ASKA-encoded gene was isolated.

In 43 cases, an ASKA-encoded ORF improved growth in the presence of a single toxin (Table 5.6). These compound-specific responses included 12 cases in which metabolic enzymes appeared to display catalytic promiscuity and/or substrate ambiguity. Examples included the carbonic anhydrase, CynT, evincing growth in the presence of enoxacin (reminiscent of the human enzyme's promiscuous esterase activity) (Gould & Tawfik, 2005) and the purine nucleotidase, YrfG, increasing fitness in the presence of a toxic analog (6-mercaptopurine). Finally, 13 of the 17 instances in which I was unable to assign a proposed mechanism of action (Table 5.4) involved uncharacterized ASKA ORFs that were isolated from a single environment (Table 5.6).

Table 5.6: Genes that conferred increased resistance to one toxic compound only.

Gene	Gene product	Associated compound	Proposed mechanism
luxS	S-ribosylhomocysteine lyase	Sulfadiazine	Biofilm formation & regulation
tabA	Toxin-antitoxin biofilm protein	D-Serine	Biofilm formation & regulation
aroL	Shikimate kinase II	D,L-Methionine hydroxamate	Catalytic promiscuity and/or substrate ambiguity
cynT	Carbonic anhydrase	Enoxacin	Catalytic promiscuity and/or substrate ambiguity
elaD	Deubiquitinase	Sodium bromate	Catalytic promiscuity and/or substrate ambiguity
iaaA	Isoaspartyl peptidase	L-Aspartic-β-hydroxamate	Catalytic promiscuity and/or substrate ambiguity
puuD	γ-glutamyl-γ-aminobutyrate hydrolase	Benzethonium chloride	Catalytic promiscuity and/or substrate ambiguity
tehB	Tellurite resistance protein	Potassium tellurite	Catalytic promiscuity and/or substrate ambiguity
tmcA	$tRNA^{Met}$ cytidine acetyltransferase	Sanguinarine	Catalytic promiscuity and/or substrate ambiguity
wrbA	NAD(P)H:quinone oxidoreductase	Tinidazole	Catalytic promiscuity and/or substrate ambiguity
yagT	Aldehyde oxidoreductase, Fe-S subunit	Furaltadone	Catalytic promiscuity and/or substrate ambiguity
yeaD	Putative mutarotase	Aztreonam	Catalytic promiscuity and/or substrate ambiguity
yrfG	Purine nucleotidase	6-Mercaptopurine	Catalytic promiscuity and/or substrate ambiguity
ysgA	Putative dienelactone hydrolase	Patulin	Catalytic promiscuity and/or substrate ambiguity
phnC	Alkylphosphonate ABC transporter, ATP-binding subunit	Sulfathiazole	Efflux/transport
ugpC	Glycerol-3-phosphate ABC transporter; ATP-binding subunit	Piperacillin	Efflux/transport
yedA	Uncharacterized inner membrane transporter	Iodoacetate	Efflux/transport
yjeH	Uncharacterized inner membrane transporter	Sodium nitrite	Efflux/transport
wbbL_1	Putative lipopolysaccharide biosynthesis glycosyltransferase	Lidocane	Envelope/capsule biosynthesis

Gene	Gene product	Associated compound	Proposed mechanism
nudB	Dihydroneopterin triphosphate pyrophosphatase	Sulfadiazine	Maintaining metabolic flux
tfaX	Uncharacterized protein from DLP12 prophage	Nitrofurantoin	Prophage gene
ybcD	Pseudogene from DLP12 prophage; predicted fragment of replication protein	Benzethonium chloride	Prophage gene
feoC	Transcriptional regulator of ferrous iron transport	Rifampicin	Regulatory effect
hda	Regulator of DnaA	Guanazole	Regulatory effect
soxS	SoxS transcriptional activator	Enoxacin	Regulatory effect
yidF	Uncharacterized protein; predicted transcriptional regulator	D-Serine	Regulatory effect
bfd	Bacterioferritin-associated ferredoxin	Penicillin G	Stress response
gmr	Cyclic-di-GMP phosphodiesterase	3,5-Dinitrobenzene	Stress response
mzrA	Modulator of EnvZ and OmpR	Vancomycin	Stress response
pheS	Phenylalanyl-tRNA synthetase α-chain	Sodium nitrite	Unknown
yeeX	Uncharacterized cytoplasmic protein	Piperacillin	Unknown
yeiW	Uncharacterized protein	Potassium chromate	Unknown
ygeN	Uncharacterized protein	Niaproof	Unknown
yghW	Uncharacterized protein	Fusaric acid	Unknown
yhbT	Uncharacterized protein	Sodium dichromate	Unknown
yhfK	Uncharacterized inner membrane protein	Josamycin	Unknown
yhiK	Uncharacterized protein	Benzethonium chloride	Unknown
yjbI	Uncharacterized protein	Blasticidin S	Unknown
yjfK	Uncharacterized protein	Harmane	Unknown
ykiA	Uncharacterized protein	Antimony (III) chloride	Unknown
ymjB	Uncharacterized protein	Niaproof	Unknown
ypjC	Uncharacterized protein	Ornidazole	Unknown
yqjD	Uncharacterized protein	2,2'-Dipyridyl	Unknown

5.2.4 THE INTERPLAY OF FITNESS AND RESISTANCE

Mutations that confer antibiotic resistance can have a variety of effects on the fitness of the cell in the absence of drug (Andersson & Hughes, 2010). Resistance mutations often bear fitness costs, although examples of cost-free resistance (Sander *et al.*, 2002) and mutual compensation (*i.e.*, simultaneous increases in resistance and fitness) have also been described (Marcusson *et al.*, 2009). The library-on-library screen revealed cases where

protein over-expression in the ASKA clones was responsible for enhancing fitness in the presence of a toxin.

It was unclear from the screen whether the plasmid-encoded ORFs were improving the general vigour of the host, or increasing resistance. To investigate this further, the relative fitness (*W*) for each of 25 strains relative to the negative control (*i.e.*, *E. coli* DH5α-E with pCA24N-NoIns) was measured. Fitness tests were conducted in the absence of toxin, but in the presence of 50 μM IPTG (see Section 5.4.3 for details). Using the same set of 25 strains, I also determined the minimum inhibitory concentrations (MICs) for 29 toxin/strain combinations (see Section 5.4.4 for details). The toxins were chosen based on their anti-bacterial properties, and as examples of each mechanistic class (as defined in Table 5.4 of Section 5.2.3). MICs were determined by either plate-based measurements (see Section 5.4.4.1 for details), or broth-based measurements (see Section 5.4.4.2 for details). An example of plate-based measurements is shown in Figure 5.1, where E-test strips containing cefuroxime were used to measure the MIC values of the negative control and the YcgZ-expressing clone.

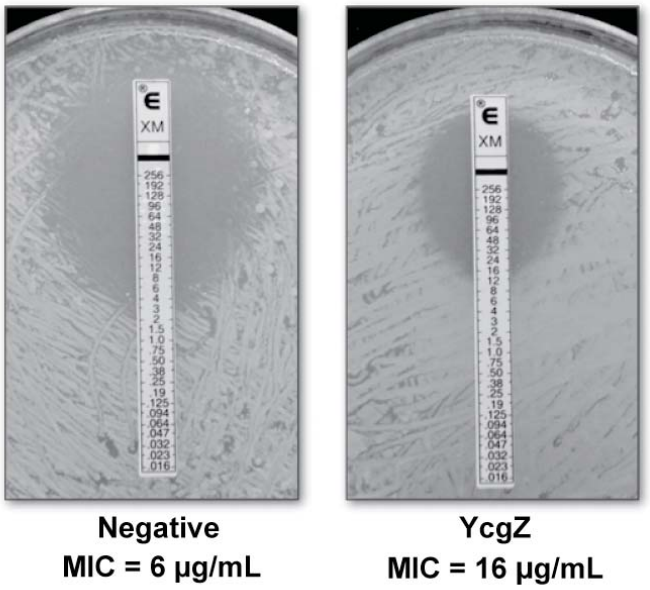


Figure 5.1: YcgZ over-expression conferred resistance to cefuroxime in *E. coli*.

Cefuroxime susceptibility in *E. coli* was tested using E-test strips (Section 5.4.4). Clone over-expressing YcgZ showed a higher resistance to cefuroxime, as evident by the smaller zone of inhibition.

In the absence of toxin, only two of the 25 ASKA strains showed significantly reduced fitness (i.e. $W < 1$; $P < 0.01$) compared to the negative control (Table 5.7). Nine ASKA strains were not significantly different from the control ($P > 0.01$), and 14 of the strains out-

competed it ($W > 1$; $P < 0.01$). In contrast, almost all of MICs determined for ASKA strains were increased compared to those that were measured for the negative control strain (Table 5.7).

By far the largest increase in MIC was for trimethoprim, where over-expression of its intracellular target, dihydrofolate reductase (*folA*), effected very high levels of resistance (at least 500-fold greater than the negative control strain). The FolA-expressing strain was also one of the least fit in the absence of antibiotic ($W = 0.84$). Expression of the multidrug transporters MdtM and Cmr led to 16-fold and 8-fold increases in the MICs for puromycin and lincomycin respectively, without imparting fitness costs in a toxin-free environment. The remainder of the selected ASKA strains exhibited modest increases (<8-fold) in MICs compared to the negative control (Table 5.7).

In general, there was a poor correlation between resistance, and fitness in the absence of toxin. In some cases, fitter clones were also more resistant. For example, the YcgZ-expressing clone had the highest relative fitness of the 25 that we measured ($W = 1.72$), and it also showed a greater increase in aztreonam resistance (5.9-fold) than the other two clones that were isolated from the aztreonam-containing PM well (Table 5.7). However, there were also cases where the opposite was true. For example, over-expression of YdaC imparted a 4-fold increase in erythromycin resistance and a small increase in fitness ($W = 1.33$), while YdfW over-expression gave a smaller increase in resistance (3-fold) but a larger increase in fitness ($W = 1.51$). Similarly, over-expressing NudB gave a higher MIC for sulfadiazine than LuxS over-expression (32 μg/mL *versus* 24 μg/mL), yet the NudB-expressing strain was considerably less fit ($W = 0.68$ *versus* $W = 1.11$).

Table 5.7: Relative fitness and drug resistance data for *E. coli* cells over-expressing ORFs from the ASKA library.

ASKA ORF	$W \pm SE^*$	Compound	MIC (μg/mL) with ASKA ORF	MIC (μg/mL) with pCA24N-NoIns	Fold increase in MIC
1. β-lactams					
bdm	$1.47 \pm 0.04^{\dagger}$	Cephalothin	16	8	2.0
bfd	$1.16 \pm 0.03^{\dagger}$	Penicillin G	48	32	1.5
rbsR	$1.40 \pm 0.04^{\dagger}$	Aztreonam	0.19	0.064	3.0
yeaD	$1.20 \pm 0.03^{\dagger}$	Aztreonam	0.094	0.064	1.5
ycgZ	$1.72 \pm 0.05^{\dagger}$	Aztreonam	0.38	0.064	5.9
ycgZ	See above	Cefuroxime	16	6	2.7
2. Aminoglycosides					
yejG	0.94 ± 0.04	Sisomicin	0.19	0.125	1.5
yejG	See above	Tobramycin	1.5	0.5	3.0

ASKA ORF	$W \pm$ SE*	Compound	MIC (μg/mL) with ASKA ORF	MIC (μg/mL) with pCA24N-NoIns	Fold increase in MIC
3. Anti-folates					
folA	0.84 ± 0.03†	Trimethoprim	>32	0.064	>500
luxS	1.11 ± 0.02†	Sulfadiazine	24	16	1.5
nudB	0.68 ± 0.04†	Sulfadiazine	32	16	2.0
4. Quinolones					
cynT	1.17 ± 0.04†	Enoxacin	0.25	0.19	1.3
marA	1.08 ± 0.02	Ciprofloxacin	0.125	0.032	3.9
galE	1.16 ± 0.01†	Nalidixic acid	96	32	3.0
yfdO	1.03 ± 0.02	Nalidixic acid	48	32	1.5
5. Tetracyclines					
bcr	1.03 ± 0.04	Oxytetracycline	6	2	3.0
6. Macrolide-Lincosamide-Streptogramin					
cmr	1.45 ± 0.02†	Lincomycin	1536	192	8.0
rbsR	See above	Lincomycin	384	192	2.0
ydfW	1.51 ± 0.12†	Spiramycin	36	24	1.5
ydfW	See above	Erythromycin	96	32	3.0
ydaC	1.33 ± 0.03†	Erythromycin	128	32	4.0
7. Glycopeptides					
mzrA	1.23 ± 0.07	Vancomycin	256	128	2.0
8. Nitrofurans					
tfaX	1.18 ± 0.02†	Nitrofurantoin	0.094	0.19	0.5
9. Rifamycins					
feoC	0.78 ± 0.07	Rifampicin	6	4	1.5
tusE	0.87 ± 0.04	Rifampicin	6	4	1.5
10. Other toxins					
mdtM	1.07 ± 0.03	Puromycin	128	8	16.0
puuD	1.12 ± 0.03†	Benzethonium chloride	6	6	1.0
ybcD	1.02 ± 0.02	Benzethonium chloride	8	6	1.3
yhiK	1.19 ± 0.01†	Benzethonium chloride	12	6	2.0

* Fitness (W) of the ASKA strain, in the absence of toxin, relative to a neutrally-marked negative control that harbours pCA24N-NoIns. Values are expressed as mean ± S.E.M. ($n = 8$).

† Significance level of $P < 0.01$ for a two-tailed test with the null hypothesis that $W = 1$, calculated using the t distribution and 7 degrees of freedom.

The host strain, *E. coli* DH5α-E, contains a chromosomal *gyrA* mutation. The *gyrA* gene encodes DNA gyrase, and mutations at this locus are known to give rise to nalidixic acid resistance. Unexpectedly, two genes, *galE* and *yfdO*, were identified in this study. Over-expression of GalE and YfdO further increased the MIC for nalidixic acid, by 3-fold and 1.5-

fold respectively (Table 5.7). In the absence of nalidixic acid, the GalE-expressing strain was slightly fitter than the control strain (W = 1.16), whereas the YfdO-expressing strain was nearly as fit as the control strain (W = 1.03). The implication is that there are alternate routes to nalidixic acid resistance in *E. coli*, and/or novel ways (possibly mediated by GalE and/or YfdO) to compensate for fitness costs that are associated with *gyrA* mutations.

In a single case, expression of an ASKA ORF (an uncharacterized prophage gene, *tfaX*) decreased the MIC for the antibiotic (nitrofurantoin) in which it was selected. The TfaX-expressing strain showed a small but statistically significant fitness improvement in the absence of nitrofurantoin (W = 1.18). This suggests that increased fitness in the absence of an antibiotic can translate into a selective advantage when the drug is introduced; however, this appears to be uncommon. It is noteworthy that no single ASKA gene was isolated from more than seven PM wells (see Table 5.5 for genes that conferred resistance to more than one toxic compound). It might be expected that one (or a handful) of general, growth-enhancing genes would have been isolated from a majority of the 237 toxin-containing environments, were such a gene present in the ASKA pool. Instead, this approach has revealed many latent resistance determinants in the *E. coli* genome.

5.3 Discussion

5.3.1 An overview of the results

In contrast to Chapter 4 where the ability to grow better in novel nutrients was selected, in this chapter I used resistance to antibiotics and toxins as a readily selectable trait. This approach is different from previous studies where roles in resistance have been inferred indirectly, based on the increased sensitivity of mutants produced by transposon inactivation or gene knockouts (Breidenstein *et al.*, 2008; Duo *et al.*, 2008; Liu *et al.*, 2010). In the first part of the experiment, I identified toxin-containing environments where the pooled ASKA clones out-grew the negative control. In those PM wells, at least one member of the pool had gained an improved ability to survive and reproduce, through the over-expression of a single, wild-type ORF. Next, serial transfer was used to identify the fittest strain(s) in the ASKA pool. This experimental design precluded the identification of ASKA clones with increased susceptibility to toxins, although these are likely to have been present in many (and perhaps all) of the PM wells.

In Chapter 4, thiamine starvation appeared to be an issue for the biased gene enrichments from the wells of PM 3 to PM 8 (Section 4.3.3). However, thiamine was not a factor in selecting ASKA ORFs from the toxin-containing environments. This was because the base medium in the wells of PM 11 to PM 20 was derived from rich nutrients (2.0 g/L of tryptone, 1.0 g/L of yeast extract and 1.0 g/L of NaCl). This formulation is identical to that

of PM 1, PM 2, PM 9 and PM 10 [(Zhou *et al*., 2003); also see Sections 4.2.3.1 and 4.2.7], and provides a source of thiamine through yeast extract.

In total, I identified 61 different ASKA ORFs that improved fitness in 86 of 237 toxin-containing environments. Eighteen of these ORFs were hits in multiple wells (Table 5.5 in Section 5.2.3); overall, I identified 115 examples of resistance. Hits were obtained from every class of toxic compound tested (Table 5.3 in Section 5.2.2). The hit rate (resistance to 36% of all toxins) was comparable with the study conducted in Chapter 4, where 35% of the environments yielded at least an enriched ASKA ORF. Consistent with the results obtained in Chapter 4, this study emphasizes just how likely evolutionary innovation can be. The results of this chapter imply that the over-expression of a randomly chosen *E. coli* protein will impart resistance to a novel toxin with a probability of $(61/5{,}272) \times (86/237) \approx 0.4\%$.

On the whole, the results supported the hypothesis that a bacterium from a non-clinical environment (in this case, a laboratory strain of *E. coli*) can nevertheless possess a significant reservoir of latent resistance proteins (Martínez, 2008). More generally, the identification of many *E. coli* proteins that have not been previously associated with antibiotic resistance suggests that resistance activities can be evolved from pre-existing proteins (Wright, 2007). Protein over-expression can provide an effective initial adaptive response for populations that are exposed to antimicrobial compounds.

5.3.2 A RESERVOIR OF UNEXPLORED TOXIN RESISTANCE

The PM plates contained four concentrations of each toxin. Therefore, a hit could have been due to an increase in resistance (*i.e.*, a strain undergoing division at a toxin concentration that was inhibitory to the other strains, including the negative control), or due to particularly rapid growth in a sub-inhibitory concentration of the toxin (*i.e.*, dividing faster than the other strains in the pool). In this respect, the library-on-library screen in this chapter mimics the dynamics of a clinical setting (Martínez *et al*., 2007; Marcusson *et al*., 2009). At an infection site, the ability to out-grow the neighbouring cells is as important as the ability to resist the actions of antibiotics. Therefore, the fitness values and the MICs of 25 ASKA strains were determined in order to elucidate the correlation between their growth rates and resistance activities (Table 5.7 in Section 5.2.4).

Almost exclusively, the proteins that were identified from the screen did facilitate increased resistance (Table 5.7 in Section 5.2.4). The majority of the measured MIC increases were modest (<8-fold); nevertheless, theoretical models have shown that even small changes such as these can be responsible for increasing the severity of bacterial infections (Drusano, 2004). In one case — cefuroxime — the expression of a small stress response protein, YcgZ, increased the MIC from 6 μg/mL to 16 μg/mL (Table 5.7 and Figure 5.1 in Section 5.2.4). The clinical breakpoint for resistance to this β-lactam is 8 μg/mL (The

European Committee on Antimicrobial Susceptibility Testing [EUCAST], 2011); that is, the non-pathogenic host strain used in this study attained clinically relevant levels of resistance, simply through over-expression of a previously uncharacterized protein. Furthermore, most strains showed no fitness cost associated with expressing an ASKA ORF in the absence of toxin (and indeed, many out-competed the negative control; Table 5.7). Strong selection pressure for cost-free resistance mutations has been observed in clinical isolates of *M. tuberculosis* (Sander *et al.*, 2002). Together, the MIC and fitness data suggest that up-regulating the expression of pre-existing, latent resistance determinants may play an important role in the emergence of drug-resistant pathogens.

Quinolone antibiotics are of particular interest in clinical settings because these drugs are fully synthetic (Heeb *et al.*, 2011). For this reason, the development of quinolone resistance in bacteria is generally assumed to be a lengthy process due to the lack of intrinsic resistance. However, almost half of the quinolone antibiotics tested in this screen yielded at least one ASKA-encoded resistance gene (Table 5.1 in Section 5.2.1.1; Table 5.3 in Section 5.2.2). In one case, *E. coli* cells over-expressing MarA were ~4-fold more resistant to ciprofloxacin than the control clone, and no biological cost was associated with MarA over-expression in the absence of the antibiotic ($W = 1.08$; $P > 0.01$). Protein over-expression has successfully provided a basal level of resistance to the cells, which could pave the way to full-blown resistance. There is only an 8-fold difference between the clinical breakpoint for ciprofloxacin resistance (1.0 μg/mL; The European Committee on Antimicrobial Susceptibility Testing [EUCAST], 2011) and the MIC of the MarA-expressing clone in this study (0.125 μg/mL).

Ciprofloxacin-resistant *E. coli* mutants were recently isolated from a microenvironment that mimicked an infected host (Zhang *et al.*, 2011). All isolates contained three single amino acid changes, and one of the substitutions was located in the transcriptional repressor of the *marRAB* operon, MarR (Cohen *et al.*, 1993). MarA, the transcriptional activator for the same operon, is an antagonist of MarR. Hence, the results of this study (based on MarA over-expression) suggested that the *marR* mutation was likely to increase the expression of the operon (including *marA*), and eventually result in bacterial resistance to ciprofloxacin. This specific example confirms that the resistance determinants (such as *marA*) isolated from the library-on-library screen can be informative in explaining the molecular origins of antibiotic resistance.

Other proteins identified from the screen were found to act through a variety of mechanisms to effect resistance (Table 5.4 in Section 5.2.3). Many of these mechanisms — such as toxin efflux/transport, stress responses and biofilm formation — were unsurprising, even if the identities of the individual resistance genes were less predictable. In 15 of 115 cases, over-expressed enzymes appeared to be acting promiscuously to impart resistance. In addition to these mechanisms, there was a prevalence of (i) prophage genes; and (ii) genes of unknown function in the list. Prophage genes are usually assumed to be cryptic and/or

defective remnants of temperate bacteriophage genomes, although a recent study has shown that genes from the CP4-57 and DLP12 prophages can play roles in biofilm development (Wang *et al.*, 2009). Here, I have reported 14 cases in which the over-expression of prophage genes (including two from DLP12) can improve growth in the presence of toxins. Twelve of these cases arose from the over-expression of only three prophage genes: *ydaC*, *ydfW* and *yfdO* (Table 5.5 in Section 5.2.3). This extends the previous finding (Wang *et al.*, 2009), demonstrating that prophage genes can modulate broad, compound non-specific cellular responses when they are activated from latency.

In contrast, 20 uncharacterized, non-prophage genes that showed specificity in their actions were identified; that is, they each imparted a growth advantage in the presence of a single toxin. These included genes for putative enzymes (*wbbL_1*, *yeaD*, *ysgA*), predicted transporters (*yedA*, *yjeH*) and a predicted transcriptional regulator (*yidF*), as well as 14 genes for which functional annotation was impossible due to insufficient homology with any gene of known function (Table 5.6 in Section 5.2.3). These results provide experimental evidence for the importance of cryptic genes as a reservoir of evolutionarily-accessible functions (Hall *et al.*, 1983).

5.3.3 Protein over-expression: a possible adaptive mechanism

In summary, I have used two tools from functional genomics (PM plates and the ASKA collection) to conduct a comprehensive search for *E. coli* proteins that can impart improved growth phenotypes in the presence of novel nutrients (Chapter 4) or toxic compounds (Chapter 5). The hit rates of Chapter 4 (improved growth in 35% of all environments) and Chapter 5 (resistance to 36% of all toxins) were considerably higher than a previous multicopy suppression screen (Patrick *et al.*, 2007), where only 20% of single-gene knockout strains were rescued by non-cognate ASKA ORFs. Hence, the catalogue of proteins identified from the library-on-library screen provides a unique picture of a bacterium's latent evolutionary potential and emphasizes the high frequency at which novel traits can evolve.

By cataloguing sources of phenotypic innovation, I have revealed the diversity of adaptive mechanisms that can be underpinned by over-expression mutations such as gene amplification. Every plasmid-carrying ASKA clone approximated a non-plasmid-containing strain in which a single chromosomal locus had been amplified. Growth under selection began when the pooled ASKA clones were used to inoculate each PM well. In order to be scored positive, a PM well had to contain cells that out-grew the negative control, in which no ASKA ORF was being expressed. Therefore, I was specifically looking for cases in which proteins possessed activities that were valuable in the presence of either a novel nutrient or a toxin, and cases in which increases in their dosage were required to uncover the latent activities. However, the library-on-library screen was not designed to consider mutations

beyond protein over-expression (or gene amplification). The next logical step will be to determine feasible mutational routes after protein over-expression, by applying selection continuously. This approach, which is underpinned by the Innovation, Amplification and Divergence (IAD) model (Bergthorsson *et al*., 2007), can potentially demonstrate the emergence of new activities from pre-existing protein scaffolds. Related questions, such as the nature of the trade-off between new and old biochemical functions, can also be addressed.

In Chapters 2 and 3, I examined the mutational routes for evolving one pre-existing protein scaffold to have the activity of another. In contrast, Chapters 4 and 5 allowed me to identify pre-exising scaffolds that could mediate the evolution of novel phenotypes. Overall, the prevalence of promiscuity in the *E. coli* proteome, as described in Chapters 4 and 5, appears to confirm that the second tenet of the IAD model (Bergthorsson *et al*., 2007) is correct. Hence, the IAD model is a feasible, likely-to-be-widespread model for the evolution of new traits.

5.4 MATERIALS AND METHODS

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

Powdered X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside; ForMedium) was dissolved in DMSO (dimethyl sulfoxide), and was always prepared as 20 mg/mL stocks. Benzethonium chloride, lincomycin (Duchefa Biochemie), puromycin (Melford), sisomicin and vancomycin (Duchefa Biochemie) were dissolved in water. Enoxacin and sulfadiazine were dissolved in 1 N NaOH. Oxytetracycline (Duchefa Biochemie) was dissolved in water : ethanol (2 : 1). Spiramycin was dissolved in ethanol. 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.

5.4.1 IDENTIFICATION OF LATENT RESISTANCE GENES

The library-on-library screen and serial enrichment method (described in Section 4.4.3 and 4.4.4, respectively) were used to identify latent resistance genes from the ASKA collection. PM 11 to PM 20 plates (Biolog Inc.) were inoculated with both the ASKA pool and the control clone, in parallel. IF-10 medium (Biolog Inc.) was used throughout the experiment (see Section 4.4.3 for more information). Two independent repeats of the complete screen were carried out.

5.4.2 NEUTRALLY MARKED *E. COLI* STRAIN

E. coli DH5α-E carries the *lacZ*ΔM15 mutation. In order to distinguish the two competing strains in each fitness assay (which will be further described in Section 5.4.3), a mini-Tn7 system (Choi *et al.*, 2005) was used to mark the negative control strain (*E. coli* DH5α-E harbouring pCA24N-NoIns) with a functional copy of *lacZ* (and a gentamicin resistance cassette, Gm^R). The marker genes were integrated at the *att*Tn7 site downstream of the chromosomal *glmS* gene. The presence of a functional copy of *lacZ* allowed blue/white colony screening, where blue colonies were the marked negative control and white colonies were the competitor.

5.4.2.1 CONSTRUCTION OF A SUICIDE PLASMID

A suicide plasmid that carries a *lacZ* fragment and does not replicate in *E. coli* was constructed. This was achieved by reassembling suitable fragments from two available plasmids, pUC18R6K-mini-Tn7T-*Gm* and pUC18-mini-Tn7T-*Gm-lacZ* (Section I.4 in

Appendix I). Both plasmids carry Gm^R and the Tn7 mini-transposons. The R6K origin of replication in the plasmid pUC18R6K-mini-Tn7T-*Gm* ensures no replication in *E. coli* strains without the *pir* gene, but the plasmid lacks a *lacZ* gene (Figure 5.2A). Replication of the pUC18R6K-mini-Tn7T-*Gm* depends on the presence of the *pir* gene. On the other hand, the pUC18-mini-Tn7T-*Gm-lacZ* plasmid carries a *lacZ* gene, but it replicates in *E. coli* due to its compatible ColE1 origin of replication (Figure 5.2A).

Therefore, a fragment containing *lacZ* (and Tn7L) was excised from pUC18-mini-Tn7T-*Gm-lacZ* by *Spe*I/*Sap*I digestion. This fragment was ligated to the pUC18R6K-mini-Tn7T-*Gm* backbone that had been digested similarly. The resulting plasmid was termed pUC18R6K-mini-Tn7T-*Gm*-lacZ (Figure 5.2A), and was maintained in *E. coli* DH5α–λ*pir* (Section I.3 in Appendix I).

5.4.2.2 *LACZ*-TAGGED CONTROL CLONE

An aliquot of *E. coli* DH5α-E was transformed with equal amounts (250 ng) of pUC18R6K-mini-Tn7T-*Gm-lacZ* and the helper plasmid, pUX-BF13 (Section I.4 in Appendix I). Expression of the transposase genes from pUX-BF13 was essential for Tn7 transposition. The transformed cells were recovered at 37°C for 3 h, before being plated on selective LB agar plates [supplemented with 15 μg/mL gentamicin and 40 μg/mL X-gal]. The prolonged recovery time was to allow expression of the transposition functions from the helper plasmid, and eventually, the transposition event. Colonies that turned blue indicated the presence of a functional *lacZ* gene. Successful integration of the ~5.2-kb $lacZ/Gm^R$ fragment in the bacterial chromosome was verified by sequencing the PCR product amplified using a Tn7R-specific primer (ecTn7R.for) and a *glmS*-specific primer (ecGlmS.rev) (Figure 5.2B; Section I.6 in Appendix I).

The $lacZ/Gm^R$-tagged strain (DH5α-E::*lacZ*) was made electrocompetent, and finally transformed with pCA24N-NoIns to create a marked negative control clone. The *lacZ*-marked negative control strain and the unmarked negative control strain were equally fit ($W = 1.00 \pm 0.01$, mean ± S.E.M.; $n = 8$ replicates; $P = 0.93$). Therefore, the *lacZ*-marked strain was used as the negative control for the pairwise fitness assays (Section 5.4.3).

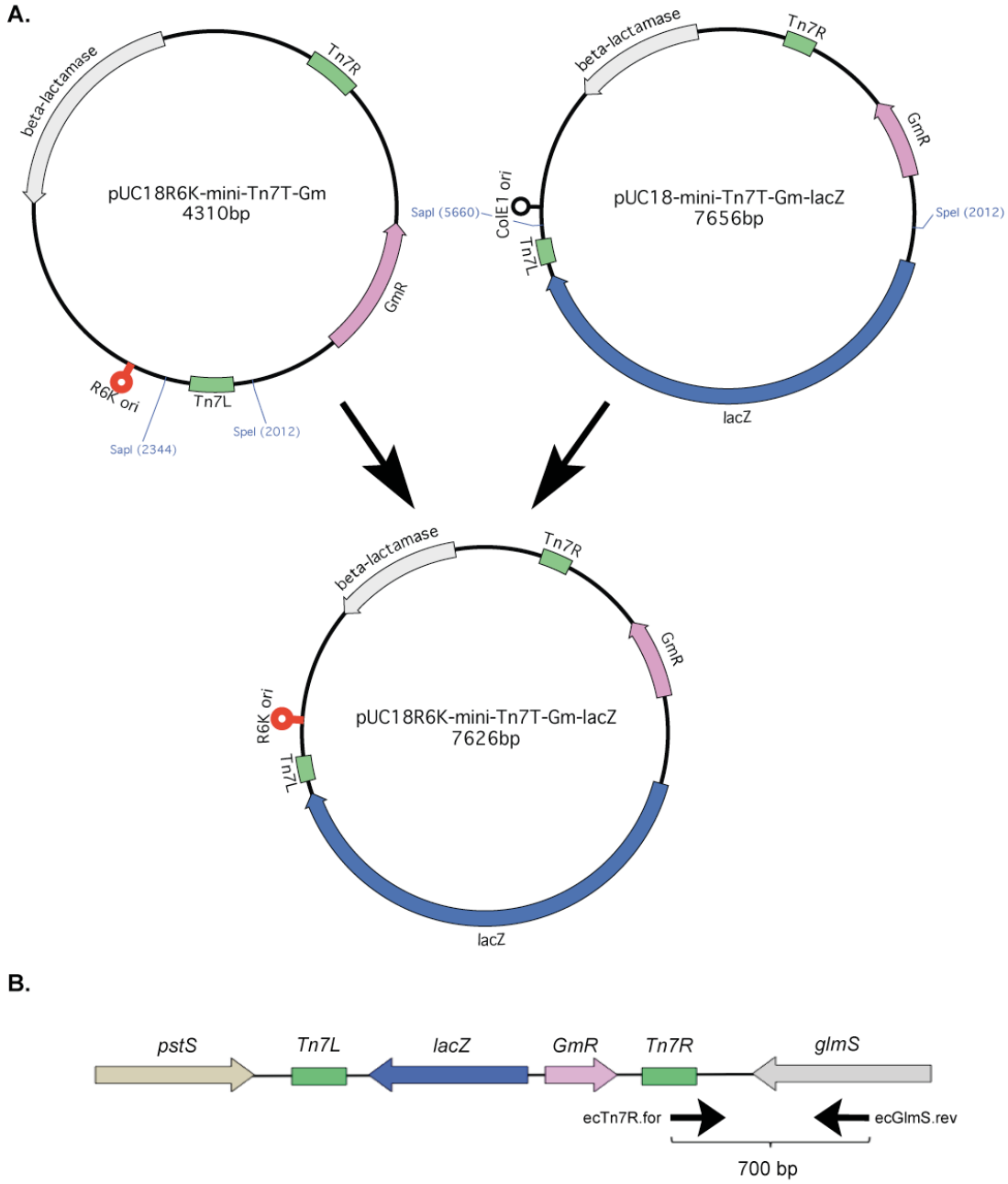


Figure 5.2: Schematic diagram of pUC18R6K-mini-Tn7T-*Gm-lacZ* and the integration region of the *lacZ*/*Gm*^{R} genes in *E. coli* chromosome.

(A) The 4.0-kb vector backbone of pUC18R6K-mini-Tn7T-*Gm-lacZ* was obtained by digesting pUC18R7K-mini-Tn7T-*Gm* with *Spe*I and *Sap*I. The Tn7L element was removed during vector linearization. The 3.6-kb fragment containing *lacZ* and Tn7L was excised from pUC18-mini-Tn7T-*Gm-lacZ* by digesting with the same combination of restriction endonucleases. The *lacZ*/Tn7L fragment was inserted into the pUC18R6K vector backbone, thereby forming a suitable suicide plasmid for *E. coli*. (B) Transposition mediated by the Tn7 elements was site- and orientation-specific. The Tn7 elements were inserted at the *att*Tn7 site (in the presence of a helper plasmid) located in the intergenic region between the *pstS* and *glmS* genes. Amplification of the intergenic region between the Tn7R element and *glmS* gene using the ecTn7R.for and ecGlmS.rev primers yielded a 700-bp PCR product.

5.4.3 Relative fitness assays

Twenty-five ASKA plasmids that conferred increased growth in the presence of antibiotics were purified from individual clones and used to transform fresh aliquots of *E. coli* DH5α-E. The fitness of each of these IPTG-induced strains, in the absence of antibiotics and in competition with the *lacZ*-marked negative control (which carried pCA24N-NoIns; Section 5.4.2), was measured using Lenski's protocol (Lenski *et al.*, 1991).

In each experiment, single colonies of the ASKA strain and of the negative control were used to inoculate separate 5 mL aliquots of the competition medium [2.0 g/L tryptone (Acumedia), 1.0 g/L yeast extract (HiMedia), 1.0 g/L NaCl, 34 μg/mL chloramphenicol, and 50 μM IPTG]. This medium was used because it accurately represents growth conditions in the PM plates (Zhou *et al.*, 2003). The cultures were grown at 37°C for 24 h, so that the two competing strains were pre-conditioned to the medium (Lenski *et al.*, 1991). Equal amounts of the two competing strains (~10^6 cfu per strain) were then mixed together in a fresh aliquot of the competition medium (5 mL), and the strains were grown in competition for another 24 h. The initial (t = 0 h) and final (t = 24 h) densities of each competitor were measured by spreading diluted aliquots of the culture on LB-chloramphenicol agar plates supplemented with X-gal (40 μg/mL). The relative fitness, *W*, of the ASKA strain was estimated as the ratio of its number of doublings in 24 h, divided by the number of doublings of the *lacZ*-marked negative control (Lenski *et al.*, 1991). Eight replicates were carried out to estimate the mean relative fitness of each ASKA strain.

5.4.4 Antibiotic susceptibility testing

The retransformed ASKA strains and the unmarked negative control strain from the fitness assays were grown in 10 mL LB-chloramphenicol to mid-log phase before IPTG (final concentration, 50 μM) was added to induce protein over-expression for 2 h. The cells were harvested by centrifugation (2,000 ×g, 15 min, 4°C), and the pellets were resuspended to OD_{600} ~0.2 in sterile water. These cell suspensions were used for MIC determination (see below). Reported MICs values were the average of at least two independent experiments.

5.4.4.1 E-tests

MICs for aztreonam, cefuroxime, cephalothin, ciprofloxacin, erythromycin, nalidixic acid, nitrofurantoin, penicillin G, rifampicin, tobramycin and trimethoprim were determined using E-test strips (AB bioMérieux). E-test strips are impregnated antibiotic gradients, which are depicted on each strip (for examples, see Figure 5.1 in Section 5.2.4). A lawn of ~10^6 cells was 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 and IPTG (50 μM). The relevant E-test strip was placed on top of the lawn

and the agar plates were incubated at 37°C for 24 h. The MIC was read off the E-test strip based on the size of the clearance zone on the bacterial lawn.

5.4.4.2 BROTH MICRODILUTIONS

MICs for benzethonium chloride, enoxacin, lincomycin, oxytetracycline, puromycin, sisomicin, spiramycin, sulfadiazine, and vancomycin were determined using the broth microdilution method (Wiegand *et al.*, 2008). Two-fold serial dilutions of an antibiotic were set up in 96-well microtiter plate. Each well was inoculated with ~10^5 cells in Mueller-Hinton broth (BD Difco™) supplemented with chloramphenicol and IPTG (50 μM). All wells contained a final volume of 100 μL. Cell growth was monitored at OD_{600} using a Synergy 2 plate reader (Biotek Instruments). The MIC was defined as the lowest concentration of an antibiotic that prevented cell growth (*i.e.*, OD_{600} <0.1).

CHAPTER 6

Concluding remarks

6.1 A LOOK BACKWARDS IN ENZYME EVOLUTION

In Chapters 2 and 3, I investigated a novel example of reciprocal promiscuity. The PLP cofactor that Alr and MetC have in common allowed the enzymes to catalyze each other's reaction (in addition to their native reaction), despite the fact that they are not homologues. These findings emphasize that promiscuity is usually a product of contingency, rather than being indicative of shared ancestry. After all, only a few key components or residues are critical for catalysis, and they are usually in the active site. For example, the promiscuous Kemp elimination activities of human serum albumin and a catalytic antibody were attributed to catalytic lysines in rudimentary active sites that were hydrophobic overall (James & Tawfik, 2001). For PurF, its promiscuous TrpF activity appeared to stem from its affinity for phosphoribosylated substrates, including 5-phosphoribosyl-1-pyrophosphate [the substrate for PurF's native reaction] and phosphoribosylanthranilate [the substrate for TrpF reaction] (Patrick & Matsumura, 2008).

In the absence of an enzyme scaffold, it is believed that the PLP cofactor can catalyze all of the enzymatic PLP-dependent reactions weakly (Eliot & Kirsch, 2004; Vacca *et al.*, 2008), due to the electrophilicity of the pyridine ring (Richard *et al.*, 2009). Recruitment of a protein scaffold would have enhanced this catalytic potential of PLP, as well as stabilizing the aldehyde group at position C4' of PLP by forming a covalent bond to a lysine in the active site. Simply put, once PLP is recruited into an active site, it can endow a protein

scaffold with new activities. It appears that at least five scaffolds have successfully recruited PLP, throughout evolutionary history (Christen & Mehta, 2001; Eliot & Kirsch, 2004). With PLP bound in their active sites, these scaffolds could evolve further into the PLP-dependent enzymes that we observe today.

However, it seems unlikely that all five scaffolds recruited PLP simultaneously. Instead, I propose that there was a single, most ancient PLP-dependent enzyme that was able to recognize a broad range of substrates and catalyze many different reactions. Presumably, this key evolutionary innovation (protein-PLP binding) would have been strongly selected, once it arose. As the primordial protein repertoire expanded, different protein scaffolds emerged. Some of these new scaffolds could also bind PLP, and this would have allowed subsequent specialization of each lineage, towards different substrate and reaction specificities. The MetC variants that I characterized in Chapter 3 provide evidence for my model. Any one of several mutations at position 113 improved the enzyme's promiscuous alanine racemase activity, while having modest effect on its native activity. Therefore, a protein such as MetC(P113S) mimics a viable, bifunctional ancestor from a time before $(\beta/\alpha)_8$-barrel proteins such as Alr had evolved to bind PLP.

In the context of longstanding evolutionary models (Section 1.1), the PLP-mediated promiscuity in Alr and MetC is not consistent with Horowitz's theory of retrogade evolution. Alr and MetC do not have the most important feature of enzymes that obey this evolutionary model — homology. Is reciprocal promiscuity in non-homologous enzymes consistent with Jensen's model? In this case, the answer is yes, and no. Consistent with his model, the promiscuous activities of Alr and MetC are analogous to the fortuitous activities of primitive enzymes. However, Jensen (as well as Horowitz) did not take the role of cofactors into account in his explanations for enzyme evolution. In his original arguments, Jensen assumed only that promiscuity is inherent in (primitive) enzymes. In light of the results presented in Chapters 2 and 3, it is apparent that the reactivity of smaller biomolecules, such as PLP, is also important to consider in evolutionary models. Indeed, these small biomolecules might have withstood harsh environments in the primitive world better than proteins, whose three-dimensional shapes must be maintained for their activities. Taken together, I propose a "cofactor-first" model for the very early evolution of metabolic networks, in which PLP (as well as other enzyme cofactors) played significant roles in driving the evolution of new enzymes.

My model is partly supported by a very recent study, which used a structural census analysis to conclude that the oldest aerobic enzyme was PLP synthase (Kim *et al.*, 2012). PLP synthase produces PLP from either pyridoxine 5'-phosphate (PNP) or pyridoxamine 5'-phosphate (PMP). The authors proposed that this enzyme appeared ~2.9 billion years ago. However, PLP-dependent enzymes are present in all three kingdoms of life, and archaea are proposed to have diverged from bacteria more than 4 billion years ago (Battistuzzi *et al.*, 2004). These results support my proposal that PLP was recruited by different scaffolds at

different times during evolution — and that horizontal gene transfer led to the widespread adoption of the resulting PLP-dependent enzymes. However, the timeline also implies that it may have been one of the PLP precursors (PNP or PMP) that was important as a promiscuous cofactor in the most ancient cells. Alternatively, PLP may have been synthesized via a different route before PLP synthase evolved.

A major question that follows from this discussion is, which lineage of PLP-dependent enzymes is the oldest? Furthermore, how would the biophysical and kinetic behaviours of this ancestral protein compare to extant PLP-dependent enzymes? A combination of phylogenetic and directed evolution studies will be useful in shedding light on the emergence of the very first PLP-dependent enzyme(s). Ancestral proteins inferred from phylogenetics analysis can be reconstructed in the laboratory, and their biochemical functions can be investigated accordingly. Recently, a 1 billion year-old dehydrogenase enzyme (Hobbs *et al.*, 2012) and a 4 billion year-old thioredoxin (Perez-Jimenez *et al.*, 2011) have been successfully reconstructed, suggesting that it may be possible to study very ancient PLP-dependent enzymes in the same way.

6.2 LOOKING FORWARDS IN ENZYME EVOLUTION

In Chapters 4 and 5, I used a library-on-library screen to probe the effects of protein over-expression on bacterial phenotypes, on a proteome-wide scale. This screen was developed to demonstrate that protein promiscuity and increased expression — the two tenets of the IAD model (Section 1.2) — could drive cellular adaptation. I found that many proteins, when over-expressed, could impart improved growth on *E. coli* that had been placed in environments containing either a novel nutrient or toxin. These findings demonstrate that protein over-expression may be a common, *de novo* adaptive response.

Protein over-expression also appears to be a feasible mechanism for cells to circumvent thiamine starvation (Chapter 4). In hindsight, this should not be too surprising, considering that promiscuous proteins can rescue many auxotrophies when they are expressed in high levels (Chapters 2 and 3). Protein over-expression did not only buffer against physiological perturbances, but also facilitate cellular adaptation to foreign environments (Chapters 4 and 5), via the (promiscuous) activities of a variety of proteins, including membrane transporters, prophage-derived proteins and regulatory elements. My overall findings demonstrated the diversity of functional innovations in a contemporary cell, and promiscuity is only a subset of this pool of innovations.

In many cases, the growth improvement conferred by each over-expressed protein was marginal. Nevertheless, it seems reasonable to suggest that these small increases in fitness can play significant roles during the initial stages of adaptation. Any increase in fitness means that a cell can produce greater numbers of viable descendants than its competitors in

a population. This improves the likelihood that descendants of that cell will acquire (rare) mutations that further increase their fitness.

Increased expression of promiscuous proteins underpins the IAD model for the evolution of new functions from pre-existing ones. Although protein over-expression was induced artificially (*i.e.*, regulated by a strong promoter) in this study, I argued that this approach mimics *bona fide* over-expression resulted from gene amplification events. For example, a recent study showed that two gene promoters were part of large amplicons (ranging from 18 to 271 kb) in *Acinetobacter baylyi* (Seaton *et al.*, 2012). Increased expression of the benzoate- and catechol-degrading enzymes resulting from these amplifications allowed the organism to grow on benzoate. Furthermore, the copy numbers of genes that are under selection can reach greater than 100 (Gaines *et al.*, 2010; Seaton *et al.*, 2012). This is comparable to the copy number of the pCA24N plasmid of the ASKA library.

Overall, the prevalence of promiscuity and its ability to drive cellular adaptation implies that the IAD model is biochemically feasible. To further demonstrate the connection between promiscuity, gene amplification and the emergence of new function, non-plasmid-containing bacterial populations could be evolved experimentally by serial passage, with increasing selection pressure. I predict that the ASKA clones identified in this thesis will be the genes that are amplified in the serial passage experiments, because of their fitness-enhancing effects.

6.3 FINAL COMMENTS

In summary, I have used both single gene/protein analyses and genome-/proteome-wide screens to address the two themes of this thesis (looking backwards and forwards in enzyme evolution). I have shown that promiscuity is common, and that proteins often have promiscuous activities that are the primary functions of non-homologous enzymes. However, *a priori*, promiscuity is mostly unpredictable. High-throughput experiments such as those I described remain an ideal way to gain broadly relevant insights into enzyme evolution.

The bacterium used throughout this work, *E. coli*, is one of the most well-studied organisms in science. Its proteome consists of over four thousand proteins, and a majority of their primary functions have been studied. However, the work described in this thesis emphasizes that much remains to be discovered, both about the primary functions of previously uncharacterized proteins, and the promiscuous activities of proteins that we thought we had already characterized. The evolutionary potential of *E. coli* appears to be vast, and this will undoubtedly be the same for other organisms on Earth.

APPENDIX I

General materials and methods

I.1 REAGENTS

Chemicals used throughout this research were purchased from Sigma-Aldrich, unless otherwise specified. Reagent purity was always analytical grade or higher. All solutions were prepared using autoclaved MilliQ® water.

I.2 GROWTH MEDIA AND ANTIBIOTICS

All *E. coli* strains were routinely grown in LB medium (ForMedium; 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) at 37°C (Bertani, 1951), with agitation at 180 rpm. Agar plates were prepared by adding bacteriological agar (ForMedium) to a final concentration of 1.5% (w/v). All growth media were supplemented with antibiotics and/or nutrients where necessary.

All general antibiotics, except chloramphenicol (Duchefa Biochemie), were bought from Melford. To make stock solutions, chloramphenicol was dissolved in ethanol, while all other antibiotics were dissolved in water. The final concentrations used were: ampicillin, 100 μg/mL; carbenicillin, 50 μg/mL; chloramphenicol, 34 μg/mL; gentamicin, 15 μg/mL; kanamycin, 30 μg/mL. All antibiotics were stored at −20°C.

I.3 BACTERIAL STRAINS

For long term storage, a single bacterial colony was propagated overnight in selective LB. Glycerol was added to a 300-µL aliquot of the saturated culture to 12.5% (v/v) final concentration, and the culture was stored at −80°C. Agar plates with cultured bacteria were stored at 4°C for no longer than one week.

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

Strain	Genotype	Reference(s)
BW25113 Δ*metC*	F⁻ Δ(*araD*–*araB*)567 Δ*lacZ*4787(::*rrnB*–3) λ⁻ *rph*–1 Δ(*rhaD*–*rhaB*)568 *hsdR*514 Δ*metC*::*Kan*^{R}	(Baba *et al.*, 2006)
BW25113 Δ*metE*	F⁻ Δ(*araD*–*araB*)567 Δ*lacZ*4787(::*rrnB*–3) λ⁻ *rph*–1 Δ(*rhaD*–*rhaB*)568 *hsdR*514 Δ*metE*::*Kan*^{R}	(Baba *et al.*, 2006)
DH5α–λ*pir*	F⁻ ϕ80*lacZ*ΔM15 Δ(*lacZYA*–*argF*)U169 *recA*1 *endA*1 *hsdR*17(r_k^-, m_k^+) *supE*44 *thi*–1 *gyrA*96 *relA*1 λ*pir*	Dr. Xue-Xian Zhang (Massey University)
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 a *Gm*^{R}-*lacZ* marker inserted at the *att*Tn7 site.	This study
K-12	Wildtype *E. coli*; F⁺ λ⁺	American Type Culture Collection (ATCC) 10798
MB2795	*araD*139 Δ(*araA*–*leu*)7697 *galU galK* Δ(*lac*)174 *rpsL thi*–1 Δ*dadX*::*frt* Δ*alr*::*frt*	Prof. Kurt Krause (University of Otago)
WMP001.1	MB2795 Δ*metC*::*Kan*^{R}	This study

I.4 PLASMIDS

All purified plasmid DNA was kept at −20°C.

Table I.2: Designation and characteristics of plasmids used in this study.

Plasmid	Description	Reference
pUC19	*Amp*^{R}; 2.7-kb control plasmid for transformation	Invitrogen
pBAD-*epMetC*(P113A)	*Amp*^{R}; plasmid-encoded *metC* variants with random mutations; used to construct a second-generation MetC library.	This study
pBAD-*epMetC*(P113Q)	*Amp*^{R}; plasmid-encoded *metC* variants with random mutations; used to construct a second-generation MetC library.	This study
pBAD-*epMetC*(P113S)	*Amp*^{R}; plasmid-encoded *metC* variants with random mutations; used to construct a second-generation MetC library.	This study
pBAD-*epMetC*(P113T)	*Amp*^{R}; plasmid-encoded *metC* variants with random mutations; used to construct a second-generation MetC library.	This study
pBAD-*metC*	*Amp*^{R}; plasmid-encoded wildtype *metC* from *E. coli*.	This study

Plasmid	Description	Reference
pBAD-*metC*(P113A)	Amp^R; plasmid-encoded *metC* with P113A mutation.	This study
pBAD-*metC*(P113A-9)	Amp^R; second-generation winner from the MetC(P113A) library; carried a A113P mutation.	This study
pBAD-*metC*(P113Q)	Amp^R; plasmid-encoded *metC* with P113Q mutation.	This study
pBAD-*metC*(P113Q-12)	Amp^R; second-generation winner from the MetC(P113Q) library; no additional mutation apart from its parental P113Q mutation.	This study
pBAD-*metC*(P113S)	Amp^R; plasmid-encoded *metC* with P113S mutation.	This study
pBAD-*metC*(P113S-2)	Amp^R; plasmid-encoded *metC* with P113S, P156L and A227T mutations; second-generation MetC winners.	This study
pBAD-*metC*(P113S-14)	Amp^R; plasmid-encoded *metC* with P113S and I251M mutations; second-generation MetC winners.	This study
pBAD-*metC*(P113S-16)	Amp^R; plasmid-encoded *metC* with P113S, K120N and P352A mutations; second-generation MetC winners.	This study
pBAD-*metC*(P113S-18)	Amp^R; second-generation winner from the MetC(P113S) library; no additional mutation apart from its parental P113S mutation.	This study
pBAD-*metC*(P113T)	Amp^R; plasmid-encoded *metC* with P113T mutation.	This study
pBAD-*metC*(P113T-20)	Amp^R; second-generation winner from the MetC(P113T) library; no additional mutation apart from its parental P113T mutation.	This study
pBAD/*Myc-His-B*	Amp^R; pBR322-derived expression vector with the PBAD promoter.	Invitrogen
pBAD/*Myc-His-lacZ*	Amp^R; a pBAD plasmid carrying the *lacZ* gene.	Invitrogen
pCA24N-*alr*	Cam^R; an ASKA plasmid carrying the *alr* gene.	(Kitagawa *et al.*, 2005)
pCA24N-*alr*(GFP−)	Cam^R; pCA24N-*alr* without the GFP tag.	This study
pCA24N-*alr*(R209E)	Cam^R; a codon mutation (R209E) in the *alr* gene; plasmid lacks the GFP tag.	This study
pCA24N-*bglA*	Cam^R; an ASKA plasmid carrying the *bglA* gene.	(Kitagawa *et al.*, 2005)
pCA24N-*cycA*	Cam^R; an ASKA plasmid carrying the *cycA* gene.	(Kitagawa *et al.*, 2005)
pCA24N-*dadX*	Cam^R; an ASKA plasmid carrying the *dadX* gene.	(Kitagawa *et al.*, 2005)
pCA24N-*epAlr-1*	Cam^R; selected winner from the Alr library; no mutation in the *alr* gene.	This study
pCA24N-*epAlr-2*	Cam^R; selected winner from the Alr library; no mutation in the *alr* gene.	This study
pCA24N-*epAlr-3*	Cam^R; plasmid-encoded *alr* with Y274F mutation; selected winner from the Alr library.	This study
pCA24N-*epAlr-11*	Cam^R; plasmid-encoded *alr* with Y274F and A335T mutations; selected winner from the Alr library.	This study
pCA24N-*epAlr-14*	Cam^R; plasmid-encoded *alr* with R266S and Y274F mutations; selected winner from the Alr library.	This study
pCA24N-*epAlr-19*	Cam^R; plasmid-encoded *alr* with M29I and Y274F mutations; selected winner from the Alr library.	This study
pCA24N-*epAlr-22*	Cam^R; plasmid-encoded *alr* with L241M, R263H and Y274F mutations; selected winner from the Alr library.	This study
pCA24N-*epAlr-24*	Cam^R; plasmid-encoded *alr* with T97S, P252R and Y274F mutations; selected winner from the Alr library.	This study
pCA24N-*epAlr-28*	Cam^R; plasmid-encoded *alr* with I242V and Y274F mutations; selected winner from the Alr library.	This study

Plasmid	Description	Reference
pCA24N-*epAlr-41*	*Cam*^{R}; plasmid-encoded *alr* with V150I and Y274F mutations; selected winner from the Alr library.	This study
pCA24N(*KanR*)-*epMetC*	*Kan*^{R}; plasmid-encoded *metC* variants with random mutations; used to construct first-generation MetC library.	This study
pCA24N(*KanR*)-*metC*(P113Q)	*Kan*^{R}; plasmid-encoded *metC* with P113Q mutation; first-generation MetC winners.	This study
pCA24N(*KanR*)-*metC*(P113A)	*Kan*^{R}; plasmid-encoded *metC* with P113A mutation; first-generation MetC winners.	This study
pCA24N(*KanR*)-*metC*(P113T)	*Kan*^{R}; plasmid-encoded *metC* with P113T mutation; first-generation MetC winners.	This study
pCA24N(*KanR*)-*metC*(P113S)	*Kan*^{R}; plasmid-encoded *metC* with P113S mutation; first-generation MetC winners.	This study
pCA24N(*KanR*)-tS1	*Kan*^{R}; derived from pCA24N-tS1(no-ext).	Dr. Monica Gerth (Massey University)
pCA24N-*metC*	*Cam*^{R}; an ASKA plasmid carrying the *metC* gene.	(Kitagawa *et al*., 2005)
pCA24N-*metC*(GFP−)	*Cam*^{R}; pCA24N-*metC* without the GFP tag.	This study
pCA24N-*metE*	*Cam*^{R}; the ASKA plasmid carrying the *metE* gene.	(Kitagawa *et al*., 2005)
pCA24N-NoIns	*Cam*^{R}; pCA24N plasmid without any gene insert; 5.2 kb.	This study
pCA24N-*trpD*(GFP−)	*Cam*^{R}; an ASKA plasmid carrying the *trpD* gene; GFP tag removed.	Dr. Wayne Patrick/ Lab stock
pCA24N-tS1(no-ext)	*Cam*^{R}; an ASKA plasmid carrying a mutated *purF* gene; GFP tag removed.	Dr. Wayne Patrick/ Lab stock
pCA24N-tS1-corr	*Cam*^{R}; an ASKA plasmid carrying the wildtype *purF* gene.	(Patrick & Matsumura, 2008)
pCA24N-*wcaL*	*Cam*^{R}; an ASKA plasmid carrying the *wcaL* gene.	(Kitagawa *et al*., 2005)
pKD46	*Amp*^{R}; temperature-sensitive plasmid expressing the arabinose-induced λ Red recombinase.	Lab stock (from the The Coli Genetics Stock Centre, Yale)
pUC18T-mini-Tn*7*T-*Gm*-*lacZ*	*Amp*^{R}, *Gm*^{R} on mini-Tn*7*T suicide vector; ori_{ColE1}; *lacZ* transcriptional fusion vector.	Dr. Xue-Xian Zhang (Massey University)
pUC18R6K-mini-Tn*7*T-*Gm*	*Amp*^{R}, *Gm*^{R} on mini-Tn*7*T suicide vector; ori_{R6K}.	Dr. Monica Gerth
pUC18R6K-mini-Tn*7*T-*Gm*-*lacZ*	*Amp*^{R}, *Gm*^{R} on mini-Tn*7*T suicide vector; ori_{R6K}; *lacZ* transcriptional fusion vector.	This study
pUX-BF13	*Amp*^{R}; ori_{R6K}; helper plasmid encoding the TnsABC+D.	Dr. Monica Gerth

I.5 ANALYTICAL SOFTWARE

MacVector (MacVector, Inc.) was used to view, align, annotate, modify and design all DNA and protein sequences. It was also used to import sequencing chromatograms for further analyses. Molecular structures were visualized and manipulated using MacPymol (Schrödinger LLC).

I.6 OLIGONUCLEOTIDES

All primers were custom synthesized by Integrated DNA Technologies (Coralville, Iowa). Upon arrival, TE buffer (10 mM Tris-HCl, pH 8.0; 1 mM EDTA) was used to resuspend lyophilized primers to a final concentration of 100 μM. Aliquots of 10 μM working stock were prepared by diluting the 100 μM master stock with sterile MilliQ® water. All primers were stored at −20°C.

Table I.3: Primers used in this study.

Primer	Sequence (5′ → 3′)	Reference
Alr.rev	GCGGCCGCATAGGCCTTAATCCACGTATTTC	This study
Alr-R209E.for	TGACTGGGTGGAACCGGGCATC	This study
Alr-R209E.rev	GATGCCCGGTTCCACCCAGTCA	This study
ecGlmS.rev	CAGTTAACTGTGCTGTTGATGC	This study
ecTn7R.for	CAGCATAACTGGACTGATTTCAG	This study
His-metC.for	ATGAACATGTCTCACCATCACCATCACCAT	This study
His-metC.rev	CGCATCTAGATTATACAATTCGCGCAAAAC	This study
KeioKan^R.out	CGGTGCCCTGAATGAACTGC	Lab stock
metC877.rev	ATGCCCAACCAGATATTTGGTG	This study
MetC(GFP−).rev	GCGGCCGCATAGGCCTTATACAATTCGCGCA	This study
metC-dwnstr.rev	GACTTTTCACAATAAAATGTCTGCAAAATTGTCCAAAAG	This study
metC-upstr.for	CGCGAATATTCATGCTAGTTTAGACATCCAGAC	This study
pBAD-60	ATCGCAACTCTCTACTGTTTCTCCATAC	Lab stock
pBAD.rev	GAAAATCTTCTCTCATCCGCCAAAAC	This study
pCA24N.for	GATAACAATTTCACACAGAATTCATTAAAGAG	(Patrick *et al.*, 2007)
pCA24N.rev	CCCATTAACATCACCATCTAATTCAAC	(Patrick *et al.*, 2007)
pCA24N.rev2	CAAATCCAGATGGAGTTCTGAGG	(Patrick & Matsumura, 2008)
pCA24N-NoIns.for*	*Phos*-GGCCTATGCGGCCGCAGTAAAG	This study
pCA24N-NoIns.rev*	*Phos*-GGCCCTCAGGGCCGGATCCGTAT	This study
pUCrev+120	GCAGTGAGCGCAACGCAATTA	Lab stock
Rev+270	TAACCGTATTACCGCCTTTGAGTGAGC	Lab stock
Univ-70	ATGTGCTGCAAGGCGATTAAGTTGGGT	Lab stock

*Primers were phosphorylated at the 5′ terminus.

I.7 AGAROSE GEL ELECTROPHORESIS

Depending on the size of DNA, gels were prepared with 0.8% (w/v) agarose for 1–10 kb, or 1.5% (w/v) agarose for 100–1000 bp. Agarose powder (Axygen) was dissolved in 1× TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA, pH 8.0). The nucleic acid stain, ethidium bromide (Bio-Rad), was added to a final concentration 0.5 μg/mL. Four volumes of each DNA sample were mixed with one volume of 5× GelPilot DNA Loading Dye (Qiagen) prior to loading into the well, except when samples were pre-mixed with special reaction buffer (see Section I.10). The samples were run alongside appropriate DNA ladders at 100 V until the bottom tracking dye had migrated through ¾ of the gel. Agarose gels were viewed using a UV transilluminator (part of the Gel Doc™ 2000 Gel Documentation System, Bio-Rad).

When DNA was to be purified from the gel (see Section I.8), 1× SYBR Safe™ Gel DNA stain (supplied as 10,000× concentrate by Invitrogen) was used instead of ethidium bromide. Such gels were viewed using a Safe Imager™ Blue Light Transilluminator (Invitrogen) to eliminate DNA exposure to UV light.

I.8 DNA EXTRACTION AND CLEAN-UP

Genomic DNA was extracted using the Wizard® Genomic DNA Purification Kit (Promega), according to the manufacturer's instructions. Plasmid DNA was routinely purified using the QIAprep Spin Miniprep kit (Qiagen). DNA from PCRs, agarose gels and ligation reactions was cleaned up with the QIAquick purification kit series (Qiagen). When a concentrated DNA sample was desired (*e.g.* a 10-μL DNA eluate), MinElute Spin columns (Qiagen) were used in place of the usual QIAprep or QIAquick Spin columns.

I.9 DNA QUANTIFICATION

Each DNA sample was diluted to 50 μL in a UVette (Eppendorf), and its concentration was determined spectrophotometrically in an Eppendorf Biophotometer. Alternatively, the amount of DNA could be estimated by comparing its band intensity with that of a known DNA marker on an agarose gel.

I.10 PCR SCREENING

Each routine PCR screen was performed as a 20-μL reaction, in 0.2 mL thin-walled tubes. Each reaction contained 1× Green GoTaq® Reaction buffer (pre-mixed with two tracking dyes; Promega), 0.25 mM dNTP mix (iNtRON Biotechnology), 1.25 U i-Taq™ DNA

polymerase (iNtRON Biotechnology), 0.5 μM of the forward and reverse primers, and DNA template (either 1 μL of saturated bacterial culture or plasmid DNA, or one colony from a cultured agar plate).

The standard thermocycling parameters were as follows: 95°C for 5 min, 35 cycles of 95°C for 1 min, 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 T_m of the two primers. All PCR reactions were carried out in an MJ Mini Thermal Cycler (Bio-Rad).

I.11 Restriction digest and Ligation

All restriction endonucleases were obtained from New England Biolabs. A typical digestion contained 1× reaction buffer, 0.1–2 μg of DNA and 2-10 U restriction endonuclease per 10 μL reaction. Unless otherwise stated, DNA was digested for 1–16 h, depending on different endonuclease stabilities as reported by the manufacturer. Other recommendations from the manufacturer included incubation temperature, buffer compatibilities (especially for double digestion), bovine serum albumin supplementation and heat inactivation.

All ligation reactions were performed in 0.2 mL thin-walled tubes. Normally, a ligation reaction contained 1× T4 DNA Ligase Buffer, 2.5 Weiss units of T4 DNA Ligase (Fermentas) per 10 μL reaction, appropriate amounts of linearized vector and digested gene insert (in 3-fold molar excess than vector). The ligation reaction was incubated at 16°C for 4–16 h, before heat-inactivation at 65°C for 10 min. Ligated products were either cleaned up (see Section I.8), or diluted at least 5-fold using sterile MilliQ® water before being used to transform *E. coli*.

I.12 Electrocompetent cells

Electrocompetent cells were prepared as described (Sambrook & Russell, 2001). An *E. coli* strain (from either freezer stock or a single colony) was grown overnight in 4 mL LB supplemented with the appropriate nutrients or antibiotics. The saturated 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) (Hanahan, 1983), supplemented with the appropriate nutrients or antibiotics. The culture was incubated at 37°C to an OD_{600} ~0.4. The cells were evenly divided into 6 × 50 mL pre-chilled centrifuge tubes, and allowed to incubate on ice for another 1 h. All following steps were carried out at 4°C.

The cells were pelleted at 2,200 ×g for 17 min, at 4°C (Heraeus Multifuge 1S-R), washed with 40 mL of pre-chilled 10% (v/v) glycerol, and re-pelleted as described. The six cell pellets were resuspended with 10% glycerol and pooled into 4 × 40 mL. The pelleting

step was repeated. The four pellets were resuspended using 10% glycerol and further pooled into 2 × 40 mL. The cells were spun down similarly, and the resulting two pellets were resuspended with 10% glycerol and pooled into 1 × 40 mL. The cells were pelleted for the final time.

The final cell pellet was weighed, 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 cell resuspension was dispensed as 50 μL aliquots, and either used immediately for electroporation (see Section I.13), or stored at −80°C until use.

I.13 TRANSFORMATION

For library construction, electrocompetent cells were prepared on the same day that they were transformed. Otherwise, frozen electrocompetent cells were thawed on ice and used for general cloning purposes.

Electroporation was carried out according to standard procedures (Sambrook & Russell, 2001). Generally, plasmid DNA or ligated products was added to a 50-μL aliquot of electrocompetent cells, and allowed to incubate on ice for 1 min. The cells were transferred to a sterile, pre-chilled electroporation cuvette with 0.1 cm gap (Bio-Rad). The cuvette was placed in the shocking chamber, and pulsed at 1.8 kV (Bio-Rad Gene Pulser II; other settings: 200 Ω and 25 μF). Specifically, the MB2795 and WMP001.1 strains of *E. coli* (Table I.1 in Section I.3) were always pulsed at 2.5 kV in electroporation cuvette with 0.2 cm gap (Bio-Rad). Time constants typically ranged from 4.5–5.5 ms for all electroporations. The electroporated cells were resuspended in 0.5 mL of pre-warmed SOC medium (SOB + 20 mM glucose) (Hanahan, 1983), and transferred to a sterile glass test tube. The cells were incubated at 37°C for 1 h. Recovered cells were then spread on selective LB agar plates containing the appropriate antibiotics and/or nutrients. Agar plates were incubated overnight at 37°C. Most *E. coli* strains (except strains MB2795 and WMP001.1; see Section I.3) prepared using the methods from Section I.12 and I.13 typically yielded transformation efficiencies of $\geq 10^8$ cfu/μg of pUC19 (Invitrogen).

I.14 DNA SEQUENCING

All amplicon and plasmid sequencing was carried out by either the Massey Genome Service (Palmerston North), or Macrogen Inc. (South Korea). Both companies utilize automated capillary-based ABI3730 DNA Analyzers (Applied Biosystems) for routine dye-terminator sequencing.

I.15 SDS-PAGE

The discontinuous buffer system (Laemmli, 1970) was used for all SDS-PAGE carried out in this study. The stock solutions for making polyacrylamide gels were 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).

All resolving gel mixtures consisted of 12% acrylamide : *N*,*N*′-methylene-bis-acrylamide (29 : 1) solution, 375 mM Tris-HCl (pH 8.8), 0.1% sodium dodecyl sulfate, 0.05% 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.

Polyacrylamide gels of 1 mm thickness were prepared using the Mini-PROTEAN gel casting assembly (Bio-Rad). The resolving gel mixture was added between the glass plates until it was ~3.5 cm from the top of the plates. The gel mixture was then layered with 1 cm of 70% (v/v) ethanol, and gel polymerization was allowed to take place. After ~30 min, the ethanol was removed, and the polymerized resolving gel was rinsed with 1 mL of 1.5 M Tris-HCl (pH 8.8). The stacking gel mixture was poured on top of the resolving gel before inserting a clean 10-well comb. Similarly, the stacking gel was allowed to polymerize for ~30 min. The polymerized gel was either used immediately, or wrapped in moistened paper towels and stored at 4°C. Stored gels were usually used within two weeks.

Protein samples 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 prior to electrophoresis. In the meantime, the polyacrylamide gel was assembled into the Mini-PROTEAN Tetra Cell electrophoresis tank (Bio-Rad). After removing the comb from the gel, the heated samples were loaded into each well. 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 until the tracking dye was 0.5 cm from the bottom of the gel. The gel was removed from the glass plates, and stained in Coomassie Blue [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 then washed extensively with water. The gel was destained with Destaining Solution (6 volumes of water : 3 volumes of methanol : 1 volume of acetic acid) for at least 1 h with gentle rocking.

[Blank page]

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: Valerie Wooi Chee Soo

Name/Title of Principal Supervisor: Dr. Wayne M. Patrick

Name of Published Research Output and full reference:

Soo, V. W. C., Hanson-Manful, P. & Patrick, W. M. (2011). Artificial gene amplification reveals an abundance of promiscuous resistance determinants in Escherichia coli. Proc Natl Acad Sci USA, 108:1484-1489.

In which Chapter is the Published Work: Chapters 4 and 5

Please indicate either:

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

  and / or

- Describe the contribution that the candidate has made to the Published Work:

  The candidate designed the research methodology, performed 90% of the research, analyzed the data, and wrote the first draft of the resulting manuscript. Contributions by other authors have been specifically acknowledged at the beginning of, and throughout, the relevant thesis chapters.

Candidate's Signature — 20 February, 2012 — Date

W. M. Patrick

Principal Supervisor's signature — 20 February, 2012 — Date

APPENDIX III

Publication arising from this work

No reprint permission is required from the publisher as part of author's rights. The article is also freely available through the journal's open access policy.

[Blank page]

Artificial gene amplification reveals an abundance of promiscuous resistance determinants in *Escherichia coli*

Valerie W. C. Soo, Paulina Hanson-Manful, and Wayne M. Patrick[1]

Institute of Natural Sciences, Massey University, Auckland 0632, New Zealand

Edited by Richard E. Lenski, Michigan State University, East Lansing, MI, and approved November 30, 2010 (received for review August 13, 2010)

Duplicated genes provide an important raw material for adaptive evolution. However, the relationship between gene duplication and the emergence of new biochemical functions is complicated, and it has been difficult to quantify the likelihood of evolving novelty in any systematic manner. Here, we describe a comprehensive search for artificially amplified genes that are able to impart new phenotypes on *Escherichia coli*, provided their expression is up-regulated. We used a high-throughput, library-on-library strategy to screen for resistance to antibiotics and toxins. Cells containing a complete *E. coli* ORF library were exposed to 237 toxin-containing environments. From 86 of these environments, we identified a total of 115 cases where overexpressed ORFs imparted improved growth. Of the overexpressed ORFs that we tested, most conferred small but reproducible increases in minimum inhibitory concentration (≤16-fold) for their corresponding antibiotics. In many cases, proteins were acting promiscuously to impart resistance. In the absence of toxins, most strains bore no fitness cost associated with ORF overexpression. Our results show that even the genome of a nonpathogenic bacterium harbors a substantial reservoir of resistance genes, which can be readily accessed through overexpression mutations. During the growth of a population under selection, these mutations are most likely to be gene amplifications. Therefore, our work provides validation and biochemical insight into the innovation, amplification, and divergence model of gene evolution under continuous selection [Bergthorsson U, Andersson DI, Roth JR (2007) *Proc Natl Acad Sci USA* 104:17004–17009], and also illustrates the high frequency at which novel traits can evolve in bacterial populations.

antibiotic resistance | evolutionary innovation | molecular evolution | protein promiscuity | phenotype microarray

All species must adapt to survive in changing environments. At the molecular level, the presence of novel nutrients and toxins can drive the evolution of proteins that recognize or metabolize them (1, 2). Duplicated copies of preexisting genes provide the primary genetic source for functional innovation (3, 4). However, the relationship between duplication and the emergence of new biochemical functions is complicated because gene duplications are generally expected to be either selectively neutral (5) or deleterious (6). Therefore, it has been difficult to systematically assess the likelihood of evolving novel traits. In the present study, we have sought to address this gap in the Darwinian paradigm.

Numerous hypotheses have been proposed to explain the fates of duplicated genes within populations (7–9). A particularly appealing model identifies a mechanism by which selection can act continuously to favor both an increase in gene dosage, and divergence of one copy from the parental gene. This model, termed adaptive radiation (10) or the innovation, amplification, and divergence (IAD) model (11), is rooted in the notion that many existing proteins display broad specificity and secondary activities in addition to the function that they evolved to carry out (12). In primordial times, this "promiscuity" would have allowed many metabolic functions to be carried out by a minimal number of multitasking proteins (13). In more specialized, modern-day proteins, promiscuity is likely to be a product of contingency: active sites typically contain a variety of reactive groups (proton donors and acceptors, metal ions, and so forth) that could (by chance) play roles in secondary reactions (14). These secondary activities are assumed to provide the necessary reservoir of selectable functions.

In the IAD model, selection is imposed on a weak, promiscuous activity. To increase fitness, duplications (and higher-order amplifications) of the promiscuous gene are selected and maintained in the population. In turn, this increase in copy number improves the likelihood of point mutations that enhance the promiscuous activity, by increasing the number of mutational targets yet still allowing at least one copy to retain the parental activity (11).

Two central tenets of the IAD model are: (*i*) that gene amplification events are common; and (*ii*) that promiscuous activities are widespread, as well as being biochemically and mechanistically diverse. Experiments with unselected bacterial populations have shown the first point to be accurate (15), with cells that bear duplications at any given locus reaching steady-state frequencies of $\sim 10^{-3}$ (16). We are interested in cataloguing the promiscuous activities within an entire proteome to understand the protein biochemistry that underlies the evolution of new phenotypes.

Multicopy suppression experiments are a powerful and tractable way of demonstrating that adaptation can begin with the increased expression of promiscuous proteins. In these experiments, conditionally auxotrophic bacterial strains are transformed with plasmid-encoded libraries of ORFs. The libraries are screened for genes that can suppress the phenotype of the original mutation, provided they are present in high copy number (17–19) or overexpressed (20, 21). Previously, we used the multicopy suppression approach in an attempt to assess the prevalence of promiscuity, on a proteome-wide scale. We made use of the ASKA (A Complete Set of *Escherichia coli* K-12 ORF Archive) library, which comprises every *E. coli* ORF cloned into the expression vector pCA24N (22). The plasmids of the library were pooled and used to transform 104 conditionally auxotrophic, single-gene deletion strains. Twenty-one of the auxotrophies were specifically suppressed by the overexpression of noncognate *E. coli* genes (23). The implication was that many proteins are multifunctional; however, the ability of proteome-wide promiscuity to drive an organism's adaptation to new environments remained unknown. The next logical step was to survey the entire *E. coli*

Author contributions: V.W.C.S. and W.M.P. designed research; V.W.C.S. and P.H.-M. performed research; V.W.C.S. and W.M.P. analyzed data; and V.W.C.S. and W.M.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

See Commentary on page 1199.

[1]To whom correspondence should be addressed. E-mail: w.patrick@massey.ac.nz.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1012108108/-/DCSupplemental.

proteome for its latent ability to confer genuinely new phenotypes (rather than to recreate old ones).

Here, we describe a comprehensive search for promiscuous proteins that can impart new phenotypes on *E. coli*. We provide experimental evidence that the overexpression of preexisting *E. coli* proteins can provide resistance to >80 antibiotics and toxins. Our results suggest that the evolution of novel traits is surprisingly likely, and that even the genomes of well-characterized bacteria harbor substantial reservoirs of latent resistance determinants.

Results

Library-on-Library Screen. We have conducted a global survey to uncover latent resistance genes in the *E. coli* genome. *E. coli* cells were transformed with the pooled plasmids of the ASKA ORF library and then used to inoculate every well of 10 phenotype microarray (PM) plates (Fig. 1 *A* and *B*). In total, these 10 PM plates contained 237 toxins (including antibacterial agents), with each toxin present at four concentrations (24). Growth in each toxin-containing well could be monitored by purple color development because of the presence of a tetrazolium indicator dye (25). A negative control (*E. coli* harboring the empty ASKA vector, pCA24N-NoIns) was used to inoculate a second set of PM plates in parallel. By comparing the growth rates of the ASKA pool and the control, we were able to screen for examples in which the isopropyl-β-D-thiogalactopyranoside (IPTG)-induced overexpression of ASKA-encoded genes conferred increased fitness (i.e., an improved capacity for survival and reproduction) in the presence of a toxin.

The ASKA expression vector, pCA24N, has the same modified pMB1 replication origin as pQE30 (Qiagen), which gives it a copy number of 300 to 400 per cell. The pCA24N backbone also encodes the *lacI*^{q} repressor, for tightly controlling the expression of each ASKA ORF. Therefore, we were specifically seeking examples where protein overexpression was essential for increased fitness. However, induction with a high concentration of IPTG (1 mM) severely inhibits the growth of 51% of ASKA strains (including almost all that overexpress membrane proteins), and in addition, another 28% show moderate growth inhibition (22). To avoid this source of bias, we used a significantly lower concentration of IPTG (50 μM) to induce protein expression. This amount is sufficient for derepression of the pCA24N T5-*lacO* promoter, but minimizes fitness differences associated with overexpression of different ORFs (20, 23).

In duplicate screens, the ASKA pool out-grew the negative control in the presence of 99 toxins (42% of those tested). In the majority of cases, these improved growth phenotypes manifested as small but reproducible changes in the intensity of tetrazolium color development within the first 48 h of incubation (Dataset S1). The ASKA pool showed enhanced growth in 50 of the 78 antibacterial agents (64%) contained within the PM plates. We also noted increased growth rates in the presence of 49 of the other 159 toxins that were tested (31%), which included antifungal and antiparasitic agents, inorganic salts, and dyes (Dataset S1).

Identification of Novel Resistance Genes. Next we sought to deconvolute the results of our library-on-library screen by isolating individual ASKA clones that were responsible for the observed improvements in growth rates. In each of the 99 conditions identified above, we used serial transfer to enrich for the fittest clones within the ASKA pool (Fig. 1*C*). After two transfers, cells from the enriched population were plated under nonselective conditions. PCR and DNA sequencing revealed the plasmid-encoded genes that had imparted improved growth.

An ASKA ORF was scored as a hit if it occurred at least twice among the eight clones sequenced from a serially passaged population. Given the large size of the ASKA pool (5,272 clones), it was highly unlikely ($P \approx 0.005$) that any two of the eight randomly chosen colonies would contain the same ASKA plasmid, unless that clone had been positively selected during serial transfer. Our approach yielded hits from 86 of the 99 conditions tested, including representatives of all of the compound classes that were tested (Table 1). For example, the PM plates contained 22 β-lactam antibiotics, and we isolated ASKA clones that improved growth in 11 of them (Table 1, row 1). Our serial enrichment strategy yielded a single hit from 60 of the 86 conditions. In 38 of these, the ASKA clone was enriched to apparent homogeneity (i.e., eight of eight sequenced clones were identical). However, serial transfer gave two different hits in 23 conditions,

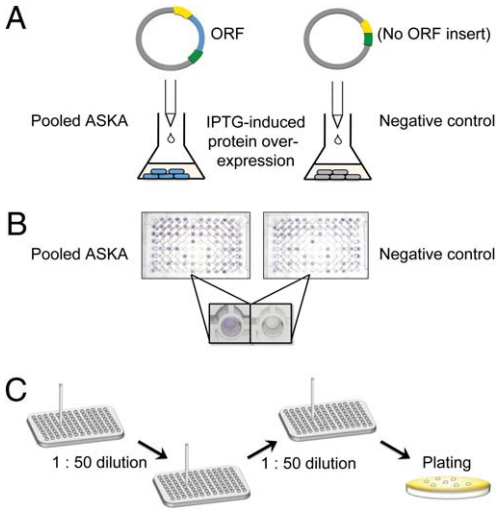


Fig. 1. The library-on-library screen and subsequent isolation of ASKA-encoded resistance genes. (*A*) *E. coli* cells harboring the pooled plasmids of the ASKA ORF collection (blue) and the negative control clone (gray) were grown to midlog phase in parallel, and protein overexpression was induced by the addition of IPTG (50 μM). (*B*) Both cultures were used to inoculate every well of two sets of PM plates. A well was scored positive when the ASKA pool out-grew the negative control, as shown by more rapid tetrazolium color development (*Inset*). (*C*) For each positive PM well, the fittest ASKA clone (or clones) was isolated by two rounds of serial transfer before an aliquot of the enriched culture was plated on nonselective medium to allow identification of the ASKA ORF (or ORFs).

Table 1. Summary of hits obtained from the library-on-library screen

Class of toxic compound	Number of positives*	Number of ASKA hits†
Antibacterials		
β-Lactams	11 (of 22)	14
Aminoglycosides	4 (of 14)	4
Anti-folates	7 (of 9)	10
Quinolones	4 (of 9)	6
Tetracyclines	3 (of 8)	4
MLS antibiotics	6 (of 7)	10
Glycopeptides	1 (of 3)	1
Nitrofurans	3 (of 3)	4
Rifamycins	1 (of 2)	2
Steroid antibacterial	1 (of 1)	1
Other toxins	45 (of 159)	59
Total	86 (of 237)	115

*The number of toxic compounds for which screening and serial enrichment yielded at least one ASKA-encoded resistance gene. The total number of toxins that were screened in each class is listed in parentheses.
†Total number of ASKA ORFs that were isolated from PM wells containing toxins of each compound class.

and three hits in another three cases (the β-lactam, aztreonam; the antiseptic, benzethonium chloride; and the toxic alkaloid, sanguinarine). Overall, therefore, we identified 115 positively selected hits from 86 toxin-containing environments (Table 1 and Dataset S1).

In 13 cases, serial transfer failed to enrich any individual ASKA clone to a frequency of two or more in the eight sequenced clones (Dataset S1). One explanation is that there were many plasmid-encoded ORFs that all imparted fitness advantages (of similar magnitudes) in those conditions. It is also possible that chromosomal mutations were responsible for improved growth, although this is unlikely when the experimental design (two independent ASKA pools, both required to out-grow negative control populations) and the small size of each input population (~50,000 cells) are considered. Alternatively, further transfers may have been required to identify the fastest-growing clones in the ASKA pool. By imparting stringent criteria for identifying novel resistance genes, we minimized the probability of isolating false-positives.

Diverse Mechanisms of Resistance. Our initial focus was on identifying proteins—and particularly enzymes—that possessed weak secondary activities. However, it rapidly became apparent that our library-on-library screen had uncovered many different classes of protein that were contributing to resistance. Therefore, we used the EcoCyc (26) and UniProt (27) databases, in addition to the references cited in Dataset S1, to assign proposed modes of action for each of the 115 cases that we had identified (Table 2).

Eighteen genes occurred multiple times and were responsible for improving growth in a total of 72 toxin-containing environments (Table S1), suggesting broad and nonspecific mechanisms. Consistent with previous results (28–31), the expression of efflux pumps and transporters was common among this subset of compound nonspecific resistance determinants. For example, expression of the Bcr multidrug transporter increased growth in the presence of three antifolate antibiotics (sulfachloropyridazine, sulfathiazole, and sulfisoxazole), two tetracyclines (oxytetracycline and rolitetracycline), and one β-lactam (cefoperazone). An uncharacterized outer-membrane β-barrel, YcbS, improved growth in the presence of four toxic compounds: a mutagen, 9-aminoacridine; an uncoupling agent, FCCP; a herbicide, methyl viologen; and an alkaloid, sanguinarine. Similarly, the MdtM and Cmr efflux pumps improved growth in the presence of five and three toxins, respectively (Table S1). The selection of these membrane proteins confirmed that we had minimized any bias associated with IPTG-induced overexpression.

We also discovered regulators, stress-response proteins, and proteins from capsule biosynthesis that each improved fitness in the presence of multiple toxins (Table S1). The only readily predictable example was the transcriptional activator MarA, which we isolated from environments containing two quinolones (ciprofloxacin and ofloxacin), a tetracycline (penimepicycline), a macrolide (tylosin), and a mutagen (proflavine). MarA is known to mediate multidrug resistance by up-regulating the efflux system and down-regulating membrane permeability (32, 33). In other cases, we found genes that had not previously been associated with resistance, including the biofilm modulation protein, Bdm, the small stress response protein, YcgZ, and the global regulator of transcription, CpdA (cAMP phosphodiesterase). The likely pleiotropic effects of these in altering the expression of downstream resistance determinants can be rationalized. Another prominent hit was the UDP-glucose 4-epimerase (GalE) ASKA strain, which was isolated from seven PM wells. GalE is involved in a variety of galactose metabolic processes, including glycolysis and the biosynthesis of lipopolysaccharide and capsule components. The host strain, *E. coli* DH5α-E, is *gal*$^-$, although the exact mutations responsible for this phenotype are not known. It is possible that the fitness effects seen when GalE was overexpressed were simply a result of complementation of the *gal*$^-$ mutation. However, if this was the case, we would have expected the GalE-expressing ASKA strain to out-grow the negative control in many more than seven of the 237 environments that we tested. Instead, GalE expression appears to mediate a suite of pleiotropic effects that are relevant for the evolution of resistance. Finally, other multiple-occurring ORFs hint at the functions of previously uncharacterized proteins (Table S1). For example, the YejG-expressing clone was isolated from PM wells containing apramycin, sisomicin, and tobramycin, suggesting that it may possess broad enzymatic activity toward these aminoglycoside antibiotics.

In 43 cases, an ASKA-encoded gene improved growth in the presence of a single toxin (Dataset S2). These compound-specific responses included 12 cases in which metabolic enzymes appeared to display catalytic promiscuity or substrate ambiguity. Examples included the carbonic anhydrase, CynT, evincing growth in the presence of enoxacin [reminiscent of the human enzyme's promiscuous esterase activity (34)] and the purine nucleotidase, YrfG, increasing fitness in the presence of a toxic analog (6-mercaptopurine). Finally, 13 of the 17 instances in which we were unable to assign a mechanism of action (Table 2) involved uncharacterized ASKA ORFs that were isolated from a single environment (Dataset S2).

Table 2. Proposed modes of action for 115 cases of resistance imparted by ASKA ORFs

Mechanism of resistance	Frequency
Efflux pump/transporter	22
Regulatory effect	19
Catalytic promiscuity and/or substrate ambiguity	15
Prophage gene	14
Stress response	13
Envelope/capsule synthesis	8
Biofilm formation and regulation	5
Overexpression of the drug's target	1
Maintenance of metabolic flux	1
Unknown	17
Total	115

Interplay of Fitness and Resistance. Mutations that confer antibiotic resistance can have a variety of effects on the fitness of the cell in the absence of drug (35). Resistance mutations often bear fitness costs, although examples of cost-free resistance (36) and mutual compensation (i.e., simultaneous increases in resistance and fitness) have also been described (37). Our library-on-library screen revealed cases where ASKA clones were responsible for enhancing fitness in the presence of a toxin. It was unclear from our screen whether the plasmid-encoded ORFs were improving the general vigor of the host or increasing resistance. To investigate this further, we measured the fitness value (W) for each of 25 strains relative to the negative-control strain (i.e., *E. coli* DH5α-E with pCA24N-NoIns). Fitness tests were conducted in the absence of toxin, but the presence of 50 μM IPTG. Using the same set of 25 strains, we also determined the minimum inhibitory concentrations (MICs) for 29 toxin/strain combinations. We chose to focus on anti-bacterial compounds and to test examples of each mechanistic class (as defined in Table 2).

In the absence of toxin, only two of the 25 ASKA strains showed significantly reduced fitness (i.e., $W < 1$; $P < 0.01$) compared with the negative control (Table 3). Nine ASKA strains were not significantly different from the control ($P > 0.01$), and 14 of the strains out-competed it ($W > 1$; $P < 0.01$). In contrast, almost all of MICs determined for ASKA strains were increased compared with those that were measured for the negative-control strain (Table 3).

By far the largest increase in MIC was for trimethoprim, where overexpression of its intracellular target, dihydrofolate reductase (*folA*), effected very high levels of resistance (at least 500-fold greater than the negative control strain). The FolA-expressing strain was also one of the least fit in the absence of antibiotic ($W = 0.84$). Expression of the multidrug transporters MdtM and Cmr led to 16-fold and 8-fold increases in the MICs for puromycin and lincomycin, respectively, without imparting fitness costs in a toxin-free environment. The remainder of the selected ASKA strains exhibited modest increases (<8-fold) in MICs compared with the negative control (Table 3).

In general, there was a poor correlation between resistance and fitness in the absence of toxin. In some cases, fitter clones were also more resistant. For example, the YcgZ-expressing clone had the highest relative fitness of the 25 that we measured ($W = 1.72$), and it also showed a greater increase in aztreonam resistance (5.9-fold) than the other two clones that were isolated from the aztreonam-containing PM well (Table 3). However, there were also cases where the opposite was true. For example, overexpression of YdaC imparted a fourfold increase in erythromycin resistance and a small increase in fitness ($W = 1.33$); YdfW overexpression gave a smaller increase in resistance (threefold) but a larger increase in fitness ($W = 1.51$). Similarly, overexpressing NudB gave a higher MIC for sulfadiazine than LuxS overexpression (32 vs. 24 $\mu g \cdot mL^{-1}$), yet the NudB-expressing strain was considerably less fit ($W = 0.68$ vs. $W = 1.11$).

In a single case, expression of an ASKA ORF (an uncharacterized prophage gene, *tfaX*) decreased the MIC for the antibiotic (nitrofurantoin) in which it was selected. The TfaX-expressing strain showed a small but statistically significant fitness improvement in the absence of nitrofurantoin ($W = 1.18$). This suggests that increased fitness in the absence of an antibiotic can translate into a selective advantage when the drug is introduced; however, this appears to be uncommon. It is noteworthy that no single ASKA gene was isolated from more than seven PM wells (*vide supra*) (Table S1). It might be expected that one (or a handful) of general, growth-enhancing genes would have been isolated from a majority of the 237 toxin-containing environ-

Table 3. Relative fitness and drug resistance data for *E. coli* cells overexpressing ORFs from the ASKA library

ASKA ORF	$W \pm$ SE*	Compound	MIC (μg/mL) with ASKA ORF	MIC (μg/mL) with pCA24N-NoIns	Fold-increase in MIC
β-Lactams					
bdm	1.47 ± 0.04†	Cephalothin	16	8	2.0
bfd	1.16 ± 0.03†	Penicillin G	48	32	1.5
rbsR	1.40 ± 0.04†	Aztreonam	0.19	0.064	3.0
yeaD	1.20 ± 0.03†	Aztreonam	0.094	0.064	1.5
ycgZ	1.72 ± 0.05†	Aztreonam	0.38	0.064	5.9
ycgZ	See above	Cefuroxime	16	6	2.7
Aminoglycosides					
yejG	0.94 ± 0.04	Sisomicin	0.19	0.125	1.5
yejG	See above	Tobramycin	1.5	0.5	3.0
Anti-folates					
folA	0.84 ± 0.03†	Trimethoprim	>32	0.064	>500
luxS	1.11 ± 0.02†	Sulfadiazine	24	16	1.5
nudB	0.68 ± 0.04†	Sulfadiazine	32	16	2.0
Quinolones					
cynT	1.17 ± 0.04†	Enoxacin	0.25	0.19	1.3
marA	1.08 ± 0.02	Ciprofloxacin	0.125	0.032	3.9
galE	1.16 ± 0.01†	Nalidixic acid	96	32	3.0
yfdO	1.03 ± 0.02	Nalidixic acid	48	32	1.5
Tetracyclines					
bcr	1.03 ± 0.04	Oxytetracycline	6	2	3.0
Macrolide-Lincosamide-Streptogramin					
cmr	1.45 ± 0.02†	Lincomycin	1,536	192	8.0
rbsR	See above	Lincomycin	384	192	2.0
ydfW	1.51 ± 0.12†	Spiramycin	36	24	1.5
ydfW	See above	Erythromycin	96	32	3.0
ydaC	1.33 ± 0.03†	Erythromycin	128	32	4.0
Glycopeptides					
mzrA	1.23 ± 0.07	Vancomycin	256	128	2.0
Nitrofurans					
tfaX	1.18 ± 0.02†	Nitrofurantoin	0.094	0.19	0.5
Rifamycins					
feoC	0.78 ± 0.07	Rifampicin	6	4	1.5
tusE	0.87 ± 0.04	Rifampicin	6	4	1.5
Other toxins					
mdtM	1.07 ± 0.03	Puromycin	128	8	16.0
puuD	1.12 ± 0.03†	Benzethonium chloride	6	6	1.0
ybcD	1.02 ± 0.02	Benzethonium chloride	8	6	1.3
yhiK	1.19 ± 0.01†	Benzethonium chloride	12	6	2.0

*Fitness (W) of the ASKA strain, in the absence of toxin, relative to a neutrally-marked negative control that harbors pCA24N-NoIns. Values are expressed as mean ± SE ($n = 8$).

†Significance level of $P < 0.01$ for a two-tailed test with the null hypothesis that $W = 1$, calculated using the t distribution and 7 df.

ments, were such a gene present in the ASKA pool. Instead, our approach has revealed many latent resistance determinants in the *E. coli* genome.

Discussion

The aim of this study was to gauge the extent to which a simple adaptive response—overexpression of a preexisting protein—could impart new phenotypes on *E. coli*. A handful of analogous examples have been discovered adventitiously (2), but at the outset of this study the likelihood (and therefore relevance) of this route to evolutionary innovation was unclear. We used resistance to antibiotics and toxins as a readily selectable trait. Our survey was the most comprehensive to date, because we used the wells of 10 PM plates as our toxin-containing environments. We also used the ASKA ORF collection to directly identify proteins that could impart resistance. This process contrasts with previous studies where roles in resistance have been inferred indirectly, based on the increased sensitivity of mutants produced by transposon inactivation or gene knockouts (28, 30, 31).

In the first part of our experiment, we identified toxin-containing conditions where the pooled ASKA clones out-grew the negative control. In those PM wells, at least one member of the pool had gained an improved ability to survive and reproduce, through the overexpression of a single, wild-type ORF. Next, serial transfer was used to identify the fittest strains in the ASKA pool. Our approach was designed to focus on monogenic examples, and to exclude multicell phenomena (such as cross-feeding between members of the pool). Our experimental design also precluded the identification of ASKA clones with increased susceptibility to toxins, although these are likely to have been present in many (and perhaps all) of the PM wells. In total, we identified 61 ASKA ORFs that improved fitness in 86 of 237 toxin-containing environments. Eighteen of these ORFs were hits in multiple wells (Table S1); overall, we identified 115 examples of resistance. Hits were obtained from every class of toxic compound tested (Table 1). The hit rate (resistance to 36% of all toxins) was even higher than in our previous study (23), where only 20% of single-gene knockout strains were rescued by noncognate ASKA ORFs. As hypothesized previously (38), we have shown that a bacterium from a nonclinical environment (in this case, a laboratory strain of *E. coli*) can nevertheless possess a significant reservoir of latent resistance determinants. More generally, our study emphasizes just how likely evolutionary innovation can be. Our results imply that the overexpression of a randomly chosen *E. coli* protein will impart resistance to a novel toxin with a probability of $(61/5{,}272) \times (86/237) \approx 0.4\%$.

The PM plates contained four concentrations of each toxin. Therefore, a hit could have been because of an increase in resistance (i.e., a strain undergoing division at a toxin concentration that was inhibitory to the other strains, including the negative control), or because of particularly rapid growth in a subinhibitory concentration of the toxin (i.e., dividing faster than the other strains in the pool). In this respect, our experiment mimics the dynamics of a clinical setting (37, 39). Almost exclusively, the proteins that we identified did facilitate increased resistance (Table 3). The majority of the measured MIC increases were modest (<8-fold); nevertheless, even small changes such as these can be responsible for increasing the severity of bacterial infections (40). In one case—cefuroxime—we showed that expression of a small stress response protein, YcgZ, increased the MIC from 6 to 16 $\mu g \cdot mL^{-1}$. The clinical breakpoint for resistance to this β-lactam is 8 $\mu g \cdot mL^{-1}$ (41); that is, our nonpathogenic host strain attained clinically relevant levels of resistance, simply through overexpression of a previously uncharacterized protein. Furthermore, most strains showed no fitness cost associated with expressing an ASKA ORF in the absence of toxin (and indeed, many out-competed the negative control) (Table 3). Strong selection pressure for cost-free resistance mutations has been observed in clinical isolates of *Mycobacterium tuberculosis* (36). Together, our MIC and fitness data suggest that up-regulating the expression of preexisting, latent resistance determinants may play an important role in the emergence of drug-resistant pathogens.

We discovered proteins acting through a variety of mechanisms to effect resistance (Table 2). Many of these mechanisms—such as toxin efflux/transport, stress responses and biofilm formation—were unsurprising, even if the identities of the individual resistance genes were less predictable. In 15 of 115 cases, overexpressed enzymes appeared to be acting promiscuously to impart resistance. In addition to these mechanisms, there was a prevalence of prophage genes and genes of unknown function in our list. Prophage genes are usually assumed to be cryptic or defective remnants of temperate bacteriophage genomes, although a recent study has shown that genes from the CP4-57 and DLP12 prophages can play roles in biofilm development (42). Here, we have reported 14 cases in which the overexpression of prophage genes (including two from DLP12) can improve growth in the presence of toxins. Twelve of these cases arose from the overexpression of only three prophage genes: *ydaC*, *ydfW*, and *yfdO* (Table S1). This result extends the previous finding (42), demonstrating that prophage genes can modulate broad, compound nonspecific cellular responses when they are activated from latency.

In contrast, we identified 20 uncharacterized, nonprophage genes that showed specificity in their actions; that is, they each imparted a growth advantage in the presence of a single toxin. These included genes for putative enzymes (*wbbL_1*, *yeaD*, *ysgA*), predicted transporters (*yedA*, *yjeH*), and a predicted transcriptional regulator (*yidF*), as well as 14 genes for which functional annotation was impossible due to insufficient homology with any gene of known function (Dataset S2). These results provide experimental evidence for the importance of cryptic genes as a reservoir of evolutionarily-accessible functions (43).

Our experiments have demonstrated the biochemical feasibility of the IAD model for the origins of new genes (11). Every plasmid-carrying ASKA clone approximated a nonplasmid-containing strain in which a single chromosomal locus had been amplified. Growth under selection began when the pooled ASKA clones were used to inoculate each PM well. To be scored positive, a PM well had to contain cells that out-grew the negative control, in which no ASKA ORF was being expressed. Therefore, we were specifically looking for cases in which proteins possessed activities that were valuable in the presence of a toxin and cases in which increases in their dosage were required to uncover the latent activities. We halted our selection experiments at this stage; that is, we applied selection, but we did not consider mutation beyond (artificial) gene amplification. The next step will be to determine feasible mutational routes for the divergence of amplified copies from their original function toward their new role in antibiotic resistance. These studies are on-going in our laboratory, and will allow related questions, such as the nature of the tradeoff between new and old enzymatic functions, to be addressed.

Concluding Remarks. We have used two tools from functional genomics (PM plates and the ASKA collection) to conduct a comprehensive search for *E. coli* proteins that can impart improved growth in the presence of toxic compounds. The resulting catalog provides a unique picture of a bacterium's latent evolutionary potential and emphasizes the high frequency at which novel traits can evolve. By cataloguing sources of phenotypic innovation, we have revealed the diversity of adaptive mechanisms that can be underpinned by overexpression mutations such as gene amplification. Our results suggest that the IAD model (11) is a biochemically and evolutionarily feasible—and perhaps dominant—mechanism for the birth of new genes under selection.

Materials and Methods

Experimental methods are summarized below. Detailed protocols are provided in the *SI Materials and Methods*.

Library-on-Library Screen. The library-on-library screening protocol is outlined in Fig. 1. Plates PM11 to PM20 were from Biolog Inc. After inoculation, the PM plates were incubated at 37 °C for 7 d and the growth in each well was scored daily. Two independent replicates of the screen were performed.

Serial Enrichment to Identify Resistance Genes. For conditions in which the ASKA pool reproducibly out-grew the negative control, fresh aliquots of the pool were now used to inoculate the corresponding wells of new PM plates. At the first sign of growth (i.e., purple color development), the culture in that well was diluted 50-fold with fresh medium and transferred to the corresponding well of a new plate. After a second transfer, cells were spread on LB-chloramphenicol plates. PCR amplification and sequencing of the ASKA ORFs from at least eight of the resulting colonies revealed each ORF that was responsible for enhanced growth.

Relative Fitness Assays. Twenty-five ASKA plasmids that conferred increased growth in the presence of antibiotics were purified from individual clones and used to transform fresh aliquots of *E. coli* DH5α-E. The fitness of each IPTG-induced strain, in the absence of antibiotics and in competition with the negative control harboring pCA24N-NoIns, was measured using Lenski's protocol (44). *E. coli* DH5α-E carries the *lacZ*ΔM15 mutation. To distinguish the two strains in each fitness assay, a mini-Tn7 system (45) was used to mark the negative control with a functional copy of *lacZ*. Control experiments showed that marking the strain had no effect on its fitness in the conditions used. The initial (t = 0 h) and final (t = 24 h) densities of each competitor could then be measured by spreading them on LB agar plates supplemented with X-gal. Each relative fitness value (W) in Table 3 is the mean of eight replicates.

Antibiotic Susceptibility Testing. MICs were determined using either E-test strips (AB bioMérieux) or a broth microdilution method (46). Reported values are the mean of at least two independent experiments.

ACKNOWLEDGMENTS. We thank Ilana Gerber, Susan Morton, and Laura Nigon for their technical assistance. We are also grateful to Monica Gerth for her advice on fitness assays and to John Roth for his comments on the manuscript. Financial support for this research was provided by the Maurice and Phyllis Paykel Trust, the Lottery Health Research Committee, the Auckland Medical Research Foundation, and the Marsden Fund.

1. Hegeman GD, Rosenberg SL (1970) The evolution of bacterial enzyme systems. *Annu Rev Microbiol* 24:429–462.
2. Mortlock RP (1984) *Microorganisms as Model Systems for Studying Evolution* (Plenum Press, New York, London).
3. Lewis EB (1951) Pseudoallelism and gene evolution. *Cold Spring Harb Symp Quant Biol* 16:159–174.
4. Ohno S (1970) *Evolution by Gene Duplication* (Springer-Verlag, New York).
5. Lynch M, Conery JS (2003) The origins of genome complexity. *Science* 302:1401–1404.
6. Wagner A (2005) Energy constraints on the evolution of gene expression. *Mol Biol Evol* 22:1365–1374.
7. Roth C, et al. (2007) Evolution after gene duplication: Models, mechanisms, sequences, systems, and organisms. *J Exp Zoolog B Mol Dev Evol* 308:58–73.
8. Conant GC, Wolfe KH (2008) Turning a hobby into a job: How duplicated genes find new functions. *Nat Rev Genet* 9:938–950.
9. Innan H, Kondrashov F (2010) The evolution of gene duplications: Classifying and distinguishing between models. *Nat Rev Genet* 11:97–108.
10. Francino MP (2005) An adaptive radiation model for the origin of new gene functions. *Nat Genet* 37:573–577.
11. Bergthorsson U, Andersson DI, Roth JR (2007) Ohno's dilemma: Evolution of new genes under continuous selection. *Proc Natl Acad Sci USA* 104:17004–17009.
12. Khersonsky O, Tawfik DS (2010) Enzyme promiscuity: A mechanistic and evolutionary perspective. *Annu Rev Biochem* 79:471–505.
13. Jensen RA (1976) Enzyme recruitment in evolution of new function. *Annu Rev Microbiol* 30:409–425.
14. O'Brien PJ, Herschlag D (1999) Catalytic promiscuity and the evolution of new enzymatic activities. *Chem Biol* 6:R91–R105.
15. Anderson P, Roth J (1981) Spontaneous tandem genetic duplications in *Salmonella typhimurium* arise by unequal recombination between rRNA (*rrn*) cistrons. *Proc Natl Acad Sci USA* 78:3113–3117.
16. Reams AB, Kofoid E, Savageau M, Roth JR (2010) Duplication frequency in a population of *Salmonella enterica* rapidly approaches steady state with or without recombination. *Genetics* 184:1077–1094.
17. Berg CM, Wang MD, Vartak NB, Liu L (1988) Acquisition of new metabolic capabilities: Multicopy suppression by cloned transaminase genes in *Escherichia coli* K-12. *Gene* 65:195–202.
18. Christ D, Chin JW (2008) Engineering *Escherichia coli* heat-resistance by synthetic gene amplification. *Protein Eng Des Sel* 21:121–125.
19. Morett E, et al. (2008) Sensitive genome-wide screen for low secondary enzymatic activities: The YjbQ family shows thiamin phosphate synthase activity. *J Mol Biol* 376: 839–853.
20. Miller BG, Raines RT (2004) Identifying latent enzyme activities: Substrate ambiguity within modern bacterial sugar kinases. *Biochemistry* 43:6387–6392.
21. Patrick WM, Matsumura I (2008) A study in molecular contingency: Glutamine phosphoribosylpyrophosphate amidotransferase is a promiscuous and evolvable phosphoribosylanthranilate isomerase. *J Mol Biol* 377:323–336.
22. Kitagawa M, et al. (2005) Complete set of ORF clones of *Escherichia coli* ASKA library (*A* Complete *S*et of *E. coli K*-12 ORF *A*rchive): Unique resources for biological research. *DNA Res* 12:291–299.
23. Patrick WM, Quandt EM, Swartzlander DB, Matsumura I (2007) Multicopy suppression underpins metabolic evolvability. *Mol Biol Evol* 24:2716–2722.
24. Bochner BR, Gadzinski P, Panomitros E (2001) Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Res* 11:1246–1255.
25. Bochner BR (2009) Global phenotypic characterization of bacteria. *FEMS Microbiol Rev* 33:191–205.
26. Keseler IM, et al. (2009) EcoCyc: A comprehensive view of *Escherichia coli* biology. *Nucleic Acids Res* 37:D464–D470.
27. The UniProt Consortium (2010) The universal protein resource (UniProt) in 2010. *Nucleic Acids Res* 38:D142–D148.
28. Breidenstein EBM, Khaira BK, Wiegand I, Overhage J, Hancock REW (2008) Complex ciprofloxacin resistome revealed by screening a *Pseudomonas aeruginosa* mutant library for altered susceptibility. *Antimicrob Agents Chemother* 52:4486–4491.
29. Nishino K, Yamaguchi A (2001) Analysis of a complete library of putative drug transporter genes in *Escherichia coli*. *J Bacteriol* 183:5803–5812.
30. Duo M, Hou S, Ren D (2008) Identifying *Escherichia coli* genes involved in intrinsic multidrug resistance. *Appl Microbiol Biotechnol* 81:731–741.
31. Liu A, et al. (2010) Antibiotic sensitivity profiles determined with an *Escherichia coli* gene knockout collection: Generating an antibiotic barcode. *Antimicrob Agents Chemother* 54:1393–1403.
32. Cohen SP, Hächler H, Levy SB (1993) Genetic and functional analysis of the multiple antibiotic resistance (*mar*) locus in *Escherichia coli*. *J Bacteriol* 175:1484–1492.
33. Ruiz C, Levy SB (2010) Many chromosomal genes modulate MarA-mediated multidrug resistance in *Escherichia coli*. *Antimicrob Agents Chemother* 54:2125–2134.
34. Gould SM, Tawfik DS (2005) Directed evolution of the promiscuous esterase activity of carbonic anhydrase II. *Biochemistry* 44:5444–5452.
35. Andersson DI, Hughes D (2010) Antibiotic resistance and its cost: Is it possible to reverse resistance? *Nat Rev Microbiol* 8:260–271.
36. Sander P, et al. (2002) Fitness cost of chromosomal drug resistance-conferring mutations. *Antimicrob Agents Chemother* 46:1204–1211.
37. Marcusson LL, Frimodt-Møller N, Hughes D (2009) Interplay in the selection of fluoroquinolone resistance and bacterial fitness. *PLoS Pathog* 5:e1000541.
38. Martínez JL (2008) Antibiotics and antibiotic resistance genes in natural environments. *Science* 321:365–367.
39. Martínez JL, Baquero F, Andersson DI (2007) Predicting antibiotic resistance. *Nat Rev Microbiol* 5:958–965.
40. Drusano GL (2004) Antimicrobial pharmacodynamics: critical interactions of 'bug and drug'. *Nat Rev Microbiol* 2:289–300.
41. The European Committee on Antimicrobial Susceptibility Testing (EUCAST).http://www.eucast.org/clinical_breakpoints.Accessed August 13, 2010.
42. Wang X, Kim Y, Wood TK (2009) Control and benefits of CP4-57 prophage excision in *Escherichia coli* biofilms. *ISME J* 3:1164–1179.
43. Hall BG, Yokoyama S, Calhoun DH (1983) Role of cryptic genes in microbial evolution. *Mol Biol Evol* 1:109–124.
44. Lenski RE, Rose MR, Simpson SC, Tadler SC (1991) Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2000 generations. *Am Nat* 138:1315–1341.
45. Choi KH, et al. (2005) A Tn7-based broad-range bacterial cloning and expression system. *Nat Methods* 2:443–448.
46. Wiegand I, Hilpert K, Hancock REW (2008) Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat Protoc* 3:163–175.

REFERENCES

Afriat, L., Roodveldt, C., Manco, G. & Tawfik, D. S. (2006). The latent promiscuity of newly identified microbial lactonases is linked to a recently diverged phosphotriesterase. *Biochemistry,* **45**:13677–13686.

Aharoni, A., Gaidukov, L., Khersonsky, O., Gould, S. M., Roodveldt, C. & Tawfik, D. S. (2005). The 'evolvability' of promiscuous protein functions. *Nat Genet,* **37**:73–76.

Allen, H. K., Moe, L. A., Rodbumrer, J., Gaarder, A. & Handelsman, J. (2009). Functional metagenomics reveals diverse β-lactamases in a remote Alaskan soil. *ISME J,* **3**:243–251.

Anderson, P. & Roth, J. (1981). Spontaneous tandem genetic duplications in *Salmonella typhimurium* arise by unequal recombination between rRNA (*rrn*) cistrons. *Proc Natl Acad Sci USA,* **78**:3113–3117.

Andersson, D. I. & Hughes, D. (2010). Antibiotic resistance and its cost: is it possible to reverse resistance? *Nat Rev Microbiol,* **8**:260–271.

Andersson, D. I., Slechta, E. S. & Roth, J. R. (1998). Evidence that gene amplification underlies adaptive mutability of the bacterial *lac* operon. *Science,* **282**:1133–1135.

Anthony, K. G., Strych, U., Yeung, K. R., Shoen, C. S., Perez, O., Krause, K. L., Cynamon, M. H., Aristoff, P. A. & Koski, R. A. (2011). New classes of alanine racemase inhibitors identified by high-throughput screening show antimicrobial activity against *Mycobacterium tuberculosis*. *PLoS One,* **6**:e20374.

Awano, N., Wada, M., Kohdoh, A., Oikawa, T., Takagi, H. & Nakamori, S. (2003). Effect of cysteine desulfhydrase gene disruption on L-cysteine overproduction in *Escherichia coli*. *Appl Microbiol Biotechnol,* **62**:239–243.

Awano, N., Wada, M., Mori, H., Nakamori, S. & Takagi, H. (2005). Identification and functional analysis of *Escherichia coli* cysteine desulfhydrases. *Appl Environ Microbiol,* **71**:4149–4152.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L. & Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol,* **2**:2006.0008.

Babbitt, P. C. & Gerlt, J. A. (1997). Understanding enzyme superfamilies. Chemistry as the fundamental determinant in the evolution of new catalytic activities. *J Biol Chem,* **272**:30591–30594.

Babtie, A., Tokuriki, N. & Hollfelder, F. (2010). What makes an enzyme promiscuous? *Curr Opin Chem Biol,* **14**:200–207.

Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S. & Milo, R. (2011). The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme paramers. *Biochemistry,* **50**:4402–4410.

Battistuzzi, F. U., Feijao, A. & Hedges, S. B. (2004). A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, prototrophy, and the colonization of land. *BMC Evol Biol,* **4**:44.

Belfaiza, J., Parsot, C., Martel, A., Tour, C. B. d. l., Margarita, D., Cohen, G. N. & Saint-Girons, I. (1986). Evolution of biosynthetic pathways: Two enzymes catalyzing consecutive steps in methionine biosynthesis originate from a common ancestor and possess a similar regulatory region. *Proc Natl Acad Sci USA,* **83**:867–871.

Berg, C. M., Wang, M.-d., Vartak, N. B. & Liu, L. (1988). Acquisition of new metabolic capabilities: multicopy suppression by cloned transaminase genes in *Escherichia coli* K-12. *Gene,* **65**:195–202.

Bergthorsson, U., Andersson, D. I. & Roth, J. R. (2007). Ohno's dilemma: Evolution of new genes under continuous selection. *Proc Natl Acad Sci USA,* **104**:17004–17009.

Bertani, G. (1951). Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J Bacteriol,* **62**:293–300.

Betz, J. L., Brown, P. R., Smyth, M. J. & Clarke, P. H. (1974). Evolution in Action. *Nature,* **247**:261–264.

Biolog Inc. (2011). Phenotype MicroArray FAQ. Retrieved January 15, 2012, from http://www.biolog.com/newsite/products/?product=Phenotype%20Microarray%20Services&view=FAQs

Blount, Z. D., Borland, C. Z. & Lenski, R. E. (2008). Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc Natl Acad Sci USA,* **105**:7899–7906.

Bochner, B. R. (2009). Global phenotypic characterization of bacteria. *FEMS Microbiol Rev,* **33**:191–205.

Bochner, B. R., Gadzinski, P. & Panomitros, E. (2001). Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Res,* **11**:1246–1255.

Breidenstein, E. B. M., Khaira, B. K., Wiegand, I., Overhage, J. & Hancock, R. E. W. (2008). Complex ciprofloxacin resistome revealed by screening a *Pseudomonas aeruginosa* mutant library for altered susceptibility. *Antimicrob Agents Chemother,* **52**:4486–4491.

Burman, W. J., Gallicano, K. & Peloquin, C. (1999). Therapeutic implications of drug interactions in the treatment of human immunodeficiency virus-related tuberculosis. *Clin Infect Dis,* **28**:419–429.

Cadwell, R. C. & Joyce, G. F. (1992). Randomization of genes by PCR mutagenesis. *PCR Methods Appl,* **2**:28–33.

Callahan, B. P. & Miller, B. G. (2007). OMP decarboxylase — An enigma persists. *Bioorg Chem,* **35**:465–469.

Casali, N. (2003). *Escherichia coli* host strains. *Methods Mol Biol,* **235**:27–48.

Cartron, M. L., Maddocks, S., Gillingham, P., Craven, C. J. & Andrews, S. C. (2006). Feo — transport of ferrous iron into bacteria. *Biometals,* **19**:143–157.

Catic, A., Misaghi, S., Korbel, G. A. & Ploegh, H. L. (2007). ElaD, a deubiquitinating protease expressed by *E. coli*. *PLoS One,* **2**:e381.

Chang, Y.-Y. & Cronan, J. E., Jr. (1999). Membrane cyclopropane fatty acid content is a major factor in acid resistance of *Escherichia coli*. *Mol Microbiol,* **33**:249–259.

Chimnaronk, S., Suzuki, T., Manita, T., Ikeuchi, Y., Yao, M., Suzuki, T. & Tanaka, I. (2009). RNA helicase module in an acetyltransferase that modifies a specific tRNA codon. *EMBO J,* **28**:1362–1373.

Chittori, S., Simanshu, D. K., Savithri, H. S. & Murthy, M. R. N. (2007). Structure of the putative mutarotase YeaD from *Salmonella typhimurium*: structural comparison with galactose mutarotases. *Acta Crystallogr D Biol Crystallogr,* **63**:197–205.

Choi, K.-H., Gaynor, J. B., White, K. G., Lopez, C., Bosio, C. M., Karkhoff-Schweizer, R. R. & Schweizer, H. P. (2005). A Tn7-based broad-range bacterial cloning and expression system. *Nat Methods,* **2**:443–448.

Chow, J. Y., Wu, L. & Yew, W. S. (2009). Directed evolution of a quorum-quenching lactonase from *Mycobacterium avium* subsp. *paratuberculosis* K-10 in the amidohydrolase superfamily. *Biochemistry,* **48**:4344–4353.

Christ, D. & Chin, J. W. (2008). Engineering *Escherichia coli* heat-resistance by synthetic gene amplification. *Protein Eng Des Sel,* **21**:121–125.

Christen, P. & Mehta, P. K. (2001). From cofactor to enzymes. The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes. *Chem Rec,* **1**:436–447.

Clausen, T., Huber, R., Laber, B., Pohlenz, H.-D. & Messerschmidt, A. (1996). Crystal structure of the pyridoxal-5'-phosphate dependent cystathionine β-lyase from *Escherichia coli* at 1.83 Å. *J Mol Biol,* **262**:202–224.

Clausen, T., Huber, R., Messerschmidt, A., Pohlenz, H.-D. & Laber, B. (1997a). Slow-binding inhibition of *Escherichia coli* cystathionine β-lyase by L-aminoethoxyvinylglycine: A kinetic and x-ray study. *Biochemistry,* **36**:12633–12643.

Clausen, T., Laber, B. & Messerschmidt, A. (1997b). Mode of action of cystathionine β-lyase. *Biol Chem,* **378**:321–326.

Cohen, S. P., Hächler, H. & Levy, S. B. (1993). Genetic and functional analysis of the multiple antibiotic resistance (*mar*) locus in *Escherichia coli*. *J Bacteriol,* **175**:1484–1492.

Conant, G. C. & Wolfe, K. H. (2008). Turning a hobby into a job: How duplicated genes find new functions. *Nat Rev Genet,* **9**:938–950.

Cooper, G. M. (2000). *The cell: A molecular approach* (2nd ed.). Sunderland, MA: Sinauer Associates.

Copley, S. D. (2003). Enzymes with extra talents: moonlighting functions and catalytic promiscuity. *Curr Opin Chem Biol,* **7**:265–272.

Copley, S. D. (2012). Toward a systems biology perspective on enzyme evolution. *J Biol Chem,* **287**:3–10.

Courvalin, P. (2008). Predictable and unpredictable evolution of antibiotic resistance. *J Intern Med,* **264**:4–16.

D'Ari, R. & Casadesús, J. (1998). Underground metabolism. *BioEssays,* **20**:181–186.

D'Costa, V. M., King, C. E., Kalan, L., Morar, M., Sung, W. W. L., Schwarz, C., Froese, D., Zazula, G., Calmels, F., Debruyne, R., Golding, G. B., Poinar, H. N. & Wright, G. D. (2011). Antibiotic resistance is ancient. *Nature,* **477**:457–461.

Daigle, D. M., McKay, G. A., Thompson, P. R. & Wright, G. D. (1999). Aminoglycoside antibiotic phosphotransferases are also serine protein kinases. *Chem Biol,* **6**:11–18.

Datsenko, K. A. & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA,* **97**:6640–6645.

Davies, J. & Davies, D. (2010). Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev,* **74**:417–433.

Denesyuk, A. I., Denessiouk, K. A., Korpela, T. & Johnson, M. S. (2002). Functional attributes of the phosphate group binding cup of pyridoxal phosphate-dependent enzymes. *J Mol Biol,* **316**:155–172.

Desai, K. K. & Miller, B. G. (2008). A metabolic bypass of the triosephosphate isomerase reaction. *Biochemistry,* **47**:7983–7985.

Di Perri, G. & Bonora, S. (2004). Which agents should we use for the treatment of multidrug resistant *Mycobacterium tuberculosis? J Antimicrob Chemother,* **54**:593–602.

Dimitrov, J. D. & Vassilev, T. L. (2009). Cofactor-mediated protein promiscuity. *Nat Biotechnol,* **27**:892.

Drusano, G. L. (2004). Antimicrobial pharmacodynamics: Critical interactions of 'bug and drug'. *Nat Rev Microbiol,* **2**:289–300.

Dunathan, H. C. (1966). Conformation and reaction specificity in pyridoxal phosphate enzymes. *Proc Natl Acad Sci USA,* **55**:712–716.

Duo, M., Hou, S. & Ren, D. (2008). Identifying *Escherichia coli* genes involved in intrinsic multidrug resistance. *Appl Microbiol Biotechnol,* **81**:731–741.

Dwivedi, C. M., Ragin, R. C. & Uren, J. R. (1982). Cloning, purification, and characterization of β-cystathionase from *Escherichia coli*. *Biochemistry*, **21**:3064–3069.

Edgar, R. & Bibi, E. (1997). MdfA, an *Escherichia coli* multidrug resistance protein with an extraordinarily broad spectrum of drug recognition. *J Bacteriol,* **179**:2274–2280.

Ejim, L. J., Blanchard, J. E., Koteva, K. P., Sumerfield, R., Elowe, N. H., Chechetto, J. D., Brown, E. D., Junop, M. S. & Wright, G. D. (2007). Inhibitors of bacterial cystathionine β-lyase: Leads for new antimicrobial agents and probes of enzyme structure and function. *J Med Chem,* **50**:755–764.

Eliot, A. C. & Kirsch, J. F. (2004). Pyridoxal phosphate enzymes: Mechanistic, structural, and evolutionary considerations. *Annu Rev Biochem,* **73**:383–415.

Ellman, G. L. (1959). Tissue sulfhydryl groups. *Arch Biochem Biophys,* **82**:70–77.

Esaki, N. & Walsh, C. T. (1986). Biosynthetic alanine racemase of *Salmonella typhimurium*: Purification and characterization of the enzyme encoded by the *alr* gene. *Biochemistry,* **25**:3261–3267.

Eydallin, G., Montero, M., Almagro, G., Sesma, M. T., Viale, A. M., Muñoz, F. J., Rahimpour, M., Baroja-Fernández, E. & Pozueta-Romero, J. (2010). Genome-wide screening of genes whose enhanced expression affects glycogen accumulation in *Escherichia coli*. *DNA Res,* **17**:61–71.

Fajardo, A. & Martínez, J. L. (2008). Antibiotics as signals that trigger specific bacterial responses. *Curr Opin Microbiol,* **11**:161–167.

Fenn, T. D., Stamper, G. F., Morollo, A. A. & Ringe, D. (2003). A side reaction of alanine racemase: Transamination of cycloserine. *Biochemistry,* **42**:5775–5783.

Firth, A. E. & Patrick, W. M. (2005). Statistics of protein library construction. *Bioinformatics,* **21**:3314–3315.

Firth, A. E. & Patrick, W. M. (2008). GLUE-IT and PEDEL-AA: new programmes for analyzing protein diversity in randomized libraries. *Nucleic Acids Res,* **36**:W281–W285.

Fong, D. H., Xiong, B., Hwang, J. & Berghuis, A. M. (2011). Crystal structures of two aminoglycoside kinases bound with a eukaryotic protein kinase inhibitor. *PLoS One,* **6**:e19589.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y.-l. & Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics,* **151**:1531–1545.

Francino, M. P. (2005). An adaptive radiation model for the origin of new gene functions. *Nat Genet,* **37**:573–577.

Gaines, T. A., Zhang, W., Wang, D., Bukun, B., Chisholm, S. T., Shaner, D. L., Nissen, S. J., Patzoldt, W. L., Tranel, P. J., Culpepper, A. S., Grey, T. L., Webster, T. M., Vencill, W. K., Sammons, R. D., Jiang, J., Preston, C., Leach, J. E. & Westra, P. (2010). Gene amplification confers glyphosate resistance in *Amaranthus palmeri*. *Proc Natl Acad Sci USA,* **107**:1029–1034.

Gerken, H., Charlson, E. S., Cicirelli, E. M., Kenney, L. J. & Misra, R. (2009). MzrA: a novel modulator of the EnvZ/OmpR two-component regulon. *Mol Microbiol,* **72**:1408–1422.

Gerlt, J. A. & Babbitt, P. C. (2001). Divergent evolution of enzymatic function: Mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu Rev Biochem,* **70**:209–246.

Glasner, M. E., Gerlt, J. A. & Babbitt, P. C. (2006). Evolution of enzyme superfamilies. *Curr Opin Chem Biol,* **10**:492–497.

Gould, S. M. & Tawfik, D. S. (2005). Directed evolution of the promiscuous esterase activity of carbonic anhydrase II. *Biochemistry,* **44**:5444–5452.

Griswold, W. R. & Toney, M. D. (2011). Role of the pyridine nitrogen in pyridoxal 5'-phosphate catalysis: Activity of three classes of PLP enzymes reconstituted with deazapyridoxal 5'-phosphate. *J Am Chem Soc,* **133**:14823–14830.

Grogan, D. W. & Cronan, J. E., Jr. (1984). Cloning and manipulation of the *Escherichia coli* cyclopropane fatty acid synthase gene: Physiological aspects of enzyme overproduction. *J Bacteriol,* **158**:286–295.

Grogan, D. W. & Cronan, J. E., Jr. (1997). Cyclopropane ring formation in membrane lipids of bacteria. *Microbiol Mol Biol Rev,* **61**:429–441.

Guzman, L.-M., Belin, D., Carson, M. J. & Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol,* **177**:4121–4130.

Hall, B. G., Yokoyama, S. & Calhoun, D. H. (1983). Role of cryptic genes in microbial evolution. *Mol Biol Evol,* **1**:109–124.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol,* **166**:557–580.

Hancock, S. M., Vaughan, M. D. & Withers, S. G. (2006). Engineering of glycosidases and glycosyltransferases. *Curr Opin Chem Biol,* **10**:509–519.

Heeb, S., Fletcher, M. P., Chhabra, S. R., Diggle, S. P., Williams, P. & Cámara, M. (2011). Quinolones: from antibiotics to autoinducers. *FEMS Microbiol Rev,* **35**:247–274.

Hegeman, G. D. & Rosenberg, S. L. (1970). The evolution of bacterial enzyme systems. *Annu Rev Microbiol,* **24**:429–462.

Heifets, A. & Lilien, R. H. (2010). LigAlign: Flexible ligand-based active site alignment and analysis. *J Mol Graph Model,* **29**:93–101.

Hejazi, M., Piotukh, K., Mattow, J., Deutzmann, R., Volkmer-Engert, R. & Lockau, W. (2002). Isoaspartyl dipeptidase activity of plant-type asparaginases. *Biochem J,* **364**:129–136.

Hemm, M. R., Paul, B. J., Schneider, T. D., Storz, G. & Rudd, K. E. (2008). Small membrane proteins found by comparative genomics and ribosome binding site models. *Mol Microbiol,* **70**:1487–1501.

Hendrickson, H., Slechta, E. S., Bergthorsson, U., Andersson, D. I. & Roth, J. R. (2002). Amplification-mutagenesis: Evidence that "directed" adaptive mutation and general hypermutability result from growth with a selected gene amplification. *Proc Natl Acad Sci USA,* **99**:2164–2169.

Henn-Sax, M., Thoma, R., Schmidt, S., Hennig, M., Kirschner, K. & Sterner, R. (2002). Two $(\beta/\alpha)_8$-barrel enzymes of histidine and tryptophan biosynthesis have similar reaction mechanisms and common strategies for protecting their labile substrates. *Biochemistry,* **41**:12032–12042.

Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene,* **77**:51–59.

Hobbs, J. K., Shepherd, C., Saul, D. J., Demetras, N. J., Haaning, S., Monk, C. R., Daniel, R. M. & Arcus, V. L. (2012). On the origin and evolution of thermophily: Reconstruction of functional precambrian enzymes from ancestors of *Bacillus*. *Mol Bio Evol,* **29**:825–835.

Holliday, G. L., Thornton, J. M., Marquet, A., Smith, A. G., Rébeillé, F., Mendel, R., Schubert, H. L., Lawrence, A. D. & Warren, M. J. (2007). Evolution of enzymes and pathways for the biosynthesis of cofactors. *Nat Prod Rep,* **24**:972–987.

Hon, W.-C., McKay, G. A., Thompson, P. R., Sweet, R. M., Yang, D. S. C., Wright, G. D. & Berghuis, A. M. (1997). Structure of an enzyme required for aminoglycoside antibiotic resistance reveals homology to eukaryotic protein kinases. *Cell,* **89**:887–895.

Horowitz, N. H. (1945). On the evolution of biochemical syntheses. *Proc Natl Acad Sci USA,* **31**:153–157.

Horowitz, N. H. & Metzenberg, R. L. (1965). Biochemical aspects of genetics. *Annu Rev Biochem,* **34**:527–564.

Hughes, A. L. (1994). The evolution of functionally novel proteins after gene duplication. *Proc R Soc Lond B,* **256**:119–124.

Imamura, R., Yamanaka, K., Ogura, T., Hiraga, S., Fujita, N., Ishihama, A. & Niki, H. (1996). Identification of the *cpdA* gene encoding cyclic 3',5'-adenosine monophosphate phosphodiesterase in *Escherichia coli*. *J Biol Chem,* **271**:25423–25429.

Inagaki, K., Tanizawa, K., Badet, B., Walsh, C. T., Tanaka, H. & Soda, K. (1986). Thermostable alanine racemase from *Bacillus stearothermophilus*: Molecular cloning of the gene, enzyme purification, and characterization. *Biochemistry,* **25**:3268–3274.

Innan, H. & Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet,* **11**:97–108.

James, L. C. & Tawfik, D. S. (2001). Catalytic and binding poly-reactivities shared by two unrelated proteins: The potential role of promiscuity in enzyme evolution. *Protein Sci,* **10**:2600–2607.

Jensen, R. A. (1976). Enzyme recruitment in evolution of new function. *Annu Rev Microbiol,* **30**:409–425.

Joyce, A. R., Reed, J. L., White, A., Edwards, R., Osterman, A., Baba, T., Mori, H., Lesely, S. A., Palsson, B. Ø. & Agarwalla, S. (2006). Experimental and computational assessment of conditionally essential genes in *Escherichia coli*. *J Bacteriol,* **188**:8259–8271.

Jürgens, C., Strom, A., Wegener, D., Hettwer, S., Wilmanns, M. & Sterner, R. (2000). Directed evolution of a $(\beta/\alpha)_8$-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc Natl Acad Sci USA,* **97**:9925–9930.

Jurgenson, C. T., Begley, T. P. & Ealick, S. E. (2009). The structural and biochemical foundations of thiamine biosynthesis. *Annu Rev Biochem,* **78**:569–603.

Kang, L., Shaw, A. C., Xu, D., Xia, W., Zhang, J., Deng, J., Wöldike, H. F., Liu, Y. & Su, J. (2011). Upregulation of MetC is essential for D-alanine-independent growth of an *alr*/*dadX*-deficient *Escherichia coli* strain. *J Bacteriol,* **193**:1098–1106.

Kelley, L. A. & Sternberg, M. J. E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc,* **4**:363–371.

Kelly, J. A., Dideberg, O., Charlier, P., Wery, J. P., Libert, M., Moews, P. C., Knox, J. R., Duez, C., Fraipont, C., Joris, B., Dusart, J., Frère, J. M. & Ghuysen, J. M. (1986). On the origin of bacterial resistance to penicillin: Comparison of a β-lactamase and a penicillin target. *Science,* **231**:1429–1431.

Keseler, I. M., Bonavides-Martínez, C., Collado-Vides, J., Gama-Castro, S., Gunsalus, R. P., Johnson, D. A., Krummenacker, M., Nolan, L. M., Paley, S., Paulsen, I. T., Peralta-

Gil, M., Santos-Zavaleta, A., Shearer, A. G. & Karp, P. D. (2009). EcoCyc: A comprehensive view of *Escherichia coli* biology. *Nucleic Acids Res,* **37**:D464–D470.

Khersonsky, O., Roodveldt, C. & Tawfik, D. S. (2006). Enzyme promiscuity: Evolutionary and mechanistic aspects. *Curr Opin Chem Biol,* **10**:1–11.

Khersonsky, O. & Tawfik, D. S. (2010). Enzyme promiscuity: A mechanistic and evolutionary perspective. *Annu Rev Biochem,* **79**:471–505.

Kim, J. & Copley, S. D. (2007). Why metabolic enzymes are essential or nonessential for growth of *Escherichia coli* K12 on glucose. *Biochemistry,* **46**:12501–12511.

Kim, J., Kerschner, J. P., Novikov, Y., Shoemaker, R. K. & Copley, S. D. (2010). Three serendipitous pathways in *E. coli* can bypass a block in pyridoxal-5'-phosphate synthesis. *Mol Syst Biol,* **6**:436.

Kim, K. M., Qin, T., Jiang, Y.-Y., Chen, L.-L., Xiong, M., Caetano-Annollés, D., Zhang, H.-Y. & Caetano-Annollés, G. (2012). Protein domain structure uncovers the origin of aerobic metabolism and the rise of planetary oxygen. *Structure,* **20**:67–76.

Kim, P.-J., Lee, D.-Y., Kim, T. Y., Lee, K. H., Jeong, H., Lee, S. Y. & Park, S. (2007). Metabolite essentiality elucidates robustness of *Escherichia coli* metabolism. *Proc Natl Acad Sci USA,* **104**:13638–13642.

Kim, Y., Wang, X., Ma, Q., Zhang, X.-S. & Wood, T. K. (2009). Toxin-antitoxin systems in *Escherichia coli* influence biofilm formation through YjgK (TabA) and fimbriae. *J Bacteriol,* **191**:1258–1267.

Kimura, M. & Ohta, T. (1974). On some principles governing molecular evolution. *Proc Natl Acad Sci USA,* **71**:2848–2852.

King, D. & Strynadka, N. (2011). Crystal structure of New Delhi metallo-β-lactamase reveals molecular basis for antibiotic resistance. *Protein Sci,* **20**:1484–1491.

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 *S*et of *E. coli K*-12 ORF *A*rchive): Unique resources for biological research. *DNA Res,* **12**:291–299.

Korea, C.-G., Badouraly, R., Prevost, M.-C., Ghigo, J.-M. & Beloin, C. (2010). *Escherichia coli* K-12 possesses multiple cryptic but functional chaperone-usher fimbriae with distinct surface specificities. *Environ Microbiol,* **12**:1957–1977.

Korea, C.-G., Ghigo, J.-M. & Beloin, C. (2011). The sweet connection: Solving the riddle of multiple sugar-binding fimbrial adhesins in *Escherichia coli*: Multiple *E. coli* fimbriae form a versatile arsenal of sugar-binding lectins potentially involved in surface-colonisation and tissue tropism. *Bioessays,* **33**:300–311.

Koul, A., Arnoult, E., Lounis, N., Guillemont, J. & Andries, K. (2011). The challenge of new drug discovery for tuberculosis. *Nature,* **469**:483–490.

Kricker, M. & Hall, B. G. (1984). Directed evolution of cellobiose utilization in *Escherichia coli* K12. *Mol Biol Evol,* **1**:171–182.

Kricker, M. & Hall, B. G. (1987). Biochemical genetics of the cryptic gene system for cellobiose utilization in *Escherichia coli* K12. *Genetics,* **115**:419–429.

Kurihara, S., Oda, S., Kumagai, H. & Suzuki, H. (2006). γ-glutamyl-γ-aminobutyrate hydrolase in the putrescine utilization pathway of *Escherichia coli* K-12. *FEMS Microbiol Lett,* **256**:318–323.

Kurokawa, Y., Watanabe, A., Yoshimura, T., Esaki, N. & Soda, K. (1998). Transamination as a side-reaction catalyzed by alanine racemase of *Bacillus stearothermophilus*. *J Biochem,* **124**:1163–1169.

Kuznetsova, E., Proudfoot, M., Gonzalez, C. F., Brown, G., Omelchenko, M. V., Borozan, I., Carmel, L., Wolf, Y. I., Mori, H., Savchenko, A. V., Arrowsmith, C. H., Koonin, E. V., Edwards, A. M. & Yakunin, A. F. (2006). Genome-wide analysis of substrate specificities of the *Escherichia coli* haloacid dehalogenase-like phosphatase family. *J Biol Chem,* **281**:36149–36161.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature,* **227**:680–685.

Lambert, M. P. & Neuhaus, F. C. (1972). Factors affecting the level of alanine racemase in *Escherichia coli*. *J Bacteriol,* **109**:1156–1161.

LeMagueres, P., Im, H., Ebalunode, J., Strych, U., Benedik, M. J., Briggs, J. M., Kohn, H. & Krause, K. L. (2005). The 1.9 Å crystal structure of alanine racemase from *Mycobacterium tuberculosis* contains a conserved entryway into the active site. *Biochemistry,* **44**:1471–1481.

Lenski, R. E., Rose, M. R., Simpson, S. C. & Tadler, S. C. (1991). Long term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am Nat,* **138**:1315–1341.

Levy, S. B. & Marshall, B. (2004). Antibacterial resistance worldwide: causes, challenges and responses. *Nat Med,* **10**:S122–S129.

Lewis, E. B. (1951). Pseudoallelism and gene evolution. *Cold Spring Harb Symp Quant Biol,* **16**:159–174.

Lewis, V. E., Donarski, W. J., Wild, J. R. & Raushel, F. M. (1988). Mechanism and stereochemical course at phosphorus of the reaction catalyzed by a bacterial phosphotriesterase. *Biochemistry,* **27**:1591–1597.

Li, J., Attila, C., Wang, L., Wood, T. K., Valdes, J. J. & Bentley, W. E. (2007). Quorum-sensing in *Escherichia coli* is signaled by AI-2/LsrR: Effects on small RNA and biofilm architecture. *J Bacteriol,* **189**:6011–6020.

Liu, A., Tran, L., Becket, E., Lee, K., Chinn, L., Park, E., Tran, K. & Miller, J. H. (2010). Antibiotic sensitivity profiles determined with an *Escherichia coli* gene knockout collection: Generating an antibiotic bar code. *Antimicrob. Agents Chemother,* **54**:1393–1403.

Liu, M., Turner, R. J., Winstone, T. L., Saetre, A., Dyllick-Brenzinger, M., Jickling, G., Tari, L. W., Weiner, J. H. & Taylor, D. E. (2000). *Escherichia coli* TehB requires S-adenosylmethionine as a cofactor to mediate tellurite resistance. *J Bacteriol,* **182**:6509–6513.

Lodha, P. H. & Aitken, S. M. (2011). Characterization of the side-chain hydroxyl moieties of residues Y56, Y111, Y238, Y338 and S339 as determinants of specificity in *E. coli* cystathionine β-lyase. *Biochemistry,* **50**:9876–9885.

Lundqvist, T., Fisher, S. L., Kern, G., Folmer, R. H. A., Xue, Y., Newton, D. T., Keating, T. A., Alm, R. A. & de Jong, B. L. M. (2007). Exploitation of structural and regulatory diversity in glutamate racemases. *Nature,* **447**:817–822.

Lynch, M. & Conery, J. S. (2003). The origins of genome complexity. *Science,* **302**:1401–1404.

Marcusson, L. L., Frimodt-Møller, N. & Hughes, D. (2009). Interplay in the selection of fluoroquinolone resistance and bacterial fitness. *PLoS Pathog,* **5**:e1000541.

Marschall, C. & Hengge-Aronis, R. (1995). Regulator characteristics and promoter analysis of *csiE*, a stationary phase-inducible gene under the control of σ^S and the cAMP-CRP complex in *Escherichia coli*. *Mol Microbiol,* **18**:175–184.

Martí-Renom, M. A., Ilyin, V. A. & Sali, A. (2001). DBAli: A database of protein structure alignments. *Bioinformatics,* **17**:746–747.

Martínez, J. L. (2008). Antibiotics and antibiotic resistance genes in natural environments. *Science,* **321**:365–367.

Martínez, J. L., Baquero, F. & Andersson, D. I. (2007). Predicting antibiotic resistance. *Nat Rev Microbiol,* **5**:958–965.

McLoughlin, S. Y. & Copley, S. D. (2008). A compromise required by gene sharing enables survival: Implications for evolution of new enzyme activities. *Proc Natl Acad Sci USA,* **105**:13497–13502.

Meselson, M. & Yuan, R. (1968). DNA restriction enzyme from *E. coli*. *Nature,* **217**:1110–1114.

Miller, B. G. & Raines, R. T. (2004). Identifying latent enzyme activities: Substrate ambiguity within modern bacterial sugar kinases. *Biochemistry,* **43**:6387–6392.

Milligan, D. L., Tran, S. L., Strych, U., Cook, G. M. & Krause, K. L. (2007). The alanine racemase of *Mycobacterium smegmatis* is essential for growth in the absence of D-alanine. *J Bacteriol,* **189**:8381–8386.

Morar, M. & Wright, G. D. (2010). The genomic enzymology of antibiotic resistance. *Annu Rev Genet,* **44**:25–51.

Morett, E., Saab-Rincón, G., Olvera, L., Olvera, M., Flores, H. & Grande, R. (2008). Sensitive genome-wide screen for low secondary enzymatic activities: The YjbQ family shows thiamin phosphate synthase activity. *J Mol Biol,* **376**:839–853.

Mortlock, R. P. (1984). *Microorganisms as model systems for studying evolution*. New York and London: Plenum Press.

National Institute of Standards and Technology (NIST) (2011). NIST Chemistry WebBook. Retrieved November 14, 2011, from http://webbook.nist.gov/chemistry/mw-ser.html

Neumann, M., Mittelstädt, G., Ionbbi-Nivol, C., Saggu, M., Lendzian, F., Hildebrandt, P. & Leimkühler, S. (2009). A periplasmic aldehyde oxidoreductase represents the first molybdopterin cytosine dinucleotide cofactor containing molybdo-flavoenzyme from *Escherichia coli*. *FEBS J,* **276**:2762–2774.

Nie, L., Ren, Y., Janakiraman, A., Smith, S. & Schulz, H. (2008). A novel paradigm of fatty acid β-oxidation exemplified by the thioesterase-dependent partial degradation of conjugated linoleic acid that fully supports growth of *Escherichia coli*. *Biochemistry,* **47**:9618–9626.

Nishino, K. & Yamaguchi, A. (2001). Analysis of a complete library of putative drug transporter genes in *Escherichia coli*. *J Bacteriol,* **183**:5803–5812.

O'Brien, P. J. & Herschlag, D. (1999). Catalytic promiscuity and the evolution of new enzymatic activities. *Chem Biol,* **6**:R91–R105.

Ohno, S. (1970). *Evolution by gene duplication*. New York: Springer-Verlag.

Ohta, T. (1988). Evolution by gene duplication and compensatory advantageous mutations. *Genetics,* **120**:841–847.

Ohtsu, I., Wiriyathanawudhiwong, N., Morigasaki, S., Nakatani, T., Kadokura, H. & Takagi, H. (2010). The L-cysteine/L-cystine shuttle system provides reducing equivalents to the periplasm in *Escherichia coli*. *J Biol Chem,* **285**:17479–17487.

Omburo, G. A., Kuo, J. M., Mullins, L. S. & Raushel, F. M. (1992). Characterization of the zinc binding site of bacterial phosphotriesterase. *J Biol Chem,* **267**:13278–13283.

Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. *Protein Sci,* **4**:2411–2423.

Palmer, D. R. J., Garrett, J. B., Sharma, V., Meganathan, R., Babbitt, P. C. & Gerlt, J. A. (1999). Unexpected divergence of enzyme function and sequence: "*N*-acylamino acid racemase" is *o*-Succinylbenzoate synthase. *Biochemistry,* **38**:4252–4258.

Patrick, W. M., Firth, A. E. & Blackburn, J. M. (2003). User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. *Protein Eng,* **16**:451–457.

Patrick, W. M. & Matsumura, I. (2008). A study in molecular contingency: Glutamine phosphoribosylpyrophosphate amidotransferase is a promiscuous and evolvable phosphoribosylanthranilate isomerase. *J Mol Biol,* **377**:323–336.

Patrick, W. M., Quandt, E. M., Swartzlander, D. B. & Matsumura, I. (2007). Multicopy suppression underpins metabolic evolvability. *Mol Biol Evol,* **24**:2716–2722.

Patrick, W. M., Weisner, J. & Blackburn, J. M. (2002). Site-directed mutagenesis of Tyr354 in *Geobacillus stearothermophilus* alanine racemase identifies a role in controlling substrate specificity and a possible role in the evolution of antibiotic resistance. *ChemBioChem,* **3**:789–792.

Patridge, E. V. & Ferry, J. G. (2006). WrbA from *Escherichia coli* and *Archaeoglobus fulgidus* is an NAD(P)H:quinone oxidoreductase. *J Bacteriol,* **188**:3498–3506.

Peisajovich, S. G. & Tawfik, D. S. (2007). Protein engineers turned evolutionists. *Nat Methods,* **4**:991–994.

Percudani, R. & Peracchi, A. (2003). A genomic view of pyridoxal-phosphate-dependent enzymes. *EMBO Rep,* **4**:850–854.

Percudani, R. & Peracchi, A. (2009). The B6 database: a tool for the description and classification of vitamin B6-dependent enzymatic activities and of the corresponding protein families. *BMC Bioinformatics,* **10**:273.

Perez-Jimenez, R., Inglés-Prieto, A., Zhao, Z.-M., Sanchez-Romero, I., Alegre-Cebollada, J., Kosuri, P., Garcia-Manyes, S., Kappock, T. J., Tanokura, M., Holmgren, A., Sanchez-Ruiz, J. M., Gaucher, E. A. & Fernandez, J. M. (2011). Single-molecule paleoenzymology probes the chemistry of resurrected enzymes. *Nat Struct Mol Biol,* **18**:592–596.

Prasad, I. & Schaefler, S. (1974). Regulation of the β-glucoside system in *Escherichia coli* K-12. *J. Bacteriol,* **120**:638–650.

Prasad, I., Young, B. & Schaefler, S. (1973). Genetic determination of the constitutive biosynthesis of phospho-beta-glucosidase A in *Escherichia coli* K-12. *J Bacteriol,* **114**:909–915.

Qiagen (2011). Plasmid applications — Growth of bacterial cultures. Retrieved September 23, 2011, from http://www.qiagen.com/plasmid/bacterialcultures.aspx

Radzicka, A. & Wolfenden, R. (1995). A proficient enzyme. *Science,* **67**:90–93.

Reams, A. B., Kofoid, E., Savageau, M. & Roth, J. R. (2010). Duplication frequency in a population of *Salmonella enterica* rapidly approaches steady state with or without recombination. *Genetics,* **184**:1077–1094.

Rêgo, A. T., Chandran, V. & Waksman, G. (2010). Two-step and one-step secretion mechanisms in Gram-negative bacteria: constrasting the type IV secretion system and the chaperone-usher pathway of pilus biogenesis. *Biochem J,* **425**:475–488.

Rice, L. B. (2009). The clinical consequences of antimicrobial resistance. *Curr Opin Microbiol,* **12**:476–481.

Richard, J. P., Amyes, T. L., Crugeiras, J. & Rios, A. (2009). Pyridoxal 5'-phosphate: electrophilic catalyst extraordinaire. *Curr Opin Chem Biol,* **13**:475–483.

Riddles, P. W., Blakeley, R. L. & Zerner, B. (1979). Ellman's reagent: 5,5'-Dithiobis(2-nitrobenzoic acid) — a reexamination. *Anal Biochem,* **94**:75–81.

Rison, S. C. G. & Thornton, J. M. (2002). Pathway evolution, structurally speaking. *Curr Opin Struct Biol,* **12**:374–382.

Robicsek, A., Strahilevitz, J., Jacoby, G. A., Macielag, M., Abbanat, D., Park, C. H., Bush, K. & Hooper, D. C. (2006). Fluoroquinolone-modifying enzyme: a new adaptation of a common aminoglycoside acetyltransferase. *Nat Med,* **12**:83–88.

Roodveldt, C. & Tawfik, D. S. (2005). Shared promiscuous activities and evolutionary features in various members of the amidohydrolase superfamily. *Biochemistry,* **44**:12728–12736.

Roth, C., Rastogi, S., Arvestad, L., Dittmar, K., Light, S., Ekman, D. & Liberles, D. A. (2007). Evolution after gene duplication: models, mechanisms, sequences, systems, and organisms. *J Exp Zool (Mol Dev Evol),* **308**:58–73.

Rubinstein, A. & Major, D. T. (2010). Understanding the catalytic specificity in alanine racemase from quantum mechanical and molecular mechanical simulations of the arginine 219 mutant. *Biochemistry,* **49**:3957–3964.

Sambrook, J. & Russell, D. W. (2001). *Molecular cloning. A laboratory manual* (3rd ed.). New York: CSHL Press.

Sandegren, L. & Andersson, D. I. (2009). Bacterial gene amplification: implications for the evolution of antibiotic resistance. *Nat Rev Microbiol,* **7**:578–588.

Sander, P., Springer, B., Prammananan, T., Sturmfels, A., Kappler, M., Pletschette, M. & Böttger, E. C. (2002). Fitness cost of chromosal drug resistance-conferring mutations. *Antimicrob Agents Chemother,* **46**:1204–1211.

Schaefler, S. (1967). Inducible system for the utilization of β-glucosides in *Escherichia coli*. *J Bacteriol,* **93**:254–263.

Schneider, G., Käck, H. & Lindqvist, Y. (2000). The manifold of vitamin B6 dependent enzymes. *Structure,* **8**:R1–R6.

Seaton, S. C., Elliott, K. T., Cuff, L. E., Laniohan, N. S., Patel, P. R. & Neidle, E. L. (2012). Genome-wide selection for increased copy number in *Acinetobacter baylyi* ADP1: locus and context-dependent variation in gene amplification. *Mol Microbiol,* **83**:520–535.

Seebeck, F. P. & Hilvert, D. (2003). Conversion of a PLP-dependent racemase into an aldolase by a single active site mutation. *J Am Chem Soc,* **125**:10158–10159.

Soo, V. W. C., Hanson-Manful, P. & Patrick, W. M. (2011). Artificial gene amplification reveals an abundance of promiscuous resistance determinants in *Escherichia coli*. *Proc Natl Acad Sci USA,* **108**:1484–1489.

Sørensen, M. A. & Pedersen, S. (1991). Cysteine, even in low concentrations, induces transient amino acid starvation in *Escherichia coli*. *J Bacteriol,* **173**:5244–5246.

Stamper, G. F., Morollo, A. A. & Ringe, D. (1998). Reaction of alanine racemase with 1-aminoethylphosphonic acid forms a stable external aldimine. *Biochemistry,* **37**:10438–10445.

Strych, U., Penland, R. L., Jimenez, M., Krause, K. L. & Benedik, M. J. (2001). Charaterization of the alanine racemases from two Mycobacteria. *FEMS Microbiol Lett,* **196**:93–98.

Sun, S. & Toney, M. D. (1999). Evidence for a two-base mechanism involving tyrosine-265 from arginine-219 mutants of alanine racemase. *Biochemistry,* **38**:4058–4065.

The Coli Genetic Stock Center. Mutation *thi*–1. Retrieved January 16, 2012, from http://cgsc.biology.yale.edu/Mutation.php?ID=4515

The European Committee on Antimicrobial Susceptibility Testing [EUCAST] (2011). Clinical breakpoints. Retrieved September 23, 2011, from http://www.eucast.org/clinical_breakpoints

The UniProt Consortium (2010). The universal protein resource (UniProt) in 2010. *Nucleic Acids Res,* **38**:D142–D148.

Thoden, J. B., Taylor Ringia, E. A., Garrett, J. B., Gerlt, J. A., Holden, H. M. & Rayment, I. (2004). Evolution of enzymatic activity in the enolase superfamily: Structural studies of the promiscuous *o*-succinylbenzoate synthase from *Amycolatopsis*. *Biochemistry,* **43**:5716–5727.

Thomas, P. W., Zheng, M., Wu, S., Guo, H., Liu, D., Xu, D. & Fast, W. (2011). Characterization of purified New Delhi metallo-β-lactamase-1. *Biochemistry,* **50**:10102–10113.

Thornberry, N. A., Bull, H. G., Taub, D., Wilson, K. E., Giménez-Gallego, G., Rosegay, A., Soderman, D. D. & Patchett, A. A. (1991). Mechanism-based inactivation of alanine racemase by 3-halovinylglycines. *J Biol Chem,* **266**:21657–21665.

Toney, M. D. (2005). Reaction specificity in pyridoxal phosphate enzymes. *Arch Biochem Biophys,* **433**:279–287.

Toney, M. D. (2011). Controlling reaction specificity in pyridoxal phosphate enzymes. *Biochim Biophys Acta,* **1814**:1407–1418.

Tracewell, C. A. & Arnold, F. H. (2009). Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Curr Opin Chem Biol,* **13**:1–7.

Tschowri, N., Busse, S. & Hengge, R. (2009). The BLUF-EAL protein YcgF acts as a direct anti-repressor in a blue-light response of *Escherichia coli*. *Genes Dev,* **23**:522–534.

Urbach, C., Fastrez, J. & Soumillion, P. (2008). A new family of cyanobacterial penicillin-binding proteins. A missing link in the evolution of class A β-lactamases. *J Biol Chem,* **283**:32516–32526.

Uren, J. R. (1987). Cystathionine β-lyase from *Escherichia coli*. *Methods Enzymol,* **143**:483–486.

Vacca, R. A., Giannattasio, S., Capitani, G., Marra, E. & Christen, P. (2008). Molecular evolution of B6 enzymes: Binding of pyridoxal-5'-phosphate and Lys41Arg substitution turn ribonuclease A into a model B6 protoenzyme. *BMC Biochem,* **9**:17.

Vedantam, G., Guay, G. G., Austria, N. E., Doktor, S. Z. & Nichols, B. P. (1998). Characterization of mutations contributing to sulfathiazole resistance in *Escherichia coli*. *Antimicrob Agents Chemother,* **42**:88–93.

Wada, M. & Takagi, H. (2006). Metabolic pathways and biotechnological production of L-cysteine. *Appl Microbiol Biotechnol,* **73**:48–54.

Wagner, A. (1998). The fate of duplicated genes: loss or new function? *BioEssays,* **20**:785–788.

Wagner, A. (2005). Energy constraints on the evolution of gene expression. *Mol Biol Evol,* **22**:1365–1374.

Wagner, A. (2008). Neutralism and selectionism: a network-based reconciliation. *Nat Rev Genet,* **9**:965–974.

Walsh, J. B. (1995). How often do duplicated genes evolve new functions? *Genetics,* **139**:421–428.

Wang, E. & Walsh, C. (1978). Suicide substrates for the alanine racemase of *Escherichia coli* B. *Biochemistry,* **17**:1313–1321.

Wang, E. A. & Walsh, C. (1981). Characteristics of β,β-difluoroalanine and β,β,β-trifluoroalanine as suicide substrates for *Escherichia coli* B alanine racemase. *Biochemistry,* **20**:7539–7546.

Wang, X., Kim, Y. & Wood, T. K. (2009). Control and benefits of CP4-57 prophage excision in *Escherichia coli* biofilms. *ISME J,* **3**:1164–1179.

Watanabe, A., Yoshimura, T., Mikami, B. & Esaki, N. (1999). Tyrosine 265 of alanine racemase serves as a base abstracting α-hydrogen from L-alanine: The counterpart residue to lysine39 specific to D-Alanine. *J Biochem,* **126**:781–786.

Weissbach, H. & Brot, N. (1991). Regulation of methionine synthesis in *Escherichia coli*. *Mol Microbiol,* **5**:1593–1597.

Wiegand, I., Hilpert, K. & Hancock, R. E. W. (2008). Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat Protoc,* **3**:163–175.

Wilmanns, M., Hyde, C. C., Davies, D. R., Kirschner, K. & Jansonius, J. N. (1991). Structural conservation in parallel β/α-barrel enzymes that catalyze three sequential reactions in the pathway of tryptophan biosynthesis. *Biochemistry,* **30**:9161–9169.

Wilmanns, M., Priestle, J. P., Niermann, T. & Jansonius, J. N. (1992). Three-dimensional structure of the bifunctional enzyme phosphoribosylanthranilate isomerase : Indoleglycerolphosphate synthase from *Escherichia coli* refined at 2.0 Å resolution. *J Mol Biol,* **223**:477–507.

Wolf, Y. I., Brenner, S. E., Bash, P. A. & Koonin, E. V. (1999). Distribution of protein folds in the three superkingdoms of life. *Genome Res,* **9**:17–26.

Wright, G. D. (2007). The antibiotic resistome: the nexus of chemical and genetic diversity. *Nat Rev Microbiol,* **5**:175–186.

Wu, D., Hu, T., Zhang, L., Chen, J., Du, J., Ding, J., Jiang, H. & Shen, X. (2008). Residues Asp164 and Glu165 at the substrate entryway function potently in substrate orientation of alanine racemase from *E. coli*: Enzymatic characterization with crystal structure analysis. *Protein Sci,* **17**:1066–1076.

Yčas, M. (1974). On earlier states of the biochemical system. *J Theor Biol,* **44**:145–160.

Yoshimura, T., Ashiuchi, M., Esaki, N., Kobatake, C., Choi, S.-Y. & Soda, K. (1993). Espression of *glr* (*murI, dga*) gene encoding glutamate racemase in *Escherichia coli*. *J Biol Chem,* **268**:24242–24246.

You, L. & Arnold, F. H. (1996). Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide. *Protein Eng,* **9**:77–83.

Yow, G.-Y., Watanabe, A., Yoshimura, T. & Esaki, N. (2003). Conversion of the catalytic specificity of alanine racemase to a D-amino acid aminotransferase activity by a double active-site mutation. *J Mol Catal B Enzym,* **23**:311–319.

Zhai, Y. & Saier, M. H., Jr. (2002). The beta-barrel finder (BBF) program, allowing identification of outer membrane beta-barrel proteins encoded within prokaryotic genomes. *Protein Sci,* **11**:2196–2207.

Zhang, Q., Lambert, G., Liao, D., Kim, H., Robin, K., Tung, C.-k., Pourmand, N. & Austin, R. H. (2011). Acceleration of emergence of bacterial antibiotic resistance in connected microenvironments. *Science,* **333**:1764–1767.

Zhou, L., Lei, X.-H., Bochner, B. R. & Wanner, B. L. (2003). Phenotype microarray analysis of *Escherichia coli* K-12 mutants with deletions of all two-component systems. *J Bacteriol,* **185**:4956–4972.