

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

# 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

© 2012  
Valerie Wooi Chee Soo  
All rights reserved.

# 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  $\beta$ -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  $\beta$ -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

---

<b>Abstract</b>	<b>iii</b>
<b>Acknowledgements</b>	<b>v</b>
<b>Abbreviations</b>	<b>xiii</b>
<b>1. Introduction</b>	<b>1</b>
1.1 Origins of new functions and metabolic pathways	2
1.2 Evolution of new functions, from the genetic perspective	5
1.3 Experimental validation of the IAD model	7
1.4 Aims of this study	9
<b>2. Cofactor-mediated promiscuity in alanine racemase</b>	<b>11</b>
2.1 Introduction	12
2.1.1 The pyridoxal 5'-phosphate-dependent enzymes	13
2.1.2 Cystathionine $\beta$ -lyase (MetC)	16
2.1.3 Alanine racemase (Alr)	18
2.1.4 Focus of this chapter	20
2.2 Results	21
2.2.1 Alr over-expression rescued <i>E. coli</i> $\Delta$ metC	21
2.2.2 <i>E. coli</i> $\Delta$ metC cells were starved from methionine	23
2.2.3 Alr over-expression did not rescue <i>E. coli</i> $\Delta$ metE	25
2.2.4 Cystathionine $\beta$ -lyase activities of MetC and Alr	28
2.2.4.1 Kinetic activities of cystathionine $\beta$ -lyase	28
2.2.4.2 Endpoint assays and mass spectrometry	30
2.2.4.3 pH effect on the cystathionine $\beta$ -lyase activity of Alr	34
2.2.5 Enzyme activities of Alr(R209E)	35
2.2.5.1 The <i>in vivo</i> alanine racemase activity of Alr(R209E)	38
2.2.5.2 The <i>in vivo</i> cystathionine $\beta$ -lyase activity of Alr(R209E)	39
2.2.5.3 The <i>in vitro</i> activity of Alr(R209E)	40



2.2.6	Improving the cystathionine $\beta$ -lyase activity of Alr via directed evolution	42
2.2.6.1	An Alr library generated by error-prone PCR	42
2.2.6.2	Selection of Alr variants with improved <i>in vivo</i> MetC activity	44
2.2.6.3	Confirmation of the improved growth phenotype of Alr mutants	47
2.2.6.4	The parental activity of the Alr mutants	50
2.2.6.5	Kinetic characterization of Alr-3	52
2.3	Discussion	55
2.3.1	Promiscuity in non-homologous enzymes	55
2.3.2	Directed evolution of Alr	57
2.3.3	Promiscuity is driven by chemistry	60
2.4	Materials and Methods	61
2.4.1	Construction of an empty ASKA vector (pCA24N-NoIns)	61
2.4.2	Construction of pCA24N- <i>alr</i> (GFP-)	62
2.4.3	Construction of pCA24N- <i>metC</i> (GFP-)	63
2.4.4	Growth complementation assays	64
2.4.4.1	Complementation of MetC deletion	64
2.4.4.2	Complementation of Alr (and DadX) deletion	64
2.4.5	Enzyme over-expression, purification & dialysis	65
2.4.6	Enzyme quantification	66
2.4.7	Steady state kinetic assays	66
2.4.7.1	Alanine racemization	66
2.4.7.2	Cystathionine $\beta$ -elimination	67
2.4.8	Endpoint assays and electrospray mass spectrometry	67
2.4.9	Construction of Alr(R209E)	68
2.4.10	Random mutagenesis of Alr	69
2.4.10.1	Construction of randomized Alr library	69
2.4.10.2	Selection for <i>alr</i> variants with higher lyase activity	70
2.4.10.3	Retransformation/Recloning and restreaking tests	71
<b>3.</b>	<b>Reciprocal promiscuity in two non-homologous enzymes</b>	<b>73</b>
3.1	Introduction	74
3.2	Results	75
3.2.1	Suppressor clones that rescue <i>E. coli</i> MB2795 ( $\Delta alr \Delta dadX$ )	75
3.2.2	Growth complementation tests	77
3.2.2.1	MetC(GFP-) complements <i>E. coli</i> ( $\Delta alr \Delta dadX$ )	77
3.2.2.2	Growth complementation of <i>E. coli</i> ( $\Delta alr \Delta dadX \Delta metC$ )	78

3.2.3	Alanine racemase activity of MetC	80
3.2.4	Directed evolution of MetC (round 1)	82
3.2.4.1	First-generation library	82
3.2.4.2	Selection of first-generation MetC variants	85
3.2.4.3	Reselection of first-generation MetC variants	85
3.2.5	Directed evolution of MetC (round 2)	91
3.2.5.1	Second-generation libraries	91
3.2.5.2	Selection of second-generation MetC variants	93
3.2.5.3	Reselection of second-generation MetC variants	94
3.2.6	The <i>in vivo</i> native activity of the selected MetC variants	102
3.2.7	Enzyme activities of MetC mutants	104
3.3	Discussion	109
3.3.1	Alanine racemase activity of MetC	109
3.3.2	<i>In vitro</i> evolution of MetC	111
3.3.3	Reciprocal promiscuity in non-homologous scaffolds	115
3.3.4	Increased protein expression as a route to fitness?	116
3.4	Materials and Methods	118
3.4.1	Identification of suppressor genes that complement alanine racemase	118
3.4.2	Construction of <i>E. coli</i> ( $\Delta alr \Delta dadX \Delta metC$ )	119
3.4.2.1	Amplification of the Kan <sup>R</sup> cassette	119
3.4.2.2	Introduction of the Kan <sup>R</sup> cassette into MB2795	119
3.4.2.3	Verification of the recombined Kan <sup>R</sup> cassette in MB2795	121
3.4.3	Growth complementation assays	121
3.4.4	Enzyme over-expression, purification & dialysis	121
3.4.5	Enzyme quantification	121
3.4.6	Steady state kinetic assays	121
3.4.7	Random mutagenesis of MetC (round 1)	122
3.4.7.1	Construction of first-generation MetC library	122
3.4.7.2	Selection for MetC variants with higher racemase activity	124
3.4.7.3	Restreaking tests	124
3.4.8	Random mutagenesis of MetC (round 2)	125
3.4.8.1	Subcloning <i>metC</i> inserts into pBAD vector	125
3.4.8.2	Construction of second-generation MetC libraries	126
3.4.8.3	Selection for MetC variants with higher racemase activity	127
3.4.8.4	Retransformation/Recloning and restreaking tests	128

<b>4.</b>	<b>Evolution of new phenotypes mediated by protein over-expression</b>	<b>129</b>
4.1	Introduction	130
4.1.1	The genetics of adaptive evolution	130
4.1.2	The evolutionary potential of a modern proteome	131
4.2	Results	132
4.2.1	Library-on-library screen (PM 1 to PM 10)	132
4.2.2	Enrichment of the fittest clone(s)	133
4.2.3	Carbon sources	135
4.2.3.1	ASKA-encoded ORFs isolated from PM 1 & PM 2	135
4.2.3.2	Glucuronamide utilization in <i>E. coli</i>	138
4.2.4	Nitrogen sources	142
4.2.5	Phosphorus & Sulfur sources	154
4.2.6	Biosynthetic nutrients	159
4.2.7	Osmolytes & pH stresses	165
4.2.8	Cases where the negative control out-grew the ASKA pool	167
4.3	Discussion	168
4.3.1	An overview of the results	168
4.3.2	Over-represented ORFs	170
4.3.2.1	Cfa	170
4.3.2.2	YcbS	171
4.3.2.3	YlcG	172
4.3.3	Thiamine auxotrophy in <i>E. coli</i> DH5 $\alpha$ -E	173
4.3.3.1	Speculation on the effects of Cfa, YcbS and YlcG	175
4.3.4	ORFs enriched from thiamine-containing environments	176
4.3.5	Protein over-expression can mediate the emergence of new phenotypes	178
4.4	Materials and Methods	179
4.4.1	Construction of the ASKA library in DH5 $\alpha$ -E	179
4.4.2	Construction of the control clone	179
4.4.3	Library-on-library screen	179
4.4.4	Isolation of winner genes via serial enrichment	180
4.4.4.1	Probability of sampling the same ORF twice or more, by chance alone	181
4.4.5	<i>E. coli</i> growth using glucuronamide as a carbon source	181
<b>5.</b>	<b>Toxin resistance mediated by protein over-expression</b>	<b>183</b>
5.1	Introduction	184
5.1.1	Molecular origins of antibiotic resistance	184

5.2	Results	186
5.2.1	Library-on-library screen (PM 11 to PM 20)	186
5.2.1.1	Cases where the ASKA pool out-grew the negative control	186
5.2.1.2	Cases where the negative control out-grew the ASKA clone	199
5.2.2	Identification of novel resistance genes	199
5.2.3	Diverse mechanisms of resistance	201
5.2.4	The interplay of fitness and resistance	205
5.3	Discussion	209
5.3.1	An overview of the results	209
5.3.2	A reservoir of unexplored toxin resistance	210
5.3.3	Protein over-expression: a possible adaptive mechanism	212
5.4	Materials and Methods	214
5.4.1	Identification of latent resistance genes	214
5.4.2	Neutrally marked <i>E. coli</i> strain	214
5.4.2.1	Construction of a suicide plasmid	214
5.4.2.2	<i>LacZ</i> -tagged control clone	215
5.4.3	Relative fitness assays	217
5.4.4	Antibiotic susceptibility testing	217
5.4.4.1	E-tests	217
5.4.4.2	Broth microdilution	218
<b>6.</b>	<b>Concluding remarks</b>	<b>219</b>
6.1	A look backwards in enzyme evolution	219
6.2	Looking forwards in enzyme evolution	221
6.3	Final comments	222
<b>Appendix I.</b>	<b>General materials and materials</b>	<b>223</b>
I.1	Reagents	223
I.2	Growth media and Antibiotics	223
I.3	Bacterial strains	224
I.4	Plasmids	224
I.5	Analytical software	226
I.6	Oligonucleotides	227
I.7	Agarose gel electrophoresis	228
I.8	DNA extraction and clean-up	228
I.9	DNA quantification	228

I.10	PCR screening	228
I.11	Restriction digest and ligation	229
I.12	Electrocompetent cells	229
I.13	Transformation	230
I.14	DNA sequencing	230
I.15	SDS-PAGE	231
<b>Appendix II. Statement of contributions</b>		<b>233</b>
<b>Appendix III. Publication arising from this work</b>		<b>235</b>
<b>References</b>		<b>243</b>

## ABBREVIATIONS

---

Alr	Alanine racemase
ASKA	A complete set of <i>E. coli</i> K-12 ORF Archive
A <sub>xxx</sub>	Absorbance measured at xxx nm
cfu	Colony-forming unit
CHES	<i>N</i> -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) <sub>6</sub>	Hexa-histidine
IPTG	Isopropyl- $\beta$ -D-1-thiogalactopyranoside
<i>lacZ</i>	$\beta$ -galactosidase
LB	Lysogeny broth
LDH	Lactate dehydrogenase
<i>m/z</i>	Mass-to-charge ratio
MetC	Cystathionine $\beta$ -lyase
MIC	Minimum inhibitory concentration
NAD <sup>+</sup>	Oxidized form of $\beta$ -nicotinamide adenine dinucleotide
NADH	Reduced form of $\beta$ -nicotinamide adenine dinucleotide
NoIns (or NoInsert)	The negative control plasmid, pCA24N-NoIns (or <i>E. coli</i> harbouring the pCA24N-NoIns plasmid)

OD <sub>600</sub>	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- $\beta$ -D-galactopyranoside