

The evolution of enzyme functions in the metallo- β -lactamase superfamily

by

Florian Baier

MSc, University Pompeu Fabra, Barcelona, 2010

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Genome Science and Technology)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

April 2017

© Florian Baier, 2017

Abstract

Enzyme superfamilies have expanded over billions of years from the descendants of a potentially single common ancestral function. Understanding the evolution of their functional diversity is central to biochemistry, molecular and evolutionary biology. The overarching question of my thesis is how enzyme promiscuity, the serendipitous ability to catalyze non-native reactions and reactions, connects enzyme functions and facilitates molecular evolution by providing evolutionary starting points towards new functions. In particular, I primarily focus on proteins across the metallo- β -lactamase (MBL) superfamily by comparing evolutionary and functional connectivity based on the functional profiling of 24 enzymes against 10 distinct hydrolytic MBL reactions. This analysis revealed that MBL enzymes are generally promiscuous, as each enzyme catalyzes on average 1.5 reactions in addition to its native one, which leads to high functional connectivity. Furthermore, the ability to promiscuously bind different metal ions, enzymatic co-factors of MBL enzymes, provide additional mechanisms whereby the function profile of some MBL enzymes can be broadened, and thus further extends the connectivity between functions. In addition, I expand and compare the analyses of function connectivity through promiscuity to three previously published superfamily-wide function profiling studies, which revealed common trends that are discussed in the context of enzyme superfamily evolution. Finally, I assess the evolvability of promiscuous enzymes to determine their potential as evolutionary starting points towards a novel function by performing a comparative laboratory evolution experiment of two related β -lactamases, NDM1 and VIM2, towards a shared promiscuous phosphonate monoester hydrolase activity. Both trajectories accumulate 13 mutations over ten rounds of directed evolution, however the mutational solutions and evolvability is strikingly different for the two enzymes. NDM1 improves catalytic efficiency by over 20,000-fold and loses much of its solubility, *i.e.* the amount of functional enzyme in the cell. Contrarily, VIM2 improves catalytic efficiency only by 60-fold, but improves solubility. Detailed structural analysis, combined with molecular dynamics simulations, reveals a molecular understanding for the observed differences in evolvability between NDM1 and VIM2. Overall, my research contributes to our understanding of enzyme evolution and will help to advance functional annotation and engineering of enzyme.

Preface

Parts of chapter one, four and six have been written together with Janine N. Copp in the laboratory of Dr. N. Tokuriki at UBC, Vancouver, Canada and published in “Baier F., Copp J.N., Tokuriki N. (2016): Enzyme superfamilies – new approaches toward systematic mapping of evolutionary sequence-function relationships. *Biochemistry*, 2016, 55 (46), 6375–6388.” J. N. Copp developed the concepts of the initial parts of the manuscript. I developed the concepts of the enzyme evolution and function connectivity network parts together with my supervisor, Dr. Nobuhiko Tokuriki. The manuscript was written and edited by all authors.

Parts of chapter two have been performed in collaboration with G. Woollard in the laboratory of Dr. Jörg Gsponer at UBC, Vancouver, Canada and published as “Baier F. and Tokuriki N. (2014): Connectivity between catalytic landscapes of the metallo-beta-lactamase superfamily. *J. Mol. Biol.*, 426 (13), 2442-2456.” G. Woollard performed hydrophobicity calculation of protein active sites, as described in chapter two in **Figure 2.10**. I performed all other experiments and wrote the manuscript together with my supervisor, Dr. Nobuhiko Tokuriki.

Parts of chapter three have been performed in collaboration with M. Solomonson in the laboratory of Dr. N. C. J. Strynadka at UBC, Vancouver, Canada, and together with J. Chen in the laboratory of Dr. N. Tokuriki at UBC, Vancouver, Canada and published as “Baier F., Chen J., Solomonson M., Strynadka N. C. J., Tokuriki N. (2015): Distinct metal isoforms underlie promiscuous activity profiles of metalloenzymes. *ACS Chem. Biol.*, 10 (7), 1684–1693.” M. Solomonson performed ICP-MS analysis for quantitative metal content analysis as described in **Table 3.2**. J. Chen worked as an undergraduate student in the laboratory under my co-supervision and he performed most of the protein purifications, enzyme kinetics and metal ion titration experiments. I performed all other experiments and wrote the manuscript together with my supervisor, Dr. Nobuhiko Tokuriki, and some help of the other co-authors.

Parts of chapter five have been performed in collaboration with N. Hong in the laboratory of Dr. Colin J. Jackson at ANU, Canberra, Australia, and A Pabis in the laboratory of Dr. S. C. L. Kamlerlin at Uppsala University, Uppsala, Sweden. N. Hong performed crystal structure analysis of NDM1-R10 and VIM2-R10 and size exclusion chromatography of NDM1 and VIM2 variants, as shown in **Figure 5.14** and **5.8**, respectively. A. Pabis performed molecular dynamics simulations of NDM1 and VIM2 structures as shown in **Figure 5.15** and **5.17**. I performed all other experiments and wrote the chapter together with my supervisor, Dr. Nobuhiko Tokuriki.

The UBC Certificates of Ethical Approval of the Tokuriki lab relevant for the research reported in this dissertation are B16-0212 (Environmental-dependant evolution), B16-0154 (Evolutionary approaches to treat and evade antibiotic resistant pathogenic bacteria) and B13-0026 (Experimental evolution of MBL).

Table of Contents

Abstract	ii
Preface	iii
Table of Contents	v
List of Tables	ix
List of Figures	x
List of Abbreviations	xii
Acknowledgements	xiv
Chapter 1: Introduction	1
1.1 Enzymes as biological catalysts.....	1
1.2 The functional diversity of enzymes.....	3
1.3 General approaches to infer and characterize the function of enzymes.....	4
1.4 Investigating evolutionary relationships within enzyme superfamilies.....	5
1.5 The evolution of enzyme functions.....	7
1.6 General concept of enzyme evolution.....	8
1.7 Prevalence of enzyme promiscuity	9
1.8 Molecular basis of enzyme promiscuity	10
1.9 Promiscuous activities as evolutionary starting points	12
1.10 Evolvability of promiscuous activities.....	15
1.11 Epistasis in enzyme evolution.....	16
1.12 The role of protein stability in enzyme evolution.....	19
1.12.1 Thermodynamic stability and enzyme evolution.....	20
1.12.2 Kinetic stability and enzyme evolution.....	21
1.13 Functional trade-offs between native and promiscuous functions.....	22
1.14 Directed evolution as a tool to investigate fundamentals of enzyme evolution	23
1.15 The metallo- β -lactamase superfamily.....	25
1.16 Aims and scope of the dissertation	27
Chapter 2: Evolutionary relationship functions and catalytic promiscuity in the metallo-β-lactamase superfamily	29
2.1 Summary	29
2.2 Introduction.....	30
2.3 Methods.....	35
2.3.1 Construction of sequence similarity networks.....	35
2.3.2 Sequence identity and structural similarity calculation and phylogeny.....	36
2.3.3 Molecular cloning.....	36
2.3.4 Protein expression and purification	37
2.3.5 Enzyme assays and kinetics	38
2.3.6 Metal removal experiments.....	39
2.3.7 Active-site cavity detection and fraction of hydrophobic residues.....	39

2.4	Results.....	40
2.4.1	Sequence relationship between functional families of the MBL superfamily	40
2.4.2	Selection of enzymes and reactions for function-profiling analysis in the MBL superfamily	45
2.4.3	Function profiling analysis of enzymes of the MBL superfamily	47
2.4.4	Function connectivity and evolutionary divergence	51
2.4.5	Relationship between function profiles and active site features.....	52
2.5	Discussion	54
Chapter 3: Metal ion cofactor mediated catalytic promiscuity of enzymes in the metallo-β-lactamase superfamilies.....		58
3.1	Summary	58
3.2	Introduction.....	59
3.3	Methods.....	62
3.3.1	Chemicals.....	62
3.3.2	Molecular cloning.....	63
3.3.3	Protein expression and purification	63
3.3.4	Preparation of apo- and metal-substituted enzymes	64
3.3.5	Enzyme assays and kinetics	64
3.3.6	Metal content analysis.....	65
3.3.7	Lysate activity analysis.....	65
3.4	Results.....	66
3.4.1	Experimental dataset.....	66
3.4.2	Reconstitution of enzymes with various metals and activity screening	68
3.4.3	MBL enzymes exhibit distinctive metal preferences for their native activity	71
3.4.4	Metal substitution alters function profiles and exposes cryptic promiscuous activities	71
3.4.5	MBL enzymes exist as heterogeneous metal isoforms in <i>E. coli</i>	74
3.4.6	Bioavailability of metals can alter function profiles.....	75
3.4.7	Reconstitution of heterogeneous metal isoforms.....	79
3.5	Discussion	80
Chapter 4: Function connectivity in enzyme superfamilies.....		84
4.1	Summary	84
4.2	Introduction.....	84
4.3	Methods.....	86
4.3.1	Datasets and network generation	86
4.4	Results and discussion	86
4.4.1	Function connectivity networks.....	86
4.4.1.1	FCN analysis of the metallo- β -lactamase superfamily	87
4.4.1.2	FCN analysis of the cytGST superfamily	89
4.4.1.3	FCN analysis of the BKACE family.....	90

4.4.1.4	FCN analysis of HAD superfamily.....	92
4.4.2	Perspectives on function connectivity through promiscuity.....	93
4.4.2.1	Indirect connectivity through intermediate functions.....	94
4.4.2.2	The scope and level of promiscuity is different for each enzyme.....	95
4.4.2.3	Function connectivity depends on cofactor availability.....	96
4.4.3	Structural features that determine function connectivity.....	98
Chapter 5:	Cryptic genetic variation affects enzyme evolvability.....	100
5.1	Summary.....	100
5.2	Introduction.....	101
5.3	Methods.....	105
5.3.1	Generation of mutagenized library.....	105
5.3.2	Generation of DNA shuffling libraries.....	105
5.3.3	Site-directed mutagenesis.....	106
5.3.4	Pre-screen on agar plates.....	106
5.3.5	Cell lysate activity screen in 96-well plates.....	106
5.3.6	Purification of Strep-tagged proteins.....	107
5.3.7	Enzyme kinetics.....	107
5.3.8	Thermostability assay.....	107
5.3.9	Protein purification for crystallization.....	108
5.3.10	Crystallization of NDM1-R10.....	108
5.3.11	Crystallization of VIM2-R10.....	109
5.3.12	Structural data collection and structure determination.....	109
5.3.13	Molecular dynamics simulation.....	110
5.4	Results.....	111
5.4.1	The selection of evolutionary starting points.....	111
5.4.2	Directed evolution strategy.....	113
5.4.3	Fitness improvement toward PMH activity.....	115
5.4.4	Different activity and solubility changes underlie fitness improvements.....	117
5.4.5	Correlation between solubility, thermostability and structural stoichiometry.....	119
5.4.6	Different mutational paths support the distinct phenotypic outcomes.....	121
5.4.7	Mutational trajectories appear deterministic for each starting point.....	122
5.4.8	The compatibility of initial mutations in different genetic backgrounds.....	125
5.4.9	The structural basis for improved PMH activity.....	128
5.4.10	Structural adaptation of NDM1-R10.....	128
5.4.11	Understanding the structural basis of PMH substrate binding of NDM1 variants.....	132
5.4.12	Structural adaptation of VIM2-R10.....	133
5.4.13	The molecular basis of mutational incompatibility.....	135
5.5	Discussion.....	137
Chapter 6:	Conclusion and future outlook.....	142
6.1	General summary and conclusion.....	142

6.2	Future outlook.....	142
6.2.1	Integrating metal ion availability and functional divergence	143
6.2.2	Detailed characterization of the 3D domain swap of VIM2	144
6.2.3	Understanding how completely different catalytic functions evolve	145
6.2.4	Exploring and annotating protein functions is far from completion.....	146
	Bibliography	148
	Appendices.....	169
	Appendix A Supplementary material for chapter two	169
A.1	Individual kinetic parameters.....	169
	Appendix B Supplementary information for chapter three.....	171
B.1	Kinetic parameters of bla-L1.....	171
B.2	Table of individual kinetic parameters of bla-VIM2	172
B.3	Table of individual kinetic parameters of mph	173
B.4	Table of individual kinetic parameters for atsA.....	174
B.5	Table of individual kinetic parameters for rbn.....	175

List of Tables

Table 2.1 Sequence and structure similarity of enzymes used in this study.....	34
Table 2.2. Sequence and structure similarity of enzymes used in this study.....	35
Table 2.3 Information on cloning and source of enzymes assayed.	37
Table 2.4 Information on enzymatic reactions and substrates.....	53
Table 3.1 Information on enzymes used in this study.	67
Table 3.2 Quantitative metal content analysis.	74
Table 5.1 General information on enzyme characterized in this study.....	104
Table 5.2 Screening, selection and mutations of directed evolution rounds.	115
Table 5.3 Crystallographic data collection and refinement statistics.	131
Table A.1 Kinetic parameters.	169
Table B.1 Kinetic parameters of bla-L1.	171
Table B.2 Kinetic parameters of bla-VIM2.....	172
Table B.3 Kinetic parameters of mph.....	173
Table B.4 Kinetic parameters of atsA.....	174
Table B.5 Kinetic parameters of rbn.....	175

List of Figures

Figure 1.1 Sequence and function diversity within selected enzyme superfamilies.....	4
Figure 1.2 Schematic representation of functional divergence within a hypothetical enzyme superfamily.	9
Figure 1.3 Simplified threshold model for promiscuous activities as evolutionary starting points.	14
Figure 1.4 General concept of mutational epistasis in proteins.	16
Figure 2.1 Structures of representative MBL superfamily members.....	32
Figure 2.2 Active site cavity of representative MBL superfamily members.....	33
Figure 2.3 Sequence similarity network of the MBL superfamily at different BLAST E-value cut-offs.....	43
Figure 2.4 Sequence similarity network of MBL superfamily members.....	43
Figure 2.5 Mapping taxonomy and sequence length on the sequence similarity network of the MBL superfamily.....	44
Figure 2.6 Substrates used in this study.....	47
Figure 2.7 Activity patterns of selected MBL superfamily members.....	49
Figure 2.8 Differences between the efficiency of native and promiscuous activities.	49
Figure 2.9 Metal chelating control experiment for nine selected enzymes.	50
Figure 2.10 Relationship of two activities, BLA and PDE, and active-site properties.....	54
Figure 2.11 Kinetic parameters of 14 selected enzymes against 5 non-MBL superfamily reactions.	57
Figure 3.1 Structures of selected enzymes and the general catalytic mechanism of MBL enzymes.....	62
Figure 3.2 Sequence relationships of selected enzymes within the MBL superfamily.	67
Figure 3.3 Enzymatic reactions and substrates used in this study.	68
Figure 3.4 Activity levels of metal-depleted apo- and metal reconstituted enzymes.	69
Figure 3.5 Function profiles of five MBL enzymes reconstituted with various metals.	70
Figure 3.6 Metal activation of apo-bla-L1 for native and promiscuous activities.....	73
Figure 3.7 The effect of metal supplementation on enzymatic activities in cell lysate.....	77
Figure 3.8 <i>E.coli</i> cell growth in the presence of various metals.	78
Figure 3.9 Expression level of the enzymes in the lysate activity experiment with different metals.	78
Figure 3.10 The effect of metal combinations on the catalytic activities of bla-L1.	80
Figure 4.1 Schematic depictions of superfamily-wide function profiling and function connectivity networks.	87
Figure 4.2 FCNs constructed based on the function-profiling analysis of the MBL and cytGST superfamilies.	88
Figure 4.3 FCN representation of the function-profiling analysis of the BKACE and HAD superfamilies.	91

Figure 4.4 FCN representation of the metal-dependent function profiles for five MBL superfamily enzymes.	97
Figure 5.1 The comparative evolution of B1 β -lactamases towards promiscuous PMH activity.	104
Figure 5.2 Sequence identity and structure similarity among selected B1 β -lactamases.	112
Figure 5.3 Sequence alignment of selected B1 β -lactamases.	113
Figure 5.4 Overview of the laboratory evolution strategy.	114
Figure 5.5 Phenotypic adaptations of NDM1 and VIM2 during the directed evolution experiment.	117
Figure 5.6 SDS-PAGE analysis of solubility of NDM1 (top) and VIM2 (bottom) variants.	117
Figure 5.7 Effect of expression temperature on cell lysate activities.	119
Figure 5.8 Size exclusion chromatography.	121
Figure 5.9 The mutations accumulated in the evolutionary trajectories of NDM1 and VIM2.	122
Figure 5.10 Improved variants isolated in two additional directed evolution experiments.	124
Figure 5.11 Epistasis analysis of trajectory mutations.	124
Figure 5.12 Fitness and solubility effect of W93G for NDM1 and VIM2.	126
Figure 5.13 Functional compatibility of initial mutations among related B1 β -lactamases.	127
Figure 5.14 The structural basis for improved PMH activity of NDM1-R10.	130
Figure 5.15 Molecular dynamics simulations of NDM1 variants.	132
Figure 5.16 The structural basis for improved PMH activity of VIM2-R10.	134
Figure 5.17 Molecular dynamics simulations of VIM2 variants.	137

List of Abbreviations

AKS	alkylsulfatase
AMPS	Alpha-, Mu-, Pi-, and Sigma-like
ARS	arylsulfatase
ATP	adenosine triphosphate
BKACE	β -keto acid cleavage enzyme
BLA	β -lactamase
cytGST	cytosolic glutathione transferase
EC	enzyme commission
EST	esterase/lipase
FCN	function connectivity network
g	gravity
HAD	haloacid dehalogenase
HSL	homoserine lactonase
ID	identification
IPTG	isopropyl β -D-1-thiogalactopyranoside
LAC	lactonase
MBL	metallo- β -lactamase
MD	molecular dynamics
MES	2-(<i>N</i> -morpholino)ethanesulfonic acid
PCE	phosphoryl-choline esterase
PDB	protein data bank
PDE	phosphodiesterase
PMH/PPP	phosphonate monoester hydrolase (or phosphonatase)
PTE	phosphotriesterase
SSN	sequence similarity network
RMSD	root mean standard deviation
ICP-MS	inductively coupled plasma mass spectrometry
OD _{xxx}	optical density at XXX nm
PCR	polymerase chain reaction
<i>p</i> NP	<i>para</i> -nitrophenol

rpm	revolutions per minute
SDS-PAGE	sodium dodecylsulfate polyacrylamide gel electrophoresis
SLG	glyoxalase II
TPN	chlorothalonil dehalogenase
Tris	2-amino-2-(hydroxymethyl)-1,3-propanediol
WT	wild type

Amino acid abbreviations:

A (or Ala)	alanine
C (or Cys)	cysteine
D (or Asp)	aspartate
E (or Glu)	glutamate
F (or Phe)	phenylalanine
G (or Gly)	glycine
H (or His)	histidine
I (or Ile)	isoleucine
K (or Lys)	lysine
L (or Leu)	leucine
M (or Met)	methionine
N (or Asn)	asparagine
P (or Pro)	proline
Q (or Gln)	glutamine
R (or Arg)	arginine
S (or Ser)	serine
T (or Thr)	threonine
V (or Val)	valine
W (or Trp)	tryptophan
Y (or Tyr)	tyrosine

Acknowledgements

First, I thank my supervisor, Dr. Nobuhiko Tokuriki, for his guidance, teaching and patience throughout my PhD and for giving me the opportunity to also follow my own research ideas. I am thankful for his encouragement and support in developing crucial scientific skills beyond bench work: critical thinking, scientific writing and presenting research at conferences. Furthermore, I acknowledge his support for a three-month visit in the lab of Dr. Colin J. Jackson lab at the Australian National University (ANU) in Canberra.

I would also like to thank my committee members Dr. Thibault Mayor, Dr. Jörg Gsponer and Dr. Stephen J. Hallam for their scientific support and critical feedback on my research and dissertation.

I am very thankful to all collaborators; without their support much of the research described in this thesis would not have been possible. In particular, I would like to thank Nansook Hong and Dr. Colin J. Jackson for crystallizing the laboratory-evolved variants and for hosting me at the Jackson laboratory at ANU and introducing me to the world of crystallography. I am also thankful to Nansook Hong for providing me a room to stay while visiting ANU. Furthermore, I am grateful to our collaborators Dr. Anna Pabis, Alexandre Barrozo and Dr. Shina Caroline Lynn Kamerlin at the Uppsala University for their molecular dynamics simulation analysis. I would also like to thank Dr. Matthew M. Solomonson from the Strynadka lab for his help with the ICP-MS analysis. I thank Geoffrey Woollard from the Gsponer lab for his help with computational structure analysis.

I am grateful to all present and past colleagues in the Tokuriki for providing help and support throughout my research. In particular, I would like to thank John Chen and Gloria Yang for the successful work together. I also thank Dr. Janine N. Copp, Dr. Dave W. Anderson and Dr. Charlotte M. Miton for the scientific discussion and proof reading.

I acknowledge financial support from the Genome Science and Technology Graduate program and through the Faculty of Graduate Studies Travel award.

Finally, I am thankful to my family, friends, roommates and all colleagues for their constant support. Thank you Nina for the great time in Cascadia.

Chapter 1: Introduction

The emphasis of this thesis is the evolution of enzymes, and in particular, how enzyme promiscuity connects functions in enzyme superfamilies and facilitates the evolution of new enzyme functions. For my research I mainly focused on enzymes of the MBL superfamily as a model system. In the first part of the introduction I will describe enzymes as biological catalysts, their current functional diversity, and the experimental approaches that have been implemented to explore functional diversity and to infer evolutionary relationships in large enzyme superfamilies. In the second part, I will focus on the general models of enzyme evolution and explain the functional, biophysical and genetic factors that affect it. I will highlight historical studies, which pioneered the field, summarize recent advances and propose open questions that still need to be addressed. I will also introduce directed evolution as a tool to address molecular evolutionary questions. Finally, I will introduce the model system of this thesis, the MBL superfamily, describing the sequence, structure and function diversity of MBL enzymes as well as describe their catalytic mechanism. Note that throughout my thesis I will mainly focus on enzymes, although most concepts apply to proteins and their functions in general.

Parts of chapter one have been written together with Janine N. Copp in the laboratory of Dr. N. Tokuriki at UBC, Vancouver, Canada and published in “Baier F., Copp J.N., Tokuriki N. (2016): Enzyme superfamilies – new approaches toward systematic mapping of evolutionary sequence-function relationships. *Biochemistry*, 2016, 55 (46), 6375–6388.”

1.1 Enzymes as biological catalysts

Metabolism, whereby organic molecules are synthesized (anabolism) and fragmented (catabolism), is a fundamental requirement of all living organism and the sum of its chemical reactions to support life (Watson et al. 2015; Klitgord & Segrè 2011). The cellular metabolism receives its energy from extracellular sources from which intracellular energy-rich molecules are generated (catabolism), such as adenosine triphosphate (ATP), that are then consumed to fuel other chemical reactions that require energy (anabolism), such as the synthesis of nucleotides, amino acids, lipids, DNA, RNA, proteins, membranes, signaling and defense molecules and many other

cellular and chemical compounds (Erecińska & Wilson 1978). The chemical reactions that fragment and synthesize a molecule are organized into metabolic pathways and their sum essentially constitutes the metabolic capabilities of an organism, which ultimately defines an organism. Although most biologically relevant chemical reactions occur spontaneously under physiological conditions, their rate is far too slow to sustain life (Wolfenden 2011). Enzymes, however, catalyze and accelerate the rate of reactions from millions of years to a biologically relevant time-scale of seconds to occur under physiological conditions. For example, the yeast orotidine-5'phosphate decarboxylase (OMP decarboxylase) is an essential enzyme in the last step of the uridine monophosphate biosynthesis pathway and a knockout of its gene causes uracil auxotrophy (B. G. Miller & Wolfenden 2002). OMP decarboxylase catalyzes the decarboxylation of orotic acid with approximately 39 turnovers per second (k_{cat}), whereas the uncatalyzed reaction would require 78 million years, which means it accelerates the reaction by seventeen orders of magnitude (B. G. Miller & Wolfenden 2002). The “catalytic efficiency”, k_{cat}/K_M ($\text{M}^{-1}\text{s}^{-1}$), of enzymes is generally described as a function of substrate turnover per time, k_{cat} (s^{-1}), and the affinity for the substrate, K_M (M) or Michaelis constant, at which the reaction rate is at half of its maximum velocity, v_{max} (Benkovic & Hammes-Schiffer 2003). All catalysts, including enzymes, enhance catalytic rates by specifically lowering the activation energy (E_A) of a chemical reaction (Wolfenden 2011; Benkovic & Hammes-Schiffer 2003). The activation energy, E_A , describes the free energy difference between the ground state of the substrate and the configuration of its highest energy, also called transition state (TS), along the reaction coordinate that yields the product(s). The catalytic power of enzymes arises from particularly stabilizing the TS and thus lowering the required E_A (Schramm 2011), through, for example, neutralizing and stabilizing unfavorable charges of the TS with complementary charged groups in the active site (Benkovic & Hammes-Schiffer 2003). Additionally, protein motion and dynamics are also crucial for enzyme catalysis, and several studies demonstrated that the rate limiting steps during enzyme catalysis can be opening and closing of the active site, substrate binding and product release, which will be described later in more detail (Campbell et al. 2016; Henzler-Wildman et al. 2007; Hammes-Schiffer & Benkovic 2006; Gobeil et al. 2014).

1.2 The functional diversity of enzymes

Enzymes have evolved to a remarkable level of functional diversity, and understanding how new enzyme functions evolve is one of the most profound questions in biological science. The initial steps toward understanding the functional diversity of enzymes were taken over 80 years ago when enzymes were first structurally and biochemically characterized (Blow 2000). The numerous publications since that time have profoundly deepened our knowledge of enzyme functional diversity and their sequence-structure-function relationships (Furnham et al. 2016; Gerlt et al. 2015; Glasner et al. 2006; Gerlt, Babbitt, et al. 2011; Furnham et al. 2012; Seibert & Raushel 2005). The ExplorEnz Database, which is the primary source of the International Union of Biochemistry and Molecular Biology (IUBMB) enzyme list, currently counts 5,787 unique enzyme functions based on the Enzyme Commission (EC) number classification system with 1,681 Oxidoreductases (EC 1), 1,724 Transferases (EC2), 1,309 Hydrolases (EC 3), 602 Lyases (EC 4), 273 Isomerases (EC 5) and 188 Ligases (EC 6). The diversity of structural folds and mechanistic features in active sites (catalytic residues and cofactors) undoubtedly contribute to the extraordinary functional repertoire observed in modern enzymes (Das et al. 2015; Todd et al. 2001; Aloy et al. 2002; Meng & Babbitt 2011; Fariás-Rico et al. 2014). In 2015, the CATH database (v4.1, 2015) classifies 308,999 structural domains that fall within 2,737 superfamilies, in which proteins share a common structural fold, sequence motif, catalytic features and common ancestry (Sillitoe et al. 2015). Even within a single superfamily, the ability of enzymes to catalyze a diverse range of chemical reactions is astonishing (Brown et al. 2006; Furnham et al. 2016). Not only do enzymes of the same superfamily often act on different substrates, with distinct size, structure and electrostatics, but often also catalyze different chemical reactions, such as hydrolase, isomerase and oxidase reactions (Brown et al. 2006; Furnham et al. 2016). Examples of functionally diverse enzyme superfamilies are the haloacid dehalogenase (HAD), enolase, cytosolic glutathione transferase (cytGST), metallo- β -lactamase (MBL) and amidohydrolase superfamilies, which catalyze a wide variety of distinct chemical reactions, spanning all six E.C. classes (**Figure 1.1**) (Meng & Babbitt 2011; Gerlt, Babbitt, et al. 2011; Seibert & Raushel 2005; Bebrone 2007). For example, enzymes of the HAD superfamily, which is one of the most studied enzyme superfamilies, have 181

different enzymatic functions assigned with 18 being oxidoreductases, 45 transferases, 89 hydrolases, 7 lyases, 10 isomerases and 8 ligases. These enzyme superfamilies provide dramatic, but not unusual, examples of evolutionarily related functional diversity (Furnham et al. 2016).

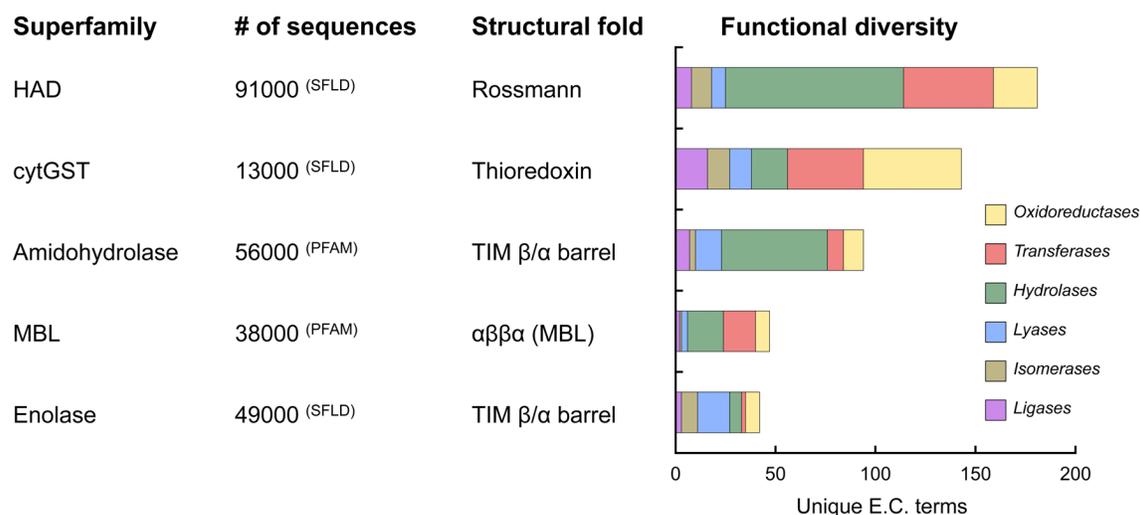


Figure 1.1 Sequence and function diversity within selected enzyme superfamilies.

Sequence diversity and structural fold information was retrieved from the SFLD (Akiva et al. 2014) and PFAM (Punta et al. 2012) databases. Function diversity was retrieved from the CATH database (Sillitoe et al. 2015). Abbreviations of superfamilies: HAD (haloacid dehalogenase), cytGST (cytosolic glutathione transferase) and MBL (metallo- β -lactamase).

1.3 General approaches to infer and characterize the function of enzymes

The diversity of enzyme functions has been uncovered by a vast number of experimental efforts. Historically, the functions of individual enzymes have been discovered via classical genetic and biochemical approaches, *e.g.* genomic context analysis, phenotypic assays that investigate gene knockouts and/or overexpression, function complementation using model organisms, enzymology and structure analysis (Blow 2000). Recently, new experimental platforms have been developed, such as microfluidics, metabolomics, activity-based proteomics and high-throughput phenotype screening, which allow large-scale functional-profiling of enzymes and discovery of new functions (Cravatt et al. 2008; Davids et al. 2013; Prosser et al. 2014; Carpenter & Sabatini 2004). For some of these platforms and in many large-scale functions-profiling studies, the enzymatic activity is assayed based on the turnover of a particular substrate, which is monitored

through a chemical or physical change. For example, Colin *et al.* recently described a high-throughput approach to discover enzymes with relatively weak promiscuous activities from a large metagenomic library using a new sensitive picodroplet approach with fluorescent ‘bait’ substrates (Colin *et al.* 2015). Assaying the turnover of most ‘natural’ or ‘native’ substrates, however, requires analytical approaches such as gas chromatography, high-performance liquid chromatography, mass spectrometry or NMR, but these methods are laborious and often not feasible in high-throughput approaches. Thus, most studies use generic substrates or coupled assays that can be detected through a color or a fluorescence change with a spectrophotometer, which also has the advantage of generally having very low detection limit, a good signal to noise ratio and inexpensive (McCall & Fierke 2000; Goddard & Reymond 2004; Reymond & Wahler 2002; Acker & Auld 2014). Although many generic substrates do not directly represent the enzymes’ native substrate, they can reveal the general chemistry and catalytic specificity of an enzyme, which can then be further investigated using physiological or a more specific set of substrates (Reymond & Wahler 2002; Goddard & Reymond 2004). Most generic substrates have a chromophore and fluorophore leaving groups, such as 6-Bromo-2-naphthol, 2-naphthol, *p*-nitrophenol, fluorescein, coumarin, 6-aminoquinoline, that are attached to various kinds of molecules, representing the native reaction of a large number of enzymes (Goddard & Reymond 2004; Reymond & Wahler 2002). Other assays are based on a coupling reaction of the leaving group or product with a chromophore (McCall & Fierke 2000; Goddard & Reymond 2004; Reymond & Wahler 2002). For example, the enzymatic cleavage of thioesters can be detected through the reduction of dithio-bis-nitrobenzoic acid (DTNB, Ellman reagent) through the formed SH-group (Riddles *et al.* 1983). The release of inorganic phosphate of phosphorylated compounds, such as phospho-sugars, nucleotides, phospholyrated-nucleoside, phosphoamino acids, can be sensed with the highly sensitive Malachite Green reagent in an endpoint assay (Baykov *et al.* 1988).

1.4 Investigating evolutionary relationships within enzyme superfamilies

The classification of enzymes into superfamilies, for which still a common evolutionary origin can be inferred, is mainly based on the structural fold, because of the extensive divergence of sequence and function within enzyme superfamilies (Glasner *et al.* 2006).

Superfamilies are further divided into functional families (or iso-functional subgroups) that perform the same metabolic function (Glasner et al. 2006). For example, enzymes of glyoxalases II family perform the same metabolic function, cleavage of S-D-lactoyl-glutathione into D-lactate and GSH, in most organisms, including bacteria, plants and animals (Suttisansanee & Honek 2011; Limphong et al. 2009; Zang 2000). Generally, enzymes of a functional family are comprised of orthologs, as their genes diverged through speciation from a common ancestral organism (Koonin 2016). Ortholog sequences, however, can have a high sequence diversity, despite their general functional conservation, which is believed to mainly arise from random mutational drift (Wagner 2008). In contrast, genes that diverged through gene duplication are called paralogs. Gene duplication is often associated with functional divergence, thus paralogs often have different metabolic functions (Koonin 2016). Most studies that analyze the relationship among protein sequences in families and superfamilies utilize phylogenetic trees (Dunwell et al. 2001). However, enzyme superfamily, or even functional families, easily exceed many thousands of sequences, and the degree of sequence similarity between enzymes in the same family and superfamily can be extremely low, with only few catalytically important residues conserved, and include large insertions and deletions (Punta et al. 2012; Gerlt, Babbitt, et al. 2011; Dunwell et al. 2001). Because phylogenetic trees are based on multiple sequence alignments, the analysis of large and very diverse sequence sets, such as that of a whole superfamily, can be computationally expensive and difficult for non-experts due its requirement for manual intervention and curation. Recently, an alternative sequence relationship characterization tool has been developed by the Babbitt group, sequence similarity networks (SSNs) (Atkinson et al. 2009; Barber & Babbitt 2012). Similar to phylogenetic trees, SSNs allow the segregation of sequences into clusters and isolation of orthologs and paralogs, but is far less computationally demanding and laborious. SSNs are constructed through independent all-versus-all pairwise sequence comparisons using protein BLAST (Basic Local Alignment Search Tool) and visualized as networks, using programs such as Cytoscape, where nodes represent a sequence and edges denote the pair-wise sequence comparisons using the BLAST *E*-value or alignment score (Altschul et al. 1990; Gerlt et al. 2015). Separation of sequence clusters is achieved by increasing or reducing the threshold at which the

pairwise sequence comparison score is visualized (Gerlt et al. 2015; Atkinson et al. 2009). By lowering the threshold, the nodes (sequences) lose their connectivity through edges and the network becomes segregated into distinct clusters, which can be continued until functional families of orthologs, *i.e.* iso-functional clusters, are separating. Obtaining iso-functional clustering can be difficult, but is facilitated by mapping attributes to each node onto the network, such as known functional information or other sequence features, such as length or organism information, obtained from literature, SwissProt and other curated databases. SSNs gained popularity in the last years and have been employed by numerous studies that investigated sequence-function relationship within protein superfamilies (Song et al. 2008; Gerlt et al. 2015; Brown & Babbitt 2012). In particular, combining sequence, function and structure information, using SSNs, provided insights into enzyme superfamily divergence (Brown & Babbitt 2014). Mapping of functional data onto SSNs also enables the identification and subsequent exploration of uncharacterized enzymes, families and subgroups within superfamilies (Pieper et al. 2009). Collaborative initiatives, such as the Structure Function Linkage Database (SFLD) and Enzyme Function Initiative (EFI), integrate sequence, structure and function information for a number of protein superfamilies using SSNs (Akiva et al. 2014; Gerlt, Allen, et al. 2011). Together, SSNs provide a strong method for characterizing the sequence and function relationships within superfamilies and, by extension, revealing their evolutionary history of functional divergence (Brown & Babbitt 2014).

1.5 The evolution of enzyme functions

In the next section, I will introduce the main scheme of this thesis: the evolution of enzyme functions. I will describe in detail the general concepts and three basic prerequisites that need to be met for the successful evolution of a new enzyme function, as proposed by Khersonsky and Tawfik (Khersonsky & Tawfik 2010). First, a promiscuous activity must provide a fitness advantage to the organism. Second, once under selection, the promiscuous function must be improvable by a few mutations without reducing the native function below a level that affects organismal fitness. Third, evolution must be completed to give rise to two functionally diverged genes (or enzymes), one maintaining the native function and one with the derived function.

1.6 General concept of enzyme evolution

The current theory of how enzymes evolve was mainly established in the 1970's and supported by subsequent work (**Figure 1.2**). In a seminal assay in 1970, John Maynard Smith famously stated that: *“It follows, that if evolution by natural selection is to occur, functional proteins must form a continuous network which can be traversed by unit mutation steps without passing through non-functional intermediates”* (Smith 1970). In other words, the functionality of enzymes must always be maintained, even during evolutionary divergence and of enzyme functions. This is because most mutations are likely to be deleterious and genes without a concrete physiological function might become non-functional. François Jacob further conceptualized this idea and postulated that: *“Evolution does not produce innovations from scratch. It works on what already exists, either transforming a system to give it a new function or combining several systems to produce a more complex one”* (Jacob 1977). In other words, evolution acts as a tinkerer (Jacob 1977). In this view, new enzyme functions generally evolve from a pre-existing enzyme repertoire, because the likelihood that a new gene/protein emerges ‘from scratch’ with the required function serendipitously is unlikely (Renata et al. 2015). Note that recent work suggests that the first primordial proteins emerged through self-assembly of short peptides, which interestingly already can exhibit latent catalytic activities (Carny & Gazit 2005; Rufo et al. 2014). Concurrently with Jacob, Roy Jensen more specifically focused on enzyme and metabolic pathway evolution and proposed that multifunctional enzymes are recruited for the evolution of new functions and metabolic pathways (Jensen 1976). More recent work shows that enzymes are often simply not perfectly specific and, in addition to their native function, exhibit promiscuous activities, which can give rise to multifunctionality (Khersonsky & Tawfik 2010; O'Brien & Herschlag 1999). In this view, upon a change of environmental conditions a previously non-essential promiscuous activity can provide a selective advantage and be crucial for survival. Selection pressure then drives the now essential promiscuous activity towards higher cellular activity levels through gene amplification and adaptive mutations for higher catalytic efficiencies. During this process, gene duplication gives ultimately rise to two functionally diverged enzymes, *i.e.* paralogs. Indeed, this classical view has been supported by several studies that investigated natural enzyme evolution (Voordeckers et al. 2012; Noor et al. 2012;

Mohamed & Hollfelder 2013; Copley 2009; Ngaki et al. 2012; R. Huang et al. 2012). For example, the evolution of xenobiotic degrading enzymes, such as organophosphate hydrolase, as well as atrazine chlorohydrolase, has been shown to arise from latent promiscuous activities (Seffernick et al. 2001; Afriat-Jurnou et al. 2012).

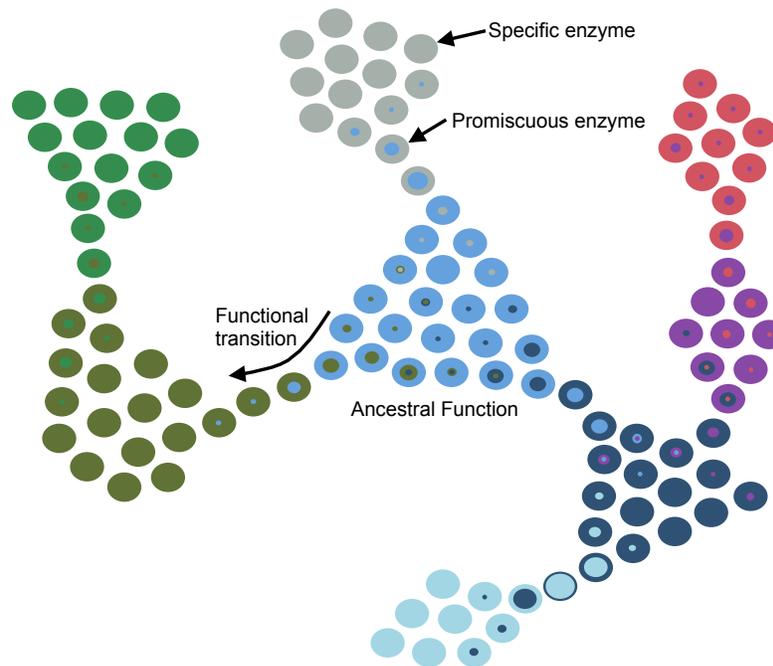


Figure 1.2 Schematic representation of functional divergence within a hypothetical enzyme superfamily.

Circles represent a single sequence (enzyme), and colors represent the native physiological function. The inner circle represents promiscuous activities. The functional divergence from a common ancestor (light blue) occurs via the recruitment of promiscuous activities and evolutionary optimization of these functions to generate new specialized enzymes (gray, deep blue, and olive). During the adaptive process or genetic drift, a new promiscuous function may subsequently arise in a derived family and lead to further expansion of the functional repertoire in the superfamily (green, pale blue, purple, and magenta).

1.7 Prevalence of enzyme promiscuity

Experimental evidence suggests that many, if not most, enzymes turn over non-native substrates (substrate promiscuity) and/or catalyze completely different reactions (catalytic promiscuity) (Hult & Berglund 2007; Pandya et al. 2014; Khersonsky & Tawfik 2010; Nobeli et al. 2009; O'Brien & Herschlag 1999). Soo *et al.* found that overexpression of 115 individual *E.coli* proteins, by screening the ASKA library (An open reading frame

collection of almost all *E. coli* K-12 genes (Kitagawa et al. 2005)), can provide *E. coli* cells the capacity to survive and respond to a large variety of (86 of 237) toxin-containing environments (Soo et al. 2011). Interestingly, further characterization of the hits revealed that in 15 of 115 cases (13%) enzyme promiscuity was responsible for the resistance. Cross-wise promiscuity, *i.e.* the native function of one enzyme is the promiscuous activity of another and vice versa, has also been observed among evolutionary related enzymes of a single superfamily. For example, enzymes of amidohydrolase superfamily share lactonase, arylesterase and phosphotriesterase activities as native and promiscuous activities, despite their high sequence divergence (Roodveldt & Tawfik 2005). A recent large-scale function profiling study assayed over 200 HAD superfamily enzymes against 167 phosphatase (98%) and phosphonate (2%) substrates and revealed a high degree of promiscuity: 75% of the enzymes turned over at least 5 different substrates (H. Huang et al. 2015). Crosswise promiscuity between phosphatases, phosphodiesterases, phosphonates and arylsulfatases of the alkaline phosphatase superfamily has also been demonstrated (Mohamed & Hollfelder 2013). Although catalytic efficiencies of promiscuous activities are generally substantially lower compared to native activities, their rate accelerations compared to the uncatalyzed reaction are still remarkable (O'Brien & Herschlag 1999). For example, a promiscuous metal-dependent enzyme from *Burkholderia caryophylli* is capable of hydrolyzing phosphomonoesters, phosphodiesteres, phosphotriesters, phosphonate monoesters, sulfate monoesters and sulfonate monoesters with rate accelerations ranging from 10^7 to as high as 10^{19} , compared to the uncatalyzed reactions (van Loo et al. 2010).

1.8 Molecular basis of enzyme promiscuity

What are the structural and mechanistic causes for enzyme promiscuity? Available mechanistic and structural studies of promiscuous activities suggest that active site features essential for the native function, or a subset of them, are coopted for promiscuous activities. The catalytic machineries of enzymes, such as nucleophilic residues, catalytic triads, metal ions and organic cofactors, provide an intrinsic reactivity to active sites that results in non-specific catalysis of other substrates and reactions (Khersonsky et al. 2006; Khersonsky & Tawfik 2010; Babbitt et al. 2010). For example, the protease chymotrypsin uses a highly reactive Ser-His-Asp catalytic triad for catalysis

of its native function as well as for promiscuous hydrolysis of amide, ester and phosphotriester compounds (O'Brien & Herschlag 1999). Furthermore, several metalloenzymes are promiscuous due to the inherent reactivity of divalent metal ions (O'Brien & Herschlag 1999; Khersonsky & Tawfik 2010; Babbitt et al. 2010; van Loo et al. 2010). As an example, several non-related lactonases (tetrahedral TS; C-O bond cleavage) promiscuously catalyze the phosphotriesterase reaction (pentavalent TS; P-O bond cleavage) and in both cases divalent metal ions bind the substrates, stabilize the negatively charged TS and activate a hydroxide ion for nucleophilic attack (Elias & Tawfik 2011). In addition, a few anecdotal cases describe that introducing different active site cofactors, such as different divalent metal ions, results in new promiscuous activities. For example, in the presence of Mg^{2+} the dihydroxyacetone kinase (DHAK) from *Citrobacter freundii* catalyzes the transfer of the γ -phosphate of ATP to dihydroxyacetone, but in the presence of Mn^{2+} the enzyme exhibits cyclase activity towards flavin adenine dinucleotide (FAD) (Sánchez-Moreno et al. 2009). Besides the intrinsic reactivity of catalytic machineries, the shape and hydrophobicity, or polarity, of enzyme active sites also determines the degree of enzyme promiscuity. Hydrophobic active sites are simply less exclusive for hydrophobic substrates, whereas more polar active sites can be more exclusive, because substrate binding depends on charge complementarity (Estell et al. 1986; Nobeli et al. 2009; Khersonsky & Tawfik 2010). Steric hindrance can simply exclude bulkier substrates, whereas large active sites can accommodate more substrates, but may yield many unproductive binding modes (Babbitt et al. 2010; Tokuriki & Tawfik 2009b). This can be best explained by more optimal structural enzyme-substrate complementarity, which yields more efficient catalysis (Benkovic & Hammes-Schiffer 2003). For example, structure analysis of the bacterial PTE with various substrates bound revealed a high active site complementarity for its native substrate paraoxon (Jackson et al. 2008; Jackson et al. 2009). In contrast, hydrolysis of phosphodiester and arylesters is also catalyzed by PTE, but at a substantially lower rate, because less interactions in the active site results in multiple productive and unproductive binding modes (Jackson et al. 2009). This concept is further supported by the directed evolution of PTE towards higher arylesterase (AE) activity, which improved the promiscuous activity by 10^5 -fold by reshaping the active site for

better complementarity and stabilization of the 2-naphthyl hexanoate substrate (Tokuriki et al. 2012). The flexibility and conformational diversity of enzyme active sites has also recently been linked to enzyme promiscuity (Tomatis et al. 2008; Zou et al. 2015; Campbell et al. 2016; Tokuriki & Tawfik 2009b). Flexible structural parts of proteins, such as active site loops and residues, sample between different conformers (structural positions), which is important for various steps of a catalytic cycle, such as substrate binding, TS stabilization and product release (Hammes-Schiffer & Benkovic 2006). Protein dynamics and motions are usually optimized for the native function, but promiscuous activities could stem from flexibility and specific conformers (Khersonsky & Tawfik 2010; Tokuriki & Tawfik 2009b). For example, resurrected Cambrian class A β -lactamases exhibit higher conformational dynamics and are more promiscuous, *i.e.* hydrolyzing a larger variety of β -lactam antibiotics, than more modern class A β -lactamases, which have more rigid active site regions and are more specific to certain antibiotics (Zou et al. 2015; Risso et al. 2013). Thus, mutations that increase flexibility of certain active site regions, and decrease other parts, can enrich and stabilize less populated conformational states which could improve promiscuous activities (Tokuriki & Tawfik 2009b; Khersonsky & Tawfik 2010). This has recently been shown through detailed structural analysis of evolutionary intermediates of the laboratory evolution of PTE towards AE activity (Campbell et al. 2016; Tokuriki et al. 2012). The initial mutation, H254R, is highly beneficial for AE activity, but its catalytic potential required stabilization of a productive rotamer and changes in mobility of active sites loops (Campbell et al. 2016). The evolutionary intermediate, which is bifunctional and highly active for PTE and AE activity, exhibited increased flexibility in active site loops, compared to both specialized enzymes. Subsequent mutations then stabilized loop regions to cancel out unproductive protein dynamics, which further improved AE activity (Campbell et al. 2016). The authors revealed the changes in dynamics occurred in a sequential order, because simultaneous destabilization/stabilization would be detrimental to protein function and stability (Campbell et al. 2016).

1.9 Promiscuous activities as evolutionary starting points

In the previous section I described many examples of enzyme promiscuity. However, what determines if a promiscuous enzyme can be a potential evolutionary starting point?

One important factor is the cellular activity level of a promiscuous activity and its contribution to organismal fitness (Soskine & Tawfik 2010). O'Brien and Herschlag proposed a threshold model in which the promiscuous activity needs to be above a certain cellular level to provide a selective advantage to the organism (**Figure 1.3**) (O'Brien & Herschlag 1999). This for example is demonstrated by the *E. coli* gamma-glutamyl phosphate reductase (ProA), which exhibits a weak (k_{cat}/K_M of $0.4 \text{ M}^{-1}\text{s}^{-1}$) promiscuous N-acetylglutamylphosphate reductase (ArgC) activity (McLoughlin & Copley 2008). The ArgC function *E. coli* is essential to produce arginine and a knockout of ArgC is lethal. However, ProA is not able to compensate an ArgC knockout and restore survival. Introduction of a single mutation into ProA, E383A, that improved ProA's ArgC catalytic efficiency by ~12-fold (k_{cat}/K_M of $4.6 \text{ M}^{-1}\text{s}^{-1}$) eventually allows ArgC deficient *E. coli* strains to grow. Interestingly, the cellular ArgC activity of ProA-E383A was further enhanced by increased expression levels due to amino acid starvation, because the native activity if ProA-E383A decreased by 2,800-fold. Thus, activity levels and the amount of enzyme in the cell are crucial for a promiscuous activity to be beneficial for an organism. However, the threshold where an activity becomes physiologically relevant can be very different for each function and depend on various factors, such as the activity level required in the cell, strength of the selection pressure imposed by the environment and protein expression or cellular location (McLoughlin & Copley 2008; O'Brien & Herschlag 1999; Soskine & Tawfik 2010). Indeed, several studies have recently shown that expression and localization were co-optimized with function during the course of evolution in order to ensure that each enzyme appears at the right time and space within the cell (Ngaki et al. 2012; Pougach et al. 2014; Voordeckers et al. 2015). Conclusively, the *in vitro* detection of a promiscuous activity and its level of catalytic efficiency might not necessarily indicate that the activity is relevant *in vivo*, and thus constitute a potentially good evolutionary starting point. The activity level of a promiscuous activity can also vary among orthologous enzymes with the same native function, due to sequence variation that is essentially neutral for the native function (Wagner 2008; Paaby & Rockman 2014; Masel & Trotter 2010; Amitai et al. 2007; Bloom et al. 2007). For example, Khanal *et. al.* investigated ProA orthologous from nine different bacterial strains and revealed that their promiscuous ArgC activity levels varies by about 50-fold

(Khanal et al. 2015). Moreover, two laboratory evolution studies subjected enzymes to “neutral drift” (accumulation of mutations under a purifying selection pressure for the native function), which altered their function profiles and introduced new promiscuous activities (Bloom et al. 2007; Amitai et al. 2007). Therefore, some enzymes might not be “evolvable”, because their promiscuous activity level is below the required threshold, but neutral genetic variation might allow enzymes to acquire new or improved promiscuous activities by chance (**Figure 1.3**) (Wagner 2008; Paaby & Rockman 2014; Masel & Trotter 2010).

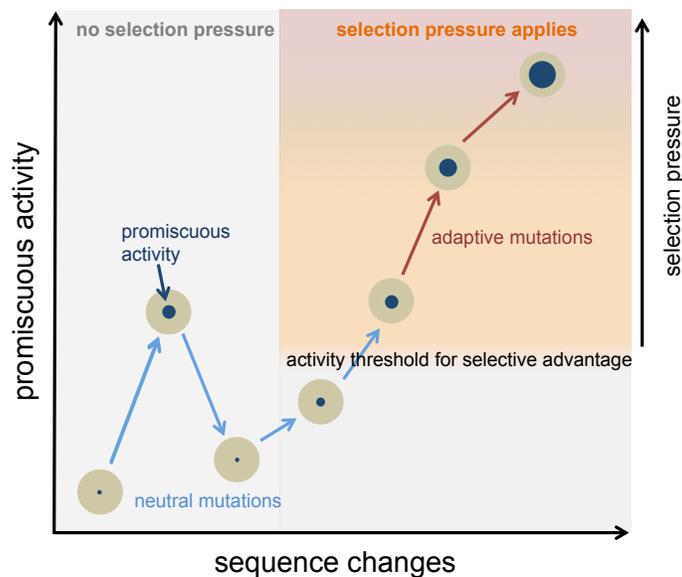


Figure 1.3 Simplified threshold model for promiscuous activities as evolutionary starting points.

Many enzymes (beige circle) exhibit promiscuous activities (inner dark blue circle), however, without any selection there is no pressure to maintain a promiscuous activity and neutral mutations push the activity above or below an activity threshold that would be required for a selective advantage. Once new selection pressure emerges (e.g. upon environmental change) the promiscuous activity would be physiologically relevant, but only advantageous when neutral mutations improved the activity above the threshold. Increase of the selection pressure leads to subsequent adaptive mutations that further improve the activity.

1.10 Evolvability of promiscuous activities

In a scenario where a promiscuous activity of an enzyme has become physiologically relevant, can it readily evolve and give rise to an enzyme with a new efficient function? In other words, what constrains the evolvability of a promiscuous activity and enzyme? I use the term evolvability here as “*the ability of a protein to adapt in response to mutation and selective pressure*” (Romero & Arnold 2009). Evolvability is crucial factor because once a promiscuous activity contributes to organismal fitness, natural selection will put pressure towards higher activity to increase organismal fitness. Several studies have shown that increases in gene and protein dosage can initially provide higher cellular activity level, but eventually higher catalytic efficiency will be indispensable, because of the cost associated with higher protein production (Sandegren & Andersson 2009). Thus, latent promiscuous activities must be evolvable through mutations, in particular because promiscuous activities are often initially very low compared to the catalytic efficiencies of native activities ($>10^3$ in k_{cat}/K_M), and thus need to be optimized to improve organismal fitness (Khersonsky & Tawfik 2010; Bar-Even et al. 2011). However, this appears not to be trivial, as suggested by many laboratory evolution studies that aimed to improve enzymes function, but do not reach the catalytic efficiency of natural enzymes (Khersonsky & Tawfik 2010; Tracewell & Arnold 2009). One reason is that beneficial mutations are very rare and quickly exhausted during adaptive evolution. It is estimated that only a small fraction (0.5– 0.01%) of random mutations improve function, whereas many are strongly deleterious (30–50%) and most comparatively neutral 50–70% on function (Romero & Arnold 2009). Additionally, many recent studies revealed that the functional effect of mutations can also be context dependent, which is described as mutational epistasis (non-additive interactions between mutations, explained in detail in the next section and **Figure 1.4**) (Starr & Thornton 2016). Indeed, mutations can have pleiotropic effects, which means they also affect other protein properties, such stability, folding or the native function. Thus, although a mutation might improve one function, its pleiotropic effects on other properties could severely impair protein and organismal fitness. In the next sections, I will describe in detail how epistasis, protein stability and functional trade-offs affect the evolvability of enzymes.

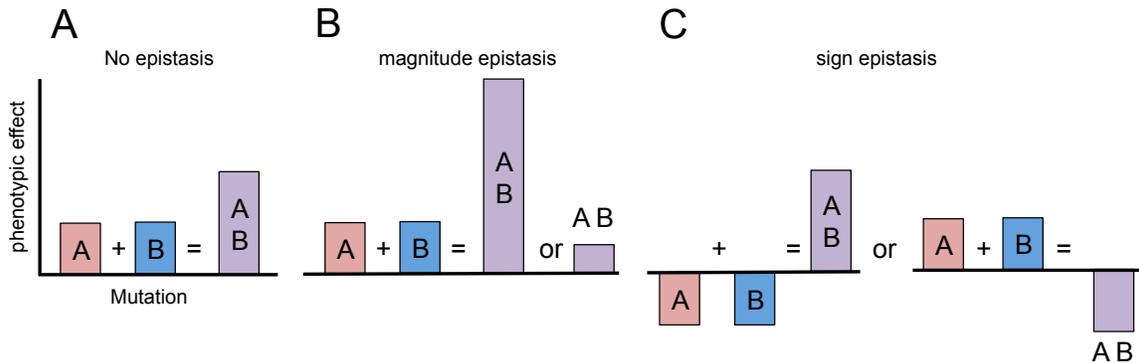


Figure 1.4 General concept of mutational epistasis in proteins.

(A) No epistasis is observed when the phenotypic effect of the double mutation AB is the sum of the individual mutations A and B, i.e. the phenotypic effect is additive. (B) When the phenotypic effect of the double mutant AB is larger or smaller than the sum of the individual mutations A and B, it is called *magnitude epistasis*. (C) In cases where the effect of individual mutations is completely inverted in the double mutant, from positive to negative or vice versa, it is called *sign epistasis*. Concept of the figure adapted from (Kaltenbach & Tokuriki 2014).

1.11 Epistasis in enzyme evolution

Enzymes adapt towards new functions through the accumulation of beneficial mutations. Natural and laboratory trajectories of enzyme evolution show simple step by step fitness improvements, which suggest that evolution is fairly deterministic and the fitness effect of mutations additive (Tracewell & Arnold 2009). However, a more detailed analysis of beneficial mutations revealed that their phenotypic is often non-additive and also depends the proteins' genetic background (Reetz 2013). This phenomenon generally describes the concept of non-additivity in biology and evolution, which is also called epistasis (Starr & Thornton 2016; P. C. Phillips 2008). The term epistasis stems from genetic studies and originally described the non-additive phenotypic outcomes of interactions between genes or alleles, but has now been adapted to many biological systems (P. C. Phillips 2008). In proteins, epistasis generally describes the non-additive effect between two or more mutations on the phenotype, such as stability or enzymatic activity, and its basic terms

are described in **Figure 1.4** (Kaltenbach & Tokuriki 2014).

Within the last decade many studies have described and discussed the importance of epistasis for molecular and protein evolution (Breen et al. 2012; Starr & Thornton 2016; P. C. Phillips 2008; Lunzer et al. 2010). Several studies revealed that epistasis restricts the accessibility of beneficial mutations, and thus constrains evolutionary trajectories. A seminal work by Weinreich et al. investigated the natural evolutionary trajectory of the TEM-1 β -lactamase towards the third-generation antibiotic cephalexin involving five mutations by generating all 120 possible mutational combinations (Weinreich 2006). Although all five mutations together lead to a 100,000 increase in resistance, only 18 of 120 possible trajectories provide a constant increase without passing through functionally impaired intermediates. Similarly, Yokoyama et al. investigated the evolution of blue-sensitivity color vision in humans from UV-sensitivity and showed that the seven-mutation trajectory was highly constrained by epistasis, 4,008 of 5,040 trajectories are impassible (Yokoyama et al. 2014). A more comprehensive analysis of several previously described evolutionary trajectories by Miton and Tokuriki revealed that most are characterized by positive epistasis, in which later mutations only become beneficial because of earlier ones, and diminishing returns epistasis, in which initial mutation provide higher fitness improvements than later one (Miton & Tokuriki 2016).

The inaccessibility of beneficial mutations cannot only constrain, but also redirect evolutionary trajectories and lead to suboptimal fitness outcomes. Salverda *et al.* investigated the repeatability of the above-described TEM-1 evolution towards cephalexin in the laboratory by performing twelve independent directed evolution experiments (Salverda et al. 2011). Although seven trajectories were similar in mutation and phenotype to the natural evolution, five lines adopted lower suboptimal outcomes, due to epistatic consequences of alternative beneficial mutations. These alternative initial mutations, which occurred stochastically, prevented the occurrence of more beneficial mutations that were obtained in more successful trajectories, and directed the evolution towards suboptimal outcomes. Similarly, neutral mutations, which occur completely stochastic, can also positively and negatively affect the evolution of a new function by introducing “epistatic ratchets”, which alter the effect of subsequent adaptive mutations and therefore the evolvability of enzymes (Kaltenbach & Tokuriki 2014). For example,

detailed analysis of the glucocorticoid receptor divergence from aldosterone to cortisol specificity was dependent on two permissive mutations, which by themselves had no apparent effect on function, but their occurrence was essential for the beneficial effect of subsequent functional mutations (Bridgham et al. 2009). Harms and Thornton further showed that such permissive mutations can be extremely rare (Harms & Thornton 2014). The authors screened a library of several thousand variants, with the functional mutations already introduced, in search of alternative permissive mutations that allowed a functional switch of the glucocorticoid receptor. Although three alternative mutations were found, none of them could have occurred during the natural evolution, because they impair the receptor's ancestral function. The results of these studies highlight the fact that epistatic interactions can be pervasive and severely affect function, which in turn makes adaptive evolution often highly constrained and unpredictable (Miton & Tokuriki 2016; Reetz 2013).

What is the molecular cause of epistasis? Epistatic effects on function are often realized by altering the position of residues that directly interact with the substrate. For example, in the laboratory evolution of the bacterial PTE towards arylesterase the initial mutation, H254R, generated a stabilizing interaction with the 2-naphthol leaving group of the substrate (Tokuriki et al. 2012). Subsequent mutations reinforced the productive rotamer position of H254R, which initially alternated between blocking the active site and interacting with the substrate. However, without H254R, the subsequent reinforcing mutations provide no functional support. In other cases, functional epistasis is caused through changes in position and flexibility of structural elements. For example, the evolution of *Bacillus cereus* β -lactamase II towards cephalixin resistance is accomplished by shift of one of the catalytic Zn^{2+} metal ions through the initial mutation G262S (Tomatis et al. 2008). A later mutation, N70S, by itself is deleterious, but together with G262S becomes beneficial through improving loop flexibility around the active site. Overall, structural and biophysical analysis of evolutionary trajectories revealed that epistasis is caused by interactions between mutations and other residues, substrates or cofactors, which can either be direct or indirect, through other residues or structural elements. In many cases epistatic interaction between mutations are caused by deleterious effects on the protein structure and folding, which will be described separately in the

following sections.

1.12 The role of protein stability in enzyme evolution

The functionality of an enzyme (and any protein) in the cell depends on its ability to fold into and maintain a stable structure. The relationship between function and stability is often referred to as protein fitness (W_p), which is proportional to catalytic activity (f) and concentration of functional protein ($[E]_0$) in the cell, as described in the simple equation:

$$W_p = [E]_0 \times f$$

Thus, if a protein is not folding properly and maintaining a stable structure the concentration of functional protein ($[E]_0$) will be too low to maintain sufficient protein fitness, which in some cases can be similar to organismal fitness, as for example in the case of antibiotic resistance enzymes (Soskine & Tawfik 2010). The role of protein stability in evolution is often been described as a threshold model (Tokuriki & Tawfik 2009c). Proteins can to some extent buffer destabilizing mutations and stay within a neutral range that grants proper structure integrity, which supports functionality, above and below a threshold. However, severely destabilizing mutations, or the accumulation thereof, can result in surpassing the lower threshold and consequently lead to loss of stability and protein fitness (Tokuriki & Tawfik 2009c). The active sites and catalytic machineries of enzymes are generally structurally very unfavorable and destabilizing, but required for function. Active sites loops must be flexible to allow substrate binding and release (Tokuriki & Tawfik 2009b; Teilum et al. 2011) Enzyme active sites also often contain catalytically important ionizable and charged residues that are shielded from water and ionic charge by hydrophobic residues, which create thermodynamically highly unfavorable conformations and protein folding issues (Tokuriki & Tawfik 2009c; Teilum et al. 2011). Therefore, mutations that improve function are destabilizing, although the intrinsic robustness of proteins usually buffers some destabilizing effects (Tokuriki et al. 2008). Furthermore, chaperones can also mask the deleterious effect of mutations by assisting protein folding (Tokuriki & Tawfik 2009a). Protein stability is divided into thermodynamic stability, the stability of the folded protein, and kinetic stability, which relates to protein folding and the activation energy barrier between folded, misfolded and unfolded states (Sanchez-Ruiz 2010).

1.12.1 Thermodynamic stability and enzyme evolution

Thermodynamic stability (ΔG) relates to the two-state equilibrium between unfolded or partially unfolded (U) with the properly folded and functional protein (N). ΔG essentially describes the Gibbs free energy (G) in kcal per mol difference between U and N . Most proteins seem to be only marginally stable with ΔG in the range of -3 to -10 kcal mol⁻¹ (DePristo et al. 2005). To put in relation, a single hydrogen bond has an energy of 2-10 kcal mol⁻¹, which means that stability changes ($\Delta\Delta G$) of even a single mutations can impair stability and proper folding (DePristo et al. 2005). However, due to the much simpler experimental measurement many studies use the heat denaturation temperature (T_m), at which the protein unfolds, as a proxy for thermodynamic stability, which generally correlate relatively well (Rees & Robertson 2001). Note that the T_m of proteins is usually shifted many degrees above the host organisms' environmental temperature (Razvi & Scholtz 2006). As described above, the relationship between protein stability and protein fitness, W_p , is often described with a threshold model (Tokuriki & Tawfik 2009c). If the stability of a protein, ΔG , is above a certain threshold level, most of the protein is properly folded inside the cell and its fitness is not impaired. However, if the stability falls below the threshold, e.g. through destabilizing mutations, most of the protein will be non-functional because of unfolding, aggregation or degradation. Thus, as long as the protein remains within the neutral stability range, protein fitness, W_p , is not affected. Interestingly, the threshold for protein fitness seems to be uncorrelated with the initial stability and is exceeded by as little as 1–3 kcal/mol for most proteins. Some studies have shown that higher initial stability can buffer the accumulation of a few destabilizing mutations, or specifically stabilize certain structural parts, and therefore facilitate enzyme evolution (Tokuriki & Tawfik 2009c). For example, a study by Bloom et al. demonstrated that a more thermostable version (T_m of 62 °C) of a cytochrome P450 enzyme exhibits more and improved catalytic activities after introducing random mutations compared to a less thermostable version of the same protein (T_m of 47 °C) (Bloom et al. 2006). Other studies demonstrated the importance of compensatory mutations and the stabilization of local structural parts for function. For example, Bloom *et al.* demonstrated that the resistance of the human N1 influenza virus to oseltamivir

through a single point mutation (H274Y) could only occur, because two previously occurring stabilizing mutations compensated its negative effect on protein stability (Bloom et al. 2010), i.e. without the two compensatory mutations the resistance mutation H274Y would not have provided a selective advantage. Overall, the maintenance of thermodynamic stability is strong requirement during protein evolution and has been incorporated into protein engineering and directed evolution schemes to improve enzyme function (Socha & Tokuriki 2013; Tokuriki & Tawfik 2009c; Tokuriki et al. 2008).

1.12.2 Kinetic stability and enzyme evolution

Kinetic stability relates to folding and unfolding properties and the energy barrier separating the folded and unfolded state of a protein (Sanchez-Ruiz 2010). As described above, thermodynamic stability assumes that the folded and unfolded or misfolded states are in equilibrium, however, most proteins, besides small fast folding proteins, do not re-fold once unfolded and thus stability of these proteins is kinetically controlled rather than by their thermodynamic stability (Sanchez-Ruiz 2010). Limited kinetic stability is often the cause of protein misfolding and aggregation and is observed in various human diseases, such as Parkinson, Alzheimer and prion diseases (Stefani 2004; Sanchez-Ruiz 2010). However, kinetic stability has received much less attention compared to thermodynamic stability, potentially because experimental evidence is more difficult to obtain, but its role for protein evolution might be equally, or even more, important. For example, the laboratory evolution of PTE towards arylesterase activity is not constrained by thermodynamic stability (T_m of $>70^\circ\text{C}$ for all variants), but through the aggregation of a folding intermediate (Wyganowski et al. 2013). Co-expression of GroEL/ES chaperone during the trajectory assisted in stabilizing, preventing aggregation and refolding of the unstable folding intermediate, whereas temporal removal of the chaperone buffering supported the occurrence of compensatory mutations that specifically stabilize the folding intermediate and allow further functional, but destabilizing, mutations to occur. Furthermore, a recent study that resurrected ancestral several bacterial RNAses H revealed that kinetic stability increased over time by decreasing unfolding rates of a folding intermediate, and thus preventing potential aggregation and misfolding (S. A. Lim et al. 2016). Conclusively, in these cases, selection pressure specifically optimized

and maintained kinetic stability, and not thermodynamic stability. Thus, the threshold requirements are different for thermodynamic and kinetic stability (Sanchez-Ruiz 2010). Kinetic stability needs to be maintained to ensure efficient folding by avoiding any folding intermediates that could lead to aggregation or misfolding and by ensuring a high-energy barrier between folded, intermediate and unfolded states (S. A. Lim et al. 2016). Overall, the evolution of new enzymes functions requires the maintenance of kinetic and thermodynamic stability above a certain threshold to support protein and organismal fitness, which however can restrict the evolution of enzyme functions.

1.13 Functional trade-offs between native and promiscuous functions

During the evolution of a new function the promiscuous activity improves through adaptive mutations, however, at the same time the native function of the enzyme is often still contributing to organismal fitness. Thus, any adaptive mutations that trade-off negatively with the native function will affect organismal fitness, and consequently lead to an evolutionary dead end (Soskine & Tawfik 2010). Strong negative trade-offs between native and new functions have been observed in laboratory and natural evolution examples. For example, in the above-described example of the *E. coli* gamma-glutamyl phosphate reductase (ProA), the mutation E383A improved its promiscuous N-acetylglutamylphosphate reductase (ArgC) activity by ~12-fold (McLoughlin & Copley 2008). However, at the same time the native activity decreased by 2,800-fold, which was partially compensated by increased expression levels due to amino acid starvation (McLoughlin & Copley 2008). Another drastic example is the natural evolution of the xenobiotic degrading enzyme atrazine chlorohydrolase (AtzA) from the melamine deaminase (TriA). Although the evolution of AtzA from TriA involves nine mutations, TriA activity is already completely abolished after only two mutations, while AtzA activity increased by 1,700-fold (Noor et al. 2012). Mechanistically, functional trade-offs between the two functions arise through different catalytic as well as substrate binding requirements in the enzymes' active site, which can be best explained by optimal structural enzyme-substrate complementarity that yields efficient catalysis (Benkovic & Hammes-Schiffer 2003). Thus, mutations, which optimize one function, may simply lead to decreased or catalytically less optimal binding and consequently impair the other

function. Other laboratory evolution examples, however, show that trade-offs are initially often very weak, with large improvements of promiscuous activities and only weak decreases of native functions (Aharoni et al. 2005). This is suggested to result from the intrinsic flexibility of proteins and their mutational robustness (Soskine & Tawfik 2010). Recently, Kaltenbach and Tokuriki updated this view and proposed that initially weak trade-offs, in these cases, are a result of artificially high selection pressure for the new function and the short evolutionary timescales, which might prevent the accumulation of deleterious mutations for the original function (Kaltenbach et al. 2016). Overall, functional trade-offs between two conflicting functions within in single enzyme might constrain the evolvability of enzymes. Note that in some cases generalistic enzymes, which exhibit high native and new activities, can also emerge as evolutionary intermediates. Nevertheless, at one point during the adaptive process, gene duplication will eventually provide a solution to the adaptive conflict or even occur before any selection applies, which gave rise to a variety of models implicating gene duplication and enzyme evolution (Innan & Kondrashov 2010). Briefly, the main models discuss if a single gene initially serves two distinct functions before gene duplication (*subfunctionalization* (Tocchini-Valentini et al. 2005)) or if gene duplication occurs first and one copy subsequently adopts a new function (*neofunctionalization* (Boucher et al. 2014)). In addition, several more detailed models have been described that distinguish between the emergence, maintenance and evolution of duplicated genes and new functions (Innan & Kondrashov 2010; Rauwerdink et al. 2016).

1.14 Directed evolution as a tool to investigate fundamentals of enzyme evolution

The first *in vitro* evolution experiment was performed on RNA molecules in the 1960's by Spiegelman and coworkers, essentially to understand what happens if “self-replicating” molecules evolve outside of biological constraints (Mills et al. 1967). With the advent of PCR, *in vitro* evolution became more popular in the 1990's and eventually was used for the directed evolution (DE) of proteins to improve function and biophysical properties for industrial and biotechnological purposes (Davids et al. 2013; Currin et al. 2015). DE mimics Darwinian evolution in a sense that a pool of random mutants of a given gene are screened or selected for a given property that improve fitness, which can

be activity for an enzymatic function or protein thermostability. ‘Screening’ means that the phenotype (parameter of selection) of each mutant in the library is measured individually and a threshold at which variants are taken to the next round can be custom-defined, which is often the most improved variant. In contrast, ‘selection’ means the screening for a phenotype is directly linked to organismal survival, *e.g.* antibiotic resistance protein or production of an essential metabolite, and circumvents measuring each mutant. This allows for much higher throughput and also directly purges out any non-functional or less fit variants (Packer & D. R. Liu 2015). In both cases, the process of mutational diversification and screening or selection can be repeated essentially indefinitely until an enzyme variant with the desired function and biophysical properties is obtained. Generally, the number of variants that can be screened depends on assay system and phenotypic readout, such as colorimetric enzyme activity assay, whereas for selection, such as antibiotic resistance, the number is significantly higher. Similar to natural evolution in an organism, the evolving activity initial needs to be above a threshold to provide a fitness advantage, which for directed evolution generally is the detection limited in the screening or survival in a selection based system. Many different DE selection and screening systems have been established that fit the need for a large diversity of enzyme functions and properties (Leemhuis et al. 2009; Packer & D. R. Liu 2015). Many examples of DE demonstrated that enzymes are often readily evolvable in the laboratory, and even allows the evolution of enzymes that catalyze non-natural reactions (Renata et al. 2015). However, many DE experiments only yield enzymes variants that are far from the catalytic efficiency of natural enzymes, which implies that the evolution of enzymes is constrained by factors such as stability and epistasis (Kaltenbach & Tokuriki 2014). Researchers, who are interested in fundamental questions of enzyme evolution and function, have discovered that DE provides a powerful experimental set up to disconnect a protein from its natural context and to explore the constraints that restrict functional adaptation on the molecular level without the biological noise of unrelated mutations, environmental fluctuations, competing organisms, etc. (Romero & Arnold 2009). Indeed, many of the above described constraints have been discovered by DE of enzymes and proteins (Romero & Arnold 2009). DE allows for a confined and controlled experimental set up with conditions and

parameters, such as mutation rate, the strength of the selection pressure, threshold of enzyme activity and fitness and selection of the organism, well defined by the researcher. Furthermore, whereas in natural evolution only the successful enzyme variant is conserved, in DE alternative or less successful solutions can also be explored and subsequently all evolutionary intermediates are available for detailed biochemical, biophysical and structural analysis. Identified mutations in a DE experiment are easily classified as beneficial, neutral or deleterious for a particular enzyme function or property and can be subsequently introduced in different combinations or variants to understand their phenotypic effect (Yuen & D. R. Liu 2007). On the other hand, DE also has its limitations in recapitulating natural evolution in the laboratory. For example, mutations that are beneficial under the defined DE conditions might alter other protein properties, such as stability, degradation, codon usage, new undesired promiscuous activities and many others, which could affect organismal fitness outside the laboratory. Thus, results of DE experiments cannot always be generalized and transferred to natural evolution, because enzymes evolve in nature under more pleiotropic constraints than present in the laboratory (Romero & Arnold 2009; Soskine & Tawfik 2010). Another major constraint of most DE experiments is the unusually high selection pressure with only the most fit variant propagated to the next generation and most other genetic variation is purged out. In contrast, in natural evolution usually a population of genetically distinct variants is maintained until a highly beneficial variant is fixed in the population (Lang & Desai 2014). Therefore, DE is a valuable tool to understand the basic principles of protein evolution and the functional effect of mutations, but has its limitations in population genetics and how mutations would eventually become fixated within a population.

1.15 The metallo- β -lactamase superfamily

The metallo- β -lactamase (MBL) superfamily represents a textbook example of a functionally diverged superfamily and will serve as my model system to study how catalytic functions evolve and diverge through promiscuous enzymes. Members of the MBL superfamily are substantially diverged in sequence as well as function (Bebrone 2007). Approximately 34,000 sequences are registered in the protein family database (Pfam) (Punta et al. 2012). The amino acid sequence identity between members can be

less than 5%, but members share structural features such as the $\alpha\beta\alpha$ -fold (MBL fold), and a mono- or bi-nuclear active site centre with a generally conserved metal binding motif (H-X-H-X-D-H) (Bebrone 2007). The first site (M1) is coordinated by three His residues; the second site (M2) consists of His and Asp residues, in addition to a bridging Asp residue that coordinates both metals. B1 and B3 β -lactamases lack the bridging Asp residue, which is substituted by a Cys or Ser (noncoordinating) residue, respectively (Bebrone 2007). The two active site metals have two essential roles during catalysis that are similar in all hydrolytic MBL enzymes: (i) activate a hydroxide ion for nucleophilic attack and (ii) stabilize the ground and transition states as Lewis acids (Karsisiotis et al. 2014). To date, at least 24 distinct functional families have been identified within the MBL superfamily, including DNA, RNA and nucleotide processing, detoxification, antibiotic resistance, quorum-quenching, and pesticide hydrolysis (Daiyasu et al. 2001; Bebrone 2007; Pettinati et al. 2016). In 1999, Aravind classified the known MBL functional families into groups based on their sequence relationships and described functions revealed that most nucleic acid related functions had a single origin, whereas other functions, such as sulfatase and β -lactamase, evolved twice independently within the MBL superfamily. Most of MBL functions involve hydrolytic reactions and target diverse substrates with different chemical properties such as phosphodiester, phosphotriester, choline-phosphoester, thiol-ester, sulphate-ester, carbon-ester and β -lactam bond (Daiyasu et al. 2001; Bebrone 2007; Pettinati et al. 2016). Other functions involve non-hydrolytic reactions such as nitric-oxidoreduction (Silaghi-Dumitrescu et al. 2005) and sulphur dioxygenation (Holdorf et al. 2012), as well as non-enzymatic functions such as binding and transport (Bebrone 2007; Puehringer et al. 2008). Despite the functional diversity, many of the hydrolytic functions are amenable to a simple characterisation by colorimetric assays using *p*-nitrophenol based compounds, Ellman's reagent or pH indicator assays (Bebrone et al. 2001; Hagelueken et al. 2006; Dong et al. 2005; Campos-Bermudez et al. 2007), which makes the MBL enzymes particularly well suited for function-profiling analysis. Thus, the MBL superfamily provides a good model system to study enzyme promiscuity and how distinct functions evolved within a structural fold.

1.16 Aims and scope of the dissertation

The overall aim of my research is to investigate constraints of enzyme evolution, using enzymes and functions of the MBL superfamily as a model system. In particular, I am investigating how enzyme promiscuity facilitates functional divergence by providing evolutionary starting points that subsequently can be evolved through adaptive mutations. I hypothesize that many enzymes will be promiscuous, and provide potential starting points, although it remains elusive how promiscuity will connect the distinct functions within entire superfamilies. However, I speculate that potentially not all functional connections through promiscuity are equally evolvable, and thus some starting points might be better than others. Each chapter addresses the aim and hypothesis from a different perspective:

In chapter two, I will describe the evolutionary relationship between functional families of the MBL superfamily in order to shed light on their evolutionary history. In addition, function-profiling analysis of 24 selected enzymes against 10 distinct MBL functions reveals promiscuity of MBL enzymes. Together, the analysis provides unprecedented insight into how functional families in the MBL superfamily are evolutionarily and functionally connected.

In chapter three, I present an investigation into how promiscuous metal ion binding *in-vitro* and in *E.coli* cells alters the scope and level of catalytic promiscuity of MBL enzymes. Further analysis reveals that the enzymes can exist as an ensemble of metal isoforms in the cells, which expands their promiscuous activities and could facilitate a functional divergence.

In chapter four, I reveal how enzyme promiscuity connects seemingly distinct functions, using the data from four family- and superfamily-wide studies, including the one described in chapter two. The data of these studies is visualized in networks, which provides a new perspective on how functions are connected functionally, through promiscuous enzymes, within enzyme superfamilies.

In chapter five, I investigate the evolvability of promiscuous activities by performing a comparative directed evolution experiment of two related enzymes towards the same activity. Subsequent biochemical, biophysical and structural analysis provides insights into the adaptive solutions of each enzyme. Together, the experiment sheds light

on how seemingly neutral sequence changes can have profound consequences on evolvability and supports a notion that contingency and stochasticity play an important role in molecular evolution.

In chapter six, I will discuss the results of each chapter in a broader perspective, suggesting future experiments expanding this work and overall directions for the field of enzyme evolution.

Chapter 2: Evolutionary relationship functions and catalytic

promiscuity in the metallo- β -lactamase superfamily

Parts of chapter two have been performed in collaboration with G. Woollard in the laboratory of Dr. Jörg Gsponer at UBC, Vancouver, Canada and published as “Baier F. and Tokuriki N. (2014): Connectivity between catalytic landscapes of the metallo-beta-lactamase superfamily. *J. Mol. Biol.*, 426 (13), 2442-2456.” G. Woollard performed hydrophobicity calculation of protein active sites, as described in chapter two in Figure 2.10. I performed all other experiments and wrote the manuscript together with my supervisor, Dr. Nobuhiko Tokuriki.

2.1 Summary

The expansion of functions in an enzyme superfamily is thought to occur through recruitment of latent promiscuous functions within existing enzymes. Thus, promiscuous activities of existing enzymes represent “functional” connections between functional families alongside their sequence and structural relationships. Such functional connectivity has been observed between individual functional families; however, little is known about how the diverse enzyme functions are connected throughout a highly diverged superfamily. Here, we describe a superfamily-wide analysis of evolutionary and functional connectivity in the metallo- β -lactamase (MBL) superfamily. We investigated evolutionary connections between functional families and related evolutionary to functional connectivity; 24 enzymes from 15 distinct functional families were challenged against 10 catalytically distinct reactions. We revealed that enzymes of this superfamily are generally promiscuous, as each enzyme catalyzes on average 1.5 reactions in addition to its native one. Thus, functions in the MBL superfamily overlap substantially; each reaction is connected on average to 3.7 other reactions whereas some connections appear to be unrelated to recent evolutionary events and occur between chemically distinct reactions. These findings support the idea that the highly distinct reactions in the MBL superfamily could have evolved from a common ancestor traversing a continuous network via promiscuous enzymes. Several functional connections (e.g., the lactonase/phosphotriesterase and phosphonate/phosphodiesterase/arylsulfatase reactions) are also observed in structurally and evolutionarily distinct superfamilies,

suggesting that these functions are generally highly connected. Additionally, our results show that new enzymatic functions could evolve rapidly from the current diversity of enzymes and range of promiscuous activities.

2.2 Introduction

New enzymatic functions are thought to evolve through the recruitment and optimization of latent promiscuous functions of existing enzymes, which led to the functional expansion of superfamilies we observe to date (Jensen 1976; O'Brien & Herschlag 1999; Khersonsky & Tawfik 2010). Hence, in addition to their sequence and structural relationships, enzyme promiscuity can provide an additional layer of evolutionary connectivity between functional families. Systematic characterisations of substrate promiscuity among homologous and reconstructed ancestral enzymes have helped to characterize how substrate specificity could have evolved within enzyme families (R. Huang et al. 2012; Voordeckers et al. 2012; Schmidt et al. 2003; Baas et al. 2013; Larion et al. 2007; Bastard et al. 2014). In addition, catalytic promiscuity, an enzyme's ability to catalyse distinct chemical reactions to their native one, also provides information about the relationship and connectivity between distinct functions of an enzyme superfamily (Khersonsky & Tawfik 2010; Weng et al. 2012; Glasner et al. 2006; Hult & Berglund 2007; Babbitt et al. 2010; Lešćić Ašler et al. 2010). For example, enzymes of related functional families can share promiscuous activities (or exhibit crosswise promiscuity), as defined by each possessing a low level of catalytic activity against the other's native reaction (Khersonsky & Tawfik 2010; Mohamed & Hollfelder 2013). Therefore, promiscuous activities of enzymes can be seen as connections between different functions of a superfamily that can be traversed by evolution, similar to the analogy of Maynard-Smith's picture of a continuous network of functional proteins (Smith 1970) (**Figure 1.2**; chapter one). In this view, new functions can evolve gradually and in a continuous manner in which all evolutionary intermediates remain functional (Kaltenbach & Tokuriki 2014). To investigate such a "functional connectivity" through enzyme promiscuity within an enzyme superfamily, function-profiling datasets of many different enzymes and functions of one superfamily need to be obtained.

However, most previous studies, as described above, have focused on the connectivity, and cross-wise promiscuity, between individual pairs of closely related homologues (>30% sequence identity) and generally included similar reactions that share chemical properties such as transition state geometry, hydrolysable bond and bond charge (van Loo et al. 2010; Aharoni et al. 2005; G. Phillips et al. 2012; Jonas & Hollfelder n.d.). Hence, it is still unclear how distinct functions in large enzyme superfamilies could have arisen through divergent evolution via promiscuous enzymes, and how these functional connections relate to historical evolutionary divergence and the chemical similarity between reactions.

In this chapter, our aim is to address these questions by performing function profiling within a functionally diverse enzyme superfamily and compare the results to the evolutionary relationship between functions. We chose the metallo- β -lactamase (MBL) superfamily as a model of a functionally diverse enzyme superfamily. Members of the MBL superfamily are substantially diverged in sequence as well as function, but are believed to have originated from a single ancestral function (Ranea et al. 2006; Aravind 1999). To date, at least 24 distinct functional families have been identified within the MBL superfamily, including DNA, RNA and nucleotide processing, detoxification, antibiotic resistance, quorum-quenching, and pesticide hydrolysis (Bebrone 2007; Daiyasu et al. 2001). Most of these functions involve hydrolytic reactions and target diverse substrates with different chemical properties such as phosphodiester, phosphotriester, choline-phosphoester, thiol-ester, sulphate-ester, carbon-ester and β -lactam bond. Other functions involve non-hydrolytic reactions such as nitric-oxidoreduction and sulphur dioxygenation, as well as non-enzymatic functions such as binding and transport (Silaghi-Dumitrescu et al. 2005; Gu et al. 2008). A comprehensive evolutionary analysis is necessary in order to understand the evolutionary and functional relationship of the highly diverged sequences and functions of the MBL superfamily. Although the functional diversity of the MBL superfamily has been described and reviewed, no comprehensive evolutionary analysis has yet been performed (Aravind 1999; Daiyasu et al. 2001; Bebrone 2007). In detail, we analysed the sequence relationship within the MBL superfamily using sequence similarity networks (SSNs), a novel pairwise sequence comparison method to analyse large and diverse datasets. We

annotated the SSNs with available functional, taxonomic and sequence length information, in order to reveal how functional families most likely diverged from each other. Furthermore, we assayed the function profile of 24 MBL members from 15 different functional families against 10 catalytically distinct reactions, with different scissile bonds (C-N, P-O, S-O, C-O and C-Cl), chemical structures, charges and sizes. We also compared general active site properties, such as active site volume and hydrophobicity, with the ability to catalyze certain reactions. Finally, we related the observed functional connections through promiscuity to evolutionary connections, obtained from the SSN analysis, and the chemical similarity between the distinct reactions.

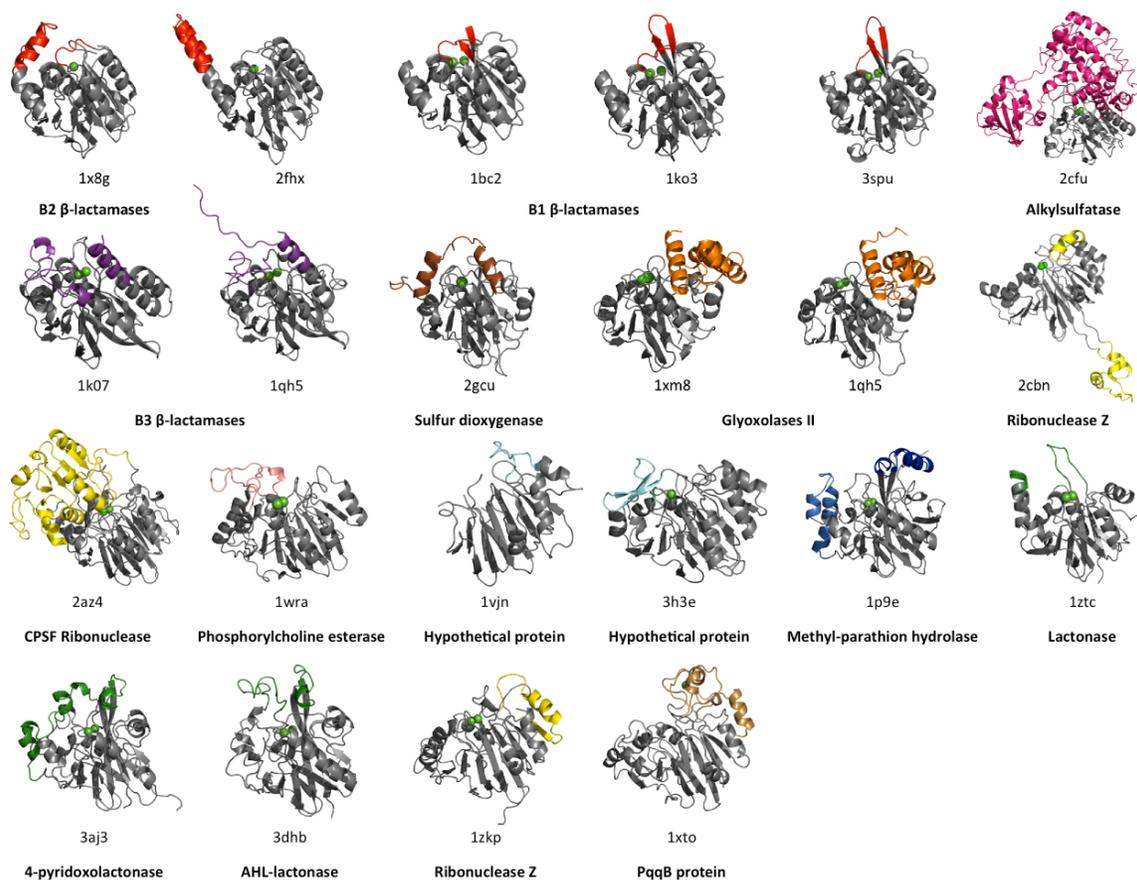


Figure 2.1 Structures of representative MBL superfamily members.

Structurally conserved backbone of the MBL superfamily shown in grey and structural changes are highlighted in color, metals are in shown as green spheres. All images have been generated with structures aligned and images taken from the same perspective.

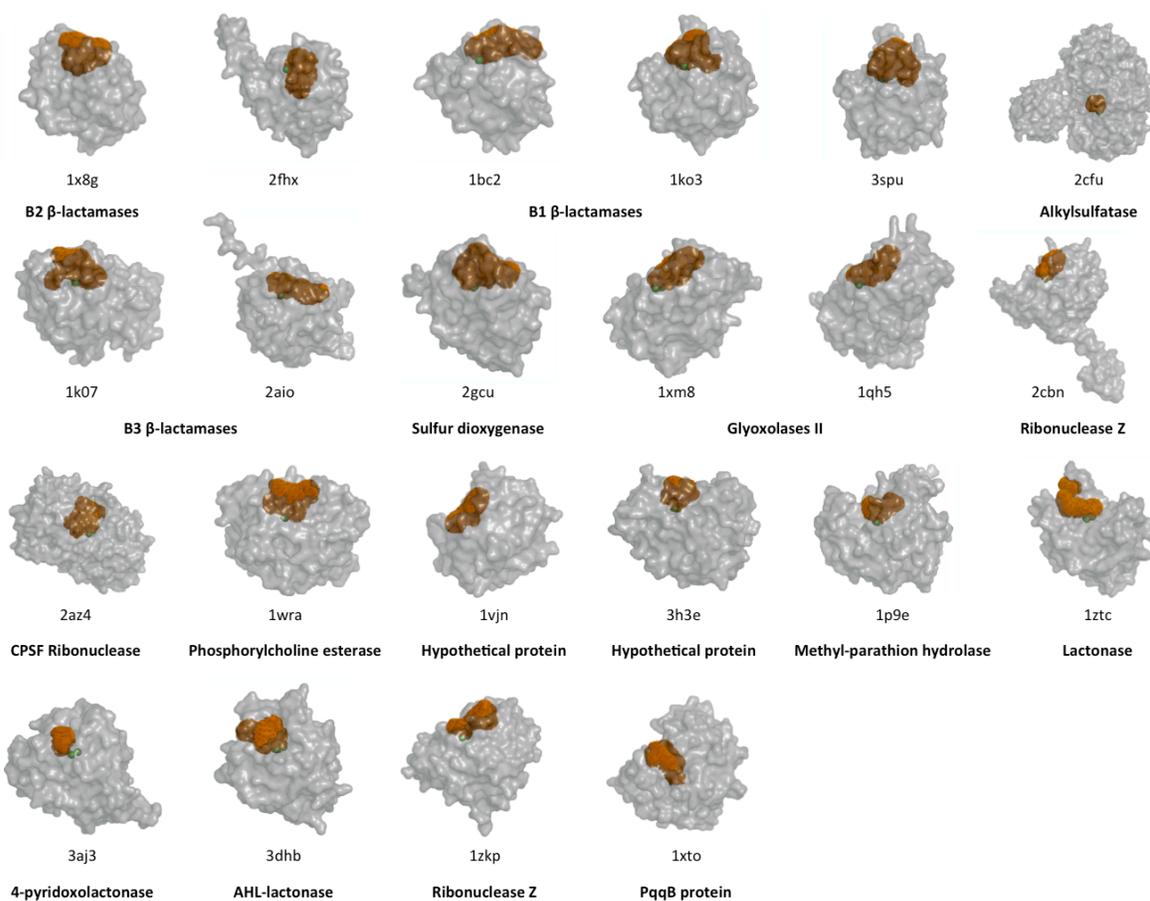


Figure 2.2 Active site cavity of representative MBL superfamily members.

Calculated active sites are shown in orange with metals in green spheres and overall enzyme surface in transparent grey. The active site cavity and its molecular volume was calculated using the ghecom server (<http://strcomp.protein.osaka-u.ac.jp/ghecom/>). All images have been generated with pymol with structures aligned and images taken from the same perspective.

Table 2.1 Sequence and structure similarity of enzymes used in this study.

PDB ID or protein name ^a	Uniprot ID	Organism of origin	Functional Annotation ^b	Amino acid residues ^c	Active site hydrophobicity ^d	Active site volume [Å ³] ^e	Native reaction ^f
1x8g	P26918	<i>A. hydrophyla</i>	B2 β-lactamase	224 (28-251)	0.26	1844	BLA
2fhx	Q8G9Q0	<i>P. aeruginosa</i>	B1 β-lactamase	245 (32-276)	0.30	1267	BLA
1bc2	P04190	<i>B. cereus</i>	B1 β-lactamase	228 (30-257)	0.27	1236	BLA
1ko3	Q9K2N0	<i>P. aeruginosa</i>	B1 β-lactamase	220 (27-266)	0.21	1039	BLA
3spu	C7C422	<i>K. pneumoniae</i>	B1 β-lactamase	270 (1-270)	0.34	1881	BLA
ChD ^a	C9EBR5	<i>P. aeruginosa</i>	Chlorothalonil dehalogenase	334 (3-336)	-	-	TPN
2cfu	Q9I519	<i>P. aeruginosa</i>	Alkylsulfatase	658 (1-658)	-	-	AKS
1k07	Q9K578	<i>F. gormanii</i>	B3 β-lactamase	264 (19-282)	0.19	1375	BLA
2aio	P52700	<i>S. maltophilia</i>	B3 β-lactamase	268 (23-290)	0.30	1051	BLA
2gcu	Q9C8L4	<i>A. thaliana</i>	Sulfur dioxygenase	244 (51-294)	0.25	2387	-
1xm8	Q9SID3	<i>A. thaliana</i>	Glyoxolase II	253 (72-324)	0.25	1036	SLG
1qh5	Q16775	<i>H. sapiens</i>	Glyoxolase II	258 (51-308)	0.27	747	SLG
AtsA ^a	P28607	<i>A. carrageenovora</i>	Arylsulfatase	305 (24-328)	-	-	ARS
2cbn	P0A8V0	<i>E. coli</i>	Ribonuclease Z	305 (1-305)	0.16	921	PDE
2az4	Q8ZZZ3	<i>S. faecalis</i>	β-CASP ribonuclease	429 (1-429)	-	-	PDE
1wra	Q8DQ62	<i>S. pneumoniae</i>	Phosphorylcholine esterase	308 (27-334)	0.18	719	PCE
1vjn	Q9WY50	<i>T. maritima</i>	Hypothetical	208 (1-208)	0.21	727	-
3h3e	Q9X207	<i>T. maritima</i>	Hypothetical	255 (1-255)	0.25	589	-
1p9e	Q841S6	<i>P. aeruginosa</i>	Methyl parathion hydrolase	331 (1-331)	0.43	787	PTE
1ztc	Q9WZZ6	<i>T. maritima</i>	Lactonase	209 (1-209)	0.39	1086	HSL
3aj3	Q988B9	<i>M. loti</i>	4-pyridoxolactonase	268 (1-268)	0.31	389	HSL
3dhh	P0CJ63	<i>B. thuringiensis</i>	AHL-lactonase	250 (1-250)	0.38	757	HSL
1zkp	Q81U06	<i>B. anthracis</i>	Putative ribonuclease	244 (1-244)	0.15	589	PDE
1xto	Q88QV5	<i>P. putida</i>	PQQ biosynthesis protein B	303 (1-303)	0.24	1957	-

^aAtsA and ChD are described by their protein names as no structural information and PDB IDs were available.

^bInformation was retrieved from the Uniprot DB, manually updated with literature information of individual proteins and our results (for details see main text).

^cAmino acid residues indicate the length of the expressed sequence. The number in brackets corresponds to the amino acid positions of the annotated sequence in the Uniprot DB.

^dActive site hydrophobicity is displayed as a fraction, calculated by dividing the number of hydrophobic residues (Phe, Met, Ile, Leu, Val, Cys, Trp, or Ala) by the total number of residues in the active site. The structures of 2cfu and 2az4 have been excluded due to the presence of an additional domain above the active site.

^eActive site volume was calculated with the ghecom server for cavity detection (www.strcmp.protein.osaka-u.ac.jp/ghecom/) and corresponds to the images in **Figure 2.2**. The structures of 2cfu and 2az4 have been excluded for this calculation due to their additional domain above the active site.

^fAnnotated native reaction of the 10 MBL superfamily reactions used in this study. Abbreviations of reactions are explained in **Figure 2.6**.

Table 2.2. Sequence and structure similarity of enzymes used in this study.

	1x8g	2fhx	1bc2	1ko3	3spu	2cfu	2gcu	1qh5	1xm8	1k07	2aio	2az4	3h3e	1zkp	1ztc	3aj3	3dhb	1p9e	1xto	2cbn	1wra	1vjn	AtsA*	ChD*
1x8g		1.3	1.1	1.0	1.3	2.9	18.1	7.4	2.7	6.6	6.6	17.7	7.9	15.6	5.3	13.7	5.3	8.5	11.2	22.0	11.1	13.4	-	-
2fhx	24.9%		1.0	0.9	1.1	4.9	18.8	13.5	2.3	11.1	9.3	18.1	5.1	16.3	6.3	6.4	12.8	13.9	12.0	19.8	11.0	14.9	-	-
1bc2	24.3%	23.6%		0.9	1.0	3.9	11.5	2.1	1.9	8.6	4.2	6.8	7.0	14.0	5.0	9.7	5.4	2.2	12.2	2.5	6.9	15.2	-	-
1ko3	23.1%	24.4%	34.4%		1.0	5.8	7.2	10.6	2.2	7.6	15.0	5.9	11.4	10.1	13.2	9.8	3.8	14.5	11.4	17.0	8.4	6.3	-	-
3spu	20.6%	19.8%	26.2%	32.5%		10.7	4.3	7.1	2.7	11.5	14.0	10.0	8.5	12.9	10.8	12.8	7.6	9.0	9.0	17.7	9.1	14.0	-	-
2cfu	5.0%	7.1%	6.6%	6.7%	5.8%		19.1	17.6	1.8	7.9	15.5	25.4	15.4	18.2	1.5	10.3	12.6	12.1	18.7	6.7	14.6	11.9	-	-
2gcu	6.6%	10.9%	10.9%	9.6%	9.6%	5.7%		1.3	1.2	3.1	4.0	14.6	15.7	7.2	4.2	8.2	3.9	5.5	9.2	5.9	13.0	12.1	-	-
1qh5	10.7%	9.9%	9.6%	8.4%	7.6%	5.8%	14.0%		0.8	2.0	2.4	19.2	9.7	9.3	5.8	9.7	5.0	12.1	9.8	19.7	14.0	9.2	-	-
1xm8	9.3%	10.7%	7.6%	8.8%	9.4%	5.9%	14.9%	25.9%		5.3	16.2	16.2	1.9	15.9	2.8	3.9	3.2	8.0	13.4	18.3	13.0	13.4	-	-
1k07	9.0%	8.2%	7.4%	9.1%	7.6%	4.9%	9.4%	5.1%	5.6%		1.2	14.8	11.5	5.7	6.1	10.0	3.6	4.6	7.4	9.5	14.1	14.0	-	-
2aio	5.9%	7.4%	5.9%	7.9%	9.3%	5.5%	6.5%	7.0%	7.8%	22.7%		9.2	2.0	7.2	11.6	3.5	3.5	5.0	14.6	15.7	12.9	13.5	-	-
2az4	7.0%	7.1%	6.7%	6.6%	7.7%	6.1%	3.9%	6.8%	6.1%	4.2%	4.5%		17.2	11.6	8.8	15.9	18.6	11.8	14.6	5.8	19.1	17.6	-	-
3h3e	7.5%	7.9%	9.8%	11.5%	8.8%	4.6%	6.2%	6.7%	6.2%	7.8%	7.0%	5.3%		5.5	8.9	3.4	6.6	11.0	11.2	12.1	12.6	13.6	-	-
1zkp	6.0%	9.2%	8.8%	8.6%	7.4%	4.9%	7.4%	6.5%	4.9%	6.5%	5.6%	7.7%	7.6%		11.6	13.0	14.4	14.3	1.2	1.7	14.4	7.0	-	-
1ztc	7.3%	8.2%	9.0%	7.2%	7.1%	3.7%	7.2%	9.3%	7.5%	7.7%	6.0%	6.6%	5.6%	5.7%		3.1	2.8	5.1	13.6	4.5	15.2	12.9	-	-
3aj3	9.2%	11.8%	11.4%	10.1%	10.0%	7.8%	10.0%	9.4%	8.8%	9.8%	6.1%	6.6%	8.4%	7.0%	13.0%		1.2	2.9	18.7	16.6	12.4	14.5	-	-
3dhb	10.3%	10.3%	11.8%	9.7%	9.2%	6.2%	9.9%	7.9%	9.4%	6.8%	5.2%	5.0%	8.3%	8.1%	14.4%	25.5%		2.6	16.6	11.8	10.9	10.5	-	-
1p9e	6.2%	11.2%	8.8%	11.7%	14.0%	8.1%	9.9%	7.2%	7.3%	6.9%	9.0%	6.6%	5.8%	5.2%	10.7%	12.7%	14.8%		12.9	1.3	11.6	0.9	-	-
1xto	5.4%	6.2%	7.5%	7.0%	5.2%	5.6%	7.7%	8.2%	6.7%	5.1%	5.1%	4.4%	4.3%	8.0%	9.0%	6.0%	5.4%	5.7%		13.9	8.6	13.8	-	-
2cbn	5.7%	7.9%	7.1%	11.2%	7.3%	3.7%	7.0%	6.5%	7.0%	8.2%	8.1%	6.5%	6.4%	7.7%	3.9%	6.6%	5.4%	8.2%	7.2%		22.9	18.3	-	-
1wra	3.6%	3.8%	2.0%	3.4%	4.0%	6.4%	3.7%	3.6%	6.7%	6.8%	6.5%	2.7%	4.3%	3.0%	4.1%	4.2%	2.3%	6.3%	3.7%	3.4%		13.8	-	-
1vjn	9.5%	9.1%	12.0%	9.7%	7.9%	3.2%	5.6%	7.0%	4.1%	6.9%	6.0%	4.7%	7.3%	9.3%	5.3%	8.8%	10.0%	4.5%	5.1%	7.0%	4.1%		-	-
AtsA*	4.8%	7.6%	5.5%	4.7%	4.3%	4.1%	5.8%	7.2%	5.2%	5.2%	6.5%	3.4%	6.2%	7.3%	4.1%	5.7%	5.5%	6.4%	6.0%	10.7%	3.1%	5.2%		-
ChD*	15.6%	11.8%	13.5%	14.3%	13.2%	8.7%	7.9%	7.8%	8.5%	7.1%	6.4%	4.5%	6.8%	5.2%	8.1%	10.0%	10.9%	9.0%	7.5%	7.8%	4.0%	5.5%	5.7%	

RMSD [Å]

Sequence Identity [%]

*For AtsA and ChD no structural information is available and therefore no structural similarity (RMSD) could be calculated.

Pairwise sequence identities were calculated from a multiple sequence alignment using ClustalW2 (standard parameters), which was then used to calculate the identities using the web based program SIAS (<http://imed.med.ucm.es/Tools/sias.html>) with gaps taken into account. To determine pairwise structural similarity we computed the root mean standard deviation (R.M.S.D.) between all structure pairs using the align command in PyMOL.

2.3 Methods

2.3.1 Construction of sequence similarity networks

The pipeline to generate sequence similarity networks was adapted from Atkinson *et al.* (Atkinson *et al.* 2009). In detail, 33,843 amino acid sequences of the metallo-β-lactamase superfamily (Pfam-IDs: PF00753, PF12706, PF13483) were retrieved from the Pfam database (Hu *et al.* 2009; Punta *et al.* 2012) on the 15th of June 2012. To facilitate further analysis, we extracted a representative set from the initial set (33,843 sequences) by applying a sequence identity threshold of 50% using CD-Hit. Subsequently, we manually added the amino acid sequences of the 24 experimentally characterized enzymes, giving a final set of 6,233 representative sequences. With these 6,233 representative sequences we performed an all versus all protein BLAST [NCBI, version +2.2.26] using an appropriate and corresponding BLAST e-value cut-off for each network. Finally, the E-values for all sequence pairs above the cut-off were imported into Cytoscape [version 2.8.3] and visualization of networks was achieved with the organic layout, in which length of connecting edges correlates with the dissimilarity of the sequences but does not represent a quantitative correlation. Sequence attributes for sequence length and taxonomic

distribution were retrieved from the Uniprot (UniProt Consortium 2015), Swiss-Prot (manually annotated and reviewed) and TrEMBL (automatically annotated) databases and uploaded into Cytoscape. Functional information was exclusively retrieved from the Swiss-Prot database (manually annotated and reviewed) and further updated with functional information from the recent literature. We would like to note that subsequent analysis of an updated Pfam release (v27.0; March 2013; the number of representative sequences increases from 6,233 to 7,463) had no consequence on the clustering pattern in the sequence similarity networks of the MBL superfamily.

2.3.2 Sequence identity and structural similarity calculation and phylogeny

Pairwise sequence identities were calculated by performing a multiple sequence alignment of all 24 characterized enzymes (**Table 2.1**) using ClustalW2 (default parameters). The alignment was then used to calculate the identities using the web-based program SIAS (<http://imed.med.ucm.es/Tools/sias.html>) with gaps taken into account (**Table 2.2**). To determine pairwise structural similarity we computed the root mean standard deviation (RMSD) between all structure pairs using the align command in PyMOL. For the structure based phylogenetic tree, the sequence alignment was performed with the Expresso algorithm (Armougom et al. 2006) and tree was built using the maximum likelihood method (100 Bootstraps) in Mega5 (Tamura et al. 2011) with default parameters.

2.3.3 Molecular cloning

The genes for 24 MBL proteins were obtained from various sources, as listed in **Table 2.3**. Encoding genes, apart from PDB ID 1wra, which was purified using a His₆-tag, were sub-cloned into a pET27(b)-Strep and MBP (maltose binding protein) tag vector, and their solubility was determined with SDS-PAGE. The most soluble construct was then used to express N-terminal Strep- or MBP-tagged fusion proteins (**Table 2.3**). The pET27(b)-Strep vector was created by inserting the Strep-tag II sequence (MASWSHPQFEKGAG) into the pET27(b) vector (Novagen), using *NdeI* and *BamHI* restriction sites. The pET27(b)-MBP vector was created by replacing the Strep-tag II

sequence with the MBP-tag sequence from the pMAL-c2e vector (NEB), using *NdeI* and *BamHI* restriction sites. All constructions were confirmed by DNA sequencing.

Table 2.3 Information on cloning and source of enzymes assayed.

PDB ID or protein name ^a	Uniprot ID	Expression tag	Gene source	Restriction sites used	ϵ [$M^{-1} \text{ cm}^{-1}$] at 280 nm
1bc2	P04190	Strep-tactin	cPCR <i>B.cereus</i> ^b	<i>NcoI/HindIII</i>	29450
1k07	Q9K578	MBP	Gene synthesis ^c	<i>NcoI/HindIII</i>	105310
1ko3	Q9K2N0	Strep-tactin	Gene synthesis ^c	<i>NcoI/HindIII</i>	35410
1p9e	Q841S6	Strep-tactin	Gene synthesis ^c	<i>NcoI/HindIII</i>	24420
1qh5	Q16775	MBP	DNASU Plasmid repository ^d	<i>NcoI/HindIII</i>	91135
1vjn	Q9WY50	Strep-tactin	DNASU Plasmid repository ^d	<i>BamHI/HindIII</i>	20065
1wra	Q8DQ62	His ₆	M. Guilmi at Inst. of Structure Biology, France ^d	-	49850
1x8g	P26918	MBP	Gene synthesis ^c	<i>NcoI/HindIII</i>	100730
1xm8	Q9SID3	MBP	Gene synthesis ^c	<i>NcoI/HindIII</i>	85510
1xto	Q88QV5	MBP	Gene synthesis ^c	<i>NcoI/HindIII</i>	116200
1zkp	Q81U06	Strep-tactin	W. Anderson at Northwestern Univ., Canada ^d	<i>NcoI/HindIII</i>	31650
1ztc	Q9WZZ6	Strep-tactin	DNASU Plasmid repository ^d	<i>NcoI/HindIII</i>	22920
2aio	P52700	MBP	Gene synthesis ^c	<i>NcoI/HindIII</i>	91135
2az4	Q82ZZ3	MBP	cPCR <i>E.faecalis</i> ^b	<i>NcoI/HindIII</i>	98795
2cbn	P0A8V0	MBP	Gene synthesis ^c	<i>NcoI/HindIII</i>	88155
2cfu	Q9I519	MBP	K. Faber at Univ. of Graz, Austria ^d	<i>NcoI/HindIII</i>	158600
2fhx	Q8G9Q0	MBP	Gene synthesis ^c	<i>NcoI/HindIII</i>	97290
2gcu	Q9C8L4	MBP	Gene synthesis ^c	<i>NcoI/HindIII</i>	78395
3aj3	Q988B9	MBP	Gene synthesis ^c	<i>NcoI/HindIII</i>	88155
3dhh	P0CJ63	Strep-tactin	Gene synthesis ^c	<i>NcoI/HindIII</i>	25940
3h3e	Q9X207	MBP	DNASU Plasmid repository ^d	<i>NcoI/HindIII</i>	85050
3spu	C7C422	Strep-tactin	Gene synthesis ^c	<i>NcoI/HindIII</i>	33690
AtsA	P28607	MBP	Gene synthesis ^c	<i>NcoI/HindIII</i>	104740
ChD	C9EBR5	MBP	Gene synthesis ^c	<i>NcoI/HindIII</i>	48025

Abbreviation: MBP, Maltose-binding protein

^aAtsA and ChD are named by their protein name as for both no structural information is available.

^bGenes amplified for subcloning from genomic DNA by polymerase chain reaction from a freshly streaked colony of the corresponding organism using gene specific primers.

^cGenes were ordered commercially synthesized from the companies BioBasic and Genewiz and subcloned into expression vectors.

^dPlasmids containing the corresponding gene were either kindly provided by the above mentioned research groups or purchased from the DNASU Arizona State University plasmid repository and subcloned into expression vectors.

Extinction coefficient were calculated for the whole fusion protein with the ProtParam from ExPasy (<http://web.expasy.org/protparam/>).

2.3.4 Protein expression and purification

All enzymes were expressed as fusion proteins in *E. coli* BL21 (DE3) cells in TB auto-induction media (EMD Millipore) supplemented with 1% w/v Glycerol, 200 μM ZnCl_2 (PDB ID 1ztc with NiCl_2) and 35 $\mu\text{g/ml}$ kanamycin. Cultures (400 ml) were inoculated from overnight cultures (Luria Broth with 40 $\mu\text{g/ml}$ kanamycin, 10 ml) and incubated initially at 30°C for 6 hours, before further incubation at 16°C for 16 hours, to express the

proteins. Cells were harvested by centrifugation at 10,000×g and pellets were frozen at -80°C for at least 2 hours. For lysis, cell pellets were resuspended in lysis buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 200 μM ZnCl₂ (Buffer A) containing 50 % B-PER protein extraction reagent (Thermo Scientific) and 10 mM imidazole (sigma) for His-tag) and incubated on ice for 1 hour. The cell lysates were clarified by centrifugation at 30,000 × g for 20 min at 4°C. Affinity tag purification of His₆-, Strep- or MBP-tag fusion proteins was performed according to the manufacturers protocol with Ni-NTA resin (Thermo Scientific), Strep-tactin resin (IBA lifesciences) and Maltose resin (NEB) respectively. Briefly, the clarified lysates were loaded on the ~2 mL resin in ~15 mL gravity columns (Biorad), which were equilibrated in Buffer A (with 10 mM imidazole for His-tag). The column were extensively washed with around 45 mL Buffer before elution with Buffer A containing either 200 mM imidazole for His-tag, 10 mM maltose for MBP-tag (sigma) or 10 mM d-desthiobiotin for Strep-tag (sigma). Elution was performed in one mL steps and the protein containing fractions were analysed with SDS-PAGE and pooled depending on the concentration and purity. The purified enzymes were concentrated to a volume of 3 ml (Microsep 10 kDa, Pall), desalted with Econo-Pac 10DG columns (Bio-Rad) and eluted in 4 ml of 50 mM Tris-HCl pH 7.5, 100 mM NaCl and 200 μM ZnCl₂. The concentration of each purified enzyme was determined by measuring absorbance at 280 nm (Table 1 includes the extinction coefficients used for each enzyme). All purifications yielded > 90% pure protein, which was verified with SDS-PAGE.

2.3.5 Enzyme assays and kinetics

The enzymatic activities for each enzyme were examined with 10 μM of the purified enzyme and 500 μM of the various substrates (TPN and Centa at 100 μM) in Buffer A supplemented with 0.2% Triton X-100. Enzymatic hydrolysis of the reactions PPP, PCE, PDE, ARS, PTE was monitored following the release of *p*-nitrophenol at 405 nm with an extinction coefficient of 18,300 M⁻¹cm⁻¹ (Goddard & Reymond 2004). The BLA reaction was measured with the substrates Centa and Imipenem (only for PDB ID 1x8g) at 405 nm and 300 nm respectively, and product formation was calculated with the extinction coefficients 6,300 M⁻¹cm⁻¹ and 9,000 M⁻¹cm⁻¹, respectively (Bebrone et al. 2001).

Hydrolysis of SLG, γ -thiobutyrolactone and thiobutylbutyrolactone, exposing a free thio-group, was followed with Ellman's reagent ((5,5'-dithio-bis-[2-nitrobenzoic acid]), Sigma) at 405 nm with an extinction coefficient of $13,600 \text{ M}^{-1}\text{cm}^{-1}$ (Riddles et al. 1983). Hydrolysis of SDS and HSL releases a proton and a chloride ion for TPN. The change in pH during hydrolysis can therefore be monitored photospectrometrically for the reactions by using a phenol red as a pH indicator (Hagelueken et al. 2006). For this assay the purified enzyme were dialysed against 1.25 mM Tris-HCl pH 7.5, 100 mM NaCl and 200 μM ZnCl_2 using the Econo-Pac 10DG columns (Bio-Rad). The reaction buffer contained 1.25 mM Tris-HCl pH 7.5, 100 mM NaCl, 50 μM phenol red and a variety of substrate concentrations. Initial rates of hydrolysis were determined by monitoring the decrease in absorbance at 560 nm at 25°C. A standard curve was prepared using HCl to calculate an extinction coefficient of $3,900 \text{ M}^{-1}\text{cm}^{-1}$. For all enzyme and substrate pairs, initial rates were determined as described above. For all activities that were at least 10-fold above the buffer only control, kinetic constants were determined as follows: initial rate measurements at various substrate concentrations were performed in duplicate and the data averaged. The Michaelis-Menten equation was then fit to the data using Kaleidograph software (Synergy).

2.3.6 Metal removal experiments

In order to confirm that the metal in the coordinating active site was responsible for activity, metals were removed and the enzymatic activities of apo and metal re-supplemented enzymes were determined. Purified enzymes were treated with 5 mM 1,2-phenanthroline and 5 mM EDTA for 18 hours at 4°C. Chelators and metals were removed by passing the samples twice through Econo-Pac 10DG desalting columns (Bio-Rad), which had been equilibrated previously with 50 mM Tris-HCl pH 7.5, 100 mM NaCl, and samples were eluted in the same buffer. In order to compare apo- versus metal-containing enzyme, 200 μM of ZnCl_2 (NiCl_2 for PDB ID 1ztc) was added to the enzyme and incubated for at least 1 hour prior to activity testing.

2.3.7 Active-site cavity detection and fraction of hydrophobic residues

Active-site of cavities of MBL enzymes were defined with the web-based program ghecom (Kawabata 2010). The active-site cavity coordinates obtained for each structure

were then used for visualization and calculation of the active-site cavity hydrophobicity. For the hydrophobicity calculation, surrounding residues were considered to be part of the active site if they met two criteria: (1) each residue has at least one atom within 6 Å of the defined cavity, and (2) the solvent accessible surface area (SASA) of the each residue was greater than 2 Å². Therefore residues that are buried deep within the core of the protein are not considered part of the active site. SASA was computed with the PyMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC) command ‘get_area’ using the default solvent radius of 1.4 Å. We then computed the fraction of hydrophobic residues as the number of Phe, Met, Ile, Leu, Val, Cys, Try, or Ala residues divided by the total number of residues in the active site. Two enzyme structures (PDB ID 2az4 and 2cfu), which have additional domains above their active sites, were excluded, because their active site cavities could not be assigned adequately.

2.4 Results

2.4.1 Sequence relationship between functional families of the MBL superfamily

To investigate the sequence relationship of MBL superfamily members, 33,843 sequences were retrieved from the Pfam database (Punta et al. 2012). Due to high divergence and quantity of the sequences, conventional approaches to characterise evolutionary relationships, such as multiple sequence alignments and phylogenetic trees, were not applicable. As an alternative, we employed sequence similarity networks to show the relationships between sequences, described as independent pairwise alignments using the BLAST E-value (Brown & Babbitt 2012; Atkinson et al. 2009). To generate sequence similarity networks, a 50% amino acid sequence identity threshold was applied to reduce the number of sequences to 6223, using the web-based sequence comparison tool CD-Hit and BLAST (Y. Huang et al. 2010; Altschul et al. 1990). Multiple sequence similarity networks were generated with differing BLAST E-value cut-offs in order to obtain optimal resolution for relationship between functional families. An E-value cut-off of $1 \times e^{-6}$ indicated the global connection between sequence clusters, but failed to generate clear separation between individual sequence clusters that represent a single functional family (**Figure 2.3 A**). A more stringent cut-off ($1 \times e^{-14}$) resulted in separation

of the sequences into distinct clusters while still maintaining connections (**Figure 2.4**). An even more stringent cut-off of $1 \times e^{-20}$ resolved large dense clusters into smaller individual clusters, but also resulted in severe loss of connectivity (**Figure 2.3 B**). Overall, ~ 70 discrete sequence clusters were observed at the most stringent cut-off. We annotated the physiological function of all sequence clusters in the sequence similarity networks for which we could find information in the literature and in annotation databases (**Figure 2.3** and **2.4**). However, the physiological functions of many clusters are unknown and are yet to be characterized (**Figure 2.3** and **2.4**). Overall, the functions of many sequence clusters are associated with nucleotides, *i.e.*, RNases, DNA repair and DNA uptake and they form several discrete sequence clusters (**Figure 2.4**). Several non-nucleotide associated functions, such as arylsulfatase and phosphoryl-choline esterase as well as enzymes involved in phosphonate metabolism, appear to have evolved from these nucleotide-associated clusters. In contrast, the majority of non-nucleotide associated functions (*i.e.* β -lactam, methyl-glyoxal, phosphotriester and lactone hydrolysis) generate a large and distinct group with resolved clusters. This suggests that nucleotide associated and non-nucleotide associated enzymes diverged early in evolution, prior to further expansion and specialization. Overall, our results provides the first comprehensive evolutionary relationship analysis in MBL superfamily using SSNs (Aravind 1999). Our results are largely consistent with previous studies that employed sequence alignment and phylogenetic analyses (Aravind 1999; Garau et al. 2005; Mir-Montazeri et al. 2011). Mapping the taxonomic distribution on the sequence similarity network reveals that most sequences in our SSN of the MBL superfamily are found in the domain of Bacteria (83%) and comparatively less in Eukarya (10%) and Archaea (6%). Several sequence clusters, such as RNases Z and CPSF as well as glyoxalases II and nitric oxide reductases, appear to be widespread among all three domains of life, bacteria, Eukaryote and Archea (**Figure 2.5 A**). On the other hand, some functions appear to be highly enriched in organisms of one domain. For example, MBL members involved in DNA repair appear only in eukaryotic genomes, and β -lactamases, alkylsulfatases and competence proteins are exclusively found in bacteria. Mapping sequence length on the sequence similarity network shows that protein size is generally maintained within a sequence cluster, but often differs between clusters. The core MBL domain generally consists of <300 amino

acids (**Table 2.1**) (Daiyasu et al. 2001), which accounts for roughly 45% of the sequences in the SSN (**Figure 2.5 B**). Thus, around 55% of the sequences have insertions or even additional domains fused to them, which suggests that altering structures, by insertions and deletions as well as fusion to an additional domain, is often associated with functional transition in the MBL superfamily (**Figure 2.5 B**) (Furnham et al. 2012).

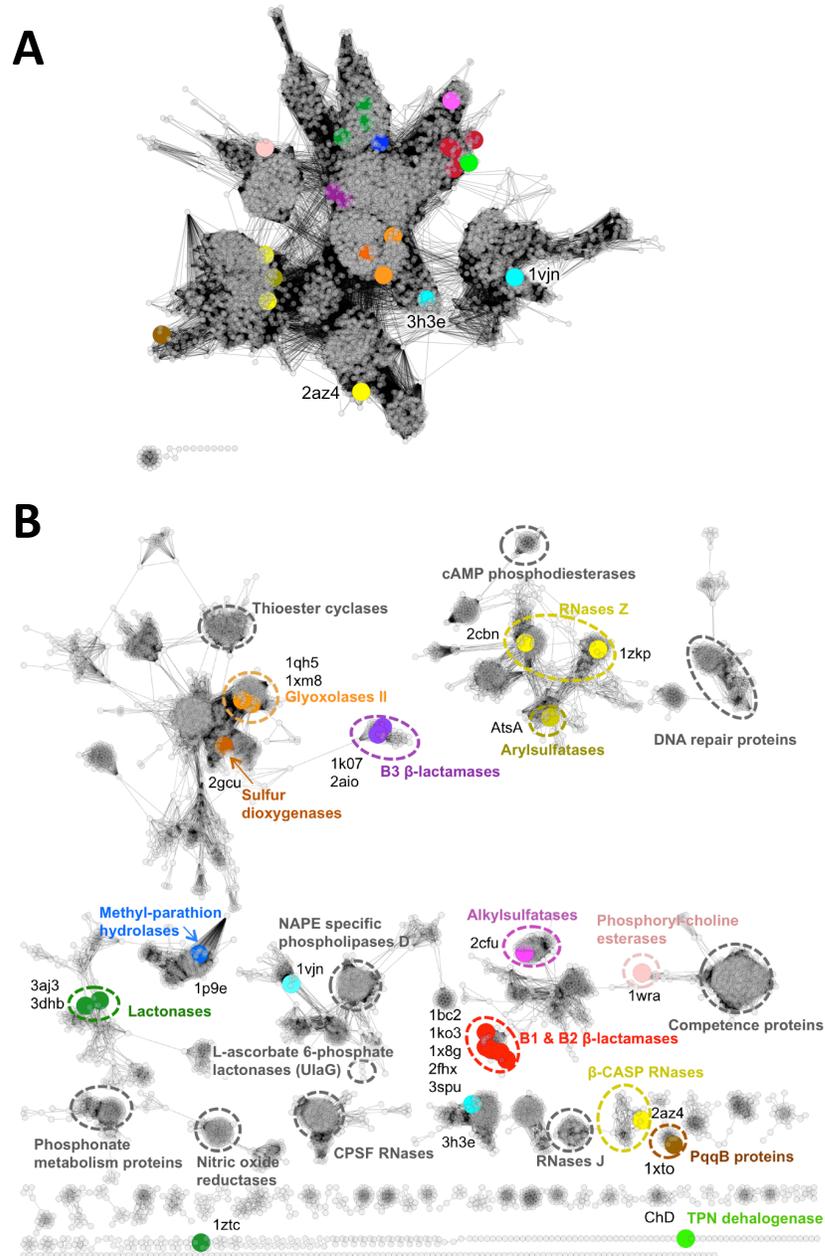


Figure 2.3 Sequence similarity network of the MBL superfamily at different BLAST E-value cut-offs.

(A) Overall connectivity of functional families (6233 sequences) at a BLAST E-value cut-offs of 1×10^{-6} . (B) Higher resolution and separation of functional families at a BLAST E-value cut-offs of 1×10^{-20} . At this cut-off functional families are mostly separated, but also some characterized members (PDB-ID 1ztc and TPN dehalogenase (ChD)) lost connectivity. Large colored nodes show sequences experimentally characterized in this study. Blue nodes with PDB IDs are hypothetical enzymes without detected and annotated native function.

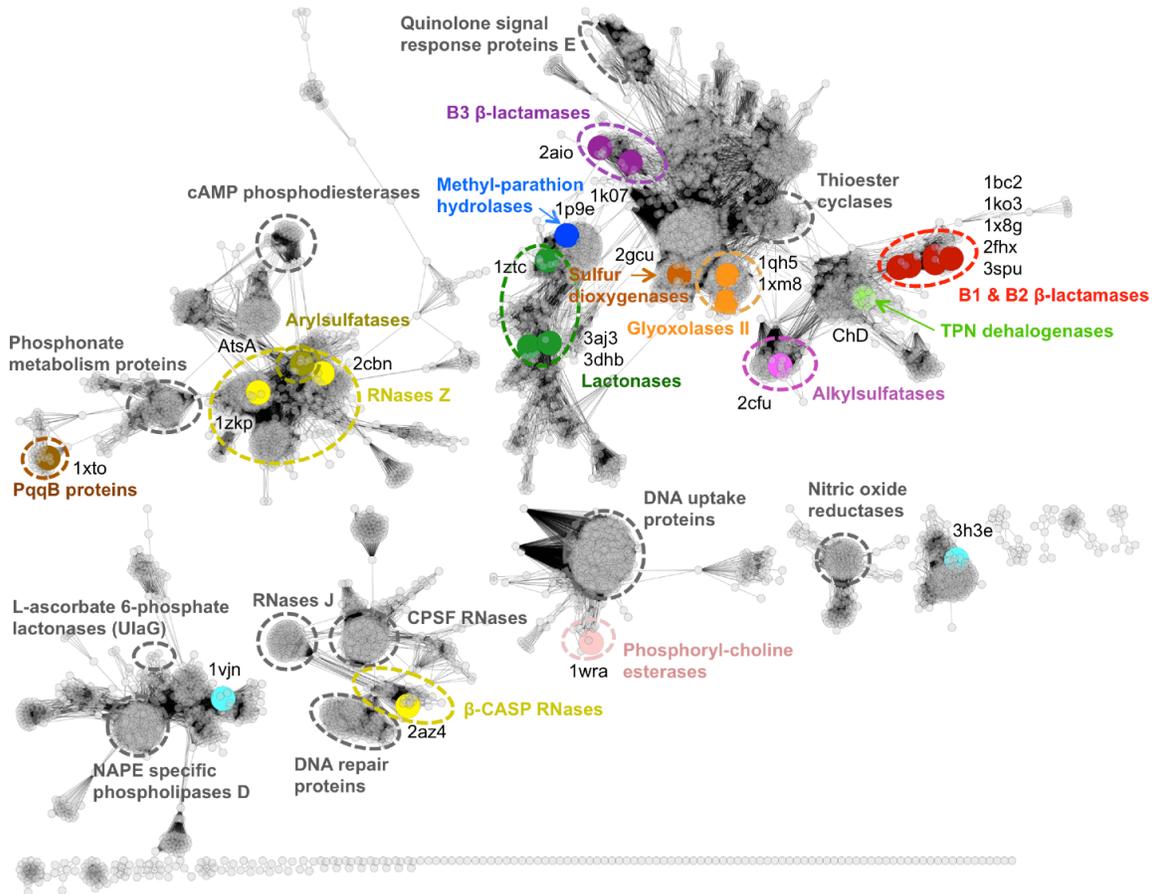


Figure 2.4 Sequence similarity network of MBL superfamily members.

6233 sequences (nodes) and lines (edges) show sequence relationship at a BLAST E-value cutoff of 1×10^{-14} . Large colored nodes show sequences that were experimentally characterized in this study with dashed colored circles indicating their approximate functional family cluster. Pale blue colored nodes (PDB ID 1vjn and 3h3e) are experimentally characterized sequences with unknown function and from clusters with unknown function. The sequence clusters of three experimentally characterized sequences and have not been functionally annotated due to the small number of functional homologues (methyl parathion hydrolase and TPN dehalogenase) and annotation ambiguity (sulfur dioxygenase). Dashed grey

circles indicate functional sequence clusters that have been experimentally characterized and reported in the literature, but have not been included in this study. For unassigned grey sequence clusters (not encircled), no confident functional information could be retrieved from the databases or literature.

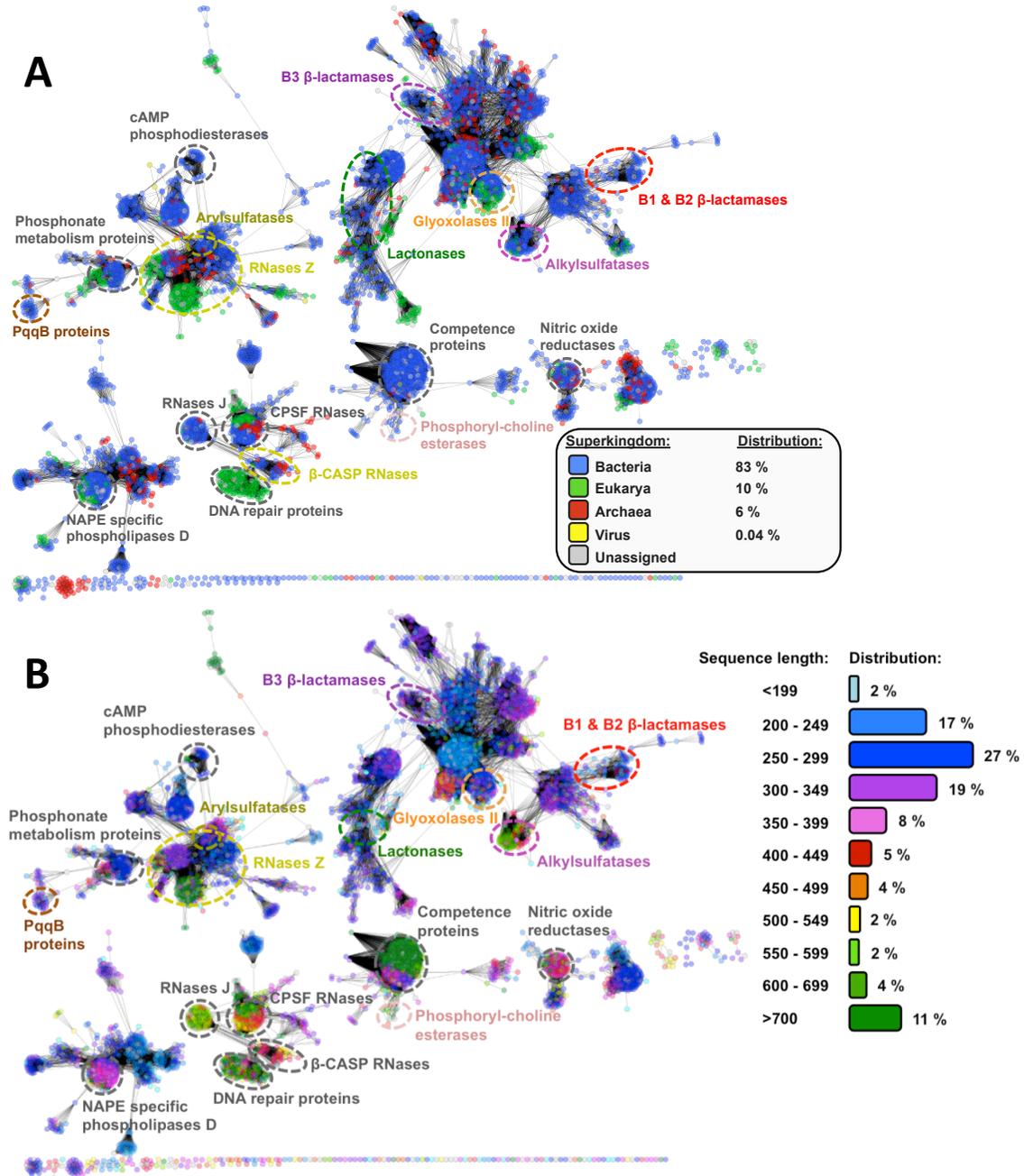


Figure 2.5 Mapping taxonomy and sequence length on the sequence similarity network of the MBL superfamily.

(A) Shows taxonomic distribution and (B) shows sequence length distribution. In both cases, the percentage of representation was calculated from annotated nodes in the SSN, with unassigned nodes excluded from calculation. Information was

retrieved from the Uniprot DB, imported into Cytoscape and mapped onto the sequence similarity network at a BLAST E-value cutoff $1 \times e^{-14}$.

2.4.2 Selection of enzymes and reactions for function-profiling analysis in the MBL superfamily

Twenty-four enzymes, which exhibit a broad range of sequence, structural and functional diversity, were chosen for exploration of catalytic promiscuity (**Table 2.1** and **2.2**; **Figure 2.1** and **2.2**). These enzymes were chosen because of previous structural and functional characterization and represent a wide range of the sequence and function space in the MBL superfamily (**Figure 2.4**). Twenty-two of these enzymes belong to 13 known functional families, including four B1 and one B2 β -lactamases as well as two B3 β -lactamases (independent from B1/B2 β -lactamases) (**Table 2.1**). The β -lactamase subclasses are separated into two distinct clusters, B1/B2 and B3 β -lactamases, which possess distinct metal binding residues and active site configuration (Bebrone 2007). We also included three lactonases (one of which, PDB ID 1ztc, had no previous assigned function, but is suggested to be a lactonase from our results) and phosphodiesterases (two RNase Zs and one, PDB ID 2az4, is suggested to be a β -CASP RNase from our results), and two glyoxalases II (**Table 2.1**). The following families were each represented by a single enzyme: methyl parathion hydrolases, phosphorylcholine esterases, chlorothalonil dehalogenases, arylsulfatases, alkylsulfatases, sulfur dioxygenases and pyrroloquinoline quinone biosynthesis proteins B (a predicted transporter) (**Table 2.1**). Two enzymes, methyl-parathion hydrolase and TPN dehalogenase degrade xenobiotic pesticides (Dong et al. 2005; G. Wang et al. 2010). Their substrates are not believed to have been present in the environment prior to the first half of the 20th century and thus are considered to have evolved very recently (Dong et al. 2005; G. Wang et al. 2010; Singh 2009; G. Wang et al. 2011). We therefore presume that both functions have not diverged enough to form individual clusters in the sequence similarity networks, but maintain their connection with the ancestral families from which they evolved (unfortunately, the function of these clusters are currently unknown (**Figure 2.4**)). Additionally, two enzymes with unknown function were also included (**Table 2.1**). The two non-annotated enzymes (PDB ID 3h3e

and 1vjn) are found in uncharacterized sequence clusters (**Figure 2.4**). All enzymes were predominantly of bacterial origin except for three eukaryotic enzymes (**Table 2.1**). Structural data was available for 22 out of the 24 enzymes described in this study, with the exceptions being the arylsulfatase and the chlorothalonil dehalogenase (**Table 2.1** and **Figure 2.1** and **2.2**).

To experimentally characterise the function profile of these 24 enzymes, we assayed their enzymatic activities against 10 catalytically distinct hydrolytic reactions, which represent native activities for 18 of the 24 enzymes (**Figure 2.6**): β -lactamase (BLA; E.C. 3.5.2.6), glyoxalase II (SLG; E.C. 3.1.2.6), phosphonate monoester hydrolase (PPP; E.C. 3.1.4.83), phosphodiesterase (PDE; E.C. 3.1.26.11), phosphotriesterase (PTE; E.C. 3.1.8.1), phosphocholine esterase (PCE; E.C. 3.1.4.1), arylsulfatase (ARS; E.C. 3.1.6.1) and alkylsulfatase (AKS; E.C. 3.1.6.-), homoserine lactonase (HSL; 3.1.1.81) and chlorothalonil dehalogenase (TPN; E.C. 3.8.1.2). For most reactions, a representative chromogenic substrate with a *p*-nitrophenol leaving group (yellow colour) was used, whereas SLG was assayed using Ellman's reagent (yellow colour) and pH indicator assay was employed for AKS, TPN and HSL reactions. Two β -lactams, Centa (chromogenic) and imipenem, were used as substrates for BLA activity, with imipenem used for B2 β -lactamase because of their low activity for the chromogenic BLA substrate, Centa (Bebrone et al. 2001). Additionally, two related organophosphate pesticides, parathion-ethyl and paraoxon (the oxidized form of parathion-ethyl), were used for PTE. The physiological functions of the sulphur dioxygenase (PDB ID 2gcu) (Holdorf et al. 2012) and the protein involved in a transport step in pyrroloquinoline quinone biosynthesis (PDB ID 1xto) (Puehringer et al. 2008) were excluded as a simple functional assay was not available. The full name, structures and scissile bond of the substrate are described in **Figure 2.6**.

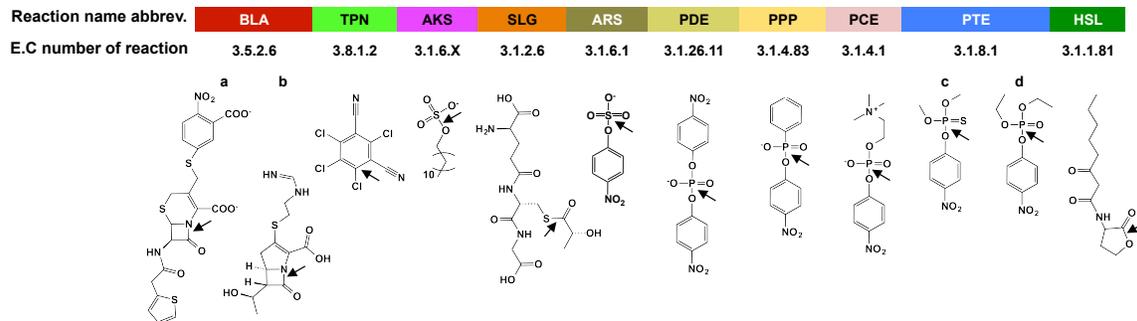


Figure 2.6 Substrates used in this study.

The three-letter code abbreviation and color scheme used in this study is shown at top of figure. Four-digit code represents E.C number classification of reactions. Arrows indicate the bonds that are broken during hydrolysis. BLA, (a) Centa and (b) Imipenem; TPN, chlorothalonil; AKS, sodium dodecyl sulfate; SLG, S-(lactoyl)- glutathione; ARS, *p*-nitrophenyl sulfate; PDE, Bis(*p*-nitrophenyl)-phosphate; PPP, *p*- nitrophenyl phenylphosphonate; PCE, *p*-nitrophenyl phosphoryl choline; PTE, (c) parathion- methyl and (d) paraoxon; HSL, N-(3-Oxoocanoyl)-DL-homoserine lactone.

2.4.3 Function profiling analysis of enzymes of the MBL superfamily

The 24 enzymes were overexpressed as affinity-tag fusion proteins in *E. coli*, purified, and their catalytic activities against each of the 10 reactions were tested using 10 μM enzyme and 500 μM substrate, but 100 μM substrate for BLA and TPN due to the sensitivity and solubility, respectively. Of the 240 assayed enzyme-substrate pairs, 56 showed activity as defined by exhibiting at least a 10-fold higher rate compared to the buffer control. Steady-state kinetic parameters were determined for these 56 reactions. The detection limit of the enzymatic activity varied depending on the substrate; reactions based on chromogenic substrates had a lower threshold ($k_{\text{cat}}/K_{\text{M}} = 10^0 \sim 10^{-2} \text{ s}^{-1}\text{M}^{-1}$) compared to those measured with a pH indicator assay ($k_{\text{cat}}/K_{\text{M}} = \sim 10^1 \text{ s}^{-1}\text{M}^{-1}$). The systematic analyses of 24 enzymes against 10 catalytically distinct hydrolytic reactions revealed widespread catalytic promiscuity in the MBL superfamily, with 36 of 56 active pairs being promiscuous activities (**Figure 2.7**). In detail, five enzymes were able to catalyse three reactions in addition to their native function and 20 of 24 enzymes catalysed hydrolysis of at least one promiscuous reaction. Notably, four proteins with native functions that were not included in our analyses (sulfur dioxygenase, pyrroloquinoline quinone biosynthesis proteins B (predicted transporter), and two proteins of unknown function) also catalysed some of the assayed reactions. We presume

that the observed activities of these enzymes are promiscuous, because their efficiency is far lower than that of native reactions (Bar-Even et al. 2011). Overall, an average of ~ 1.5 non-native reactions were carried out per enzyme. In respect to the catalytic efficiency, the enzymes were highly specialised toward their native reactions; the efficiency of promiscuous activities is on average 10^4 -fold lower than that of the native activities (median k_{cat}/K_M for native and promiscuous activities is $1.9 \times 10^4 \text{ M}^{-1}\text{s}^{-1}$ versus $4.3 \text{ M}^{-1}\text{s}^{-1}$, **Figure 2.8**). Interestingly, the differences in catalytic efficiency are largely manifested in turnover rates (median k_{cat} for native and promiscuous activities is 10 s^{-1} versus 0.008 s^{-1}) whereas K_M differs by only few fold (median K_M for native and promiscuous activities is 0.46 mM^{-1} versus 7.5 mM^{-1}). The effect on k_{cat} is suggested to arise from poor positioning and non-productive binding modes of the non-native substrate in the enzyme active sites (Babtie et al. 2010; Khersonsky & Tawfik 2010).

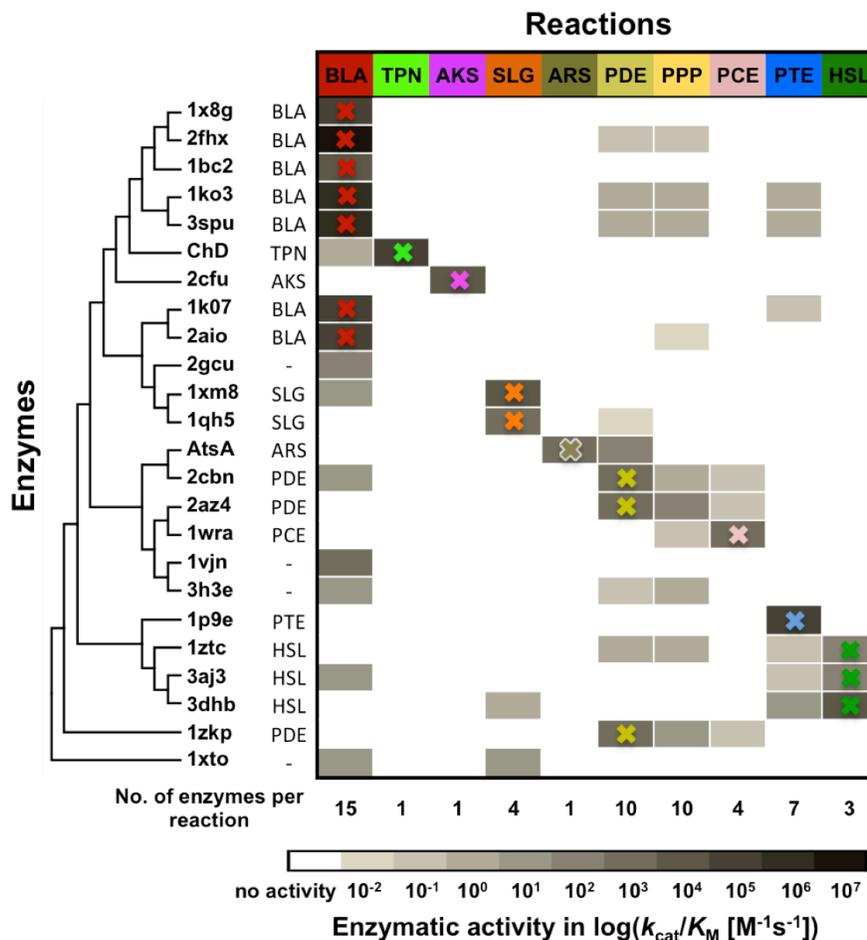


Figure 2.7 Activity patterns of selected MBL superfamily members.

Heat map of catalytic efficiencies of 24 enzymes against 10 reactions demonstrates their high degree of catalytic promiscuity and specialization for their corresponding native activity. The level of enzymatic activity for each enzyme/reaction pair is shown in catalytic efficiency (k_{cat}/K_M) with darker shading indicating higher activity, white means no activity. A colored cross indicates the native activity for each enzyme. Enzymes were arranged based on their phylogenetic relationship calculated with a structure based sequence alignment as described in the methods. The native reactions for 20 out of 24 enzymes are shown next to the protein name or PDB ID. Two enzymes (PDB ID 1vjn and 3h3e) have no annotated function and no native reaction could be identified. The individual kinetic parameters are listed in **appendix A1**.

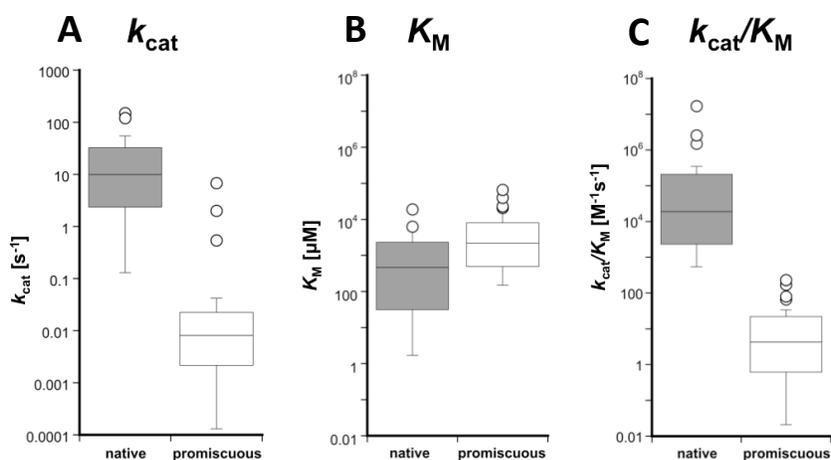


Figure 2.8 Differences between the efficiency of native and promiscuous activities.

Catalytic parameters (A) k_{cat} , (B) K_M and (C) k_{cat}/K_M of native and promiscuous activities are displayed as boxplots with standard deviation and median. The ratios of median for native versus promiscuous activity are 1200 for k_{cat} , 5 for K_M and 4400 for k_{cat}/K_M . A total of 20 native and 36 promiscuous activities were used for the analysis. The individual kinetic parameters are listed in **appendix A1**.

Stringent controls must be in place in order to reliably detect weak promiscuous activities, and several lines of evidence confirm that the observed activities were not artefacts caused by contaminants. First, purifying 24 enzymes under the same conditions (differing only according to their affinity tags) served as an internal control, as only a subset of these enzymes showed promiscuous activities for a certain reaction, while others showed no detectable activity. Second, metal-chelating experiments for representative enzymes revealed that all catalytic activities were abolished in apo-enzymes, and recovered by reconstituting the enzyme with metal ions (**Figure 2.9**).

Third, the K_M value, which is unique for each enzyme-substrate pair, differed substantially between enzymes. Multiple independent purifications of the same enzyme resulted in identical activity patterns. Finally, the risk of cross contamination between purified enzymes was eliminated by the use of freshly prepared columns and affinity chromatography resin.

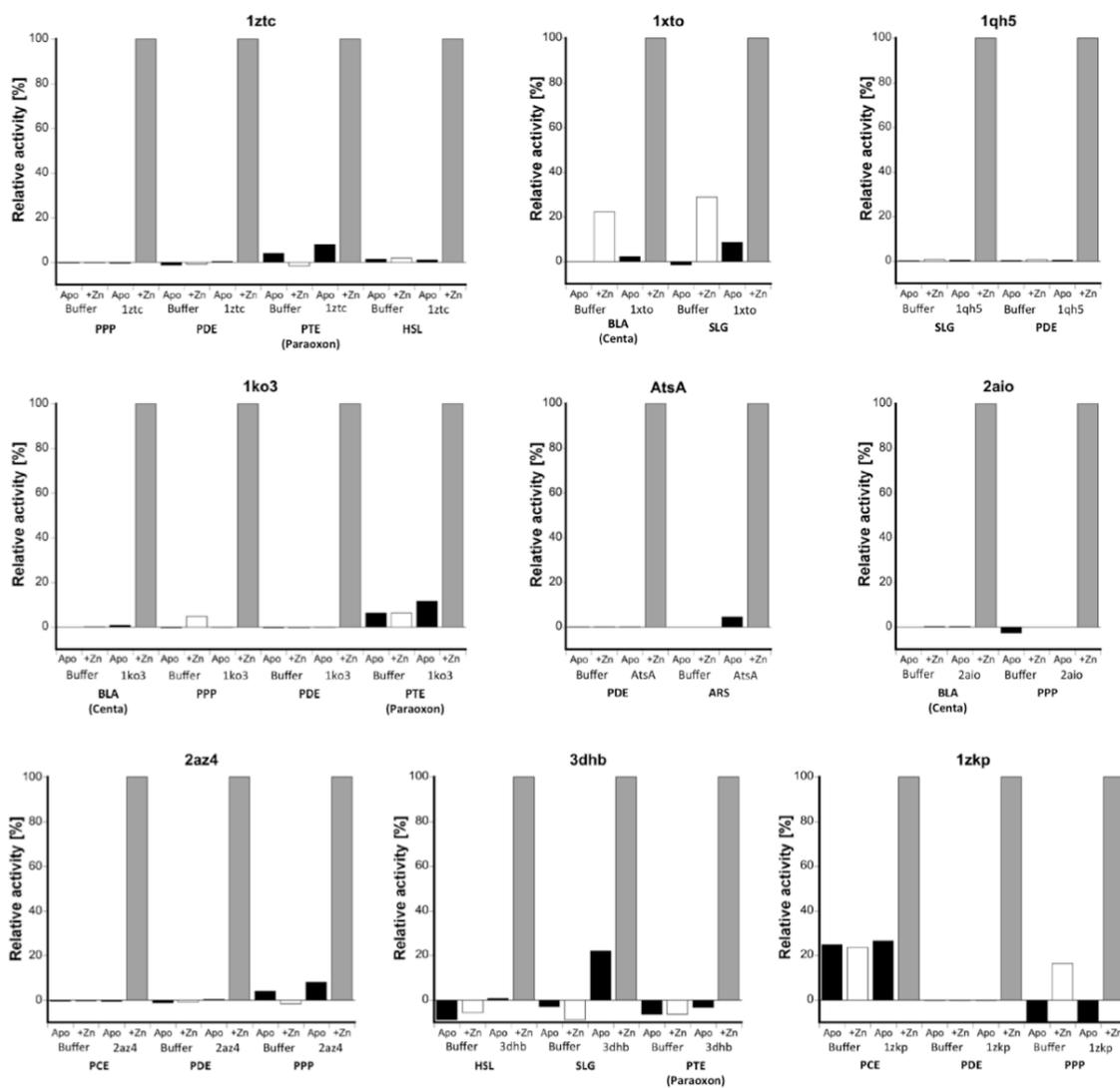


Figure 2.9 Metal chelating control experiment for nine selected enzymes.

Relative activities are based on metal reconstituted enzyme, which were used as a reference to calculate activity of the apo-enzyme and background reactions in buffer with and without metal supplied. Activities were calculated from initial rate enhancements [nM/s] at a single substrate concentration.

2.4.4 Function connectivity and evolutionary divergence

The function profiling analysis of 24 enzymes and 10 reactions revealed many functional connections through enzyme promiscuity between distinct enzyme functions of the MBL superfamily (**Figure 2.7**). Comparing these connections to the sequence relationship analysis suggest that some connections are related to the evolutionary divergence of functional families (**Figure 2.4**). For example, the arylsulfatase, AtsA, promiscuously hydrolyzes PDE in addition to its native ARS reaction (**Figure 4.2 A**). The arylsulfatase is localized in close proximity to the RNase Z cluster (the native PDE) in the sequence similarity network described in chapter two (**Figure 2.4**). Thus, function connectivity between ARS and PDE is potentially linked to evolutionary divergence of arylsulfatases and ribonucleases in the MBL superfamily. A similar pattern holds for PDE and PPP; the RNase Z cluster is adjacent to the phosphonate metabolic enzyme cluster (the native PPP) in the SSN (**Figure 2.4**), and two RNase Z enzymes catalyze PPP promiscuously (**Figure 2.7**). Interestingly, similar crosswise shared promiscuous activities between PDE, PPP and ARS have also been observed in the alkaline phosphatase superfamily (Mohamed & Hollfelder 2013), supporting the functional relationship between them. We suspect that the chemical similarity between the substrates of PDE and PPP (P-O bond, anionic substrates, similar size and structure) most likely causes their commonly shared promiscuous activities beyond evolutionary relationship of functional families. Indeed, most enzymes, which catalyze PDE (8 out of 10), can also promiscuously catalyze PPP. Another paired example of potentially evolutionary related promiscuity is PTE and HSL. The sequence cluster of the methyl-parathion hydrolase (the native PTE enzyme) is localized next to the lactonase cluster (**Figure 2.4**), and all three lactonases (HSL) exhibited promiscuous PTE activity. However, the methyl-parathion hydrolase did not exhibit HSL activity in this analysis (only when the active site metal ions are replaced with Ni^{2+} , as described in chapter three (**Figure 3.5**)). Shared catalytic activities between HSL and PTE have been also observed in two different enzyme superfamilies, the amidohydrolase superfamily and paraoxonase family (Elias & Tawfik 2011). Although the PTE and HSL reactions do not have the same hydrolyzed bound (P-O bond versus C-O bond) and transition-state geometry (trigonal bipyramidal versus tetrahedral), it has been suggested that a similar binding

mode of both reaction intermediates of N-acyl homoserine lactone and paraoxon causes the extensive connectivity between the two reactions (Elias & Tawfik 2011). Nevertheless, many functional connections in the MBL superfamily are largely supported by promiscuous activities that appear unrelated to direct evolutionary divergence. For example, native β -lactamases showed various levels of promiscuous PDE, PPP and PTE activities (**Figure 2.7**). However, the native functional families of these activities are not closely located to either β -lactamase cluster, thus β -lactamases are unlikely to have evolved or diverged from a PDE, PPP or PTE enzymes. Moreover, these three reactions possess distinct chemical properties to BLA in terms of type and charge of functional groups and transition-state geometry (**Figure 2.4**). Similarly, in addition to the shared HSL and PTE activities, one lactonase showed activity toward SLG, one for BLA and one for PPP and PDE (**Figure 2.7**). The two native glyoxalase II enzyme homologues (native SLG) also exhibit different promiscuous activities, one having BLA activity and the other one having PDE activity (**Figure 2.7**). Again, these promiscuous activities are not associated with functional families that are in close proximity in the sequence similarity network and have different chemical properties and therefore are likely to have occurred serendipitously.

2.4.5 Relationship between function profiles and active site features

Several reactions are catalysed by a number of enzymes with diverse sequences and native functions. For example, along with seven native β -lactamases of two distinct clusters, eight enzymes with different native functions were able to hydrolyse BLA (**Figure 2.7**). They are widespread in sequence space and not closely located to native β -lactamases in the sequence similarity networks (**Figure 2.4**). Interestingly, structural characterisation of active sites shows that the BLA reaction can be realized within active sites that display large a large variation in volume and hydrophobicity (**Figure 2.10**). Although Centa (a chromogenic β -lactam substrate) is hydrolysed non-enzymatically in the buffer at a relatively high rate ($k_{\text{uncat}} = 10^{-5} \text{ s}^{-1}$; **Table 2.4**), the catalytic proficiency conveyed by these promiscuous enzymes is still significant ($10^6 - 10^9 \text{ M}^{-1}$). However, a high frequency of promiscuous activities is not restricted to reactions with high k_{uncat} rates. Three reactions (PTE, PDE and PPP) with low uncatalysed rates ($k_{\text{uncat}} = 10^{-7} \sim 10^{-11} \text{ s}^{-1}$; **Table 2.4**) were also catalyzed by several enzymes with distinct native functions

and widespread in sequence space. Our structure characterization showed that diverse active sites, varying in volume and hydrophobicity, are able to catalyse the widespread PDE reaction. We also observe that native PDE and PCE (both are charged substrates) enzymes have generally more hydrophilic active sites, however no strong tendency was observed for enzymes that catalyse PDE promiscuously (**Figure 2.10**). Beyond our analysis of active site volume and hydrophobicity, we were unable to identify any specific active site features, which might contribute to the occurrence of a promiscuous activity, even between relatively close clusters that share promiscuous activities.

Table 2.4 Information on enzymatic reactions and substrates.

Reaction	Substrate	k_{uncat} [s^{-1}]	Molecular mass [Da]	Bond hydrolyzed	Bond charge	Overall charge	logP	Transition state geometry
BLA	Centa	6×10^{-5}	536	C-N	0	-1	1.85	tetrahedral
TPN	Chlorothalonil	n.d.	266	C-Cl	0	0	3.04	tetrahedral
AKS	Sodium dodecyl sulfate	n.d.	288	S-O	-1	-1	3.86	trigonal bipyramidal
SLG	S-lactoyl-glyoxal	7×10^{-6}	379	C-S	0	0	-2.39	tetrahedral
ARS	<i>p</i> -nitrophenyl sulfate	1×10^{-9}	218	S-O	-1	-1	0.92	trigonal bipyramidal
PDE	Bis-(<i>p</i> -nitrophenyl)-phosphate	2×10^{-8}	340	P-O	-1	-1	1.56	trigonal bipyramidal
PPP	<i>p</i> -nitrophenyl phenylphosphonate	2×10^{-11}	279	P-O	-1	-1	2.23	trigonal bipyramidal
PCE	<i>p</i> -nitrophenyl phosphoryl choline	2×10^{-8}	304	P-O	-1	0	-2.45	trigonal bipyramidal
PTE	Parathion-methyl	1×10^{-7}	263	P-O	0	0	3.84	trigonal bipyramidal
PTE	Paraoxon	5×10^{-8}	275	P-O	0	0	2.01	trigonal bipyramidal
HSL	N-(3-Oxoocanoyl)-DL-homoserine lactone	n.d.	241	C-O	0	0	0.19	tetrahedral

Abbreviation: n.d, not determined; abbreviations of reactions are listed in **Figure 2.6**.

Rate constants for PPP, ARS and PTE (Paraoxon) were obtained from (van Loo et al. 2010).

Other rate constants were calculated from background rates measured at pH 7.5 25°C in the presence of 100 mM NaCl, 50 mM Tris/HCl and 200 μ M ZnCl₂.

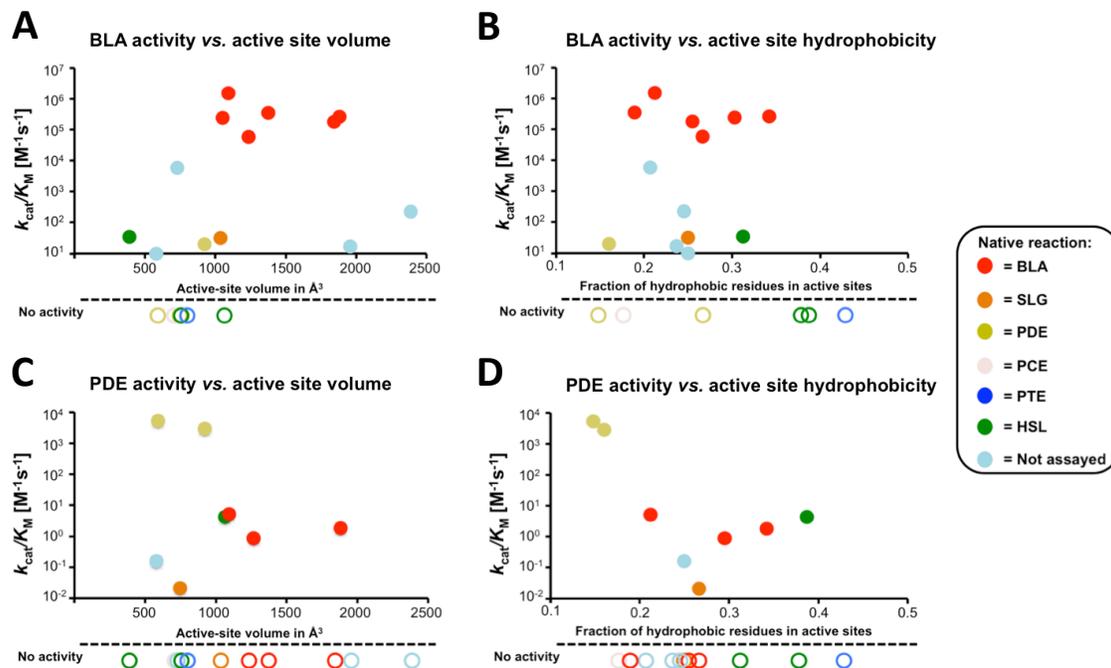


Figure 2.10 Relationship of two activities, BLA and PDE, and active-site properties.

The active site volume and fraction of hydrophobic residues in the active sites of 20 enzymes were plotted against BLA (A and B) and PDE (C and D) activities. Each enzyme is shown as a filled circle (if active) or an open circle (if not active) and colored by its native activity. The native reactions of four enzymes have not been assayed (light blue color), two enzymes (PDB ID 1vjn and 3h3e) have no identified native reaction and for two enzymes (PDB ID 2gcu and 1xto) the native reactions have not been assayed in this study. Four enzymes were excluded from the calculation, two enzymes with additional domains attached to the active site (PDB ID 2az4 and 2cfu) and the two enzymes without structural information (Gene ID AtsA and ChD). Surface representations of active site cavities for all enzymes are visualized in **Figure 2.2** and values for all enzymes for active site hydrophobicity and volume are listed in **Table 2.1**.

2.5 Discussion

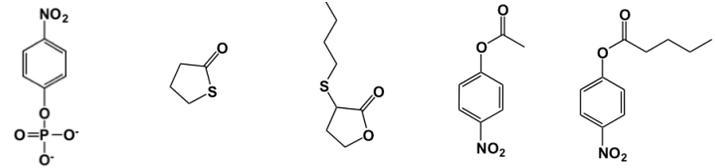
What is the molecular basis for the observed catalytic promiscuity and connectivity between the enzyme functions? Most members of the MBL superfamily possess a binuclear active site centre (generally two Zn^{2+} ions), which plays an essential role in catalysis by activating a water molecule for a nucleophilic attack as well as stabilising the charge of the ground and/or transition state. The same mechanistic feature seems to be used for promiscuous activities, because metal ions appear to be critical for both native and promiscuous activities (**Figure 2.9**). Not surprisingly, several promiscuous enzymes

from other superfamilies are also metal-dependent, thus the plasticity of active site metal ions may facilitate the introduction of promiscuous activities in general (van Loo et al. 2010; Ben-David et al. 2012; Ben-David et al. 2013; Aharoni et al. 2005; Nielsen et al. 2011; Sánchez-Moreno et al. 2009). As indicated by our study and others, lower catalytic efficiencies of promiscuous activities are mostly manifested in k_{cat} (**Figure 2.8**). Thus, we speculate that for promiscuous reactions, the substrates may be oriented sub-optimally relative to the activated water molecule, leading to >1,000-fold lower turnover rates compared to native reactions (Khersonsky & Tawfik 2010; Babbitt et al. 2010). However, to understand how enzymes are able to catalyse distinct chemical reactions within one active site, more detailed mechanistic and structural studies are required. Laboratory experiments aiming to evolve the observed promiscuous activities in the MBL superfamily might reveal how enzymes adapt to the new a function and its mechanistic requirements (Tracewell & Arnold 2009; Afriat-Jurnou et al. 2012; Tokuriki et al. 2012; Meier et al. 2013).

The existence of different promiscuous activity patterns in homologous enzymes with the same native function, demonstrates an important role of genetic diversity in the innovation and evolution of new functions (Wagner 2008). For example, the seven native β -lactamases showed distinct promiscuity patterns: some can hydrolyse only their native reaction (BLA), while others can carry out three additional reactions. Such variation has also been observed in neutral drift experiments which demonstrated that the level of promiscuous activities can differ significantly between variants with the same native activity (Bloom et al. 2007; Amitai et al. 2007). Similarly, a single activity was observed in enzymes with different native functions, indicating that new catalytic activities could arise from various sources. For example, the eight enzymes that promiscuously hydrolyse BLA are widespread in sequence space. Such promiscuous activities could provide an immediate and evolvable organismal advantage against β -lactam antibiotics in an environment where antibiotics are present (Soo et al. 2011; J. Davies & D. Davies 2010). In fact, evolution has independently developed β -lactamases in the MBL superfamily twice (B1/B2 and B3 clusters), in addition to the non-metal serine β -lactamase, which possesses a distinct fold and have catalytic serine instead of metal ions in the active site, but also hydrolyse β -lactam antibiotics (Bebrone 2007; Dellus-Gur et al. 2015). Similarly,

efficient phosphotriesterases, which hydrolyse organophosphate pesticide, have evolved among various superfamilies, including the MBL and amidohydrolase superfamilies, and from distinct ancestral functions such as lactonase, esterase, peptidase and exonuclease (Singh 2009). Thus, a diverse repertoire of existing enzymes and the range of latent promiscuous activities that stems from these enzymes potentially provides a large reservoir of evolutionary starting points. In turn, when generating novel enzymes in the laboratory, enzyme engineers may explore more genetic diversity, such as homologous enzymes from different organisms, in order to identify better starting points to engineer new enzymes with a desired function (O'Loughlin 2006).

Our study comprises a comprehensive superfamily-wide analysis of catalytic promiscuity. Nevertheless, we explore only a tiny subset of sequences and functions in the MBL superfamily. Thus, it is likely that the results in this study are just the “tip of the iceberg”. Indeed, the examination of several non-native MBL reactions, which do not define catalytic activities of any known MBL member (phosphatase, arylesterase and thiolactonase), revealed that several enzymes from our dataset are able to promiscuously catalyze these reactions (**Figure 2.11**). Without a doubt, more comprehensive analyses are likely to elucidate an even higher degree of enzyme promiscuity and functional connectivity. Such efforts may, in turn, help to improve functional assignment and can lead to the discovery of yet unidentified functions within a superfamily and novel enzyme properties for useful applications (H. Huang et al. 2015; Bastard et al. 2014; Mashiyama et al. 2014).



PDB ID or protein name	Subfamily	pNP-phosphate	γ -thiobutyro lactone	Thiobutylbutyrolactone	pNP-acetate	pNP-butyrate
		k_{cat}/K_M [$M^{-1}s^{-1}$]				
1xm8	Glyoxolase II enzyme	n.d.	850 \pm 220	1700 \pm 600	n.d.	n.d.
1qh5	Glyoxolase II enzyme	n.d.	150 \pm 30	16 \pm 4	1.5 \pm 0.2	0.4 \pm 0.1
1zkp	RNase Z	n.d.	n.d.	n.d.	n.d.	n.d.
2cbn	RNase Z	n.d.	n.d.	n.d.	0.6 \pm 0.2	0.3 \pm 0.04
3aj3	Lactonase	n.d.	280 \pm 40	80 \pm 20	3 \pm 0.8	4 \pm 0.3
3dhb	Lactonase	n.d.	33 \pm 12	26 \pm 5	1.1 \pm 0.5	4 \pm 0.6
1p9e	Methyl parathion hydrolase	n.d.	n.d.	n.d.	2 \pm 0.6	9 \pm 1
2cfu	Alkylsulfatas	n.d.	n.d.	n.d.	8 \pm 2	4 \pm 0.6
1xto	PqqB protein	n.d.	70 \pm 12	120 \pm 25	17 \pm 3	0.2 \pm 0.05
1wra	Phosphorylcholine esterase	n.d.	n.d.	n.d.	6 \pm 0.6	0.9 \pm 0.4
3h3e	hypothetical	n.d.	n.d.	n.d.	0.5 \pm 0.1	0.3 \pm 0.1
2az4	beta-CASP Rnase	n.d.	n.d.	n.d.	n.d.	n.d.
1vjn	hypothetical	n.d.	n.d.	n.d.	n.d.	n.d.
AtsA	Arylsulfatase	0.6 \pm 0.1	n.d.	n.d.	0.8 \pm 0.2	0.9 \pm 0.1

Figure 2.11 Kinetic parameters of 14 selected enzymes against 5 non-MBL superfamily reactions.

*p*NP stands for *para*-nitrophenol and n.d. for not detected. The means and standard deviation of the kinetic parameters were calculated from at least three independent measurements.

Chapter 3: Metal ion cofactor mediated catalytic promiscuity of enzymes in the metallo- β -lactamase superfamilies

Parts of chapter three have been performed in collaboration with M. Solomonson in the laboratory of Dr. N. C. J. Strynadka at UBC, Vancouver, Canada, and together with J. Chen in the laboratory of Dr. N. Tokuriki at UBC, Vancouver, Canada and published as “Baier F., Chen J., Solomonson M., Strynadka N. C. J., Tokuriki N. (2015): Distinct metal isoforms underlie promiscuous activity profiles of metalloenzymes. *ACS Chem. Biol.*, 10 (7), 1684–1693.” M. Solomonson performed ICP-MS analysis for quantitative metal content analysis as described in **Table 3.2**. J. Chen worked as an undergraduate student in the laboratory under my co-supervision and he performed most of the protein purifications, enzyme kinetics and metal ion titration experiments. I performed all other experiments and wrote the manuscript together with my supervisor, Dr. Nobuhiko Tokuriki, and some help of the other co-authors.

3.1 Summary

Within a superfamily, functionally diverged metalloenzymes often favor different metals as cofactors for catalysis. One hypothesis is that incorporation of alternative metals expands the catalytic repertoire of metalloenzymes and provides evolutionary springboards towards new catalytic functions. However, there is little experimental evidence that incorporation of alternative metal ions changes the function profile of metalloenzymes. Here, we systematically investigate how metals alter the function profiles of five functionally diverged enzymes of the metallo- β -lactamase (MBL) superfamily. Each enzyme was reconstituted *in vitro* with six different metals, Cd^{2+} , Co^{2+} , Fe^{2+} , Mn^{2+} , Ni^{2+} and Zn^{2+} , and assayed against eight catalytically distinct hydrolytic reactions (representing native functions of MBL enzymes). We reveal that each enzyme metal isoform has a significantly different activity level for native and promiscuous reactions. Moreover, metal preferences for native versus promiscuous activities are not correlated, and in some cases are mutually exclusive, *i.e.* only particular metal isoforms disclose cryptic promiscuous activities, but at the expense of the native activity. For example, the L1 B3 β -lactamase displays a 1000-fold catalytic preference for Zn^{2+} over

Ni²⁺ for its native activity, but exhibits promiscuous thioester, phosphodiester, phosphotriester and lactonase activity only with Ni²⁺. Furthermore, we find that the five MBL enzymes exist as an ensemble of various metal isoforms in *Escherichia coli*, and this heterogeneity results in an expanded function profile compared to a single metal isoform. Our study suggests that promiscuous activities of metalloenzymes can stem from an ensemble of metal isoforms in the cell, which could facilitate the functional divergence of metalloenzymes.

3.2 Introduction

Metalloenzymes depend on metal ions as co-factors to catalyze chemical reactions, and many of which require a specific metal ions to confer efficient catalytic activity. However, our knowledge of why and how enzymes select particular metals for their native function is still limited (Foster et al. 2014; Valdez et al. 2014). The evolution of metal preferences can partially be explained by bioavailability (Waldron & Robinson 2009; Waldron et al. 2009; Carter et al. 2011; Xu et al. 2008). Yet, some enzymes are only catalytically active with less bioavailable metals in the environment and cell, such as Ni²⁺ and Co²⁺ (Kobayashi & Shimizu 1999; Ragsdale 2009; Waldron et al. 2009). Furthermore, evolutionarily related enzymes of the same superfamily, including the amidohydrolase (Seibert & Raushel 2005), cupin (Dunwell et al. 2001), enolase (Gerlt & Babbitt 2001), vicinal-oxygen-chelate (Gerlt & Babbitt 2001) and MBL superfamily (Bebrone 2007; Daiyasu et al. 2001), utilize different metals for distinct catalytic functions, suggesting that metal preferences for catalysis were related to functional divergence.

How is the evolution of metal preferences associated with the functional divergence of metalloenzymes? A simple hypothesis is that promiscuous binding of alternative metals could alter the function profile of metalloenzymes and promote catalysis of non-native promiscuous activities. If the increase of the new metal-dependent promiscuous activity provides a fitness advantage to the organism, mutations could further enhance the new activity and give rise to an enzyme with a novel function and metal preference (Khersonsky & Tawfik 2010). Several studies have demonstrated that binding of an alternative metal can alter the activity level of a metalloenzyme towards non-native substrates (catalysis of the same chemical reaction with structurally distinct

substrates). For example, B1 and B3 β -lactamases of the MBL superfamily change their substrate specificity toward various β -lactam antibiotics depending on the active site metal (Badarau & Page 2006; Hu, Periyannan, et al. 2008; Hu et al. 2009). In addition to these examples of metal-dependent changes in substrate specificity, there are reported examples of metal substitution resulting in metal-dependent catalysis of catalytically distinct reactions (catalytic promiscuity). For example, the dihydroxyacetone kinase (DHAK) from *Citrobacter freundii* catalyzes the transfer of the γ -phosphate of ATP to dihydroxyacetone in the presence of Mg^{2+} , but exhibits cyclase activity towards flavin adenine dinucleotide (FAD) in the presence of Mn^{2+} (Sánchez-Moreno et al. 2009). The *N*-succinyl-L,L-diaminopimelate desuccinylase (DapE) from *Salmonella enterica* is Zn^{2+} -dependent. However, DapE exhibits promiscuous aspartyl dipeptidase activity if substituted with Mn^{2+} (Broder & C. G. Miller 2003). The human carbonic anhydrase, which catalyzes the hydration of carbon dioxide to carbonate using Zn^{2+} , is transformed into an epoxide synthase (Fernández-Gacio et al. 2006) and a bicarbonate-dependent peroxidase when reconstituted with Mn^{2+} (Okrasa & Kazlauskas 2006), or a reductase with Rh^{2+} (Jing et al. 2009). Furthermore, an enzyme within the cupin superfamily yields two different oxidation products of the acireductone substrate with different metals (Dai et al. 1999). Substitution with Ni^{2+} yields 1,3-oxygenlytic reaction activity whereas Fe^{2+} yields 1,2-oxygenlytic reaction activity (Dai et al. 1999). Together, these studies indicate that substitution of active site metals has the potential to alter the function profile of metalloenzymes. However, the focus of these studies was on individual enzyme and function pairs, and no study performed a comprehensive characterization of how metal substitution affects the function profile, including several native and promiscuous activities, of functionally distinct enzymes of an enzyme superfamily. Studying the effect of metal ion substitution on enzyme function profiles could reveal hidden evolutionary and functional connections that are only observable in the presence of certain metal ions and might yield insights how metal different metal ion requirements evolved in a superfamily and might.

To address these questions, we focus on the metal utilization of (MBL) enzymes, which catalyze a wide variety of hydrolytic reactions using different metal ions as cofactors (Bebrone 2007; Daiyasu et al. 2001). MBL enzymes share a common $\alpha\beta\alpha$ -fold

(MBL-fold) and a conserved active site metal binding motif (**Figure 3.1 A**) (Bebrone 2007). In general, two divalent metals bind in the active site; the first site (M1) is coordinated by three His residues, the second site (M2) consists of His and Asp residues, in addition, a bridging Asp residue that coordinates both metals (**Figure 3.1 B**) (Karsisiotis et al. 2014). B1 and B3 β -lactamases lack the bridging Asp residue, which is substituted by a Cys or Ser (non-coordinating) residue, respectively (Karsisiotis et al. 2014). The two active site metals have two primary roles in catalysis that are similar in all hydrolytic MBL enzymes: (i) activate a hydroxide ion for nucleophilic attack and (ii) stabilize the ground and transition state as Lewis acids (**Figure 3.1 C**) (Karsisiotis et al. 2014). While many MBL enzymes are described to be Zn^{2+} -dependent enzymes, several are most catalytically efficient with other metals such as Co^{2+} , Ni^{2+} , Mn^{2+} and Fe^{2+} (Bebrone 2007; Badarau & Page 2006; Limphong et al. 2009; Dong et al. 2005; D. Liu et al. 2008; Holdorf et al. 2012; Condon & Gilet 2011). For example, the *Escherichia coli* ribonuclease Z is most efficient in its Co^{2+} isoform (Dutta & Deutscher 2009). The human and plant glyoxalase II prefers a heterogeneous Zn^{2+}/Fe^{2+} center for the catalysis, but is also active with Ni^{+} and Co^{2+} (Broder & C. G. Miller 2003; Limphong et al. 2009). Moreover, Mn^{2+} is optimal for the *E. coli* L-ascorbate-6-P lactonase (UlaG) (Badarau & Page 2006; Garces et al. 2010; Hu, Periyannan, et al. 2008; Hu et al. 2009) and the *Pseudomonas aeruginosa* quinolone signaling response protein (Yu et al. 2009).

In the previous chapter, we reported the systematic characterization of 24 members of the MBL superfamily for distinct MBL reactions and demonstrated that the enzymes exhibit various promiscuous activities (Baier & Tokuriki 2014). This promiscuity resulted in a relative high functional connectivity between the investigated reactions. Thus, enzymes of the MBL superfamily represent an excellent model to investigate the effect of metal substitutions on the function profile of metalloenzymes due to the variety of reactions and distinct metal preferences. The aim of this chapter is to reveal how metal ion availability and substitution affects enzyme function-profiles, and thus functional connectivity through promiscuity. In detail, we assayed five functionally distinct MBL enzymes, each substituted with six metal ions, against eight MBL reactions, which provides a comprehensive dataset of 240 combinations. We also measure the metal content and function profiles of the five enzymes directly purified from *Escherichia coli*

and compare the data to the single metal isoform. Furthermore, we investigate the effect of metal ion supplementation to the expression media, which discloses if available metal ions are incorporated and can alter function profiles. Finally, we reconstitute combinations of heterogeneous metal isoforms and assay their function profile towards native and promiscuous activities.

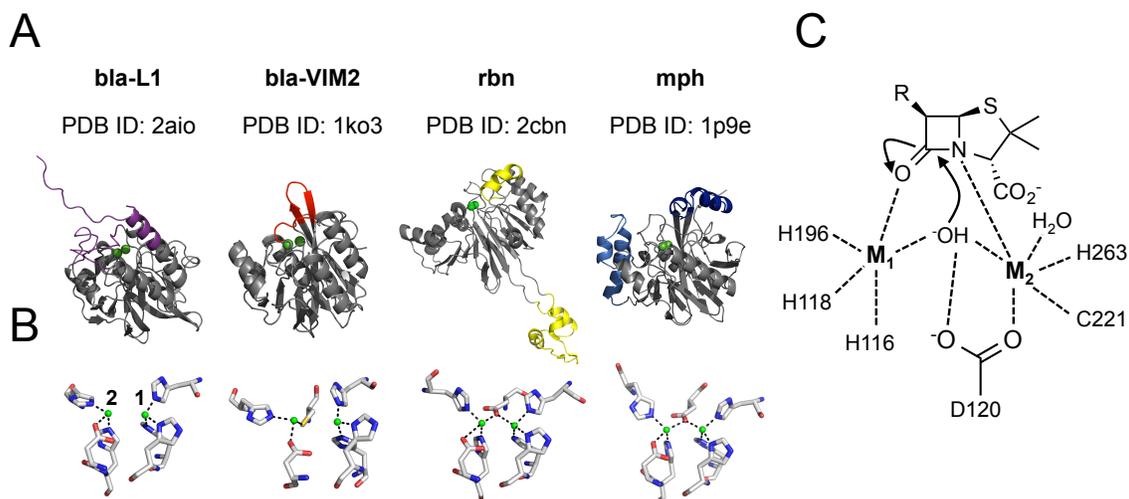


Figure 3.1 Structures of selected enzymes and the general catalytic mechanism of MBL enzymes.

(A) Cartoon representation of structures. The conserved backbone structure of the MBL superfamily is shown in grey and metals are shown as green spheres. Unique structural features for each enzyme are highlighted in color. (B) Metal binding coordination in the active site. Metal positions, M1 and M2 site, are indicated for bla-L1. (C) The proposed catalytic mechanism of B1 β -lactamases. Both M1 and M2, metals are involved in orientating and activating the hydroxide ion required for nucleophilic attack on the substrate. In particular, M1 polarizes the carbonyl group of the β -lactam ring of the substrate and, by forming an oxyanion hole, stabilizes the transition state. M2 is suggested to position the substrate, polarize the amide bond and stabilize the leaving group through interaction with the β -lactam nitrogen (Karsisiotis et al. 2014). The residues are numbered according to the B1 β -lactamase family (Bebrone 2007).

3.3 Methods

3.3.1 Chemicals

The substrates for the reactions ARS, EST, PCE, PDE, PTE and SLG, as listed in **Figure 3.3**, as well as the Ellman's reagent DTNB (5,5'-dithio-bis-[2-nitrobenzoic acid]) were purchased from Sigma-Aldrich. Centa was purchased from EMD Millipore and TBBL (5-

(thiobutyl) butyrolactone) was kindly provided by Dan S. Tawfik. All other reagents and materials were purchased as indicated.

3.3.2 Molecular cloning

The genes were cloned into the pET27(b)-Strep or -MBP (maltose binding protein) N-terminal tag vector using *NcoI* and *HindIII* restriction sites (Fermentas). The pET27(b)-Strep vector was created by inserting the Strep-tag II sequence (MASWSHPQFEKGAG) into the pET27(b) vector (Novagen), using *NdeI* and *BamHI* restriction sites. The pET27(b)-MBP vector was created by replacing the Strep-tag II sequence with the MBP-tag sequence from the pMAL-c2e vector (New England Biolabs), using *NdeI* and *BamHI* restriction sites. All DNA constructs were confirmed by DNA sequencing. *Mph* and *bla-VIM2* were over-expressed as Strep-tag fusion proteins and *bla-L1*, *atsA* and *rbn* as MBP-tag fusion proteins.

3.3.3 Protein expression and purification

All enzymes were transformed and over-expressed in *E. coli* BL21 (DE3) cells. 800 ml of LB media (Fisher BioReagents) supplemented with 40 µg/ml kanamycin (Fisher BioReagents) was inoculated with 20 ml overnight culture. Cells were further grown at 30°C for 3~4 hours until reaching an OD₆₀₀ of 0.6-0.8. Protein expression was induced by adding 0.8 mM IPTG and cultures were incubated at 20°C for 16 hours. Cells were harvested by centrifugation at 10,000 × g and pellets were frozen at -80°C for at least 2 hours before lysis. For lysis, cell pellets were resuspended in lysis buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, containing 50 % B-PER protein extraction reagent (Thermo Scientific), 100 µg/ml lysozyme (EMD millipore) and 1 U/ml of benzonase (Novagen)) and incubated on ice for 1 hour. Cell lysates were clarified by centrifugation at 30,000×g for 20 min at 4°C. Affinity tag purification of the Strep- or MBP-tag fusion proteins was performed according to the manufacturers' protocol with Strep-tactin resin (IBA lifesciences) and Maltose resin (New England Biolabs), respectively. Proteins were eluted in elution buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl) supplemented with 2.5 mM Desthiobiotin (Sigma-Aldrich) for Strep-tag and 10 mM maltose (Sigma-Aldrich) for MBP-tag purifications. The fractions containing proteins were pooled and

concentrated to a volume of 3 ml (Microsep 10 kDa, Pall) and then desalted using Econo-Pac 10DG columns (Bio-Rad). The concentration of each purified enzyme was determined by measuring absorbance at 280 nm using the proteins' specific extinction coefficient (Baier & Tokuriki 2014). All purifications yielded >90% pure protein, which was verified with SDS-PAGE.

3.3.4 Preparation of apo- and metal-substituted enzymes

To generate metal free apoenzymes, the purified proteins were subjected to three rounds of chelation (>8 hours for each round) with a metal chelator cocktail (5 mM 1,10-phenanthroline (Sigma-Aldrich) and 10 mM EDTA (Fisher BioReagents)) and subsequent removal of chelator and metal using Econo-Pac 10DG columns (Bio-Rad). Metal removal and subsequent analysis was performed in 20 mM HEPES pH 7.5 with 100 mM NaCl. All buffers were treated with Chelex 100 (Bio-Rad) to reduce metal contamination. Metal reconstitution was performed by incubating the apoenzymes with 200 μ M of Cd(II)Cl₂, Co(II)Cl₂, NH₄Fe(II)SO₄, Mn(II)Cl₂, Ni(II)Cl₂ and Zn(II)Cl₂ (all Sigma-Aldrich), respectively, for at least 1 h prior to activity measurement.

3.3.5 Enzyme assays and kinetics

BLA activity was assayed at 10 μ M enzyme and 50 μ M substrate concentrations in 50 mM Tris-HCl pH 7.5, 100 mM NaCl supplemented with 0.2% Triton X-100 and the corresponding metal at 100 μ M. The other seven catalytic activities were assayed at 10 μ M enzyme and 500 μ M substrate concentrations with the same buffer conditions. The enzymatic reactions PCE, PDE, ARS, PTE and EST were monitored following the release of *p*-nitrophenol at 405 nm and molar product formation was calculated with an extinction coefficient of 18,300 M⁻¹cm⁻¹. The BLA reaction was monitored at 405 nm for the Centa substrate with the extinction coefficient 6,300 M⁻¹cm⁻¹ (Bebrone et al. 2001). Hydrolysis of SLG and LAC, which expose a free thio-group upon hydrolysis, was followed with Ellman's reagent at 412 nm using an extinction coefficient of 14,150 M⁻¹cm⁻¹ (Riddles et al. 1983). Kinetic constants were determined as follows: initial rate measurements at various substrate concentrations were performed in duplicate and the

data averaged. The Michaelis-Menten equation was then fitted to the data using Kaleidograph software (Synergy).

3.3.6 Metal content analysis

For metal content analysis, enzymes were cultured in standard LB media. After purification, as described above, the enzymes were desalted twice using Econo-Pac 10DG columns (Bio-Rad) in 20 mM HEPES pH 7.5 with 100 mM NaCl. Metal content was measured using an inductively coupled plasma mass spectrometer (NexION 300D ICP-MS, PerkinElmer Life Sciences) and the data was analyzed with NexION software. A calibration standard (CAT# IV-STOCK-4, Inorganic Ventures) containing metals of interest (Cd, Co, Mg, Mn, Ni, Fe, Zn) was diluted with internal standard solution, 10 µg/L Sc and 1% nitric acid (CAT# IV-ICPMS-71D, Inorganic Ventures), and used to generate standard curves that ranged from 1 to 100 µg/L for each metal. Proteins were digested at 115°C in closed vessels with concentrated trace metal-grade nitric acid for 24 hours and for a further 24 hours in nitric acid with hydrogen peroxide, followed by complete evaporation. Dried protein samples were resuspended in internal standard solution (10 µg/L Sc and 1% nitric acid). Data were collected using standard mode, except for Fe, which was detected in dynamic reaction cell mode using ammonia gas.

3.3.7 Lysate activity analysis

Individual wells of a 96-well plate containing 400 µl of LB media supplemented with 40 µg/ml kanamycin were inoculated with 20 µl of overnight culture and incubated at 30°C for 2 hours. Metals were added to a final concentration of 100 µM, and cultures were incubated for 1 hour. Protein expression was induced with 1 mM IPTG for 3 hours. Cells were harvested by centrifugation at 4,000 × g and pellets were frozen -80°C for at least 30 min. For lysis, cell pellets were resuspended in lysis buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, containing 0.1 % Triton X-100, 100 µg/ml lysozyme and 1 U/ml of benzonase) and incubated at 25°C with shaking at 1200 rpm for 1 hour. The cell lysates were clarified by centrifugation at 4,000×g for 20 min at 4°C. Clarified lysates were diluted in order to obtain linear initial rates and measured against a single substrate concentration (50 µM for BLA and 500 µM for all other substrates). For this analysis the

lysates were diluted 1000-fold for the enzymes bla-VIM2, bla-L1 and rbn, 100-fold for atsA and 10-fold for mph.

3.4 Results

3.4.1 Experimental dataset

We selected five enzymes of the MBL superfamily for our analysis: the *Salmonella maltophilia* L1 B3 β -lactamase (bla-L1) (Avison et al. 2001; Spencer et al. 2005), the *P. aeruginosa* VIM2 B1 β -lactamase (bla-VIM2) (Poirel et al. 2000; Garcia-Saez et al. 2008), the *E. coli* ribonuclease Z (rbn) (Vogel 2002; Kostecky et al. 2006), the *Alteromonas carragenovora* arylsulfatase (atsA) (Barbeyron et al. 1995) and the *P. aeruginosa* methyl-parathion hydrolase (mph) (Dong et al. 2005) (**Table 3.1**). Rbn, mph, and atsA (no structure available) have a typical MBL metal binding motif, whereas the two β -lactamases, bla-L1 and bla-VIM2, differ in the M2 site and lack the bridging Asp residue (**Figure 1 B**). The five enzymes represent different functional families and are highly diverged in sequence, with pairwise identity of less than 12% (**Table 2.2**). While bla-L1 and bla-VIM2 possess the same physiological function, they belong to distantly related evolutionary families (**Figure 3.2**). We chose eight hydrolytic reactions for the function profiling (**Figure 3.3**): arylsulfatase (ARS; E.C. 3.1.6.1), β -lactamase (BLA; E.C. 3.5.2.6), esterase/lipase (EST; E.C. 3.1.1.1), N-acyl homoserine lactonase (LAC; E.C. 3.1.1.81), phosphorylcholine esterase (PCE; E.C. 3.1.4.1), phosphodiesterase (PDE; E.C. 3.1.26.11), phosphotriesterase (PTE; E.C. 3.1.8.1) and glyoxalase II (SLG; E.C. 3.1.2.6). Four of these reactions are native reactions of the five selected enzymes (**Table 2.1**): BLA for bla-L1 and bla-VIM2, PDE for rbn, PTE for mph and ARS for atsA. The remaining four reactions are native activities of alternative members of the MBL superfamily and were observed as promiscuous activities in chapter two (**Figure 2.7**). The eight substrates differ in terms of their scissile bonds (C-N (BLA), C-O (EST and LAC), C-S (SLG), P-O (PCE, PDE and PTE) and S-O (ARS)) their chemical properties such as hydrophobicity, steric property, overall charge, transition state geometry, and their leaving group (**Figure 3.3**). PTE and SLG represent the enzymes' natural substrate, whereas the other reactions were measured with generic chromogenic substrates.

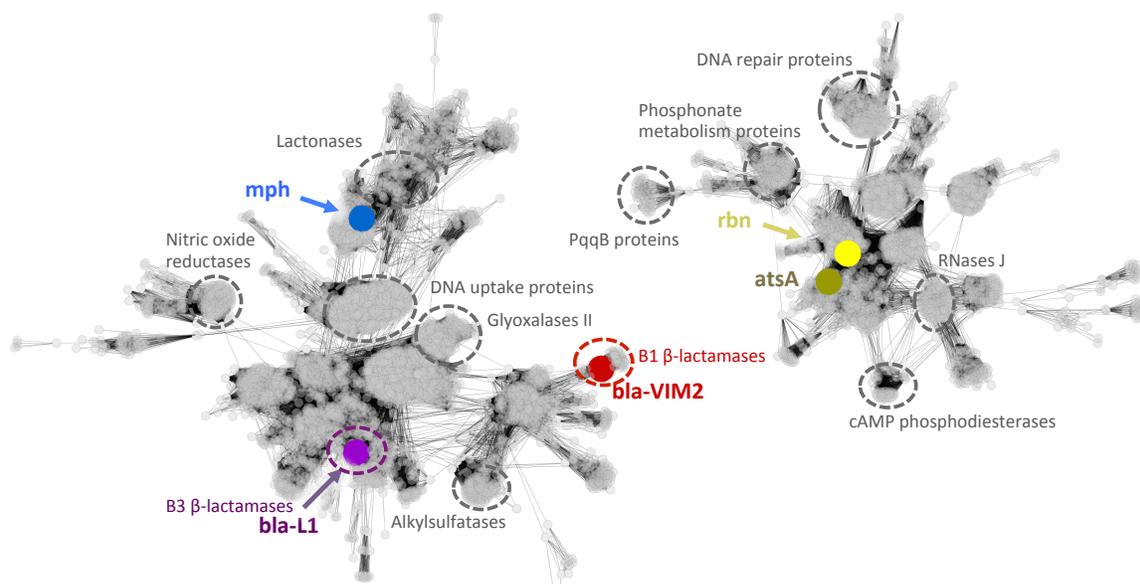


Figure 3.2 Sequence relationships of selected enzymes within the MBL superfamily.

7488 sequences (nodes) and lines (edges) show sequence similarity of functional clusters within the MBL superfamily at a BLAST E-value cutoff of $1 \times e^{-14}$. Large colored nodes indicate the enzymes that were experimentally characterized in this study. Dashed grey circles indicate several functional sequence clusters that have previously been experimentally characterized and reported in the literature, but have not been included in this study. The network was generated as described in chapter 2, using an initial set of 49,879 amino acid sequences of the metallo- β -lactamase superfamily (Pfam-IDs: PF00753, PF12706, PF13483) retrieved from the Pfam database on 25th of March 2014 and applying a sequence identity cutoff of 50% using CD-Hit (Y. Huang et al. 2010).

Table 3.1 Information on enzymes used in this study.

Enzyme	Full enzyme name	Uniprot ID	Organism of origin	Residues ^a	Molecular function	Native reaction ^b
bla-L1	L1 B3 β -lactamase	P52700	<i>Salmonella maltophilia</i>	268 (23-290)	β -lactam hydrolysis	BLA
bla-VIM2	VIM2 B1 β -lactamase	Q9K2N0	<i>Pseudomonas aeruginosa</i>	220 (27-266)	β -lactam hydrolysis	BLA
rbn	Ribonuclease BN	P0A8V0	<i>Escherichia coli</i>	305 (1-305)	tRNA processing	PDE
mph	Methyl-parathion hydrolase	Q841S6	<i>Pseudomonas aeruginosa</i>	331 (1-331)	Methyl-parathion hydrolysis	PTE
atsA	Arylsulfatase	P28607	<i>Alteromonas carrageenovora</i>	305 (24-328)	Desulfatation of polysaccharides	ARS

^aNumber of residues expressed excluding the N-terminal affinity tag. The number in brackets indicate the expressed residues of each sequence as annotated in the Uniprot database (www.uniprot.org).

^bAbbreviation of reactions used are explained in **Figure 3.3**.

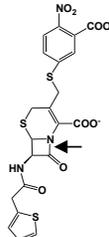
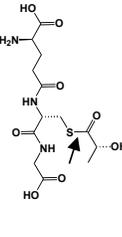
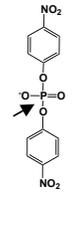
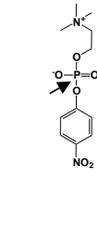
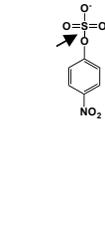
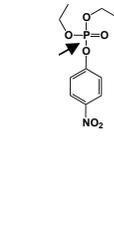
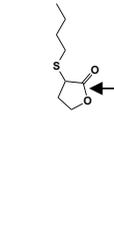
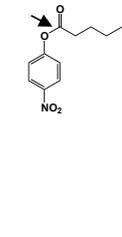
Reaction	β -lactamase	Glyoxalase II	Phosphodiesterase	Phosphorylcholine-esterase	Aryl-sulfatase	Phosphotriesterase	N-Acylhomoserine lactonase	Esterase/Lipase
Reaction abbreviation	BLA	SLG	PDE	PCE	ARS	PTE	LAC	EST
E.C. number of reaction	3.5.2.6	3.1.2.6	3.1.26.11	3.1.4.1	3.1.6.1	3.1.8.1	3.1.1.81	3.1.1.1
								

Figure 3.3 Enzymatic reactions and substrates used in this study.

BLA: Centa; SLG: S-(lactoyl)-glutathione; PDE: Bis-(*p*-nitrophenyl)-phosphate; PCE: *p*-nitrophenyl-phosphoryl-choline; ARS: *p*-nitrophenyl-sulfate; PTE: Paraoxon; LAC: 5-(thiobutyl)-butyrolactone; EST: *p*-nitrophenyl-butyrate. Three-letter code abbreviation and color scheme for each reaction is used throughout the study. The four-digit code represents the E.C. number classification of each reaction. Arrows indicates the bond that is broken during hydrolysis

3.4.2 Reconstitution of enzymes with various metals and activity screening

All enzymes were over-expressed and purified from *E. coli* using Streptavidin-Biotin (IBA) affinity chromatography. Metal-free apoenzymes were prepared by three iterative cycles of chelating and desalting. Subsequently, the apoenzymes were incubated with six different metals (Cd^{2+} , Co^{2+} , Fe^{2+} , Mn^{2+} , Ni^{2+} and Zn^{2+}) to generate metal reconstituted holoenzymes. The metal-free apoenzymes were confirmed by comparison to holoenzymes' activity levels, with apoenzymes activities being below 0.05% when compared to the holoenzymes (**Figure 3.4**). In total, 240 conditions (5 enzymes \times 6 metals \times 8 reactions) were screened. The steady-state kinetic parameters (k_{cat} , K_{M} and $k_{\text{cat}}/K_{\text{M}}$) were determined for enzyme/metal/reaction combinations that showed detectable activity, as defined by a 10-fold higher rate compared to the background reaction (**Figure 3.5**).

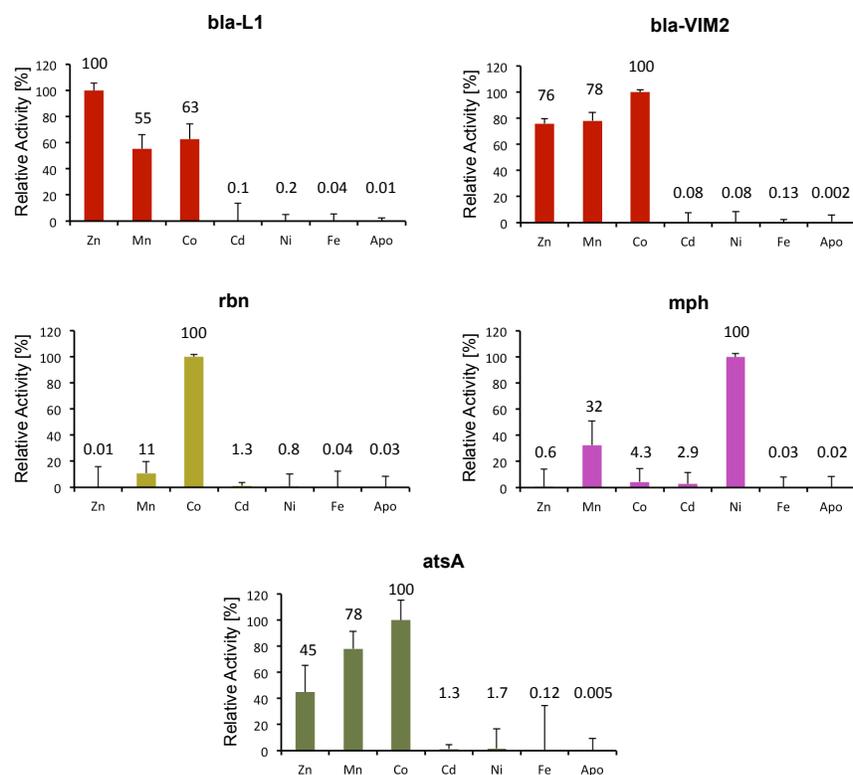


Figure 3.4 Activity levels of metal-depleted apo- and metal reconstituted enzymes.

The highest activity was set to 100 [%] for each enzyme. The initial velocity for bla-L1 with Zn^{2+} was 56 nM/s and for bla-VIM2 with Co^{2+} was 10 nM/s for BLA with 5 nM enzyme. For rbn the initial velocity with Co^{2+} for PDE was 413 nM/s with 5 nM enzyme. The initial velocity of mph with Ni^{2+} for EST was 6 nM/s with 5 nM enzyme. For atsA with Co^{2+} the initial velocity for ARS was 88 nM/s with 50 nM enzyme. Error bars indicate standard deviation of the mean from triplicate measurements.

The detection limit for enzymatic activities varied slightly depending on the substrate ($k_{cat}/K_M = 10^{-1} \sim 10^{-3} \text{ s}^{-1}M^{-1}$), because of the non-enzymatic hydrolysis (k_{uncat}) of the substrates ($10^{-5} \text{ s}^{-1} \sim 10^{-9} \text{ s}^{-1}$) (Table 2.4). The rate of non-enzymatic hydrolysis of the substrates was not altered by the presence of the investigated metals in the buffer. Furthermore, several lines of evidence support that the observed activities are not artefacts caused by other protein or metal contaminants. First, all five enzymes exhibit completely different metal-dependent function profiles and serve as internal controls, eliminating the possibility of non-specific enzyme and metal contaminations (Figure

3.5).

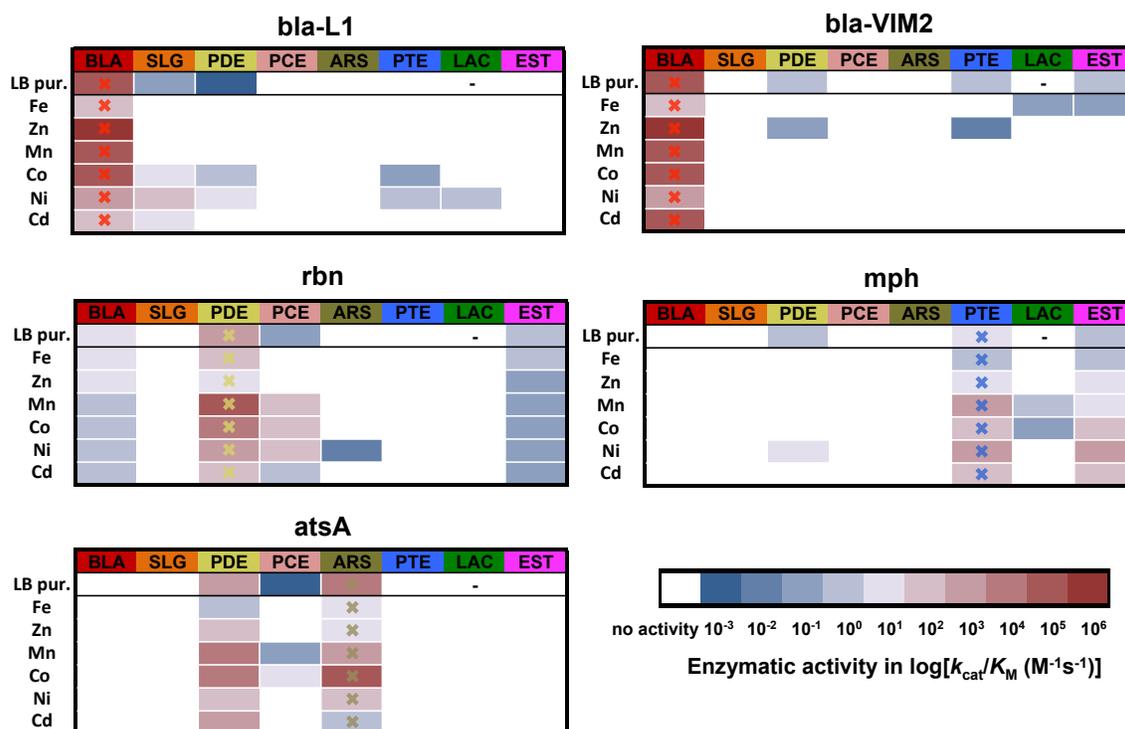


Figure 3.5 Function profiles of five MBL enzymes reconstituted with various metals.

Heat maps show catalytic efficiencies (k_{cat}/K_M) for each enzyme and metal combination. *LB pur.* indicates kinetic parameters were obtained from untreated enzymes purified from *E. coli* cells cultured in standard LB media for which the metal content is shown in **Table 3.2**. The listed metals indicate that the generated apoenzyme was reconstituted *in vitro* with the respective metal. The three-digit code represents reactions as described in **Figure 3.3**. Colored crosses indicate the native activity for each enzyme. The dash for LAC indicates that the kinetic parameters were not determined. The kinetic parameters are listed in **appendix B**.

Second, the K_M value, which is unique for each enzyme/metal/substrate combination, differed substantially between enzymes and metal isoforms. Third, multiple independent purifications resulted in identical function profiles and the use of freshly prepared affinity chromatography columns eliminated the risk of cross contamination between enzymes. Moreover, we note that no glassware was used to prepare buffers and solutions and all buffers were chelated to remove trace metals.

3.4.3 MBL enzymes exhibit distinctive metal preferences for their native activity

All five enzymes catalyzed their native reaction with every metal investigated. However, the catalytic efficiencies ($k_{\text{cat}}/K_{\text{M}}$) varied by up to 10,000-fold depending on the incorporated metal (**Figure 3.5**). The two β -lactamases, bla-L1 and bla-VIM2, show the highest catalytic efficiency for their native BLA in the presence of Zn^{2+} (bla-L1 with $k_{\text{cat}}/K_{\text{M}} = 3.7 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$ and bla-VIM2 with $k_{\text{cat}}/K_{\text{M}} = 2.0 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$). Furthermore, their metal preferences for BLA have a similar order: $\text{Zn}^{2+} > \text{Co}^{2+} \approx \text{Mn}^{2+} > \text{Ni}^{2+} \approx \text{Cd}^{2+} \approx \text{Fe}^{2+}$ for bla-L1 and $\text{Zn}^{2+} \approx \text{Co}^{2+} \approx \text{Mn}^{2+} \approx \text{Cd}^{2+} > \text{Ni}^{2+} > \text{Fe}^{2+}$ for bla-VIM2 (**Figure 3.5**). We denote “>” for over 10-fold and “ \approx ” for less than 10-fold difference in $k_{\text{cat}}/K_{\text{M}}$ between metal isoforms. Rbn (native PDE) shows a different metal preference: the highest catalytic activity is obtained with Mn^{2+} ($k_{\text{cat}}/K_{\text{M}} = 4.0 \times 10^5 \text{ M}^{-1}\text{s}^{-1}$), and follows the order of $\text{Mn}^{2+} \approx \text{Co}^{2+} > \text{Ni}^{2+} \approx \text{Cd}^{2+} \approx \text{Fe}^{2+} > \text{Zn}^{2+}$ (**Figure 3.5**). Mph (native PTE) preferentially catalyzes PTE with Ni^{2+} ($k_{\text{cat}}/K_{\text{M}} = 3.6 \times 10^3 \text{ M}^{-1}\text{s}^{-1}$) and Mn^{2+} ($k_{\text{cat}}/K_{\text{M}} = 2.9 \times 10^3 \text{ M}^{-1}\text{s}^{-1}$), and displays the metal preference order: $\text{Ni}^{2+} \approx \text{Mn}^{2+} \approx \text{Cd}^{2+} > \text{Co}^{2+} \approx \text{Zn}^{2+} > \text{Fe}^{2+}$ (**Figure 3.5**). AtsA (native ARS) exhibits the highest catalytic activity with Co^{2+} ($k_{\text{cat}}/K_{\text{M}} = 1.3 \times 10^5 \text{ M}^{-1}\text{s}^{-1}$), and follows $\text{Co}^{2+} > \text{Mn}^{2+} > \text{Ni}^{2+} \approx \text{Zn}^{2+} \approx \text{Fe}^{2+} \approx \text{Cd}^{2+}$ (**Figure 3.5**). Taken together, the investigated MBL enzymes possess substantially different metal preferences for their native activities.

3.4.4 Metal substitution alters function profiles and exposes cryptic promiscuous activities

All five enzymes exhibit promiscuous activity towards several non-native reactions (**Figure 3.5**). The metal preferences for promiscuous activities differ significantly from the native one, and metal substitution significantly change the function profiles of the investigated enzymes (**Figure 3.5**). As described above, bla-L1 catalyzed its native BLA reaction most efficiently and with high specificity in the presence of Zn^{2+} , but we did not detect any promiscuous activity with Zn^{2+} (**Figure 3.5**). However, reconstitution of bla-L1 with Cd^{2+} , Co^{2+} or Ni^{2+} revealed several cryptic promiscuous activities (**Figure 3.5**). The Co^{2+} reconstituted bla-L1 showed promiscuous activities to SLG, PDE, PTE and Cd^{2+} -bla-L1 displayed activity for SLG. Four promiscuous activities, SLG, PDE, PTE

and LAC, are observed with Ni^{2+} and thus Ni^{2+} -bla-L1 exhibits a highly promiscuous function profile. The $k_{\text{cat}}/K_{\text{M}}$ values of all five reactions of Ni^{2+} -bla-L1, native and promiscuous, are within three orders magnitude (from $k_{\text{cat}}/K_{\text{M}} = 3$ to $1,700 \text{ M}^{-1}\text{s}^{-1}$; **Figure 3.5**). We also performed metal titration experiments for bla-L1 with Zn^{2+} for its native BLA and with Ni^{2+} for promiscuous PDE, PTE and SLG activities. All activities are fully activated with around 2 equivalent of metals to the apoenzyme, which confirmed that these activities are activated by the examined metals (**Figure 3.6**). Bla-VIM2 exhibits promiscuous activities towards PDE, PTE, LAC and EST, which were dependent upon the incorporated metal (**Figure 3.5**). PDE and PTE activities are only observed in the presence of Zn^{2+} , which is the preferred metal for the native BLA activity. However, the promiscuous activities are $>10^7$ -fold lower compared to the native activity. LAC and EST are catalyzed only in the presence of Fe^{2+} , the least preferred metal for the native BLA. Thus, the Fe^{2+} isoform of bla-VIM2 represents a generalist enzyme with a promiscuous function profile and modest catalytic efficiencies; the differences in the three activities, native and promiscuous, are within three orders magnitude but the native BLA is 10,000-fold lower with Fe^{2+} than with Zn^{2+} . In addition to its native PDE activity, rbn exhibited four promiscuous activities: PCE, BLA, EST and ARS (**Figure 3.5**). PCE was detected in the presence of Mn^{2+} , Co^{2+} , Ni^{2+} and Cd^{2+} , with the same metal preference trend observed for PDE, but activities are 10 to 1000-fold lower. ARS activity was only detected in the presence of Ni^{2+} . BLA and EST are catalyzed by rbn in the presence of any of the six metals, and both showed only marginal preference for Zn^{2+} and Fe^{2+} . Mph promiscuously catalyzed EST, LAC and PDE in addition to its native PTE (**Figure 3.5**). EST was catalyzed by mph when reconstituted with any of the six investigated metals and follows a similar trend of metal preference as for its native activity. The catalytic efficiency for EST is only 2-fold lower than PTE in the presence of Ni^{2+} . Hence, the EST and PTE activities are correlated in terms of catalytic efficiency and metal preference in mph. The other two promiscuous activities were more dependent on particular metal isoforms; mph could only catalyse PDE with Ni^{2+} and LAC with Mn^{2+} and Co^{2+} . AtsA showed two promiscuous activities, PDE and PCE (**Figure 3.5**). PDE was catalysed in the presence of any of the six metals and follows the metal preference trend observed for the native ARS activity. PCE was catalysed only in the presence of Co^{2+} and Mn^{2+} and these metals were

also preferred for ARS. Overall, metal substitution significantly alters the function profile of the five investigated MBL enzymes. However, we observed no trend that suggests that certain metals promote the catalysis of a particular reaction. For example, all five investigated enzymes catalyse PDE, although with different metal preferences. PDE is preferentially catalyzed with Mn^{2+} in the case of the native phosphodiesterase *rbn*, but Ni^{2+} confers PDE for *mph* and *bla-L1*, whereas only Zn^{2+} supports PDE for *bla-VIM2*. Thus, we speculate that the metal preference for a particular reaction is highly associated with the active site architecture of each enzyme.

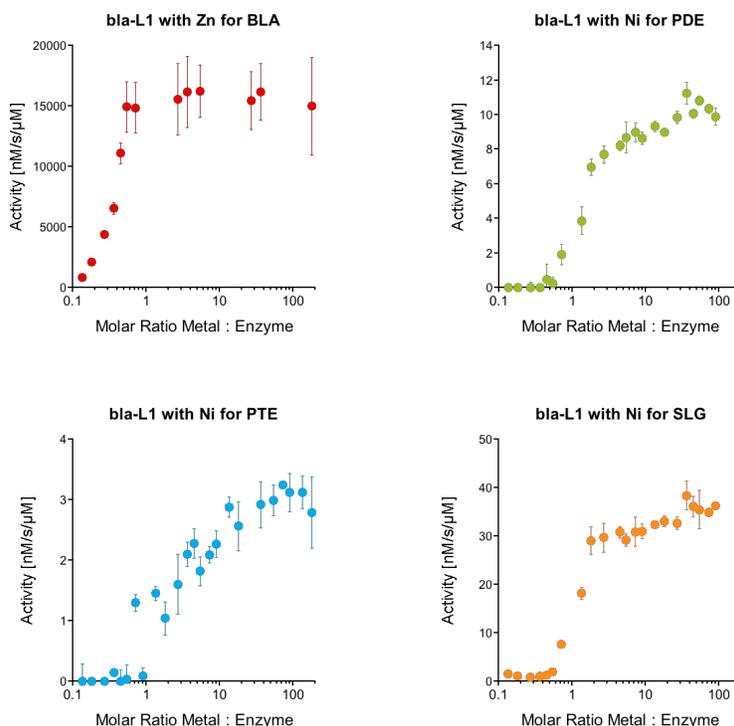


Figure 3.6 Metal activation of apo-bla-L1 for native and promiscuous activities.

Purified apo-bla-L1 was subjected to three additional passes through Econo-Pac 10DG columns to remove any residual metal and chelator. The apo-bla-L1 was incubated at 1 μM with various Zn^{2+} and Ni^{2+} concentrations ranging from 0 to 200 μM in 20 mM HEPES (pH 7.5) and 100 mM NaCl for 16 hours at 4°C. For BLA activity measurement the enzyme was diluted to 50 nM. Activity levels were normalized to 1 μM enzyme. Substrate concentrations were 100 μM for BLA and 500 μM for PDE, PTE and SLG. Titrations were performed in triplicate and the values were averaged. Error bars represent standard deviation.

3.4.5 MBL enzymes exist as heterogeneous metal isoforms in *E. coli*

To what degree do the enzymes selectively incorporate metals that confer highest activity for the native reaction in the cellular environment? Using ICP-MS (Inductively Coupled Plasma Mass Spectrometry), we examined the metal content of the MBL enzymes purified from *E. coli* cultured in standard LB media (untreated enzymes). ICP-MS data showed that all enzymes co-purify with a mixture of different metals, indicating that they most likely exist as an ensemble of various metal isoforms in the cell (**Table 3.2**).

Table 3.2 Quantitative metal content analysis.

Enzyme	<i>mol eq metal/enzyme</i> ^a					Total
	Fe	Mg	Mn	Ni	Zn	
bla-L1	0.5 ± 0.008	<i>n.d.</i>	0.01 ± 0.0002	0.01 ± 0.0002	0.3 ± 0.006	0.8
bla-VIM2	0.2 ± 0.002	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	0.1 ± 0.003	0.3
rbn	0.6 ± 0.002	<i>n.d.</i>	0.1 ± 0.001	<i>n.d.</i>	0.2 ± 0.005	0.9
mph	1.3 ± 0.03	0.4 ± 0.009	0.02 ± 0.0005	0.03 ± 0.0009	0.2 ± 0.004	1.9
atsA	2 ± 0.02	0.3 ± 0.006	0.2 ± 0.003	<i>n.d.</i>	0.1 ± 0.003	2.6

^a*mol eq metal/enzyme* indicates molar ratio of metal per enzyme.

n.d. means no significant (<0.005) amount detected.

Errors indicate the standard deviation of the mean from triplicate measurements.

Regardless of the metal preference for the native activity, considerable amounts of Fe²⁺ (0.2 to 2 equivalent amount per enzyme) and Zn²⁺ (0.1 to 0.3 equivalent) were bound to all enzymes, which is consistent with the abundance of these metals in the media: yeast extract and tryptone, the two major components of LB media, generally contain a higher amount of Mg²⁺, Fe²⁺, Zn²⁺ and Cu²⁺ than other metals.(Bovallius & Zacharias 1971; Grant & Pramer 1962) However, metals such as Mn²⁺ and Ni²⁺ also bind some of the enzymes (**Table 3.2**). In addition, we determined the function profile of the untreated enzymes by measuring their catalytic efficiencies. The untreated enzymes were less efficient in terms of catalytic efficiency but more promiscuous than when reconstituted with a single metal. For example, the two β-lactamases, bla-L1 and bla-VIM2, exhibited a 10-fold lower activity for BLA compared to the Zn²⁺-isoforms (**Figure 3.5**), which can be explained by the relatively low zinc content (0.3 and 0.1 equivalent; **Table 3.2**). However, both enzymes exhibit several metal-dependent promiscuous activities that are not conferred by the Zn²⁺-isoform. Bla-L1 shows promiscuous SLG and PDE activities (**Figure 3.5**), which may be caused by existence of the Ni²⁺-isoform, as some trace of

nickel (0.01 equivalent) was detected (**Table 3.2**). Bla-VIM2 exhibited PDE, PTE and EST promiscuous activities in addition to its native BLA (**Figure 3.5**). PDE and PTE activities are most likely to be dependent on the Zn²⁺-bla-VIM2, whereas EST activity is only conferred by Fe²⁺-bla-VIM2. Thus, enzymes incorporating Zn²⁺ or Fe²⁺ seem to be responsible for these promiscuous activities. The MBL enzymes are bi-metal enzymes, hence three configurations may explain the observed function profile: (i) individual enzymes are incorporating either 2 × Zn²⁺ or 2 × Fe²⁺, (ii) a mix of both metals occur in a single enzyme (different metal in M1 and M2 site, respectively) or alternatively, (iii) a combination of (i) and (ii) occurring in the same enzyme purification. Heterogeneous metal centres, *i.e.* configuration (ii), have been described for B1 and B3 β-lactamases and glyoxalases II of the MBL superfamily (Zang 2000; Hu et al. 2009; H. Yang et al. 2014). In addition, mph exhibits Ni²⁺-dependent PDE and atsA exhibited weak PCE activity, which is only detectable with the Mn²⁺- and Co²⁺-isoforms (**Figure 3.5**). It should be noted that the metal content obtained with ICP-MS analysis might not perfectly reflected the metal composition found in the active site of the enzymes. Some metal trace may be caused by non-specific binding to the enzyme, and on the other hand, some metal binding could be lost during the sample preparation, such as protein purification and buffer exchange with desalting columns. Moreover, heterologous overexpression of the enzymes in *E. coli* cultured in a synthetic media (LB) may not represent the metal content of the enzymes in the natural environment. Nonetheless, the results indicate that the metalloenzymes could exist as heterogeneous metal isoforms in the cell and the heterogeneity can result in relatively efficient, albeit promiscuous, function profiles.

3.4.6 Bioavailability of metals can alter function profiles

Next, we investigated whether environmental changes in metal concentration can alter the bioavailability of metals in the cell and subsequently change the abundance of particular metal isoforms and activity level of the enzymes. The MBL enzymes were over-expressed in *E. coli* in LB media supplemented with additional metals (at 100 μM). The cells were harvested, lysed in metal free buffer, and the enzymatic activity of the clarified lysate was measured. The native activities of all five enzymes changed significantly when different metals were supplied to the media (**Figure 3.7**). We confirmed that the

supplemented metals neither affect the growth rate of *E.coli* cells in our experiment (**Figure 3.8**) nor the level of the protein expression (**Figure 3.9**). Furthermore, the *E. coli* cell lysate itself showed no significant activity for the investigated reactions (**Figure 3.7**). The activity trends observed in the lysate experiments roughly follows the preference observed in the *in vitro* metal reconstitution experiment, *e.g.*, the supplement of Mn^{2+} and Co^{2+} substantially increases the native activity of *rbn*, *mph* and *atsA*, indicating that the supplemented metals are incorporated in the MBL enzymes and alter their activity level. It should be noted that four out of five enzymes are not native *E. coli* enzymes and they are over-expressed in an artificial environment (LB media). Nonetheless, these results imply that the enzymes have not evolved to be absolutely specific to the metal that confers the most efficient catalytic activity; they also bind alternative, available metals in the cell, which in turn substantially depends on metal bioavailability.

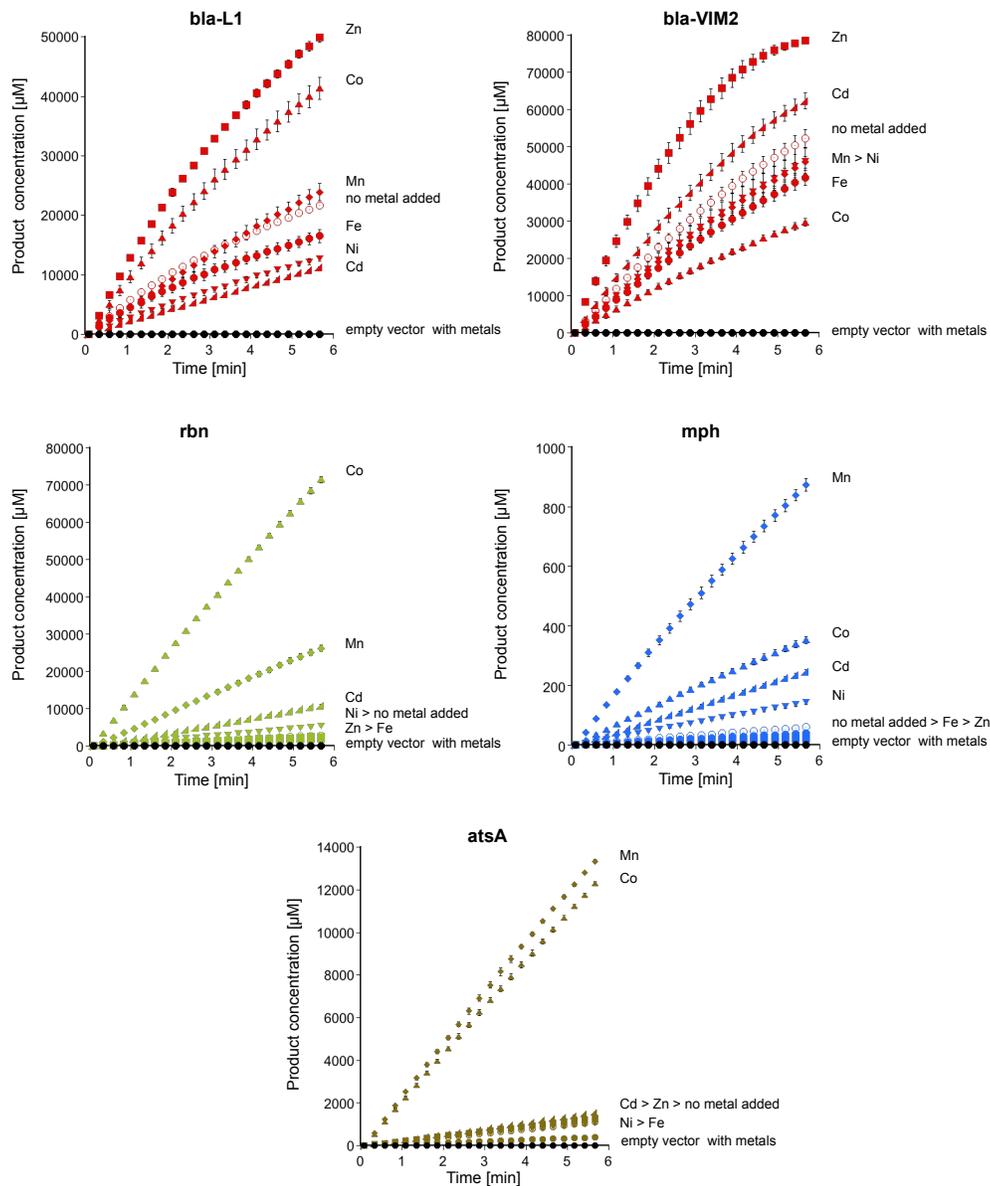


Figure 3.7 The effect of metal supplementation on enzymatic activities in cell lysate.

The enzymes were over-expressed in *E. coli* BL21 (DE3) cells in media supplemented with the respective metal. Native activity of each enzyme was measured in cell lysate together with an empty vector control with all respective metals. Experiments were performed in triplicate and the values were averaged. Error bars represent standard deviation. Growth curves and protein expressions (for each enzyme) for all metals are shown in **Figure 3.8** and **3.9**, respectively.

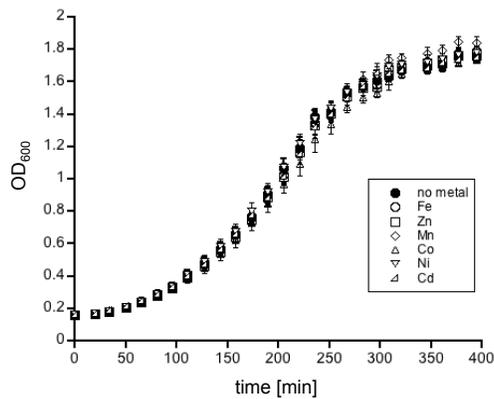


Figure 3.8 *E.coli* cell growth in the presence of various metals.

Growth curve with various metals at 100 μ M of *E.coli* BL21 (D3) cells containing an empty vector (pET28a). Cultures of 200 μ l were inoculated with 10 μ l overnight culture and grown at 30°C. Growth curves were performed in triplicate and the values were averaged. Error bars represent standard deviation.

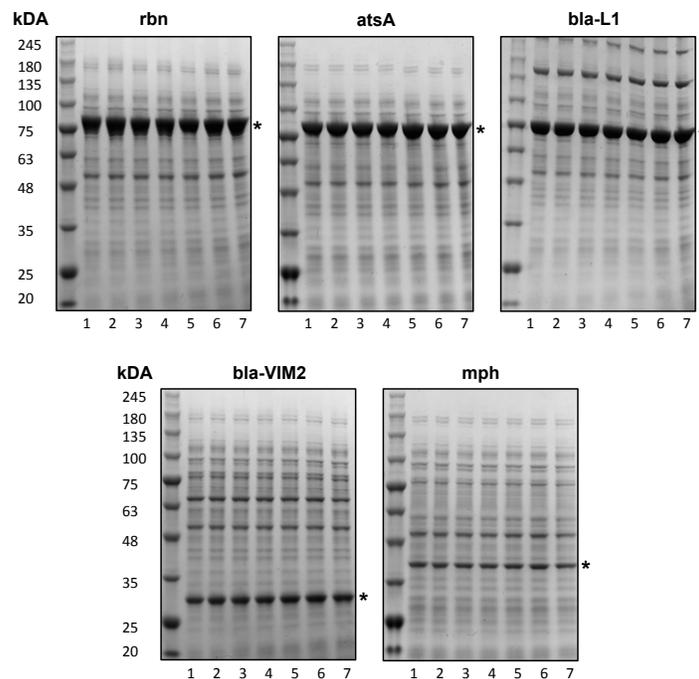


Figure 3.9 Expression level of the enzymes in the lysate activity experiment with different metals.

The soluble fraction for each enzyme and metal condition was loaded in equal volume (5 μ L) onto the gel. 1: Fe²⁺; 2: Zn²⁺; 3: Mn²⁺; 4: Co²⁺; 5: Ni²⁺; 6: Cd²⁺; 7: no metal added. Asterisks indicate the corresponding protein band of the fusion enzyme. Note that rbn, atsA and bla-L1 are expressed as MBP-tag (maltose binding protein) fusions, whereas bla-VIM2 and mph as Strep-tag fusion.

3.4.7 Reconstitution of heterogeneous metal isoforms

The lysate activity assay described above was not sensitive enough to quantify promiscuous activities due to low activity and high background levels in the cell lysate. Thus, we performed an *in vitro* metal reconstitution experiment with various combinations of two metals (Cd^{2+} , Co^{2+} , Fe^{2+} , Mn^{2+} , Ni^{2+} and Zn^{2+}). For this analysis, we selected bla-L1, which prefers Zn^{2+} for its native activity by over 1000-fold activity compared to Ni^{2+} but exhibits PDE only in the presence of Ni^{2+} , Co^{2+} and Cd^{2+} (**Figure 3.5**). Apo-bla-L1 was incubated with 21 combinations of two metals at equal concentrations and activities towards BLA and PDE reactions were subsequently measured (**Figure 3.10**). The enzyme efficiently hydrolyses BLA in the presence of Zn^{2+} regardless of the presence of other metals, which indicates that Zn^{2+} is preferentially incorporated and provides the highest activity. In contrast, promiscuous PDE activity is almost exclusively observed when Ni^{2+} is present in the combination; the only exception is the combination of $\text{Fe}^{2+}/\text{Ni}^{2+}$, which is inactive for PDE. The general binding preference of bla-L1 for Zn^{2+} and Fe^{2+} is also consistent with the metal content analysis using ICP-MS (**Table 3.2**). In addition, Zn^{2+} and Ni^{2+} seem to have similar relative affinities when added in combination to apo-bla-L1, as the combination of $\text{Zn}^{2+}/\text{Ni}^{2+}$ exhibits high activity for both the native and promiscuous activities (**Figure 3.10**). $\text{Zn}^{2+}/\text{Ni}^{2+}$ reconstituted bla-L1 displays 70% BLA activity relative to the most efficient combination and 50% PDE activity relative to the pure Ni^{2+} enzyme (**Figure 3.10**). Hence, various co-existing metal isoforms of bla-L1 ($\text{Zn}^{2+}/\text{Zn}^{2+}$, $\text{Ni}^{2+}/\text{Ni}^{2+}$, and perhaps $\text{Zn}^{2+}/\text{Ni}^{2+}$) were formed, resulting in high activity for both BLA and PDE.

bla-L1 with BLA

	Fe	Zn	Mn	Co	Ni	Cd
Fe	2 ± 0.3	73 ± 6	<1	3 ± 0.2	<1	<1
Zn		83 ± 14	72 ± 9	100 ± 14	70 ± 5	63 ± 13
Mn			4 ± 0.2	6 ± 2.4	<1	<1
Co				9 ± 0.1	3 ± 0.6	<1
Ni					<1	<1
Cd						<1

bla-L1 with PDE

	Fe	Zn	Mn	Co	Ni	Cd
Fe	<1	<1	<1	<1	2 ± 0.1	<1
Zn		<1	<1	<1	51 ± 0.2	<1
Mn			<1	2 ± 0.1	52 ± 1	<1
Co				5 ± 0.1	71 ± 0.2	<1
Ni					100 ± 1	49 ± 0.2
Cd						<1

Figure 3.10 The effect of metal combinations on the catalytic activities of bla-L1.

Apo-bla-L1 was incubated with various combinations of two metals (each metal was in 20-fold excess to enzymes) and assayed against the native BLA and promiscuous PDE activity. The highest activity (1600 nM/s for BLA and 180 nM/s for PDE) was set at 100% in each experiment. For activity measurement of BLA and PDE enzyme concentrations were 20 nM and 1 μM, respectively. Color shading (light for 1-9%, medium for 10-99% and dark for 100%) is used to highlight activity levels. Errors indicate standard deviation of the mean from triplicate metal incubations.

3.5 Discussion

Promiscuous activities of enzymes towards non-native substrates or reactions serve as evolutionary starting points towards new catalytic functions (Pandya et al. 2014; Khersonsky & Tawfik 2010). Several mechanisms that promote enzyme catalytic promiscuity have been proposed in the last decade, including conformational diversity, alternative mechanistic features and alternative binding modes of substrates in enzyme active sites (Babtie et al. 2010; Khersonsky & Tawfik 2010). Indeed, our previous study indicated that most MBL enzymes are promiscuous even with one kind of metal (Zn^{2+}) in the active site (Baier & Tokuriki 2014). However, our observations in this study suggest that binding of alternative metals significantly expands the catalytic repertoire of metalloenzymes, and thus provides further evolutionary connections between enzymatic functions.

Metal incorporation is commonly determined by a balance between metal bioavailability in the cell and relative affinity to the metals of the enzyme. Metal bioavailability could change significantly depending on environment and physiological condition of the cell whereas metal homeostasis largely controls the metal concentration in the cell. If certain metal isoforms cause a disadvantage to the organism, *e.g.*, enhancing harmful promiscuous activities, the enzyme would have evolved to incorporate favorable metals in order to discriminate against detrimental activities. In some cases, accessory proteins such as metallochaperones support incorporation of a particular metal by preferentially delivering metals from uptake systems (Waldron & Robinson 2009). The incorporation of alternative metals might be tolerated to some extent, as they may not be deleterious to the organism. For example, if metals that cause enzyme inactivity were incorporated by 50% of a metalloenzyme population, it would result only in an overall 2-fold reduction in catalytic rate, which generally would not affect organismal fitness significantly (Soskine & Tawfik 2010). Hence, heterogeneous metal isoform populations in the cellular milieu might be common for many, if not most, metalloenzymes (Imlay 2014; Cotruvo & Stubbe 2012; Waldron et al. 2009; Foster et al. 2014). In turn, different metal isoforms of enzymes may exhibit alternative activities that can confer an advantage to the organism during an environmental change and subsequent adaptive evolution.

It has been recognized that promiscuous activities play an essential role in functional divergence by providing an evolutionary starting point for new functions (Khersonsky & Tawfik 2010; Pandya et al. 2014; Brown & Babbitt 2014). Thus, promiscuous activities can be seen as evidence of evolutionary connectivity between enzymatic functions or potential for future evolutionary pathways (Mohamed & Hollfelder 2013; Baier & Tokuriki 2014; Afriat et al. 2006; Tokuriki et al. 2012). Our observations indicate that metal isoform heterogeneity can expand the repertoire of promiscuous activities and thus enhance the evolvability of enzymes. *S. enterica* *N*-succinyl-L,L-diaminopimelate desuccinylase (DapE) represents a case in which metal isoform heterogeneity supports bi-functionality in the cell. The Zn²⁺-isoform of DapE is natively involved in lysine biosynthesis but Mn²⁺-DapE exhibits promiscuous aspartyl dipeptidase activity (Broder & C. G. Miller 2003). Broder *et al.* showed that over-expression of DapE in *E. coli* can compensate for an aspartyl dipeptidase knockout and

therefore DapE must exist as both Zn²⁺ and Mn²⁺-isoforms within the cell (Broder & C. G. Miller 2003). In a scenario where a metal-dependent promiscuous activity provides a selective advantage for the organism, gene duplication and adaptive mutations may promote the evolution of a new enzyme in order to enhance the promiscuous activity. Eventually, a newly evolved enzyme may possess a metal preference that is different from the ancestral enzymes via the acquisition of mutations that increase affinity to a particular metal. Although no explicit case of a metal-dependent functional evolution has been reported to date, the diversity of metal preferences that are observed within many enzyme superfamilies suggests that such a scenario is likely (Seibert & Raushel 2005; Dunwell et al. 2001; Gerlt & Babbitt 2001).

What is the underlying molecular basis for metal-dependent promiscuous activities? Members of the MBL superfamily share a similar catalytic mechanism for hydrolysis in which the active site metal ions lower the pK_a of a water molecule for a nucleophilic attack on the scissile bond as well as stabilize the charge of the ground and/or transition state (**Figure 3.1 C**) (Karsisiotis et al. 2014). Metal ions have distinct physical and chemical properties such as radius, electronegativity and preferred coordination state (Rulísek & Vondrásek 1998). These properties could affect the structure of the active site, the position of the nucleophile (hydroxide ion) and/or the reactivity of catalytic machinery (Valdez et al. 2014). The Kamerlin research group at the Uppsala University studied the molecular mechanism and metal ion selectivity of MPH for PTE (native activity) and EST (promiscuous activity) activity, through computational calculations based on an empirical valence bond approach (Purg et al. 2016), using the experimental data of this chapter as an experimental support (Baier et al. 2015). The results show that both reactions, PTE and EST, appear to proceed through nucleophilic attack by a metal activated terminal hydroxide ion for MPH. Furthermore, the large active site volume of MPH allows the enzyme to accommodate the geometric constraints for in-line nucleophilic attack for PTE and an approximately 81° angle of attack for EST, even though this requires very different binding modes for the two substrates (Purg et al. 2016). The study also examined the metal ion dependent activity patterns of MPH with five different divalent transition metal ions, namely Co²⁺, Fe²⁺, Mn²⁺, Ni²⁺ and Zn²⁺. The computational calculations were able to reproduce experimental metal-ion-dependent

activity patterns for both substrates, and demonstrate that the origin of this effect appears to be primarily differences in the electrostatic properties of the metals themselves, coupled with very subtle changes in substrate and transition state geometries, which affect charge distributions at the transition state and corresponding transition state stabilization, rather than large rearrangements of metal ion coordination or active site architecture (Purg et al. 2016). While subtle, these differences can nevertheless be sufficient to make the difference between whether a cryptic promiscuous activity is exposed with a particular metal ion or not. Therefore, substrate- function profiles can be altered through judicious selection of different metal ions in the catalytic center.

In addition, previous studies suggested that differences in metal ion coordination geometry introduces or eliminates additional water molecule(s) in the active site, which could be the cause for some of the observed metal ion dependent function profiles (Rulísek & Vondrásek 1998). For example, *E. coli* glyoxalase I is active when Ni²⁺, Co²⁺ and Cd²⁺ is incorporated but not active with Zn²⁺ (Clugston et al. 2004). A change in coordination geometry is postulated to produce these contrasting activation profiles; from tetrahedral (Zn²⁺) to octahedral (Ni²⁺, Co²⁺ and Cd²⁺) coordination (He et al. 2000). In our study bla-L1 is highly specific and efficient for BLA with Zn²⁺, which predominantly adopts a tetrahedral coordination. In contrast, bla-L1 exhibits promiscuous activities (SLG, PDE, PTE and LAC) with Co²⁺, Ni²⁺ and Cd²⁺, which are normally coordinated in octahedral geometry. We speculate that each enzyme-metal ion-reaction pair has different requirements and further detailed mechanistic studies of metal ion mediated promiscuity are needed to elucidate the underlying mechanisms. Interestingly, several directed evolution experiments of metalloenzymes that enhance a promiscuous activity resulted in displacement of the metal position in the active site. These results indicate that subtle change in the active site metal (position or nature of the metal) can cause significant change in the function profile. In turn, the phenomenon we describe here may be a useful tool for protein engineers to explore the effect of metal substitutions, as it could provide a significant increase of target function(s) (Pordea & Ward 2009; Pordea 2015).

Chapter 4: Function connectivity in enzyme superfamilies

4.1 Summary

The sequence and functional diversity of enzyme superfamilies have expanded through billions of years of evolution from a common ancestor. Understanding how protein sequence and functional “space” have expanded, at both the evolutionary and molecular level, is central to biochemistry, molecular and evolutionary biology. Integrative approaches that examine protein sequence, structure, and function have begun to provide comprehensive views of the functional diversity and evolutionary relationships within enzyme superfamilies. In this chapter, I analyze five studies, including the one described in chapter two and three, which performed large-scale function profiling of functionally diverse enzyme superfamilies, which together revealed the native and promiscuous activities of hundreds of enzymes. Using function connectivity networks, we visualize the connection and relationship between functions as a result of enzyme promiscuity. Several intriguing insights from this recent body of work and the network analysis emerge. First, promiscuous activities are prevalent among extant enzymes. Second, despite the high function connectivity through promiscuity, many functions are not directly connected, but through ‘intermediate functions’, *i.e.* two functions are not connected by the promiscuity of a single enzyme, but through at least one other function and two other enzymes. Third, function connectivity appears to be related to evolutionary divergence in some instances, but generally appears to be due to substrate and reaction similarity. Finally, I will discuss how structural, environmental and genetic factors can influence function connectivity and ultimately enzyme evolution.

4.2 Introduction

Sequence classification, together with the mapping of biochemical and structural properties, can reveal the evolutionary relationships between functional families and help to elucidate the history of functional divergence. In addition, “promiscuous functions” of enzymes can reveal “function connectivity” and provide additional information about evolutionary relationships (Baier & Tokuriki 2014; Mohamed & Hollfelder 2013; Afriat-Jurnou et al. 2012; Roodveldt & Tawfik 2005). In the classical model of enzyme evolution new functions generally emerge by exploitation and optimization of

promiscuous activities of existing enzymes (Jensen 1976; O'Brien & Herschlag 1999). This suggests that enzyme promiscuity can act as a potentiating factor, revealing potential evolutionary trajectories toward novel specialized functions. Indeed, it has been shown that activities toward ancestral substrates are sometimes maintained in extant proteins as promiscuous functions and at the same time, members of progenitor enzyme families often exhibit derived activities as promiscuous functions (Voordeckers et al. 2012; R. Huang et al. 2012; Mohamed & Hollfelder 2013). Thus, both the historical evolution and potential future evolution of an enzyme superfamily can be, in a sense, reflected in the promiscuous functions of its existing members (Voordeckers et al. 2012; Noor et al. 2012; Mohamed & Hollfelder 2013; Copley 2009; Ngaki et al. 2012; R. Huang et al. 2012; Baier & Tokuriki 2014). Recently, several groups, including our group (described in chapter 1) have conducted enzyme characterizations that went beyond the conventional one-enzyme-at-a-time level, and instead performed large-scale function-profiling, in which a diverse set of enzymes belonging to the same enzyme superfamily is assayed against a set of substrates in an “all *versus* all” manner (Baier & Tokuriki 2014; Mashiyama et al. 2014; Bastard et al. 2014; H. Huang et al. 2015). However, the function-profile data is usually represented in a heatmap format, which, however, does not demonstrate how functions are connected via promiscuous enzymes, and thus makes it difficult to analyze and interpret the large amount of data of these analyses. The aim of this chapter is to characterize and interpret function-profiling data of in “function connectivity networks” (FCNs), in which functions are connected through promiscuous enzymes that at least perform two functions, *e.g.* one native and one promiscuous. The resulting networks are analyzed in terms of their overall connectivity and topology, and common features are extracted. The results are discussed in the context of enzyme evolution and compared to the evolutionary relationship of functional families and the chemical similarity between reactions. We also demonstrate how different activity levels thresholds, the environment, in this case metal co-factor availability and neutral genetic diversity, alter connectivity and topology of FCNs.

For our analysis we will focus on four examples of comprehensive large-scale function-profiling studies of the MBL superfamily (described in chapter two and three) (Baier & Tokuriki 2014; Baier et al. 2015), the cytGST superfamily (Mashiyama et al. 2014), the

HAD superfamily (H. Huang et al. 2015) and the β -keto acid cleavage enzyme (BKACE) family (Bastard et al. 2014). The cytGST and MBL superfamilies consist of functional groups that catalyze diverse chemical reactions, and thus the studies explored “catalytic promiscuity” of selected members of each superfamily. In the case of the BKACE family and HAD superfamily, the work primarily focused on substrate promiscuity. Note that in this chapter we refer to different “functions” as either different chemical reactions, which can include multiple substrates for one reaction, or activity toward different substrates, as used and classified in the original work.

4.3 Methods

4.3.1 Datasets and network generation

The original datasets for the analysis in this chapter are from various sources. The original datasets for the analysis of the MBL superfamily are described in chapter two and chapter three. The original dataset of the cytGST (Mashiyama et al. 2014), HAD superfamilies (H. Huang et al. 2015) and BKACE family (Bastard et al. 2014) are obtained from published articles as described and cited in each section. FCNs were generated and visualized using Cytoscape (Shannon et al. 2003) by importing the enzyme-substrate pairwise activity data.

4.4 Results and discussion

4.4.1 Function connectivity networks

To visualize and compare functional connections through promiscuous activities, we generated FCNs using the function-profiling data presented in the original publications. FCNs visualize the relationship between enzymes that perform different functions as a result of enzyme promiscuity. The general concept of FCNs is visualized and further described in **Figure 4.1**. Briefly, if a single enzyme catalyzes two different functions, then those two functions will be connected and cluster in the proximity of each other in a FCN. By contrast, two functions will not be connected and remain separated in the FCN if there is no enzyme that is capable of performing both functions, but can be indirectly connected through an “intermediate” function, which connects to both functions.

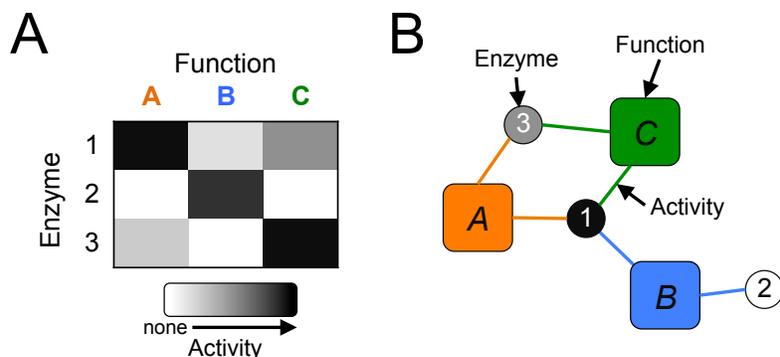


Figure 4.1 Schematic depictions of superfamily-wide function profiling and function connectivity networks.

(A) Selected enzymes from a superfamily are screened against a set of enzymatic functions to reveal native (high activity, dark shading) and promiscuous activities (lower activities, lighter shading). The activity data are typically represented in a table or heatmap format. (B) Same activity data visualized as a “function connectivity network” (FCN) using Cytoscape (Shannon et al. 2003). Enzymes (circles) are connected to a function (large square) if they exhibit activity (edges). In figures 5 and 7, multiple substrates are used to assay some of the enzymatic functions; in this case, an edge is drawn when an enzyme is active toward at least one substrate. Enzyme nodes are qualitatively shaded depending on the number of activities they catalyze, from white being specific for one function to black being highly promiscuous (at least seven functions). Enzymes and functions cluster qualitatively depending on their connectivity, with highly interconnected nodes clustering together.

4.4.1.1 FCN analysis of the metallo- β -lactamase superfamily

The metallo- β -lactamase (MBL) superfamily encompasses a large enzyme superfamily, with over 25,000 sequences from all domains of life distributed across at least 24 iso-functional subgroups that are engaged in a diverse set of cellular functions, *e.g.*, DNA, RNA and nucleotide processing, detoxification, antibiotic resistance, and quorum-quenching (Bebrone 2007; Daiyasu et al. 2001; Baier & Tokuriki 2014). In chapter 2, I described the systematic function-profiling characterization of 24 enzymes that belong to 15 different functional families against 10 distinct hydrolytic reactions (**Figure 2.9**) (Baier & Tokuriki 2014). The enzymes were highly specialized toward their native reactions ($k_{\text{cat}}/K_M \geq 10^4 \text{ M}^{-1}\text{s}^{-1}$). However, most MBL enzymes exhibited some degree of promiscuous activity, albeit with comparatively low catalytic efficiency ($k_{\text{cat}}/K_M \leq 10^2 \text{ M}^{-1}\text{s}^{-1}$; **Figure 2.10**). Overall, 18 out of the 24 enzymes catalyzed multiple reactions and, on average, each enzyme catalyzed 2.5 out of the 10 reactions tested. The prevalence of

promiscuity in this superfamily means there is high connectivity between functions, with 9 out of 10 functions being connected in the FCN (**Figure 4.2 A**). Interestingly, we observe a sub-clustering of enzymatic reactions in the FCN with reactions that are related to substrates with a negative charge at the scissile bond (PDE, PPP, PCE and ARS) and no charge at the scissile bond (TPN, BLA, SLG, LAC and PTE), which form isolated clusters (**Figure 4.2 A** and **Figure 2.6**). A similar clustering of sequences (enzymes) associated with reactions generally using either anionic or neutral substrates was also observed in SSNs and phylogenetic analysis, which suggests that the FCN clustering could to some extent reflect the evolutionary history of functional divergence (**Figure 2.5**) (Baier & Tokuriki 2014; Aravind 1999).

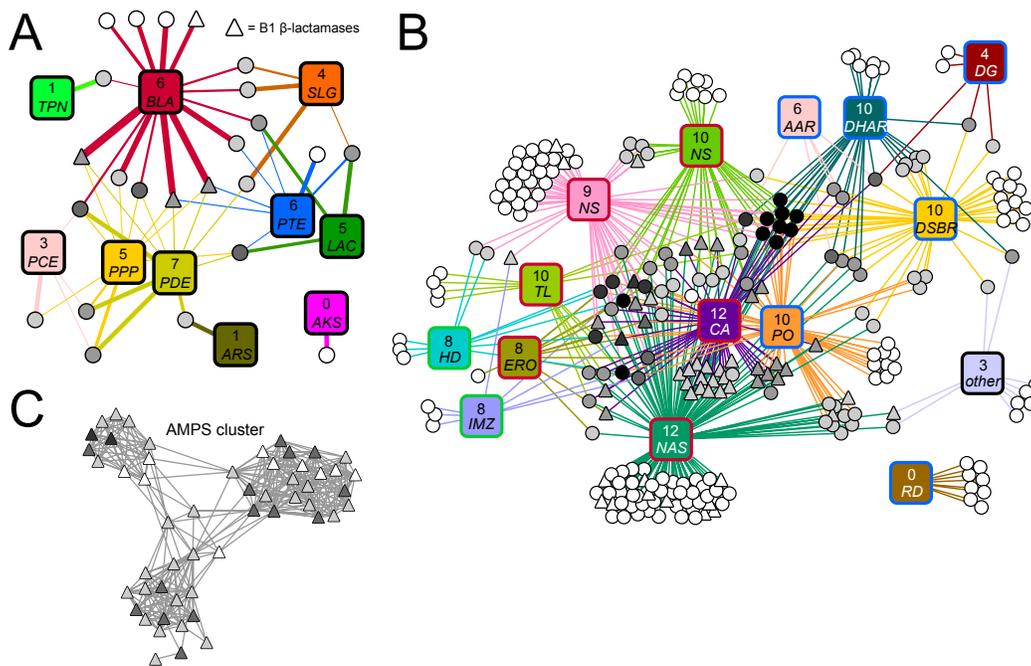


Figure 4.2 FCNs constructed based on the function-profiling analysis of the MBL and cytGST superfamilies.

(A and B) Large square nodes represent functions, and small nodes represent enzymes, and edges indicate the existence of enzymatic activity. The numbers within each function node indicates how many other functions are directly connected via promiscuous enzymes. Enzyme nodes are quantitatively shaded depending on the number of activities they catalyze, from white being specific for one function (connected to one function) to black being highly promiscuous (connected to ≥ 7 functions). (A) FCN of the MBL superfamily represents the function-profiling analysis of 24 enzymes against 10 enzymatic functions (using 12 different substrates) which represent native functions of 10 subgroups: BLA, β -lactamase; TPN, chlorothalonil dehalogenase; AKS, alkylsulfatase; SLG, glyoxalase II; ARS, arylsulfatase; PDE, phosphodiesterase; PPP, phosphonate

monoester hydrolase; PCE, phosphorylcholine-esterase; PTE, phosphotriesterase; LAC, homoserine lactonase. Thickness of the edge represents the level of enzymatic activity. (B) FCN of the cytGST superfamily represents the function-profiling analysis of 256 enzymes catalytic activity for 15 enzymatic functions using 175 different substrates. Triangle nodes represent enzymes of the AMPS sequence cluster, which is shown in C. The border color of square nodes indicates enzymatic function: glutathione *oxidized* (blue border), glutathione *consumed* (red border) or *not consumed* (green border). Enzymatic functions: NAS, nucleophilic aromatic substitution; NS, nucleophilic substitution; CA, conjugate addition; nucleophilic addition; ERO, epoxide ring opening; TL, Thiolysis; IMZ, isomerization; HD, hydrolytic dehalogenation; DSBR, disulfide bond reductase; PO, peroxidase; RD, reductive dehalogenation; DG, deglutathionylation; DHAR, dehydroascorbate reductase; AAR, alkylarsenate reductase. The numbers above each enzymatic function indicate how many other functions are directly connected via a promiscuous enzyme. (C) SSN of the 54 AMPS cluster proteins that are included in B, generated using the Enzyme function initiative (EFI) enzyme similarity tool (Gerlt et al. 2015) and visualized in Cytoscape (Shannon et al. 2003).

4.4.1.2 FCN analysis of the cytGST superfamily

The cytosolic glutathione transferase (cytGST) superfamily contains over 10,000 known sequences (Mashiyama et al. 2014). Its most common functions involve utilizing glutathione as a cofactor to metabolize endogenous compounds, detoxify chemicals, and prevent oxidative stress (Mashiyama et al. 2014; Armstrong 1997). CytGSTs structurally consist of two domains: a smaller N-terminal thioredoxin-like fold domain that binds glutathione and a larger C-terminal domain is responsible for substrate recognition and binding (Armstrong 1997). The superfamily catalyzes a diverse range of functions (over 140 E.C. terms) that can be categorized mechanistically into three groups based on the usage of glutathione: '*oxidized*', '*consumed*' or '*not consumed*' (Mashiyama et al. 2014). Furthermore, the three groups can be sub-classified into 15 functions according to their reaction chemistry (Mashiyama et al. 2014). Mashiyama *et al.* recently performed a large-scale function-profiling analysis and characterized 82 enzymes in this superfamily for 15 different functions, which were assayed using 175 different substrates with several chemically similar substrates representing one function (Mashiyama et al. 2014). The authors also combined available literature data of an additional 174 enzymes, and thus provide a function-profiling analysis for 256 enzymes. From this data, we see a variable level of promiscuity amongst the enzymes in this superfamily with 53% of enzymes are

specific and catalyze only one reaction (136 of 256) whereas ~7% of the enzymes (17 of 256) catalyze more than six reactions (Mashiyama et al. 2014). The cytGST FCN revealed that almost all functions, 14 of 15, are connected, with only RD (*reductive dehalogenation*) being completely separated (**Figure 4.2 B**). Interestingly, the clustering of individual functions relates to glutathione utilization, *i.e.*, ‘*oxidized*’, ‘*consumed*’ and ‘*not consumed*’. Glutathione *oxidized* and *not consumed* are completely separated, and only indirectly connected through *consumed* (**Figure 4.2 B**). In contrast, glutathione *consumed* connects more extensively than others; in particular, two glutathione *consumed* reactions, CA (*conjugate addition*) and NAS (*nucleophilic aromatic substitution*) are positioned as central hubs that, respectively, associate with 12 other reactions via promiscuous enzymes (**Figure 4.2 B**).

4.4.1.3 FCN analysis of the BKACE family

Until recently, only one enzymatic function had been assigned to what is now known as the BKACE family: the condensation of β -keto-5-amino-hexanoate and acetyl-CoA to produce aminobutyryl-CoA and acetoacetate within the lysine fermentation pathway (Bellinzoni et al. 2011). Formerly the family was referred to as the DUF849 family, which is comprised of 922 sequences and structurally adopts a canonical TIM β/α barrel fold (Bellinzoni et al. 2011). In 2014 Bastard *et al.* performed a systematic characterization to uncover and assign enzymatic functions to other members of this enzyme family (Bastard et al. 2014). Based on the biochemical and structural analyses of the sole previously characterized enzyme, the authors selected and characterized activity with 16 other β -keto acid substrates for 124 uncharacterized enzymes from the DUF849 family. Of the 124 enzymes, 80 were active for at least one substrate, and 15 of 17 substrates were catalyzed by at least one enzyme. These proteins led them to identify the family as one of β -keto acid cleavage enzymes, and thus, the family has now been named the “ β -keto acid cleavage enzyme” or BKACE family. This systematic function profiling revealed that over 60% of the enzymes (50 of 80) that were tested displayed activity for more than one β -keto acid substrate, whereas 40% of enzymes (30 of 80) were specific to one substrate. The FCN of the BKACE family leads to conclusions similar to those of the FCNs of the MBL and cytGST superfamilies (**Figure 4.3 A**), in which all the functions

that were tested, which correspond to different substrates of β -keto acid cleavage reaction, are connected through promiscuous enzymes. Similar to the other FCNS, we also observe that substrates with distinct chemical features *e.g.*, anionic, cationic, nonionic polar and apolar, form separate clusters. In general, nonionic polar and apolar substrates are more connected through promiscuous enzymes, as highlighted by the β -keto-hexanoate substrate (S11, **Figure 4.3 A**), which is connected to all other substrates. In contrast, charged substrates (green border, cationic S1 and S2; red border, anionic S3 and S4, **Figure 4.3 A**) form distinct clusters, which are located at the peripheries of the FCNs, and are overall less connected.

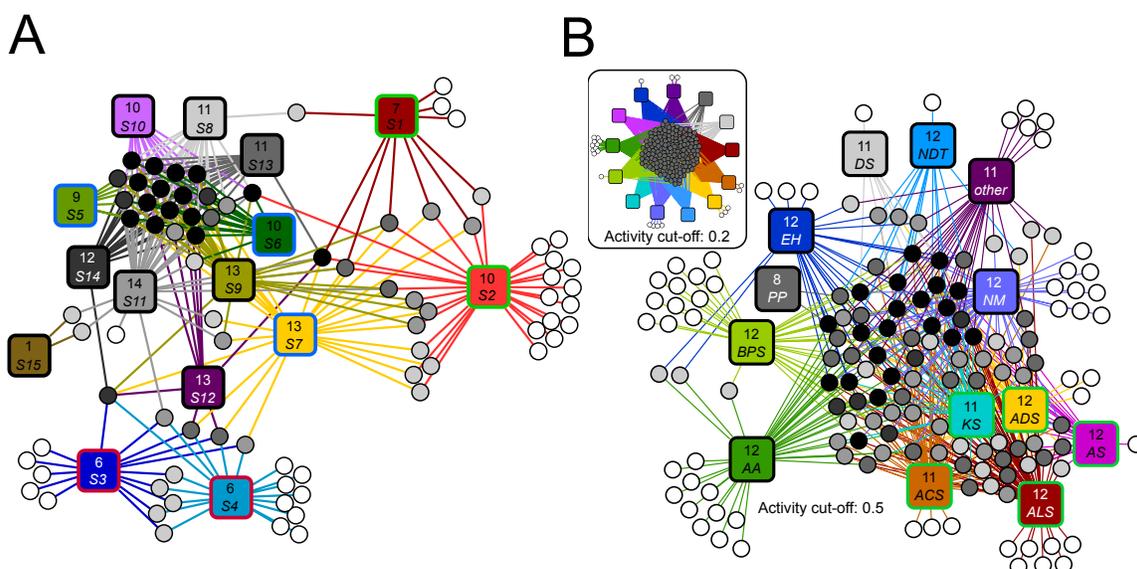


Figure 4.3 FCN representation of the function-profiling analysis of the BKACE and HAD superfamilies.

(A and B) Large square nodes represent functions, small nodes represent enzymes, and edges indicate the existence of enzymatic activity. The numbers within each function node indicates how many other functions are directly connected via promiscuous enzymes. Enzyme nodes are quantitatively shaded depending on the number of activities they catalyze, from white being specific for one function (connected to one function) to black being highly promiscuous (connected to ≥ 7 functions). (A) FCN of the BKACE family, which represents the function-profiling analysis of 80 enzymes against 15 substrates. (Bastard et al. 2014) Substrates are catalyzed by condensation with acetyl-CoA to produce a CoA ester and acetoacetate. Border color of square nodes indicates substrate property: anionic (green border), cationic (red border), nonionic polar (blue border) or apolar (black border). Substrates (only for forward reaction) are indicated by: S1, S-KAH; S2, Dehydrocarnitine; S3, β -keto-adipate; S4, β -ketoglutarate; S5, 3,5 dioxohexanoate; S6, 5-hydroxy- β -keto-hexanoate; S7, 6-acetoamido- β -keto-hexanoate; S8, β -ketopentanoate; S9, β -ketoisocaproate; S10, (E)- β -keto-hex-4-enoate; S11, β -

ketoheptanoate; S12, 7-methyl- β -ketoct-6-enoate; S13, β -ketoctanoate; S14, β -ketododecanoate; S15, benzoylacetate. **(B)** FCN of the HAD superfamily, which represents the function-profiling analysis of 204 enzymes against 167 substrates, which were classified into 13 different substrate classes: PP, phosphonates; DS, disaccharides; ALS, alcohol sugars; ACS, acid sugars; ADS, aldol sugars; NDT, nucleotide di- and triphosphates; NM, nucleotide monophosphates; KS, ketose sugars; BPS, bisphosphate sugars; AA, amino acids; AS; amine sugars; EH, easily hydrolyzed; others (2, 7, 8 and 9 carbon sugars). (H. Huang et al. 2015) Green border color of square nodes indicates monophosphate sugar substrate classes. The numbers within each substrate class node indicate how many other substrate classes are directly connected via promiscuous enzymes. The activity cut-offs correspond to the corrected absorbance of the end point assay employed by Huang *et al.* and considered $OD_{650} = >0.2$ for their analysis. Note that the FCN with the $OD_{650} = 0.5$ cut-off contains only 145 enzymes, because activities of 59 enzymes were below the more stringent 0.5 cut-off.

4.4.1.4 FCN analysis of HAD superfamily

The HAD superfamily is comprised of 120,000 sequences, which share a Rossmann fold “core” domain and a fused “cap” domain (H. Huang et al. 2015; Burroughs et al. 2006). The majority of HAD enzymes require Mg^{2+} and a conserved active site aspartic acid for catalysis (Burroughs et al. 2006). Although the superfamily is named after the haloacid dehalogenase (C-Cl bond cleavage) enzyme, the majority of enzymes are phosphate hydrolases such as phosphoesterases, ATPases, sugar phosphomutases, and other phosphatases and phosphonatas involved in P-O and P-C bond cleavage (Burroughs et al. 2006). The cellular functions of HAD phosphatases include primary metabolism of amino acids and sugars, secondary metabolism, dNTP pool regulation, cellular housekeeping, and nutrient uptake (Burroughs et al. 2006). Huang and coworkers measured the function profile of more than 200 enzymes against a diverse library of 167 phosphatase (98%) and phosphonate (2%) substrates (grouped into 13 substrate classes) (H. Huang et al. 2015). The authors considered an enzyme-substrate pair as active if a 30-min incubation of the reaction mixture (5 μ M enzyme and 1 mM substrate) provides $OD_{650} = >0.2$, using a chemical dye to detect the release of inorganic phosphate. With these criteria, most of the tested enzymes were highly promiscuous with 75% of the enzymes acting on more than five substrates and 23% being active for more than 41 substrates (and up to 143 substrates) (H. Huang et al. 2015). Not surprisingly, the FCN of the HAD superfamily generated from this dataset revealed very dense connectivity

between functions (substrate classes) and complete connection between all 13 functions, in which each function is connected to every other functional class (small box in **Figure 4.3 B**). We believe that the high connectivity is associated with the extensive number of experimental characterizations (>200 enzymes against 167 substrates). In addition, the set of substrates is less diverse compared to the other studies; they generally involve the same chemical reactions (P-O (98%) and P-C (2%) bond cleavage), and thus almost exclusively differ by their leaving group. To obtain a better separation and clustering of functions (substrate classes), we reanalyzed the data using a more stringent cut-off of $OD_{650} = 0.5$ (the highest OD_{650} observed in the assay was ~ 1.0), which gave rise to a clustering of the distinct functions in the FCN (**Figure 4.3 B**). Thus, employing an activity cut-off, similar to the BLAST E-value cut-off in SSNs, can be a useful method to observe relationships between functions and enzymes in FCNs. With this more stringent cut-off we observe that all five monophosphate sugar substrates (acid, alcohol, aldose, amine and ketose sugars) cluster together and are fairly separated from other functions. On the other hand, phosphonate substrates, which involve cleavage of P-C bonds instead of P-O (all other substrates), are the least connected substrate class (**Figure 4.3 B**). Hence, similar to the other FCNs, the substrates cluster mainly depending on their chemical properties such as scissile bond (P-C bond vs. P-O) or overall structure (monophosphate sugar substrates).

4.4.2 Perspectives on function connectivity through promiscuity

The FCNs provide intriguing insights into the function connectivity within superfamilies through enzymes with promiscuous activities, which we discuss in the following sections. There are common features and general trends that can be extracted from these function-profiling analyses. However, we would like to note some caveats in our analyses: First, each study features a different range of functions and sequences that were tested, *i.e.*, the number and type of functions (chemical reactions, substrate classes, or substrates) and number and sequence divergence of enzymes; Second, the assays for each study were different, and had different levels of sensitivity, which can also affect how many functions per enzyme are revealed. For example, more sensitive assays could identify activities that are below the detection limit of the employed assays, resulting in more

promiscuous activities being identified. Hence, lowering the threshold for assaying enzymatic activity or expanding the scope of chemical reactions and substrates assayed would reveal more promiscuous activities in enzymes that currently appear highly specific.

4.4.2.1 Indirect connectivity through intermediate functions

Despite the overall function connectivity, many functions are only indirectly connected through ‘intermediate functions’, *i.e.* two functions are not connected by the promiscuity of a single enzyme, but through at least one other function and two other enzymes. The most prominent example is seen in the BKACE family, where cationic and anionic substrates have no direct connections, but are only connected through neutral substrates (**Figure 4.3 A**). It is rational that an enzyme adapted to a cationic substrate would possess a negatively charged active site cavity that would not be appropriate for anionic substrates, and *vice versa* (Bastard et al. 2014). Enzymes that have evolved toward neutral substrates, however, may be able to promiscuously act on both anionic and cationic substrates to some extent. Similar trends are observed in the other superfamilies that we examined (**Figure 4.2 and 4.3**). Therefore, the emergence of a new function may only be associated with distinct functions within a superfamily, and its evolution could be restricted to particular progenitor enzymes. In other words, functional expansion from a common ancestor may be constrained to one or a few specific trajectories in order to reach functions that are chemically distant from the ancestral functions. In contrast, however, some functions are frequently observed as being linked by promiscuous enzymes, and thus are more likely to be connected to many other functions (**Figure 4.2 and 4.3**). These functions might easily evolve from various progenitor enzymes within a superfamily. Indeed, there are notable examples of convergent evolution in the literature: β -lactamase activity has arisen at least twice within the MBL superfamily (Aravind 1999; Bebrone 2007). In the HAD superfamily phosphomutase activity evolved independently more than once (Burroughs et al. 2006). Additional cases of convergent evolution have been observed in many other superfamilies including phosphatidylinositol-phosphodiesterases (Furnham et al. 2012), enolases (Brown & Babbitt 2014) and Zn-

dependent peptidases (Makarova & Grishin 1999).

4.4.2.2 The scope and level of promiscuity is different for each enzyme

Many of the enzymes assayed across these studies were promiscuous, and some exhibited activity toward a remarkable variety of substrates (dark and black enzyme nodes in the FCNs, **Figure 4.2** and **4.3**). In contrast, some enzymes were specific to only one function (white enzyme nodes in the FCNs; **Figure 4.2** and **4.3**). Interestingly, however, is that the range and type of promiscuity also varies substantially amongst enzymes within the same functional family. Different members of the same functional family are, generally, orthologous enzymes that play the same physiological role in different species, but that possess diverse sequences due to speciation and genetic drift (Gabaldón & Koonin 2013). These sequence changes are relatively neutral in terms of the native catalytic function, but seem to alter the level and scope of promiscuous functions, and thereby result in “cryptic genetic variation” (Wagner 2008; Paaby & Rockman 2014; Masel & Trotter 2010; Amitai et al. 2007; Bloom et al. 2007). For example, among five B1 β -lactamases from the MBL superfamily, two enzymes catalyze only their native reaction, while each of the remaining three enzymes are capable of up to three additional reactions (**Figure 4.2 A**). Additionally, members of the AMPS (Alpha-, Mu-, Pi-, and Sigma-like) subgroup of the cytGST superfamily also exhibit different levels of promiscuity; some AMPS enzymes catalyze one reaction, whereas others catalyze up to six reactions (**Figure 4.2 B**). Interestingly, specificity and promiscuity levels are not correlated with particular sequence clusters within the AMPS subgroup, but are instead scattered throughout (**Figure 4.2 C**). A similar consequence of neutral genetic variation has also been observed in laboratory evolution studies where enzymes were subjected to “neutral drift” (accumulation of mutations under a purifying selection pressure for the native function), which produced notable changes in their activity promiscuous functional profiles (Bloom et al. 2007; Amitai et al. 2007). Thus, cryptic genetic variation can result in sequences that already have higher activities towards a new substrate, and thus support a role for neutral genetic diversity in driving the innovation and evolution of new enzyme functions (Wagner 2008; Paaby & Rockman 2014; Masel & Trotter 2010). This observation also indicates, that some sequences are less “evolvable” compared to others, because the

enzymes that do not exhibit enough promiscuous activity to provide a selective advantage when the environment changes to favor a new function (McLoughlin & Copley 2008).

4.4.2.3 Function connectivity depends on cofactor availability

Function profiles and enzyme activity levels can vary depending on the environment and thus affect the topology of FCNs and ultimately enzyme evolution. A simple example is that chemical engineers used “*enzyme condition promiscuity*” to achieve catalysis under water-limited environments with hydrolytic enzymes; such environs favour ester synthesis instead of hydrolysis (Hult & Berglund 2007). Similarly, other environmental changes can promote the appearance of new promiscuous functions. For example, a temperature shift can induce changes in substrate specificity in a bacterial thymidine kinase for various nucleoside analogues (Lutz et al. 2007). Cofactor exchange, *e.g.* metal ions, flavins and hemes, can cause condition-specific promiscuity, and so lead to different function profiles depending on the conditions under which the enzymes were assayed (Sánchez-Moreno et al. 2009; Baier et al. 2015; Nielsen et al. 2011; Badarau & Page 2006; Dimitrov & Vassilev 2009; Nobeli et al. 2009). For example, a recent study by Lapalikar *et al.* showed that the substitution of F₄₂₀-dependent reductases with FMN instead of F₄₂₀ enables them to catalyze both oxidation and reduction of the same substrate (Lapalikar et al. 2012). Furthermore, a recent study by Reynolds *et al.* evolved a cytochrome P₄₅₀ enzyme to incorporate a non-proteinogenic cofactor, iron deuteroporphyrin IX, which enables the enzyme to catalyze the non-natural carbenoid-mediated olefin cyclopropanation reaction, which is not observed with the native cofactor heme (Reynolds et al. 2016). In the MBL superfamily, most enzymes are assumed to incorporate Zn²⁺ in the active site, however some have been shown to prefer Fe²⁺, Ni²⁺, Mn²⁺ and Co²⁺ (Baier et al. 2015; Silaghi-Dumitrescu et al. 2005; Hu, Gunasekera, et al. 2008; Yu et al. 2009; Hu et al. 2009). We previously characterized the effect of exchanging the active site metal ions of five enzymes from the MBL superfamily (6 different metals and 8 different reactions). This systematic analysis revealed that the function profile of these enzymes, and thus their function connectivity, varies significantly across different environments (in this case environments with different metal ion availabilities; **Figure 4.4**) (Baier et al. 2015). Interestingly, individual metal

ions display limited connectivity; however, when the function profiles of all metals are superimposed, all reactions are connected (**Figure 4.4**). Although the concentration of metal ions in the cellular milieu is generally regulated, metalloenzymes do not always achieve perfect metallation with a specific, most active metal ion(s), but rather exist as various metal isoforms in the cell (Carter et al. 2011; Xu et al. 2008; Culotta et al. 2006; Foster et al. 2014; Clugston et al. 2004; Waldron & Robinson 2009). Also, environmental variation in metal availability or cellular stress can lead to severe mismetallation of metalloenzymes (Imlay 2014; Carter et al. 2011; Xu et al. 2008; Culotta et al. 2006). Thus, co-factor depend function profile changes that were observed in these *in vitro* studies may also reflect functional variation in the natural environment. To date, there is no direct evidence to support that environmental-dependent functional promiscuity fosters the expansion of enzyme superfamily diversity. The diversity of cofactor utilities observed in enzymes within a single enzyme superfamily, however, may imply that very evolutionary scenario (Goldman et al. 2016). For example, a recent study by Goldman *et al.* suggests that the functional diversity of TIM barrel enzymes stems from the incorporation of different cofactors, including FMN, NAD, NADP, and various metal ions (Goldman et al. 2016). Interestingly, Ahmed *et al.* described new subgroups of split β -barrel fold enzymes, which display surprising cofactor diversity, including the binding and utilization of F420, FMN, FAD and heme (Ahmed et al. 2015). Interestingly, some of these enzymes promiscuously bind multiple cofactors with considerably high affinity *in vitro* (Ahmed et al. 2015), and could suggest they exist in various co-factor forms *in vivo*.

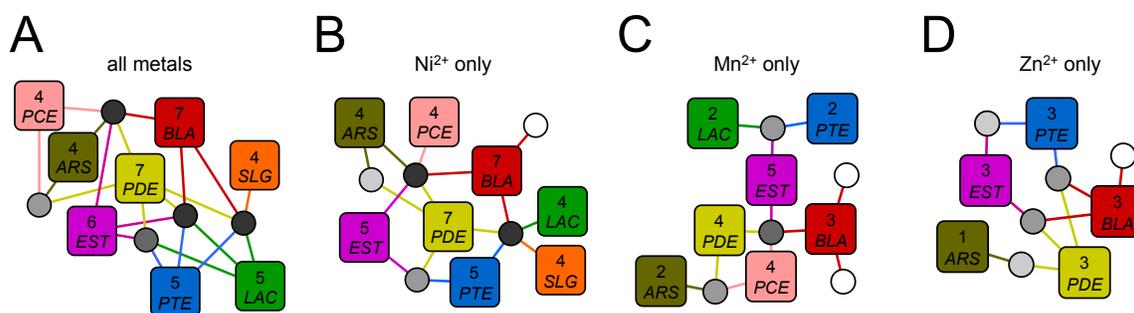


Figure 4.4 FCN representation of the metal-dependent function profiles for five MBL superfamily enzymes.

(A, B, C and D) FCNs represent the function-profiling analysis of 5 enzymes against 8 enzymatic functions with 6 different metal ions (Fe^{2+} , Zn^{2+} , Co^{2+} , Cd^{2+} , Ni^{2+} , Mn^{2+}). Large square nodes represent functions, small nodes represent

enzymes, and edges indicate the existence of enzymatic activity. Enzyme nodes are quantitatively shaded depending on the number of activities they catalyze, from white being specific for one function (connected to one function) to black being highly promiscuous (connected to ≥ 7 functions). The numbers within each function node indicates how many other functions are directly connected via a promiscuous enzyme. **(A)** FCN of the function profiles of all six reconstituted metal isoforms. **(B)** FCN of function profiles of the Ni^{2+} reconstituted metal isoforms. **(C)** FCN of function profiles of the Mn^{2+} reconstituted metal isoforms. **(D)** FCN of function profiles of the Zn^{2+} reconstituted metal isoforms; the corresponding FCNs of Co^{2+} , Cd^{2+} and Fe^{2+} are not shown due to space limitations. Enzymatic functions: BLA, β -lactamase; SLG, glyoxalase II; ARS, arylsulfatase; PDE, phosphodiesterase; PCE, phosphorylcholine-esterase; PTE, phosphotriesterase; LAC, lactonase; EST, esterase/lipase.

4.4.3 Structural features that determine function connectivity

In general, the members of a single enzyme superfamily share the same structural fold and mechanistic features in the active site (Gerlt & Babbitt 2001). Similarly, it has been shown that the native and promiscuous activities within a single enzyme are catalyzed in the same active site (Tokuriki et al. 2012). Elucidating the molecular and structural basis for enzyme chemical and substrate specificity, however, remains extremely challenging. For example, in our previous study of 24 MBL enzymes, we found that many enzymes, despite large differences in active site shape, volume and hydrophobicity, could catalyze the same reactions, such as the native functions of phosphodiesterases or β -lactamases (**Figure 2.11**) (Baier & Tokuriki 2014). The only tendency we found was that enzymes that hydrolyze charged substrates as their native function possess more hydrophilic active sites (**Figure 2.11**). Similarly, Bastard *et al.* also found a structure-function correlation for enzymes of the BKACE superfamily. In particular, they found that enzymes with more hydrophobic and non-charged active sites turn over hydrophobic and non-charged polar substrates. In contrast, enzymes that have negatively or positively charged residues in the active site turn over generally positively and negatively charged substrates, respectively. Members of the HAD superfamily are structurally classified depending on the cap domain above the active site, as being either C0 (minimal cap) or C1 and C2 (large cap). Huang *et al.* found that C1 and C2 enzymes exhibit generally more promiscuity compared to C0 enzymes, and concluded that the insertion of additional cap domains led to extended function profiles of HAD enzymes (H. Huang et al. 2015). For

the cytGST superfamily, Mashiyama *et al.* could not identify any specific active site property or feature that is associated with the catalysis of particular reactions (Mashiyama *et al.* 2014). Besides these rough tendencies, no study has successfully provided explicit molecular explanations for different types of enzyme functionality across a superfamily, let alone for the existence of promiscuous activities. Thus, further efforts to characterize structural and functional details, including structural dynamics (Gobeil *et al.* 2014; Campbell *et al.* 2016; Babbitt *et al.* 2010), are necessary for us to advance our understanding of the structure-function relationships that determine the catalytic function that different enzymes perform (Carvalho *et al.* 2014) .

Chapter 5: Cryptic genetic variation affects enzyme evolvability

Parts of chapter five have been performed in collaboration with N. Hong in the laboratory of Dr. Colin J. Jackson at ANU, Canberra, Australia, and A Pabis in the laboratory of Dr. S. C. L. Kamlerlin at Uppsala University, Uppsala, Sweden. N. Hong performed crystal structure analysis of NDM1-R10 and VIM2-R10 and size exclusion chromatography of NDM1 and VIM2 variants, as shown in **Figure 5.14** and **5.8**, respectively. A. Pabis performed molecular dynamics simulations of NDM1 and VIM2 structures as shown in **Figure 5.15** and **5.17**. I performed all other experiments and wrote the chapter together with my supervisor, Dr. Nobuhiko Tokuriki.

5.1 Summary

How does genetic variation among evolutionary and functionally related enzymes affect their evolvability and evolutionary adaptation towards a new function? We address this question by evolving in parallel two related β -lactamases, NDM1 and VIM2, towards a shared promiscuous phosphonate monoester hydrolase activity. We observed striking differences in their response to adapt to the new function over ten rounds of directed evolution. NDM1 adapted to higher “fitness” by improving catalytic efficiency by 20,000-fold (k_{cat}/K_M), but partially lost its solubility, *i.e.*, the amount of functional enzyme in the cell. In contrast, VIM2 only exhibited a 60-fold increase in catalytic efficiency, but improved its solubility, partially through dimerization. Furthermore, a total of 26 mutations were fixed in both trajectories, which however are strikingly different for each enzyme. The mutational pathways are also incompatible between both trajectories. For example, a single initial mutation, W93G, improved the PMH activity of NDM1 by 300-fold, while the same mutation is detrimental for the new function when introduced to VIM2 (by 10-fold). A detailed structural analysis, coupled to molecular dynamics simulations, provides a molecular basis for the observed differences in phenotypic adaptation, evolvability and mutational incompatibility between NDM1 and VIM2. Our results demonstrate that seemingly neutral mutations can have profound consequences on evolvability and evolutionary outcomes during adaptation.

5.2 Introduction

The mutational robustness of proteins, in which most mutations do not affect structural integrity and physiological function, leads to a large degree of genetic variation among orthologous proteins (Tokuriki & Tawfik 2009c; Wagner 2008; Paaby & Rockman 2014). Although such genetic variation is neutral with respect to the protein's native function, it can alter other properties, such as latent promiscuous functions, that are not under immediate selection pressure (Aharoni et al. 2005; McGuigan & Sgrò 2009; Paaby & Rockman 2014). For example, laboratory neutral diversification, with purifying selection for the native function, of protein and RNA enzymes introduced cryptic genetic variation that yielded variants that exhibit new promiscuous functions (Bloom et al. 2007; Amitai et al. 2007; Hayden et al. 2011). Upon a change in selection pressure, *e.g.* through environmental or genetic perturbations, this promiscuous functions, can become essential for organism survival and provide an adaptive advantage. Thus, cryptic genetic variation can produce genotypes that are phenotypically pre-adapted to new circumstances. In many instances, however, several genotypes can share the same promiscuous activities and thus would be theoretically equally evolvable. For example, as described in chapter two, several B1 β -lactamases exhibit promiscuous phosphodiester, phosphotriester and phosphonate activity at a relatively similar level in addition to their native function. Hence, the question arises if these pre-adapted sequences would evolve similarly under the same selection pressure towards improving the latent promiscuous function? In other words, does cryptic genetic variation also affect evolutionary trajectories and the evolutionary outcome? This is an important question especially for the evolution of enzymes, because promiscuous activities are often initially very low compared to the catalytic efficiencies of native activities, which are generally above a k_{cat}/K_M of $10^3 \text{ M}^{-1}\text{s}^{-1}$ in, and thus need to be optimized to provide an organismal advantage (Khersonsky & Tawfik 2010). Therefore, for the evolution of a new enzyme function, two main prerequisites need to be met. First, the promiscuous activity needs to be above a certain level to confer an initial selective advantage to the organism (O'Brien & Herschlag 1999). Second, the promiscuous activity must be improvable through mutations, as improvements will enhance the selective advantage and consequently organismal fitness. We use term evolvability here as “*the ability of a protein to adapt in response to*

mutation and selective pressure” (Romero & Arnold 2009). Theoretical studies using RNA folding as a proxy for selective advantage have shown that the evolvability of various genetic backgrounds could be different due to mutational epistasis, *i.e.*, the functional effect of a mutation depends on the genotypic background in which it occurs (Schaper et al. 2012). In recent years, the prevalence of epistasis and the importance of the starting genetic background during evolutionary adaptation have been demonstrated in *E. coli* (Khan et al. 2011), RNA virus (Burch & Chao 2000), bacteriophage λ (Meyer et al. 2012) and ribozyme populations (Hayden et al. 2011). In proteins, epistasis has been shown to constrain evolutionary trajectories towards new functions, as in the case of the TEM-1 β -lactamase towards a third-generation antibiotic cephalosporin (Weinreich 2006), the evolution of receptor specificity (Bridgham et al. 2009) or the evolution of influenza virus proteins (Gong et al. 2013). Recently, some studies have shown that the functional effect of mutations differs significantly among related proteins with the same phenotype. For example, Parera *et al.* demonstrated that the effect of a single substitution, A156T, on 56 distinct variants of the hepatitis C virus NS3 protease, varied from being deleterious to beneficial (Parera & Martinez 2014). In the case of a promiscuous activity, a recent study by Khanal *et al.* observed that a single mutation, E382G, improved a promiscuous NAGSA dehydrogenase activity by 50- to 770-fold among nine gamma-glutamyl phosphate reductase (ProA) orthologs (Khanal et al. 2015). Thus, epistasis could strongly impact the evolvability of some enzymes more than others, towards the same new phenotype (Harms & Thornton 2013). However, to date, no study experimentally addressed whether related enzymes variants, *e.g.*, close homologs such as orthologous enzymes, would differ in their evolvability and mechanisms of adaptation towards a new function (Hartl 2014). Will they plateau at different activity levels? Will they acquire unique or common mutations and adopt similar structural solutions? The aim of this chapter is to investigate the evolvability of orthologous enzymes towards a shared promiscuous activity and expose their potential as evolutionary starting points. Ideally, in such an experiment two or more enzymes would be subjected to the same mutation and selection conditions in a highly controlled experimental set up.

Directed evolution is a powerful tool to address evolutionary questions that enables us to perform experiments in a highly controllable setup and allows subsequent

characterization of evolutionary intermediates with the aim of unveiling the molecular details underlying evolutionary transitions (Romero & Arnold 2009). Furthermore, directed evolution in a comparative set up allows the identification and characterization of unsuccessful evolutionary solutions, whereas in nature only successful evolutionary outcomes are observable (Schulenburg et al. 2015).

Here, we conducted an empirical test of evolvability by performing comparative directed evolution towards a shared promiscuous phosphonate monoester hydrolase (PMH) activity, starting from two orthologous B1 β -lactamases which share 33 % sequence identity and a high structural similarity ($C\alpha$ RMSD 1 Å) (**Figure 5.1**). Both enzymes are subjected to ten rounds of directed evolution, for which the most improved variant of each round served as the starting point for the next round. The resulting evolutionary trajectories are compared in their level of adaptation and all evolved variants are characterized in detail to reveal the underlying differences in molecular adaptations. In particular, we examined mutational, catalytic activity, protein solubility and stability changes of each variant. We also addressed the repeatability of the evolution and tested the functional effect of individual mutations in different genetic backgrounds. Finally, we performed structural and molecular dynamics simulation to elucidate the molecular basis of adaptation and the compatibility of mutations.

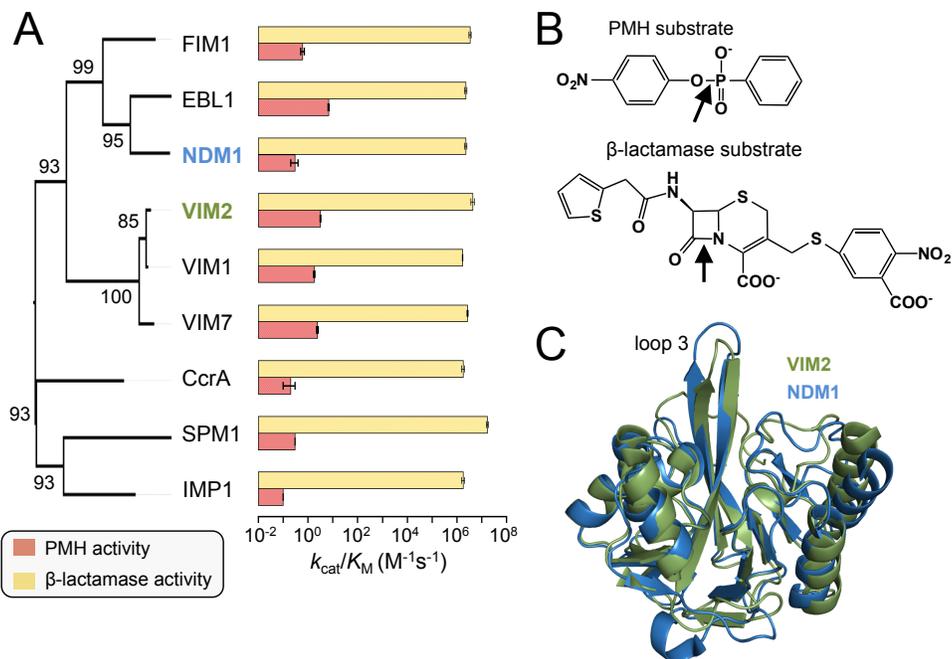


Figure 5.1 The comparative evolution of B1 β -lactamases towards promiscuous PMH activity.

(A) Catalytic efficiencies (k_{cat}/K_M) of nine B1 β -lactamases for native β -lactamase and promiscuous PMH activity. The phylogenetic relationship is shown on the left with bootstrap values at each node. Error bars of catalytic efficiencies represent standard deviation of triplicate measurements. (B) Chemical structure of the chromogenic β -lactamase substrate (CENTA) and PHM substrate (*p*-nitrophenol phenylphosphonate) used in this study to assay enzymatic activity. The arrows indicate the bond broken during catalysis. (C) Structural overlay ($C\alpha$ RMSD 1 Å) of the two B1 β -lactamases NDM1 (blue, PDB ID: 3spu) and VIM2 (green, PDB ID: 1ko3) selected for the directed evolution experiment.

Table 5.1 General information on enzymes characterized in this study.

Enzyme name	Uniprot ID	Organismal source	PDB ID code	Structural resolution
FIM1	K7SA42	<i>Pseudomonas aeruginosa</i>		
EBL1	Q2N9N3	<i>Erythrobacter litoralis</i>		
NDM1	C7C442	<i>Klebsiella pneumonia</i>	3spu	2.1
VIM2	Q9K2N0	<i>Pseudomonas aeruginosa</i>	1ko3	1.9
VIM1	Q9XAY4	<i>Pseudomonas aeruginosa</i>		
VIM7	Q840P9	<i>Pseudomonas aeruginosa</i>	2y87	1.9
CcrA	P25910	<i>Bacteroides fragilis</i>	1znb	1.9
SPM1	Q8G9Q0	<i>Pseudomonas aeruginosa</i>	2fhx	1.9
IMP1	Q79MP6	<i>Pseudomonas aeruginosa</i>	1ddk	3.1

5.3 Methods

5.3.1 Generation of mutagenized library

Random mutant libraries were generated with error-prone PCR using nucleotide analogues (8-Oxo-2'-deoxyguanosine-5'-Triphosphate (8-oxo-dGTP) and 2'-Deoxy-P-nucleoside-5'-Triphosphate (dPTP); TriLink). Two independent PCR reactions were prepared, one with 8-oxo-dGTP and one with dPTP. Each 50 μ L reaction contained 1 \times GoTaq Buffer (Promega), 3 μ M MgCl₂, 1 ng template DNA, 1 μ M of primers (forward (T7 promoter): taatacgaactcactataggg; reverse (T7 terminator): gctagttattgctcagcgg), 0.25 mM dNTPs, 1.25 U GoTaq DNA polymerase (Promega) and either 100 μ M 8-oxo-dGTP or 1 μ M dPTP. Cycling conditions: Initial denaturation at 95°C for 2 minutes followed by 20 cycles of denaturation (30 seconds, 95°C), annealing (60 seconds, 58°C) and extension (70 seconds, 72°C) and a final extension step at 72°C for 5 minutes. Subsequently, each PCR was treated with *Dpn* I (Fermentas) for 1 h at 37°C to digest the template DNA. PCR products were purified using the Cycle Pure PCR purification kit (E.N.Z.A) and further amplified with a 2 \times Master mix of Econo TAQ DNA polymerase (Lucigen) using 10 ng template from each initial PCR and the same primers at 1 μ M in a 50 μ L reaction volume. Cycling conditions: Initial denaturation at 95°C for 2 minutes followed by 30 cycles of denaturation (30 seconds, 95 °C), annealing (20 seconds, 58°C) and extension (70 seconds, 72°C) and a final extension step at 72°C for 2 minutes. The PCR products were purified and cloned as described above. This protocol consistently yielded 1-2 amino acid substitutions per gene.

5.3.2 Generation of DNA shuffling libraries

The staggered extension process (StEP) protocol was used to recombine multiple mutants (Zhao et al. 1998). Plasmids of variants were mixed in equimolar amounts to 500 ng of total DNA and used as a template for the StEP reaction. Cycling conditions: Initial denaturation at 95°C for 5 minutes followed by 100 cycles 95 °C for 30 s followed by 58 °C for 5 s. PCR product was treated with *Dpn* I (Fermentas) for 1 h at 37°C to digest the template DNA. PCR products were purified using the Cycle Pure PCR purification kit and further amplified with a 2 \times Master mix of Econo TAQ DNA polymerase. Libraries were cloned into the pET29(b) vector as described above.

5.3.3 Site-directed mutagenesis

Single-point mutant variants were constructed by site-directed mutagenesis as described in the QuikChange Site-Directed Mutagenesis manual (Agilent) using specific primers. All variants contained only the desired mutation, which was confirmed by Sanger DNA sequencing (Genewiz).

5.3.4 Pre-screen on agar plates

Libraries in pET29-pMBP were electroporated into *E. coli* BL21 (DE3) cells and incubated for 1 h at 37°C prior to plating. For low antibiotic prescreen, the transformants were plated on agar plates (150 mm diameter) containing 4 µg/mL ampicillin, 0.1 mM IPTG, 200 µM ZnCl₂ and 40 µg/mL kanamycin, yielding >500 colonies. The minimum inhibitory concentration (MIC) of ampicillin for *E. coli* cells expressing NDM1 and VIM2 is approximately 256 µg/mL, whereas for the *E. coli* cells alone it is <2 µg/mL. Subsequently, surviving colonies were directly picked from plates for rescreen in 96-well plates. For direct PMH prescreen, transformation reactions were plated on six agar plates (150 mm diameter) containing 40 µg/mL kanamycin, such that each plated contained between 400-2000 colonies. Colonies were replicated onto nitrocellulose membrane (BioTrace NT Pure Nitrocellulose Transfer Membrane 0.2 µm, PALL Life Sciences), which was then placed onto LB agar plates containing 1 mM IPTG, 200 µM ZnCl₂ and 40 µg/mL kanamycin for overnight protein expression at room temperature. After expression, the membrane was placed into an empty petri dish and the cells were lysed by alternating incubations at -20°C and 37°C three times for 10 min each. To assay activity, 25 mL of 0.5% Agarose in 50 mM Tris-HCl buffer pH 7.5 containing 200 µM ZnCl₂ and 250 µM *p*-nitrophenyl phenylphosphonate (Sigma) was poured onto the membrane. Colonies with active enzymes developed a yellow color due to the hydrolyzed substrate. The most active colonies (~200 variants) were directly picked from plates for screening in 96-well plates.

5.3.5 Cell lysate activity screen in 96-well plates

To test the fitness and solubility of variants, individual wells of a 96-well plate containing 400 µl of LB media supplemented with 40 µg/ml kanamycin were inoculated with 20 µl of overnight culture and incubated at 30°C for 3 hours. Protein expression was

induced by adding IPTG to a final concentration of 1 mM and further incubation at 30°C (20°C and 37°C for testing temperature effect on expression) for 3 hours. Cells were harvested by centrifugation at 4,000×g for 10 min and pellets were frozen -80°C for at least 30 min. For lysis, cell pellets were resuspended in lysis buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 200 μM ZnCl₂, containing 0.1 % Triton X-100, 100 μg/ml lysozyme and 1 U/ml of benzonase) and incubated at 25°C with shaking at 1200 rpm for 1 hour. The cell lysates were clarified by centrifugation at 4,000×g for 20 min at 4°C. Clarified lysates were diluted (1000-fold for β-lactamase activity, 2-fold for phosphonate hydrolase activity) in order to obtain linear initial rates and measured against a single substrate concentration (90 μM for β-lactamase activity and 500 μM for phosphonate hydrolase activity).

5.3.6 Purification of Strep-tagged proteins

All variants were cloned as described above, transformed, overexpressed in *E. coli* BL21 (DE3) cells and purified as described in chapter two and in (Baier & Tokuriki 2014).

5.3.7 Enzyme kinetics

The kinetic parameters and activity levels of purified of enzyme variants were obtained as described previously (Baier & Tokuriki 2014). Briefly, phosphonate monoester hydrolysis (PMH) was monitored following the release of *p*-nitrophenol at 405 nm with an extinction coefficient of 18,300 M⁻¹ cm⁻¹ (Baier & Tokuriki 2014). The β-lactamase activity was monitored at 405 nm for the hydrolysis of the Centa substrate, and molar product formation was calculated with the extinction coefficient of 6,300 M⁻¹ cm⁻¹ (Bebrone et al. 2001).

5.3.8 Thermostability assay

The thermal stability of variants was measured with a thermal shift assay as described previously (Wyganowski et al. 2013). Briefly, enzyme variants (2 μM) were mixed with 5 × SYPRO Orange dye (Invitrogen) in a 20 μl reaction and heated from 25 to 95 °C in a 7500 Fast Real-Time PCR system (Applied Biosystems). Measurements were conducted in triplicate and unfolding was followed by measuring the change in fluorescence caused by binding of the dye (excitation, 488 nm; emission, 500–750 nm). The melting

temperature (T_m) is calculated from midpoint of the denaturation curve and values were averaged.

5.3.9 Protein purification for crystallization

The NDM1 and VIM2 protein variants were fused to a N-terminal His₁₀-tag containing a TEV (Tobacco Etch Virus nuclear-inclusion-a endopeptidase) cleavage site between the protein and the His₁₀-tag. Proteins were expressed in *E. coli* BL21 (DE3) cells in TB medium (400 ml) supplemented with 1% glycerol, 50 μ g/mL kanamycin and 200 μ M ZnCl₂. Cells were grown at 30 °C for 6 hours. The temperature was lowered to 22 °C and the cells were incubated for a further 16 hours and harvested by centrifugation for 15 minutes at 8,500 \times g (R9A rotor, Hitachi), then resuspended in buffer A (50 mM HEPES pH 7.5, 500 mM NaCl, 20 mM imidazole, and 200 μ M ZnCl₂) and lysed by sonication (OMNI sonic ruptor 400). Cellular debris was removed by centrifugation at 29,070 \times g for 60 minutes (R15A rotor, Hitachi). The supernatant was loaded onto a 5 ml Ni-NTA superflow cartridge (Qiagen) followed by extensive washing with buffer A prior to elution of proteins in buffer B (50 mM HEPES pH 7.5, 500 mM NaCl, 500 mM imidazole, and 200 μ M ZnCl₂). Protein-containing fractions were analyzed by SDS-PAGE (Bolt Mini Gels, Novex). The buffer B containing the proteins was exchanged to TEV reaction buffer (50 mM Tris-HCl pH 8.0, 0.5 mM EDTA, 1 mM DTT, and 150 mM NaCl) using HiPrep 26/10 desalting column (GE healthcare). 20% TEV protease (w/w) was added and incubated at 4 °C for 4 days. The TEV reaction buffer was exchanged to buffer A before TEV protease and His-tag containing debris were removed by Ni-NTA superflow column (5 mL, Quiagen). His-tag cleaved protein was then concentrated using a 10 kDa molecular weight cut-off MWCO ultrafiltration membrane (Amicon, Millipore) and loaded on HiLoad 16/600 Superdex 75 pg column (GE Healthcare). Proteins were eluted into crystallization buffers (described below).

5.3.10 Crystallization of NDM1-R10

NDM1-R10 protein in crystallization buffer (20 mM HEPES pH 7.5, 2 mM β -mercaptoethanol, 150 mM NaCl, and 100 μ M ZnCl₂) was concentrated to 15 mg/mL using a 10 kDa molecular weight cut-off ultrafiltration membrane (Amicon, Millipore)

and crystallized by the hanging drop method. The hanging drops were prepared by mixing protein solution (1 μL) and well solution (2 μL). Crystals appeared after two weeks in 0.1 M MES (pH 6.75) and 1.3 M MgSO_4 at 18 °C and continued to grow. Crystals were soaked in cryoprotectant solution for 30 seconds (precipitant, and 25% glycerol), and flash cooled in liquid nitrogen. The MES bound crystal diffracted to 1.67 Å at beam line MX1 at the Australian Synchrotron. The product bound structure was obtained by soaking the crystal in precipitant solution, containing 15 mM substrate for 3 minutes to 30 minutes before soaking in cryoprotectant solution and flash cooling in liquid nitrogen. The crystals diffracted to 1.68-2 Å at a beam line MX2 (0.9537 Å) at the Australian Synchrotron.

5.3.11 Crystallization of VIM2-R10

The first size exclusion peak of VIM2.R10 protein (dimeric fractions), in buffer containing 50 mM HEPES pH 7.5, 150 mM NaCl, and 200 μM ZnCl_2 , was concentrated to 2.6 mg/mL using a 10 kDa molecular weight cut-off ultrafiltration membrane (Amicon, Millipore) and crystallized by the hanging drop method. The drops were prepared by mixing a protein solution (2 μL), well solution (4 μL), 1 mM TCEP, and 2.5 mM PNPP. Crystals appeared after two weeks in 0.1 M HEPES (pH 7.5) and 1.2 M sodium citrate at 18 °C. Crystals were soaked in cryo-protectant solution for several minutes (precipitant, and 10% glycerol), and flash cooled in liquid nitrogen. The crystal diffracted to 2 Å at beam line MX1 (0.9537 Å) at the Australian Synchrotron.

5.3.12 Structural data collection and structure determination.

The crystallographic data were collected at 100 K at the Australian Synchrotron. Data were processed using XDS. Scaling was performed using Aimless in the CCP4 program suite. Resolution estimation and data truncation were performed by using overall half-dataset correlation $\text{CC}(1/2) > 0.5$. Molecular replacement was used to solve all structures with MOLREP using the structures deposited under PDB accession codes 3SPU and 1KO3 as starting models for NDM1 and VIM2, respectively. The model was refined using phenix.refine and Refmac v5.7 in CCP4 v6.3 program, and the model was subsequently optimized by iterative model building with the program COOT v0.7.

5.3.13 Molecular dynamics simulation

Molecular dynamics simulations VIM2 and NDM1 variants were performed using the Q simulation package (Marelius et al. 1998) and the OPLS-AA force field (Jorgensen & Maxwell 1996). The OPLS-AA compatible parameters for *p*-nitrophenyl phenylphosphonate (PMH substrate) were generated using MacroModel version 10.3 (Schrödinger LLC, v. 2014-1). Partial charges for PPP were calculated using the RESP procedure (Cieplak et al. 1995), with the use of Antechamber (AmberTools 12) (J. Wang et al. 2006) and Gaussian09 (Revision C.01 (Frisch et al. 2009)). The structure of VIM2-WT (PDB ID 4PVO) and NDM1-WT (PDB ID 4HL2) were obtained from the Protein Data Bank, and structure of NDM1-R10 with the PMH product bound (PDB ID 5K4M) was obtained as described above. The structures of single W93G mutants of both VIM2 and NDM1 were generated by manually mutating respective tryptophan residues to glycine of the WT structures. In the simulations of VIM2 chain B of the PDB structure was used. The PMH substrate was placed manually in the active site of the enzymes based on the position of the PMH product found in the crystal structure of NDM1-R10. The Zn²⁺ ions were described using a tetrahedral dummy model based on the dummy model originally described by Åqvist and Warshel (Åqvist & Warshel 1989). The model was built by placing four dummy atoms in a tetrahedral geometry around a central metal particle, and parametrised to reproduce the experimental solvation free energy and solvation geometry of the zinc ion (for the description of analogous octahedral dummy model and parameterization procedure see (Duarte et al. 2014)). All simulations were performed using surface-constrained all-atom solvent (SCAAS) boundary conditions (G. King & Warshel 1989) with a spherical droplet of water with a radius of 24 Å centered on the bridging hydroxide ion, containing all crystallographic water molecules, complemented with TIP3P water molecules (Jorgensen & Chandrasekhar 1983). All protein atoms and water molecules within 85% of the sphere were allowed to move freely with no restraints, atoms in the last 15% of the sphere were subject to 10 kcal mol⁻¹ Å⁻² positional restraints, and all atoms outside this sphere were subjected to 200 kcal mol⁻¹ Å⁻² positional restraints to maintain them at their crystallographic positions. Protonation states of all ionizable residues within the inner 85% of the simulation sphere were assigned using PROPKA 3.1 (Søndergaard et al. 2011; Olsson et al. 2011) and the

protonation states of histidine side chains were determined by visual inspection of the surrounding hydrogen-bonding pattern of each residue. All ionizable residues outside of the 85% of the sphere were kept in their uncharged forms to avoid simulation artifacts. All systems were initially equilibrated with $200 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ positional restraints over the total timescale of 95 ps, during which the alternating heating, cooling and reheating was performed to release steric clashes and equilibrate the positions of the solvent molecules and hydrogen atoms, and to reach the target simulation temperature of 300K. The initial equilibration was completed by performing 10 ns of simulation at 300K, which was followed by 100 ns production simulation, subject to further analysis. During the final stage of the equilibration and production simulation, weak $0.5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ restraints were applied on the PMH substrate in order to keep it within the simulation sphere, and $1.0 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ restraints were applied on the metal ions, side chains of the metal ligands and the bridging hydroxide ion to assure proper coordination geometry of the metal centers. Apart from the very initial stages of equilibration, the time step of 1 fs was used throughout the simulations. The results of the simulations were analyzed using: VMD (Humphrey et al. 1996), POVME (Durrant et al. 2011; Durrant et al. 2014) and Gromacs package (Abraham et al. 2015).

5.4 Results

5.4.1 The selection of evolutionary starting points

We previously showed that some metallo- β -lactamases (MBLs) exhibit a promiscuous PMH activity, which differs from the native β -lactamase reaction by its scissile bond (P-O vs. C-N), transition state geometry (trigonal bipyramidal vs. tetrahedral) as well as in the overall substrate shape and size (**Figure 5.1**) (Baier & Tokuriki 2014). PMH activity among MBLs can vary by up to a 100-fold ($k_{\text{cat}}/K_{\text{M}}$ ranges between 10^{-1} and $10^1 \text{ M}^{-1}\text{s}^{-1}$, **Figure 5.1 A**) and is 10^5 -fold to 10^7 -fold lower than the native activity. In contrast, the level of native β -lactamase activity is fairly similar among MBLs ($k_{\text{cat}}/K_{\text{M}} = \sim 10^6 \text{ M}^{-1}\text{s}^{-1}$, **Figure 5.1 A**). MBLs are found in various bacteria and are highly genetically diverged with pairwise sequence identities as low as 24%, despite their functional and structural similarity (**Figure 5.2** and **Table 5.1**). Thus, the genetic diversity among MBLs provides a great opportunity to test the evolvability of distinct sequences towards the same

promiscuous activity under highly controlled experimental conditions. From this set of MBLs, we selected NDM1 and VIM2, which have a pairwise amino acid sequence identity of 35% (**Figure 5.2**), as starting points for a comparative directed evolution experiment to examine their evolvability towards PMH activity. The rationale for our selection was that for both enzymes detailed functional and structural information was available, which facilitates subsequent characterization and interpretation of results (Garcia-Saez et al. 2008; D. King & Strynadka 2011). Despite genetic variation, the two enzymes exhibit the same level of β -lactamase activity, protein solubility, thermostability and an overall structural similarity (R.M.S.D. of 1.03 Å, **Figure 5.2** and **5.5**). Notably, the initial k_{cat}/K_M for PMH activity is 10-fold higher for VIM2 compared to NDM1 (**Figure 5.5 B**).

	FIM1	EBL1	NDM1	VIM2	VIM1	VIM7	CcrA	SPM1	IMP1
FIM1	-	-	-	-	-	-	-	-	-
EBL1	51%	-	-	-	-	-	-	-	-
NDM1	50%	63%	-	1.03	-	0.97	1.13	1.05	1.14
VIM2	45%	41%	44%	-	-	0.36	1.07	1.03	1.11
VIM1	44%	40%	43%	94%	-	-	-	-	-
VIM7	42%	38%	42%	84%	84%	-	0.85	0.94	1.15
CcrA	43%	41%	41%	45%	44%	44%	-	0.97	1.25
SPM1	30%	28%	28%	34%	32%	35%	33%	-	1.13
IMP1	35%	34%	36%	39%	38%	38%	42%	34%	-

Figure 5.2 Sequence identity and structure similarity among selected B1 β -lactamases.

Pairwise sequence identities were calculated from a multiple sequence alignment using ClustalW2 (standard parameters), which was then used to calculate the identities using the web based program SIAS (<http://imed.med.ucm.es/Tools/sias.html>) with gaps taken into account. To determine pairwise structural similarity we computed the root mean standard deviation (RMSD) between all structure pairs using the align command in PyMOL.

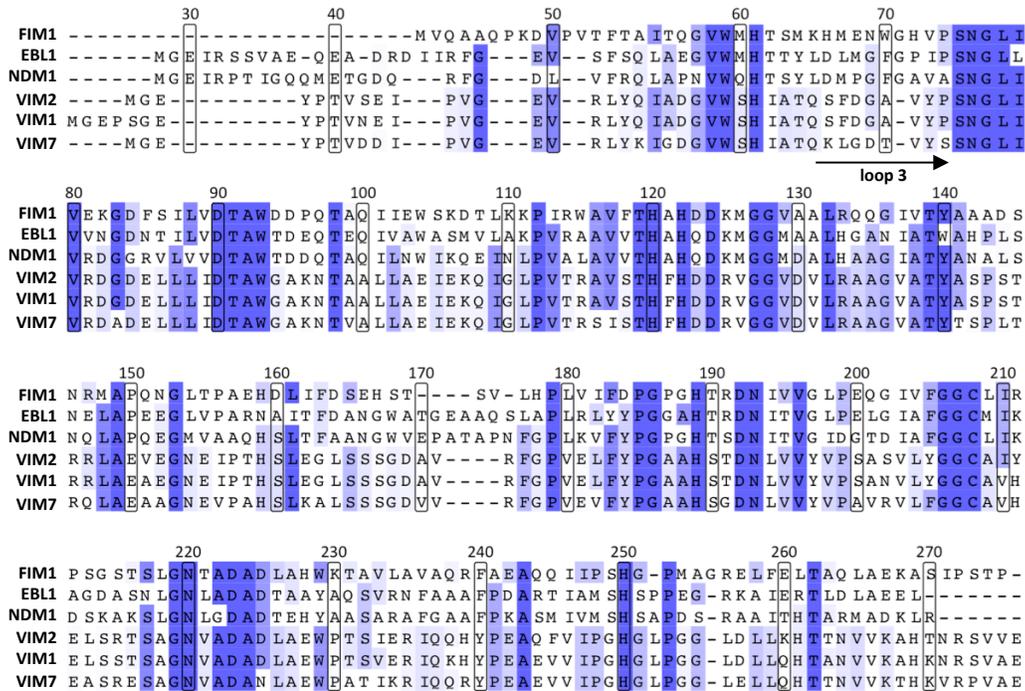


Figure 5.3 Sequence alignment of selected B1 β-lactamases.

Conserved residues are colored in shades of blue. Residue numbering is based on NDM1 (PDB-ID 3spu).

5.4.2 Directed evolution strategy

The enzymes were subjected to the same directed evolution procedure (Figure 5.4). Briefly, the enzymes were fused to a maltose binding protein (MBP)-tag for periplasmic expression in *E. coli*. Randomly mutagenized libraries of the enzymes were generated by error-prone PCR, resulting in 1~2 amino acid substitutions per gene. Due to low PMH activity of the initial variants, a direct screening of PMH activity on agar plate did not enable us the detection of positive clones. Therefore, we employed a β-lactam (ampicillin) antibiotic preselection (purifying selection for the enzymes' native activity) for the first eight rounds of the laboratory evolution experiment. The ampicillin concentration was ~64-fold below the enzymes' minimum inhibitory concentration (MIC), thus the process purges out non-functional variants but still retains variants that are mildly compromised in their native activity (Bebrone 2007). Colonies grown on the

prescreening plates were picked into 96-well plates (in total 396 variants for each enzyme per round), regrown, lysed and screened for PMH activity. The most improved variants were isolated, sequenced and used as templates for the next round of evolution. For the last two rounds (R9 and R10) the antibiotic prescreening was replaced with a direct colorimetric PMH activity screening on agar plates, allowing a direct screening of 2,000-3,000 variants per round. Overall, during the ten rounds of directed evolution a total of more than 5,200 ($8 \times 400 + 2 \times 2000$) functionally enriched variants were screened for each enzyme. Additionally, three rounds of DNA shuffling were performed for each enzyme to recombine beneficial mutations when several improved variants were identified (**Table 5.2**).

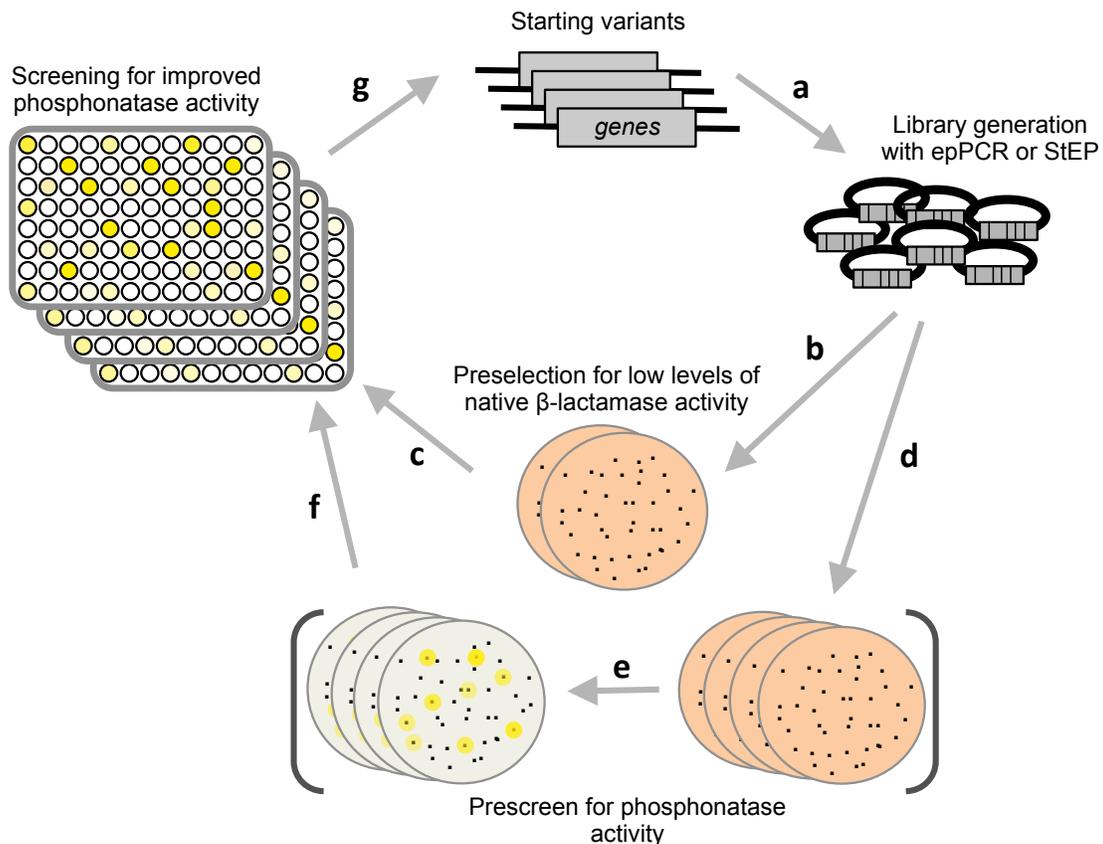


Figure 5.4 Overview of the laboratory evolution strategy.

(a) Starting variants were mutagenized using error-prone PCR or StEP (staggered extension process) recombination and ligated into a vector containing a N-terminal MBP-tag and periplasmic expression signal peptide. The resulting library was then transformed into *E. coli* BL21 cells for selection. (b) In the first eight rounds the library was plated on agar plates containing 4 µg/ml ampicillin in order to preselect for functional variants. (c) Subsequently, surviving colonies were directly picked

from plates for rescreen in 96-well plates. **(d)** In the last two rounds a direct prescreen for phosphonate was applied and colonies were replicated onto a nitrocellulose membrane. **(e)** After expression and cell lysis the phosphonate of the variants was assayed and active variants could be identified through the occurrence of a yellow color. **(f)** The most active colonies (~200 variants) were directly picked from plates for screening in 96-well plates. **(g)** The most improved variant(s) served as the starting point for the next round of directed evolution.

Table 5.2 Screening, selection and mutations of directed evolution rounds.

Round	Prescreen	Variants screened	NDM1 mutations	VIM2 mutations
1	4 µg/ml ampicillin	384	W93G, N166T	V72A (S)
2	4 µg/ml ampicillin	384	K211R, G222D (S)	D223A, F67L
3	4 µg/ml ampicillin	384	Q151R	S202R
4	4 µg/ml ampicillin	384	S251F	T64A (S)
5	4 µg/ml ampicillin	384	M154V, D96A (S)	G35R, V41A
6	4 µg/ml ampicillin	384	D223E	N154T
7	4 µg/ml ampicillin	384	N103K	T191P, T263S
8	4 µg/ml ampicillin	384	A233V	E150K
9	PMH	2000-3000	L49P	V46D
10	PMH	2000-3000	V88M	D68N (S)

(S) = DNA Shuffling was performed in this round

5.4.3 Fitness improvement toward PMH activity

In our directed evolution experiment, we defined “enzyme fitness” as the level of enzymatic activity in *E. coli* cell lysate, a function of catalytic efficiency and soluble expression, which is the direct measure of our selection screening. Initially, VIM2 showed 4-fold higher enzyme fitness compared to NDM1 (**Figure 5.5 A**). For both enzymes, fitness steadily improved over the ten rounds of directed evolution (**Figure 5.5 A**). Furthermore, both trajectories exhibit diminishing returns in fitness improvements, *i.e.*, large improvements in initial rounds are gradually diminishing in later rounds until no significant improvement can be obtained anymore, which is consistent with other exhaustive laboratory evolution experiments (Tokuriki et al. 2012; Kaltenbach et al.

2015; Noor et al. 2012; Miton & Tokuriki 2016). However, despite these similarities, the two trajectories show substantial differences in their initial improvements as well as the level of fitness they reached (**Figure 5.5 A**). Interestingly, despite the fact that NDM1 was the lower starting point, NDM1-R10 (the final variant of the NDM evolution after ten rounds of directed evolution) reached a 16-fold higher fitness level compared to VIM2-R10. Overall, NDM1-R10 exhibits a 3600-fold increase in fitness, whereas VIM2-R10 only improved by 50-fold, which is a 70-fold difference (**Figure 5.5 A**). Together, over the ten rounds of directed evolution the fitness improvement for PMH activity varies significantly between NDM1 and VIM2, despite the identical directed evolution conditions.

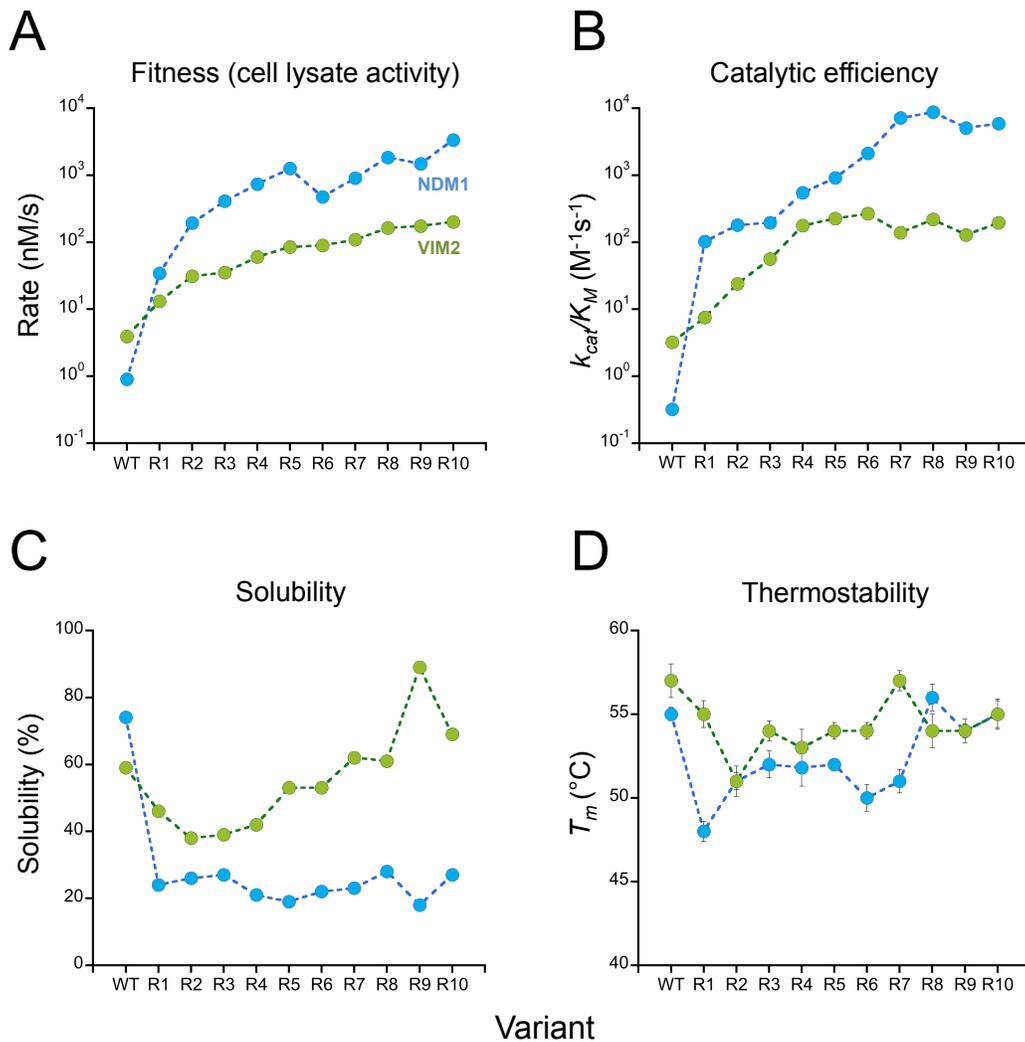


Figure 5.5 Phenotypic adaptations of NDM1 and VIM2 during the directed evolution experiment.

(A) Fitness improvement of NDM1 (blue) and VIM2 (green) (phosphonatase activity in cell lysate), which represents the selection criteria in the directed evolution experiment. WT indicates the wild-type enzyme, and R1 to R10 represents the isolated variant of each round. Values represent the average of three independent experiments. (B) Catalytic efficiencies (k_{cat}/K_M) of purified variants for the phosphonatase activity. (C) The soluble expression of variants as determined by SDS-PAGE analysis (Figure 5.6). (D) Thermostability of purified variants as calculated from the midpoint of the thermal denaturation curve in a thermal shift assay. Error bars represent standard deviation from three independent assays.

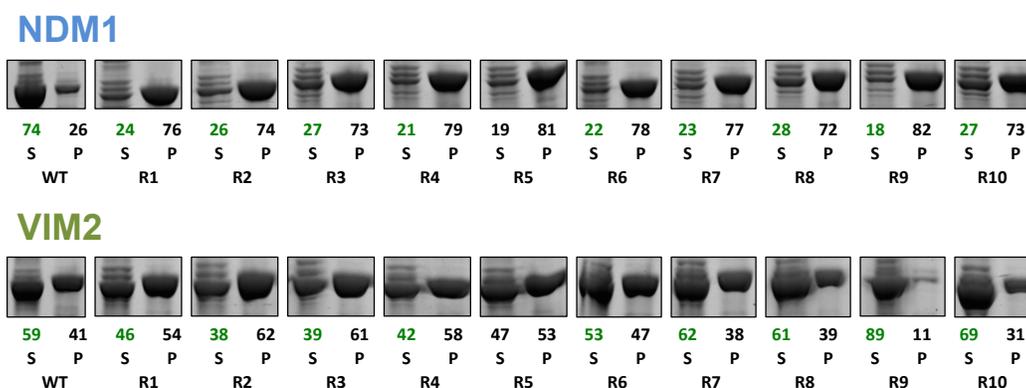


Figure 5.6 SDS-PAGE analysis of solubility of NDM1 (top) and VIM2 (bottom) variants.

The soluble and insoluble pellet fractions of cell lysates (S (in green) and P, respectively) were analyzed by SDS-PAGE. The percentage of protein in the soluble and insoluble fraction was determined by the relative intensities of the supernatant and pellet bands of the protein variant.

5.4.4 Different activity and solubility changes underlie fitness improvements

Enzyme fitness, or enzymatic activity in cell lysate in our system, is largely associated with two phenotypic parameters: catalytic efficiency (k_{cat}/K_M) and the amount of functional and soluble enzyme in the cell lysate ($[E]$) (Bershtein et al. 2012; Tokuriki & Tawfik 2009c). To reveal how both parameters change during the evolution of our two enzymes targets, we measured k_{cat}/K_M and soluble expression $[E]$ of all variants (Figure 5.5). We find that the fitness improvement of NDM1 underlies a k_{cat}/K_M increase of 20,000-fold, from 0.3 to 5900 $M^{-1}s^{-1}$ (Figure 5.5 B). Yet, the increase occurred at the expense of soluble protein expression (Figure 5.5 C and 5.6). In particular, a strong

function-solubility tradeoff was observed in round one. The solubility of NDM1 decreased from 74% to 24%, and the solubility remained at ~20% throughout the trajectory, while $k_{\text{cat}}/K_{\text{M}}$ gradually increased over the remaining rounds. Similarly in the VIM2 trajectory, $k_{\text{cat}}/K_{\text{M}}$ mainly contributed to the increase in enzyme fitness (50-fold), which improved in the first four rounds by 60-fold (from 3.2 to 200 $\text{M}^{-1}\text{s}^{-1}$), but eventually stagnated after round 5 (**Figure 5.5 B**). Function-stability tradeoffs are also observed in the first two rounds of the VIM2 trajectory (from 59% to 38%). In contrast to NDM1 however, VIM2 further improved fitness after round 5 by regaining and increasing solubility instead of $k_{\text{cat}}/K_{\text{M}}$ (**Figure 5.5 C and 5.6**). Together, NDM1-R10 reached a 30-fold higher $k_{\text{cat}}/K_{\text{M}}$ for PMH activity compared to VIM2-R10 (5900 vs 200 $\text{M}^{-1}\text{s}^{-1}$) despite the fact that VIM2 possessed 10-fold higher activity level prior to the evolution experiment. Thus, the $k_{\text{cat}}/K_{\text{M}}$ improvements for PMH activity are overall 300-fold higher for NDM1 compared to VIM2, which further highlights the substantial difference in evolvability between these two enzymes (**Figure 5.5 B**). Notably, the $k_{\text{cat}}/K_{\text{M}}$ of NDM1-R10 (5900 $\text{M}^{-1}\text{s}^{-1}$) reached the level of two phosphonate hydrolases from the alkaline phosphatase superfamily, *RIPMH* and *BcPMH*, $k_{\text{cat}}/K_{\text{M}}$ of 5300 and 15000 $\text{M}^{-1}\text{s}^{-1}$, respectively, which suggests that NDM1-R10 potentially reached the level of natural PMH enzymes.

We also monitored the evolution of the native β -lactamase activity and found that it was only marginally affected in both trajectories, at least when using the generic β -lactamase substrate CENTA (**Figure 5.7**). Thus, the two enzymes retained a generalist phenotype; in particular NDM1-R10, which exhibits a $k_{\text{cat}}/K_{\text{M}}$ of over 10^4 $\text{M}^{-1}\text{s}^{-1}$ for β -lactamase and PMH activity. The most likely reason for retaining the β -lactamase activity could be that we employed a low level ampicillin selection during the first 8 rounds of the evolution in order to enrich for catalytically functional variants, although the two later rounds with direct PMH screening (no ampicillin selection) did not result in a loss of β -lactamase activity. Taken together, the distinct fitness improvements between NDM1 (3600-fold) and VIM2 (50-fold) are combination of activity and solubility changes, but each enzyme adapted in a different way. NDM1 improves fitness by catalytic activity at the expense of protein expression, whereas VIM2 improves fitness through a combination of catalytic efficiency and solubility.

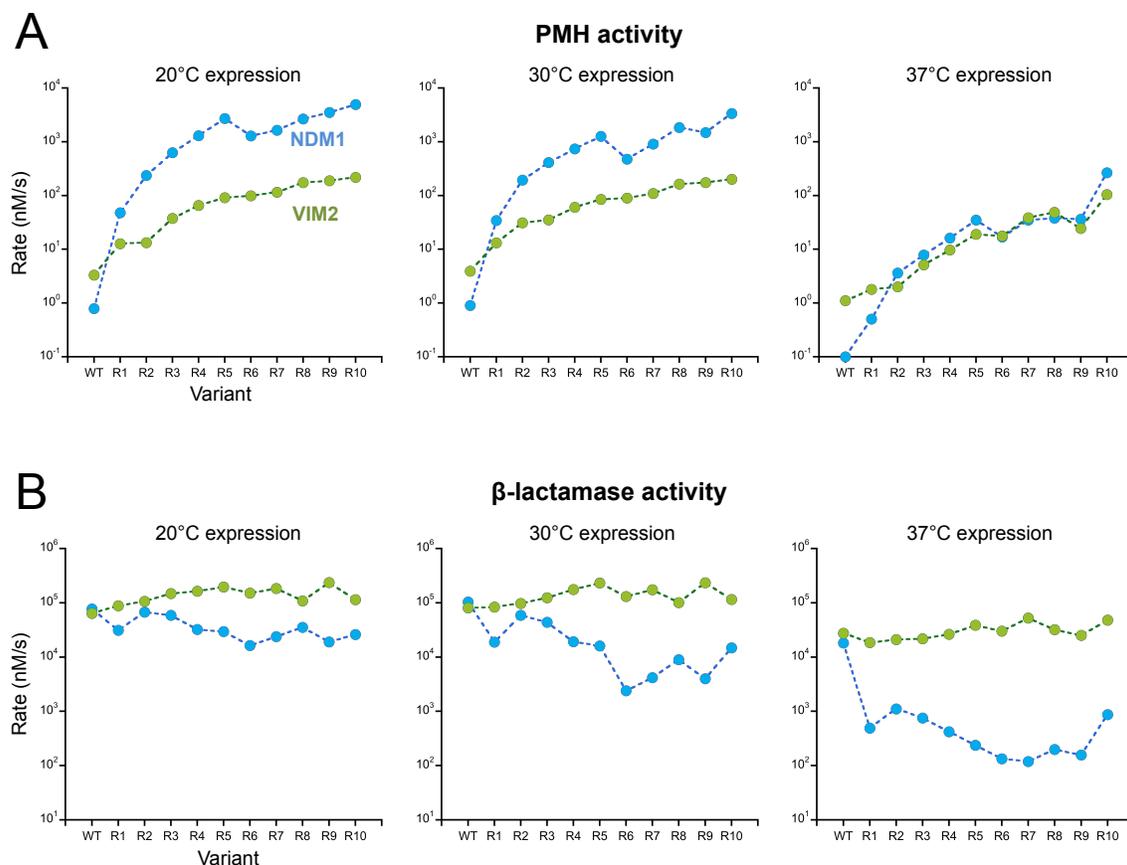


Figure 5.7 Effect of expression temperature on cell lysate activities.

Variants of NDM1 and VIM2 were expressed in 96-well plates at the indicated temperature and tested for (A) PHM and (B) β -lactamase activity. Cell lysate preparation and activity measurement were performed identical for all expression temperatures as described in material and methods.

5.4.5 Correlation between solubility, thermostability and structural stoichiometry

To examine whether changes in protein solubility [E] were associated with changes in thermostability (T_m), we measured the T_m of all variants. We found that the decrease in [E] in the initial rounds has a similar trend with a decrease in T_m for NDM1 and VIM2 (Figure 5.5 C and D). In later rounds however, the T_m does not correlate with [E]. The T_m of VIM2 variants remains largely unchanged in later rounds, whereas [E] increased (Figure 5.5 C and D). On the contrary, NDM1 increased T_m after round 2, yet [E] remains unchanged (Figure 5.5 C and D). Thus, we speculate that changes in solubility are likely to be associated with protein folding (or kinetic stability), rather than the T_m of the folded protein (Wyganowski et al. 2013). To test this hypothesis, we decreased

(20°C) and increased (37°C) the temperature of expression (30°C during the evolution) for all variants and assayed their fitness (cell lysate activity) for PMH and β -lactamase activity, which together provides a good proxy for expression and solubility (**Figure 5.7**). The experiment revealed that higher expression temperatures significantly decreased the fitness levels of NDM1 variants, whereas VIM2 variants are more robust to a temperature change during expression (**Figure 5.7**).

To determine whether there is a change in stoichiometry (monomer to dimer transition) associated with the observed change in solubility, we performed size exclusion chromatography for the wild-type and most evolved variants (**Figure 5.8**). Previous studies suggest that VIM2 exists as a monomer in solution, whereas NDM1 can also partly exist as a dimer (Garcia-Saez et al. 2008; D. King & Strynadka 2011). Interestingly, although size exclusion peaks for VIM2-WT, NDM1-WT and NDM1-R10 all showed a single monomeric peak, the chromatogram of VIM2-R10 showed an additional peak, corresponding to a dimeric form of VIM2-R10, as estimated by a comparison with a molecular weight standard (**Figure 5.8**). The observation that VIM2 changes from a pure monomer to a monomer/dimer mixture in solution, might partially explain its increase in solubility and robustness to higher expression temperatures during the trajectory, as dimerization has also been previously observed during directed evolution experiments and linked to improved solubility and stability (Fraser et al. 2016; Qu et al. 2000; Thoma et al. 2000; Bershtein et al. 2012).

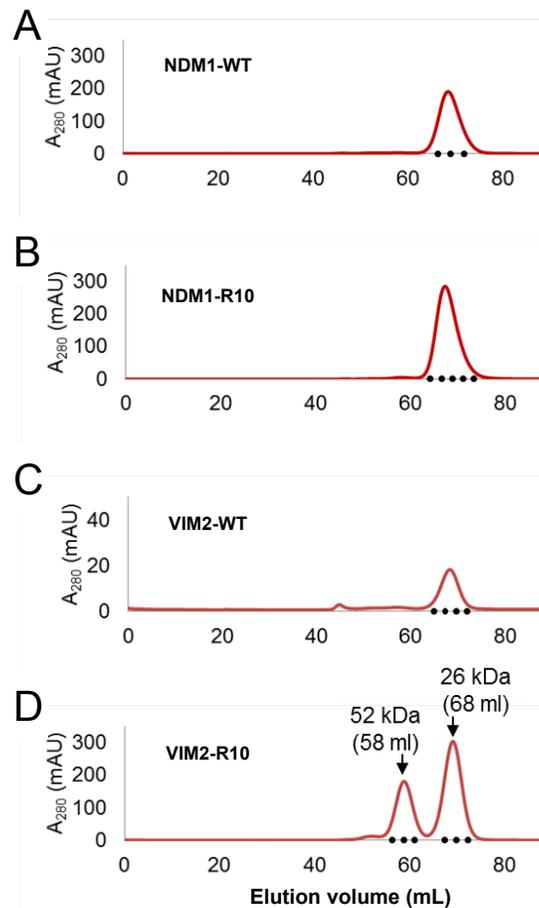


Figure 5.8 Size exclusion chromatography.

Size exclusion chromatograms of NDM1 (A), NDM1-R10 (B), VIM2 (C) and VIM2-R10 (D) are shown (Superdex 75, GE Healthcare). Protein sizes were identified based on the calibration curve on the manufacturer's instruction (GE Healthcare) for Gel filtration calibration kits LMW (low molecular weight). N. Hong performed this experiment and prepared the figure.

5.4.6 Different mutational paths support the distinct phenotypic outcomes

To investigate the mutational solutions of each trajectory we sequenced all variants and mapped the mutations onto the wild-type structures of NDM1 and VIM2, respectively. Each trajectory accumulated 13 mutations over ten rounds of directed evolution (**Figure 5.9 A** and **Table 5.2**), with most mutations located around the respective active sites (**Figure 5.9 B** and **C**). However in each trajectory, the mutations are confined to different structural areas (**Figure 5.9**). For VIM2, 6 out of the 13 mutations occurred within or next to loop 3 (**Figure 5.9 C**). On the contrary, NDM1 obtained only one mutation near

determined the fitness and mutations of the most improved variants, *i.e.* we “replayed” the evolution. Overall, the additional screenings revealed similar trends to the initial library, with NDM1 variants generally displaying substantially higher improvements than VIM2 variants (**Figure 5.10**). Furthermore, initial mutations fixed in both prior trajectories were repeatedly identified among the most improved variants: W93G (R1) for NDM1 and V72A (R1) and F67L (R2) for VIM2 (**Figure 5.10**). In order to investigate why later mutations were not isolated in these additional screenings, which were also beneficial in the trajectory, we measured the effect of mutations obtained in round 2 to 4 on the background of the wild-type enzymes. We focused only on the mutations up to round 4, because the activity improvement of later mutations is relatively low (**Figure 5.11**). Interestingly, the positive effect of mutations is more pronounced within the trajectory, and less in the background of the wild type enzymes, in particular for NDM1, which most likely explains why we isolated only initial mutations in the additional screening experiments (**Figure 5.10**). Note, that such epistatic interactions and dependencies among mutations have been observed in several enzyme evolution studies (Miton & Tokuriki 2016). Taken together, both adaptive trajectories are likely to occur deterministically and repeatedly, because of, first, limited availability of functional mutations from the starting points and, second, initial mutations permitting the later fixation of beneficial mutations. Finally, we would like to note that the observed repeatability might also partially arise from the functional constraints imposed by the antibiotic resistance prescreen. However, both enzymes have the same constraints and, nevertheless, the mutational solutions and fitness improvements are different for both trajectories, which will be further explored in the next section.

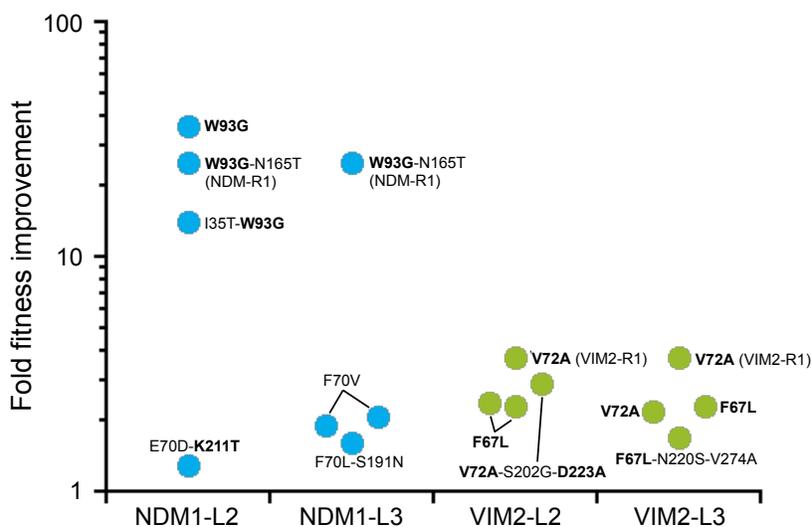


Figure 5.10 Improved variants isolated in two additional directed evolution experiments.

Two additional independent libraries were generated for NDM1 and VIM2 and screened for improved PMH activity in the cell lysate fitness. The three most improved variants of each library were sequenced and their fitness improvement measured. Mutations that occurred in the respective trajectory are highlighted in bold. The previously selected variants of the completed directed evolution experiments are shown for reference. Note that for NDM1-R1 the mutation N165T had no functional effect, and thus is not highlighted in bold.

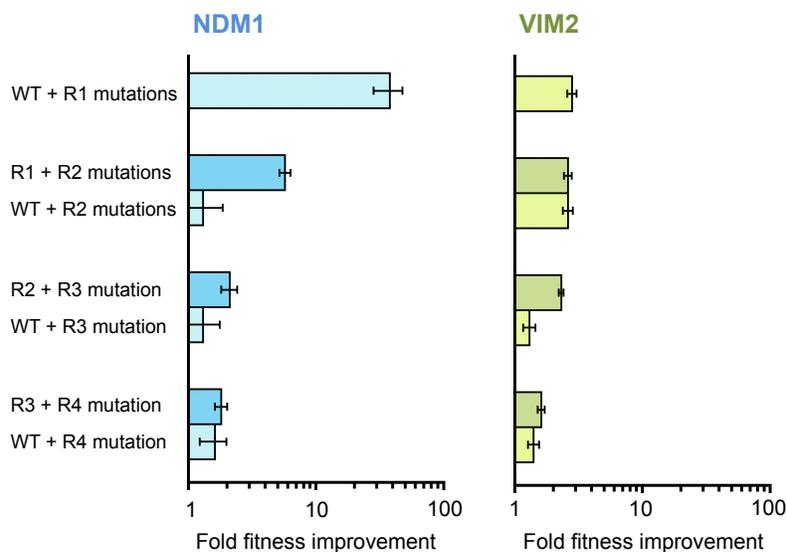


Figure 5.11 Epistasis analysis of trajectory mutations.

The mutations occurring in the trajectory of NDM1 (left) and VIM2 (right) were introduced into the respective wild type (WT) sequence and the change in fitness was compared to the fitness improvement observed in the trajectory. Errors bars represent the propagated standard deviation from three independent experiments.

5.4.8 The compatibility of initial mutations in different genetic backgrounds

We further investigated the reason for the two enzymes to adopt distinct mutational and phenotypic solutions, instead of seeing the fixation of identical mutations in both trajectories. For example, it was puzzling to observe that, although residue W93 is largely conserved among B1 β -lactamases including VIM2, this position was only mutated in the NDM1 trajectory (providing a 300-fold increase in k_{cat}/K_M) but never occurred on the VIM2 background. To confirm this observation, we introduced the mutation W93G in the VIM2 background and tested its “compatibility” and effect on fitness. Unlike in the NDM1 background, W93G decreases VIM2’s activity by 10-fold, albeit surprisingly not affecting its solubility (**Figure 5.12 A and B**). We also tested four other hydrophobic residues (Ala, Val, Leu or Phe) at position W93 in VIM2, which however also had a deleterious effect on fitness similar to W93G (**Figure 5.12 C**). Thus, W93G is functionally incompatible with the VIM2 background, which explains why it is not occurring in the directed evolution experiment. But how does the W93G mutation affect the PMH activity in other B1 β -lactamases? To address this, we introduced the mutation in four related B1 β -lactamases, FIM1, EBL1, VIM1 and VIM7, and assayed their PMH and β -lactamases activities. Because of the low PMH activity of some enzymes variants in cell lysate, all variants were purified and their catalytic activities (at a fixed enzyme and substrate concentration) were measured. Our results indicate that, similar to the NDM1 background, W93G improves the PMH activity of FIM1 and EBL1 (which share 45% and 56% sequence identity to NDM1, respectively) albeit to a lesser extent (**Figure 5.13 A**). On the other hand, W93G decreased PMH activity for VIM1 and VIM2 by 7-fold and 10-fold, respectively (**Figure 5.13 A**). However, in the case of VIM7 (80% sequence identity to VIM2) PMH activity increased by 3.2-fold with W93G (**Figure 5.13 A**). Our results indicate that the highly beneficial character of the W93G mutation for PMH activity is not accessible for VIM1 and VIM2, and provides different improvements for other homologs. Thus, its functional effect is highly contingent on the genetic background on which it occurs. We also tested the mutational effect of W93G on the native β -lactamase activity, which was deleterious for all VIM variants we tested (\sim 10-fold decrease), while it was slightly beneficial (\sim 2-fold increase) for NDM1 and EBL1 and neutral for FIM1.

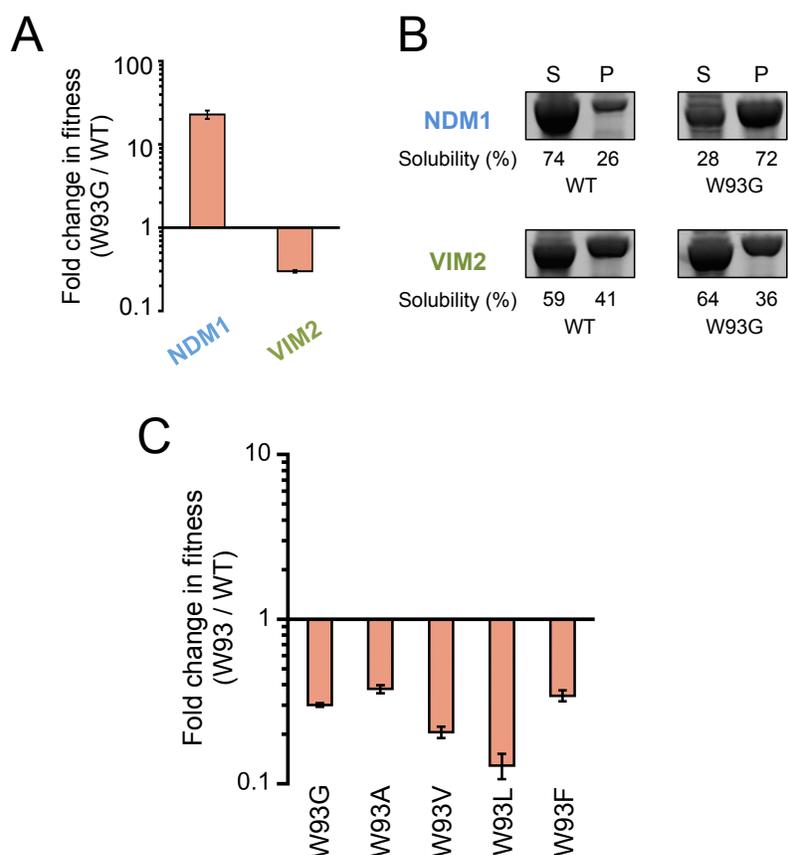


Figure 5.12 Fitness and solubility effect of W93G for NDM1 and VIM2.

(A) Fold change in fitness (PMH activity in cell lysate) of W93G mutants compared to WT variants. (B) Change in solubility of W93G mutants compared to WT variants. The soluble and insoluble fractions of cell lysates (S and P, respectively) were analyzed by SDS-PAGE and the relative intensities of the supernatant and pellet bands were used to calculate the percentage of solubility. (C) Fold change in fitness of various hydrophobic residues at position W93 in VIM2.

Furthermore, we examined the effect of the initial mutation of VIM2, which occurred in round one, V72A, on the six B1 β -lactamases (Figure 5.13 B). Note that the position is not conserved among B1 β -lactamases (Figure 5.3). Thus, when the wild type amino acid is not V72 we introduced the original and mutant VIM2 amino acid (V and A at 72), and assessed the activity of each variant in order to calculate the effect on activity. The functional effect of V72A for PMH and β -lactamase activity is far less pronounced compared to W93G and neutral to most enzymes besides VIM1 and VIM2, which improved PMH activity by only 2-fold (Figure 5.13 B). Thus, the initial mutation of

VIM2, V72A, is beneficial to neutral for most homologs, but the improvements are only marginal.

Finally, assessing changes in T_m of the mutant and wild-type variants revealed a common trend between increase in PMH activity and decrease in thermostability (**Figure 5.13 C**). For example, W93G improves the PMH activity of NDM1 and FIM1, but decreases their T_m by $>7^\circ\text{C}$ (**Figure 5.13 C**). On the other hand, the three VIM enzymes and EBL1 showed only modest decrease ($<3^\circ\text{C}$) in T_m with W93G (**Figure 5.13 C**). In the case of the VIM2 mutation, V72A, thermostability was only slightly impaired for VIM1 and VIM7 with a 3.9°C and 2.7°C decrease, respectively (**Figure 5.13 D**). However, at this point it is difficult to rationalize the biophysical and structural consequence of the W93G on the protein structures.

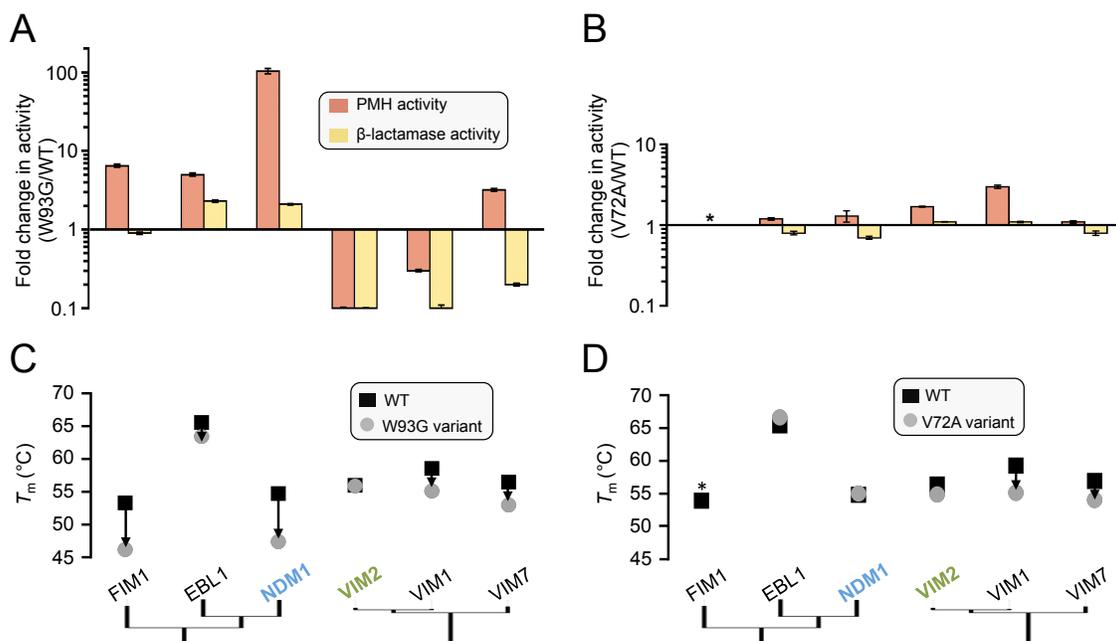


Figure 5.13 Functional compatibility of initial mutations among related B1 β -lactamases.

Fold change in PMH and β -lactamase activity of (A) W93G and (B) V72A mutant variants compared to wild type levels. Activity levels of purified enzymes were measured at one enzyme ($1\ \mu\text{M}$ for PMH and $1\ \text{nM}$ for β -lactamase activity) and substrate ($500\ \mu\text{M}$ for PMH and $100\ \mu\text{M}$ for β -lactamase activity) concentration. Change in thermostability of (C) W93G and (D) V72A (grey circles) mutant variants compared to the wild type (black square). Thermostability of purified variants, which was calculated from the midpoint of the thermal denaturation curve in a thermal shift assay. Asterisks indicate that the fold change in activity could not be determined, because one of the variants was not soluble. The phylogenetic

relationship of the enzymes is shown at the bottom. Errors bars represent the propagated standard deviation calculated from three replicate measurements.

5.4.9 The structural basis for improved PMH activity

To obtain insights into the structural changes that occurred during the trajectories, our collaborators in the Jackson research group at ANU in Canberra, Australia, solved the crystal structures of the most evolved variants, NDM1-R10 and VIM2-R10. The wild-type structures of NDM1 (PDB ID 3SPU) and VIM2 (PDB ID 1KO3) have been previously published (Garcia-Saez et al. 2008; D. T. King et al. 2012). The overall scaffold of VIM2 and NDM1 is composed of two linked half β/α barrels where the core beta sheets of each half barrel are sandwiched to fold a single monomeric enzyme (**Figure 5.1**). We solved the structure of NDM1-R10, in complex with the phenylphosphonic acid product in the active site and in the apo form (MES bound in the active site) to a resolution of 1.7 Å (**Figure 5.14**). As described earlier, size exclusion chromatography showed monomeric peaks for VIM2-WT, NDM1-WT and NDM1-R10, whereas the size exclusion chromatogram of VIM2-R10 showed a monomeric and a dimeric peak (**Figure 5.8**). For VIM2-R10, monomeric and dimeric fractions were screened separately for crystallization. However, the monomeric fraction never crystallized, but the dimeric fraction produced crystals and diffracted to a resolution of 2.2Å (**Figure 5.16**). Although VIM2-R10 crystals were also soaked in a substrate-containing solution, no substrate or product was identified in the structures. Data collection and refinement statistics for NDM1-R10 and VIM2-R10 crystals are presented in **Table 5.3**.

5.4.10 Structural adaptation of NDM1-R10

To investigate the structural basis of the improved PMH activity of NDM1-R10, we compared the structural changes between NDM1-R10 and NDM1-WT (**Figure 5.14 A**). A total of 13 mutations accumulated, including five located in the active site (W93G, D223E, K211R, G222D, S251F), four mutations not directly but in the vicinity of the active site (A233V, L49P, M154V, V88M), and four mutations on the surface (Q151R, D96A, N103K, N166T) (**Figure 5.9**). The overall structure and metal ion positions

between NDM1-R10 and NDM1-WT remained very similar (R.M.S.D. < 1 Å, **Figure 5.14 A**). However, a closer analysis disclosed that two major structural changes are likely associated with the large increase in catalytic efficiency of NDM1-R10. First, W93G alone (round 1) causes a 300-fold $k_{\text{cat}}/K_{\text{M}}$ increase, removed the steric hindrance between the side chain of W93 and the phenyl ring of the PMH substrate (**Figure 5.14 B**). At the same time, W93G caused a major displacement of loop 3 (L3) towards the active site (~6 Å), which appears to create a complementary pocket for substrate/product binding underneath loop 3 (**Figure 5.14 B and C**). Second, several mutations were involved in remodeling and stabilizing the active site loop L10, which may have further improved complementary binding of the PMH substrate (**Figure 5.14 C**). Four mutations occurred on or near L10 (G222D, K211R, D223E, S251F), which introduced new hydrogen bonds and potentially stabilized conformations around L10 (**Figure 5.14 D**). In detail, G222D (round 2) in L10 generated two new hydrogen bonds with the side chain of N220, which caused its rotation by ~180° degrees. In addition, K211R (round 2) forms a new hydrogen bond with the backbone carbonyl of S217 on L10. The side chain conformation of K211R is potentially stabilized, and thus its interaction with S217, by the subsequent mutation S251F (round 4) on L12. Taken together, our structure analyzes of NDM1-R10 revealed that the activity improvement of NDM1 is primarily due to optimizing the active site complementarity for the PMH substrate, through alterations of the active site loops L3 and L10. Nevertheless, the catalytic center of NDM1-WT and NDM1-R10 is retained, which is consistent with the fact that β-lactamase activity remained almost unchanged over the trajectory (**Figure 5.7**).

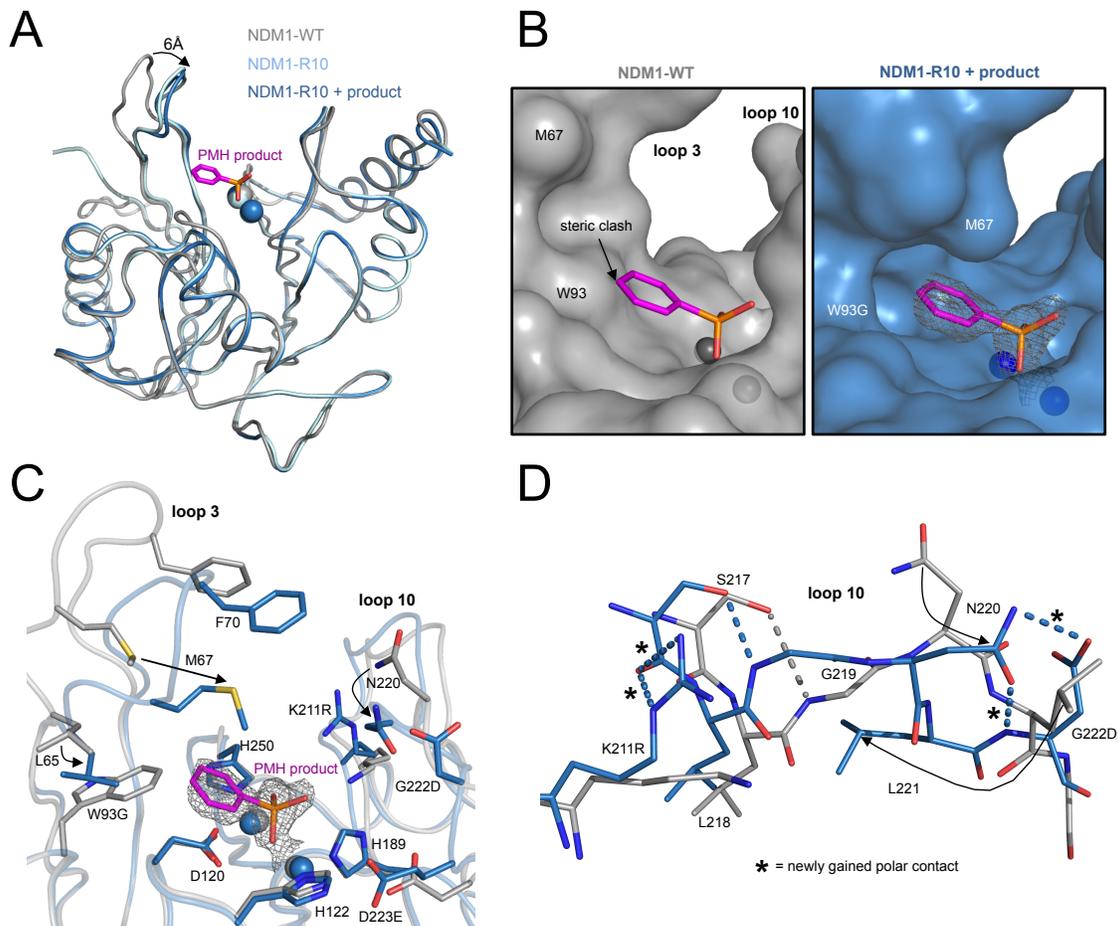


Figure 5.14 The structural basis for improved PMH activity of NDM1-R10.

(A) Comparison of the C- α backbone of NDM1-WT (gray, PDB ID 3spu) and NDM1-R10 in the apo (cyan, PDB ID 5JQJ) and PMH product (magenta sticks) complexed form (blue, PDB ID 5K4M). The active site metal ions are shown as spheres and colored according to the structure. (B) Surface views of the active site of NDM1-WT (left, grey) and NDM1-R10 with the PMH product bound (right, blue). The product binding in NDM1-WT is based on the complexed NDM1-R10 structure and was generated by superimposing both structures in PyMol. Electron density ($2F_o-dF_c$) of the PMH product is contoured at 1σ . (C) Active site superposition of NDM1-WT (gray, PDB ID 3spu) and NDM1-R10 (blue, PDB ID 5K4M) with the PMH product bound (magenta). Electron density ($2F_o-dF_c$) of the PMH product is contoured at 1σ . Arrows indicate repositioning of active site residues. (D) Comparison of polar contacts in loop 10 between NDM1-WT (gray) and NDM1-R10 (blue). Polar contacts were identified and visualized using PyMOL. NDM1-R10 gains four new polar contacts (indicated by asterisk) in loop 10, which potentially rigidifies its position. N. Hong performed the experiment and solved the crystal structures of NDM-R10.

Table 5.3 Crystallographic data collection and refinement statistics.

Variant	NDM1-R10	NDM1-R10 (product)	VIM2-R10
PDB ID	5JQJ	5K4M	-
Wavelength (Å)	0.9537	0.9537	0.9537
No. copies in an asymmetric unit	1	1	4
Resolution range (Å)	34.45-1.67 (1.73-1.67)	39.13-1.98 (2.05-1.98)	37.89-2.19 (2.27-2.19)
Space group	C 2 2 2 ₁	C 2 2 2 ₁	C 1 2 1
Unit cell (Å, °)	37.80 137.72 77.46 90 90 90	37.82 138.16 78.20 90 90 90	128.60 41.67 156.76 90 99 90
Total reflections	46608 (3668)	29257 (2838)	80087 (7307)
Unique reflections	23318 (1837)	14613 (1421)	41963 (4025)
Multiplicity	2.0 (2.0)	2.0 (2.0)	1.9 (1.8)
Completeness (%)	97.22 (77.72)	99.72 (99.37)	97.82 (96.08)
Mean I/sigma(I)	17.93 (1.89)	11.13 (5.78)	11.02 (1.95)
Wilson B-factor (Å ²)	19.27	14.37	32.1
^a R-merge	0.029 (0.46)	0.029 (0.08)	0.050 (0.50)
^b R-meas	0.04	0.04	0.071
^c CC1/2	0.999 (0.588)	0.999 (0.978)	0.997 (0.575)
^d CC*	1.00 (0.861)	1.00 (0.994)	0.999 (0.854)
R-work	0.144 (0.240)	0.142 (0.150)	0.210 (0.223)
R-free	0.187 (0.287)	0.208 (0.198)	0.284 (0.278)
Number of non-hydrogen atoms	2052	1982	6945
macromolecules	1758	1749	6623
Ligands	28	46	26
Water	266	187	296
Protein residues	230	230	886
RMS(bonds) (Å)	0.02	0.017	0.037
RMS(angles) (°)	2.08	1.95	2.12
Ramachandran preferred (%)	98.05	96.63	91
Ramachandran allowed (%)	0.98	2.4	-
Ramachandran outliers (%)	0.98	0.96	3.4
Clashscore	3.95	5.67	14.27
Average B-factor (Å ²)	24	16.8	38
macromolecules	22.1	15.4	38
Ligands	36.8	29.7	52.1
Solvent	35	26.6	36.2

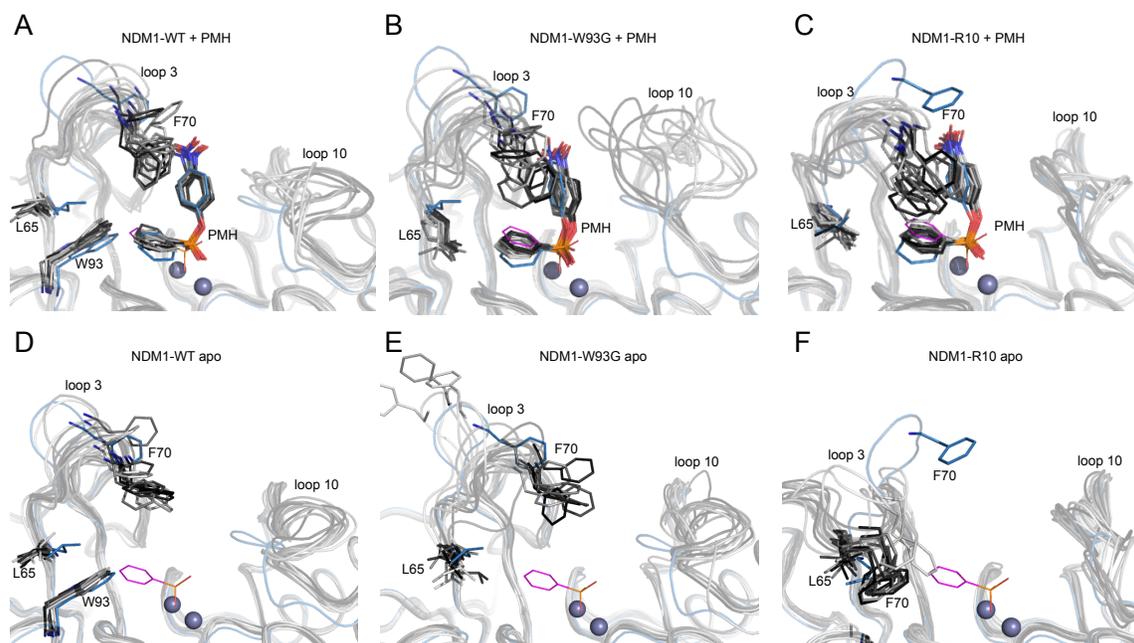


Figure 5.15 Molecular dynamics simulations of NDM1 variants.

(A – F) Overlay of MD simulation snapshots obtained from 100 ns trajectories, with the energy minimized structures shown in blue (0 ns) and snapshots in light gray (50 ns) to black (100 ns) every 5 ns. (A – C) MD simulations of NDM1 variants with the PMH substrate bound. Substrate binding is based on the position of the PMH product (magenta lines) in the complexed structure of NDM1-R10 (PDB ID 5K4M). (D – F) MD simulations of NDM1 variants in the apo form without the substrate bound, with the PMH product (magenta lines) shown for reference. The NDM1-W93G mutant was generated *in silico* based on the wild-type structure as described in the material and methods. A. Pabis performed the MD simulation experiment.

5.4.11 Understanding the structural basis of PMH substrate binding of NDM1 variants

To further understand the structural basis of mutations for PMH activity in NDM1 variants we performed molecular dynamics (MD) simulations in collaboration with the Kamerlin research group. We selected three variants to be computationally analyzed, NDM1-WT (PDB ID 4HL2), NDM1-W93G (mutant was generated *in silico*, based on PDB ID 4HL2) and NDM-R10 (PDB ID 5K4M). For all three variants, 100 ns MD simulations with and without the PMH substrate were performed (Figure 5.14). The initial positioning of the substrate was based on the product state structure of NM1-R10 (Figure 5.14). During the simulations the substrate remained stable in all three variants at

the same position as the product in NDM1-R10 structure (**Figure 5.15**). However, the results of the MD simulations of NDM1-WT and NDM-W93G reveal that the hindrance between the active site and the substrate is indeed resolved by W93G (**Figure 5.15 A and B**). Also, W93G induces a higher flexibility for L3, which seems to allow F70 to interact with the substrate. This interaction appears to be more pronounced in NMD-R10. In detail, L3 moved further inward and F70 appears to form π - π stacking interactions with the substrate the *p*-nitrophenol-leaving group and the phenyl substituent of the substrate. In addition, loop 10 became more rigid and N220 seems to interact with the *p*-nitrophenol-leaving group of the substrate. In the MD simulations without the substrate docked (**Figure 5.15 D – F**), we observe a similar behavior of the active site loops; L3 with F70 moves further inward and L1 becomes more rigid in NDM1-R10. Conclusively, the structural basis of NDM1's improvement is a combination of improved substrate complementarity as well optimization of active site loop dynamics. Similar structural adaptations have been observed in previous enzyme evolution and engineering studies (G. Yang et al. 2016; Tokuriki et al. 2012; Gobeil et al. 2014; Tomatis et al. 2008).

5.4.12 Structural adaptation of VIM2-R10

To investigate the structural basis of evolutionary adaption of VIM2, we compared the structural changes of VIM2-R10 to VIM2-WT. As described earlier, VIM2-R10 exists as monomer and dimer in solution but we were only able to crystallize the dimer fraction (**Figure 5.8**). We would like to note that at this point, however, we are not certain if both or only the monomeric or the dimeric form is catalytically active. Interestingly, the crystal structure analysis of VIM2-R10 revealed a 3D domain swapped dimer (Y. Liu & Eisenberg 2002), in which half of the structure is exchanged symmetrically between two entangled subunits (**Figure 5.16 A and B**). In detail, the two half β/α barrels of chain A were disassembled and each domain swapped its half β/α barrel with another half β/α barrel of chain B. We hypothesize that the 3D domain swapping was initiated by mutation D223A (round 2). In VIM2-WT D223 forms a hydrogen bond with H122, which is a metal coordinating residue at the α -site and is located at the margin of the linker between the two subunits. The mutation D223A eliminates this interaction and thus increases the flexibility of this region (**Figure 16 D**). We suspect that N154T also

contributes, due to the loss of a salt-bridge with backbone oxygen of F121, which is also located near the linker region. However, we believe that these are not the only mutations that caused and support the 3D domain swapping, and thus we require further biochemical and structural analyses to reveal detailed insights in the molecular basis of the structural rearrangement.

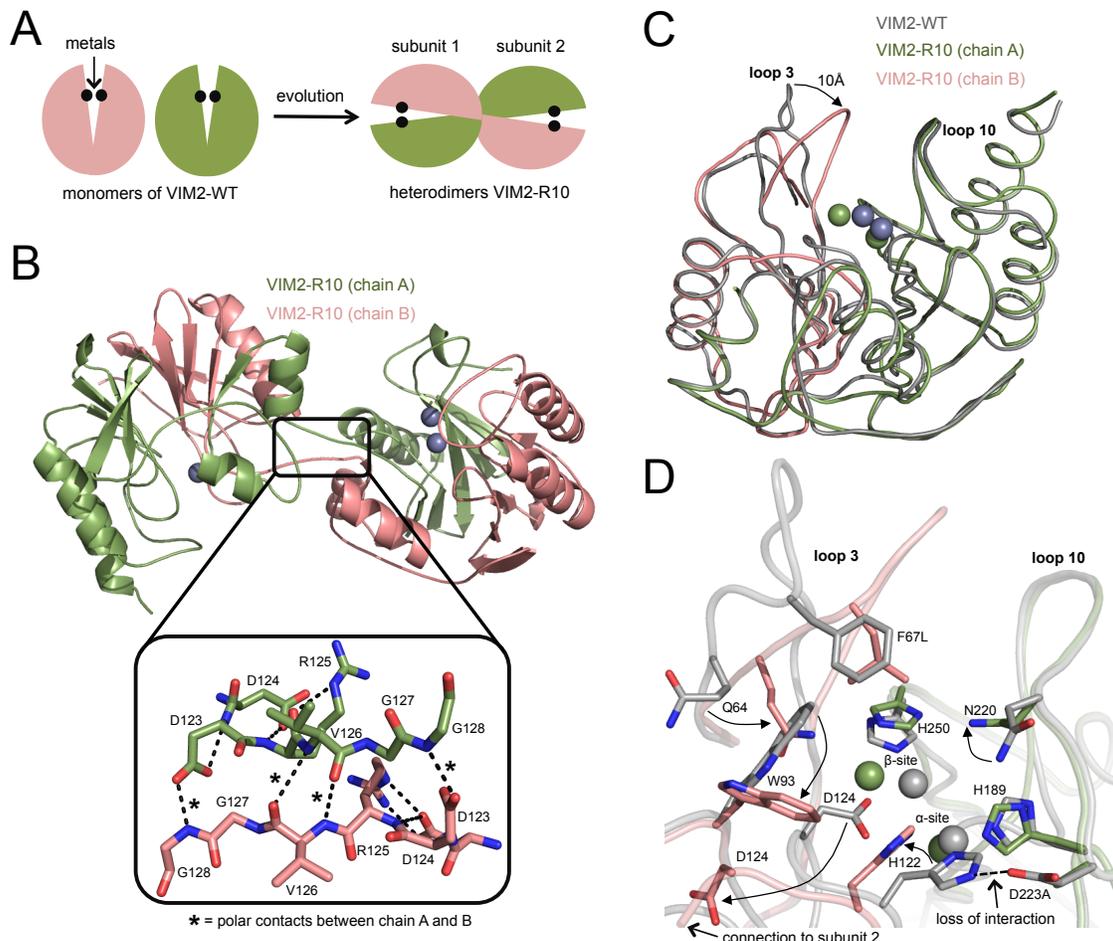


Figure 5.16 The structural basis for improved PMH activity of VIM2-R10.

(A) Schematic presentation of 3D domain swapping as it occurred during the evolution from VIM2-WT to VIM2-R10. VIM2-WT exists in a monomeric form (left), whereas VIM2-R10 swapped domains between chain A and B that form a heterodimer with two structurally identical subunits. Active site metals are presented as black spheres. (B) Structural representation of the VIM2-R10 domain-swapped heterodimer (chain A in green and chain B in salmon). The close up view shows the polar contacts in the linker region between both subunits (residues 123 to 128). Polar contacts were identified and visualized using PyMOL. (C) Comparison of the C- α backbone of VIM2-WT (gray, PDB ID 1ko3) and VIM2-R10 (green chain A and salmon chain B). The active site metal ions are shown as spheres (gray for VIM2-WT and green for VIM-R10). (D) Active site superposition of VIM2-WT

(gray, PDB ID 1ko3) and VIM2-R10 (green, chain A and salmon, chain B). Arrows indicate repositioning of active site residues.

In addition to the 3D domain swap of VIM2-R10, we also observe significant structural changes around the active site and the catalytic metal center (**Figure 16 C and D**). First, the metal coordination changed dramatically, which is mainly caused by an opening of the chains and 3D domain swapping. This altered drastically the conformation and position of the metal coordinating residues D124 and H122 (**Figure 5.16 D**). The rotation of D124 away from metal center caused a reduced occupancy of the Zn^{2+} at the β -site. In our structure analysis, only subunit 1 clearly showed binding of Zn^{2+} in the β -site, despite addition of Zn^{2+} during the purification process. In addition, the conformation of H122, which coordinates the Zn^{2+} in the α -site, caused a 1.5 Å movement of the Zn^{2+} into a more buried area and changed the Zn-Zn distance from 4.2 Å in to 5.2 Å. Second, we observe a structural change of the active site residues and loops (**Figure 5.16 D**). Specifically, the active site residue W93 adopted an alternative conformation by rotating $\sim 90^\circ$ from its original position compared to VIM2-WT, which is potentially caused by drastic repositioning of D124, as W93 in VIM2-R10 partially occupies its position. Furthermore, similar to NDM1-R10, the active site loop L3 moved inwards by around 10Å and caps the active site. We suspect that the L3 movement is due to the several mutations in L3 itself (V41A, V46D, T64A, S66P, F67L, V72A) and W93 rotation, which generate space underneath L3 (**Figure 5.16 D**). However, because we were unable to obtain any substrate or product density in the active site of VIM2-R10, it is difficult to rationalize these structural changes to PMH activity improvements as we could for NMD1. Furthermore, the activity improvements of VIM2-R10 are far lower than that of NDM1-R10 (60 vs. 20,000-fold in k_{cat}/K_M respectively), and thus are less pronounced and more difficult to rationalize structurally. Thus, although our analysis provided a first glimpse, more detailed structural, biochemical and mechanistic analyses are required to gain detailed insights into functional changes of VIM2-R10 regarding PMH activity and its 3D domain swapping.

5.4.13 The molecular basis of mutational incompatibility

Finally, we were interested in understanding the molecular basis underlying the

mutational incompatibility of W93G on the VIM2 background. In collaboration with the Kamerlin research group at the Uppsala University we conducted MD simulations of VIM2-WT (PDB ID 4PVO) and VIM2-W93G (the mutation was introduced *in silico* based on the wild-type structure) with and without the PMH substrate. In the MD simulation of VIM2-WT with the PMH substrate bound the enzyme seemed not to have a steric hindrance in positioning the substrate in the active site, and indeed is similar to NDM1-W93G (**Figure 5.17**). The reason for this is that W93 in VIM2-WT adopts a different orientation compared to NDM1-WT, which could explain the initially 10-fold higher catalytic efficiency of VIM2-WT over NDM1-WT (**Figure 5.17**). The MD simulation of VIM2-W93G did not show a significant change in substrate binding compared to the simulation of VIM2-WT, besides that L3 became more mobile, and thus does not explain why VIM2-W93G exhibits 10-fold lower PMH (and β -lactamase) activity compared to VIM-WT. In addition, we performed MD simulations without the substrate and examined how the protein structure of VIM2-WT could be affected by W93G. MD simulations in the apo form showed an increase in flexibility and drastic repositioning of L3 and L10 in the W93G mutant compared to VIM2-WT. In detail, a collapse of L3 and L10 caused residues of both loops to occupy parts of the active site, with Q64 occupying the position of W93 and F67 moving to the position of the PMH product, thus most likely blocking substrate accessibility to the active site. Our results are consistent with the fact that VIM2-W93G exhibits a 10-fold lower PHM and β -lactamase activity. In contrast, apo MD simulations of NDM1-WT, NDM1-W93G and NDM1-R10 did not show a reduced substrate accessibility of active site (**Figure 5.15**). Taken together, our results suggest that W93G has different structural and functional consequences in VIM2 and NDM1, which ultimately led to a different evolvability of each enzyme towards PMH activity.

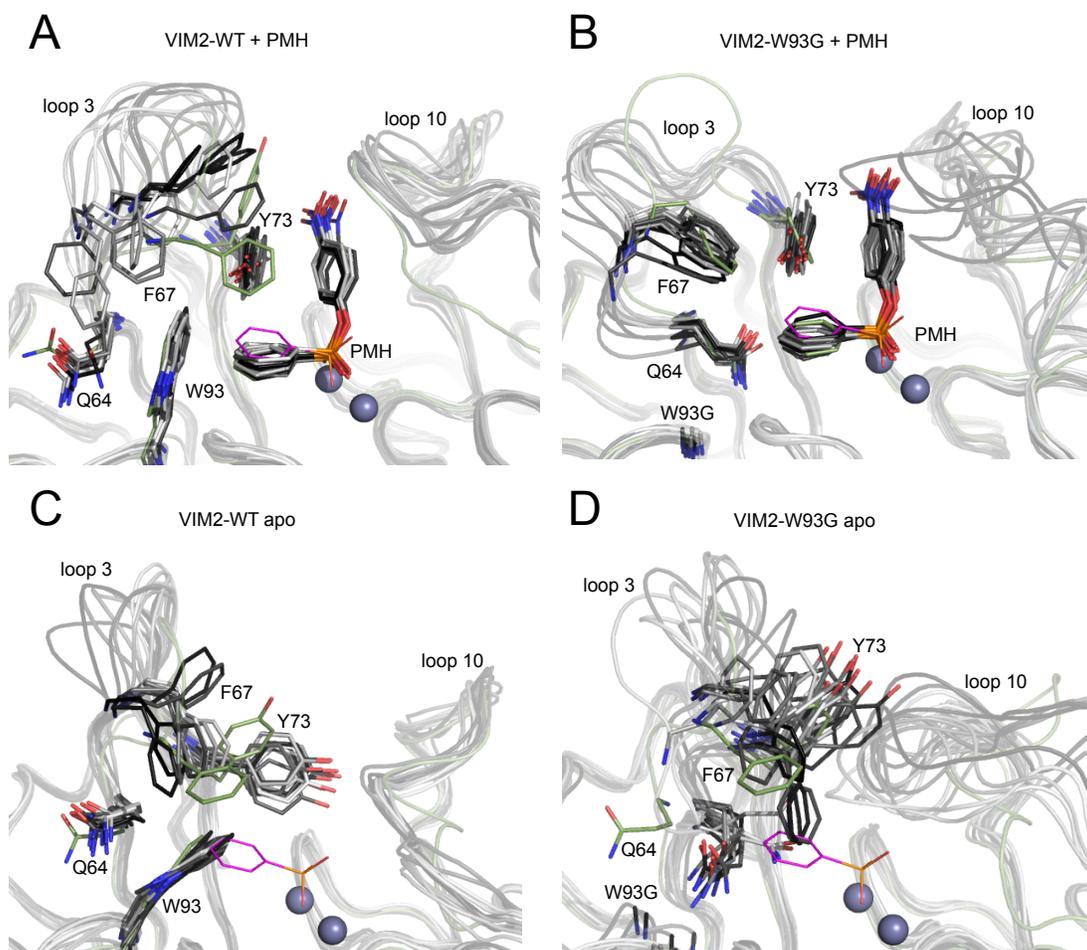


Figure 5.17 Molecular dynamics simulations of VIM2 variants.

(A – D) Overlay of MD simulation snapshots obtained from 100 ns trajectories, with the energy minimized structures shown in green (0 ns) and snapshots in light gray (50 ns) to black (100 ns) every 5 ns. (A and B) MD simulations of VIM2 variants with the PMH substrate bound. Substrate binding is based on the position of the PMH product (magenta lines) in the complexed structure of NDM1-R10 (PDB ID 5K4M). (C – D) MD simulations of VIM2 variants in the apo form without the substrate bound, with the PMH product (magenta lines) shown for reference. The VIM2-W93G mutant was generated *in silico* based on the wild-type structure as described in material and methods.

5.5 Discussion

Our comparative experimental evolution demonstrated that evolvability of two related enzymes, NDM1 and VIM2, under identical selection conditions, yielded strikingly different outcomes in terms of activity improvement and mutational and structural solutions. Overall, NDM1, which was the initially lower starting point, evolved a 30-fold higher catalytic efficiency with a k_{cat}/K_M of $8700 \text{ M}^{-1}\text{s}^{-1}$ (vs. $270 \text{ M}^{-1}\text{s}^{-1}$ for VIM2-R10). In

addition, VIM2 improved fitness through regained solubility, likely achieved by structural dimerization, whereas the structural adaptation of NDM1 relied on optimizing the substrate binding, at the expense of solubility. Despite our limited directed evolution set-up (< 400 variants screened for PMH activity per round and enzyme), for which we employed a pre-selection using ampicillin antibiotic resistance (native function), several outcomes indicate that our screening system was sufficient in identifying the most beneficial mutations for each enzyme. First, the direct prescreening with PMH activity, which expanded our screening capacity up to ~2,000 variants per round, did not yield any further improvement in the last two rounds. Second, the replicate experiment of library screening from the wild-type enzymes repeatedly isolated variants with the same mutations and fitness improvements. Third, introducing the highly beneficial W93G mutation of NDM1 into VIM2 resulted in a loss of PMH activity. Thus, we are confident that evolution from both NDM1 and VIM2 is likely to be highly deterministic and repeatable. Such determinism and repeatability in evolutionary trajectories has been observed in other proteins evolution studies (Dickinson et al. 2013), *e.g.* of antibiotic resistance (Weinreich 2006), pesticide degradation (Noor et al. 2012), altitude adaptation (Tufts et al. 2015), amino acid synthesis (Lunzer et al. 2005), fluorescence color (Field & Matz 2010) and glucocorticoid receptor evolution (Harms & Thornton 2014). Thus, our results further support the notion that protein evolution is largely deterministic from a given sequence starting point.

However, beyond single enzyme determinism, our results suggest that different sequences, despite structural and functional similarity, will adopt unique phenotypic and mutational solutions that are incompatible with other sequences. In particular, the highly beneficial mutation W93G on NDM1 is deleterious on the background of VIM2, which consequently had to adopt a different, but far less efficient, molecular mechanism of evolution. Hence, initial sequence differences can cause mutational incompatibility among orthologous enzyme and lead to contingency in protein evolution, where the presence of particular, evolvable, genotypes is stochastic because of neutral drift. For example, if only VIM2 (or VIM1 or VIM7) is available as an evolutionary starting point for PMH activity, adaptation would be slow and yield less fit progeny compared to a population that has NDM1 as an evolutionary starting point. Therefore, selecting from a

population with high neutral genetic variation could potentially accelerate evolutionary adaptation, as the change of finding evolvable genotypes is more likely. Indeed, such an observation on the population level has been revealed in a study by Hayden *et al.*, which shows that a ribozyme population with neutral genetic diversity evolves more rapidly towards a new function compared to a single genotype (Hayden *et al.* 2011). The reason for this is that some genotypes of the neutral network were pre-adapted for the new activity that was only selected after neutral genetic diversification, and were therefore more evolvable towards it. In conclusion, the evolution of a new enzyme function depends on the presence of evolvable (protein) sequences, but their occurrence as well as their spatial and temporal presence can be highly stochastic, which leads to unpredictability in molecular evolution (Harms & Thornton 2014; Meyer *et al.* 2012; Harms & Thornton 2013; Miton & Tokuriki 2016).

What are the structural and biophysical factors that restricted the evolvability of VIM2? The main biophysical factors that have been described to influence the evolvability of proteins is thermodynamic stability (Bloom *et al.* 2006) and kinetic stability (solubility) (Tokuriki & Tawfik 2009c; Wyganowski *et al.* 2013). In this model, mutations that alter function often trade off with stability, and thus high stability could allow the tolerance of crucial function-changing mutations. However, VIM2 has a slightly higher initial thermostability and solubility compared to NDM1, and both actually increased for VIM2 during the trajectory. Thus, thermodynamic and kinetic stability (solubility) constraints are unlikely to restrict VIM2 in its evolvability. In fact, introducing the initial mutation W93G of NDM1 into VIM2 does not impair its thermostability and solubility, but simply decreases its PMH activity. On the other hand, testing the effect of W93G in other B1 β -lactamases reveals a positive effect on PMH activity in all NDM1 related sequences (and VIM7), which however trades off with thermostability and solubility. Structural and MD simulation results point to a structural cause of incompatibility and restricted evolvability of VIM2. In NDM1, W93G optimizes the substrate complementarity of the active site, however in VIM2 W93G appears not to be able to further optimize binding. In contrast, W93G in VIM2 result in a very high flexibility of active site loops and consequently a reduced substrate accessibility of the

active site. Thus, we believe that slight differences in the structure between NDM1 and VIM2, and flexibility of local structural elements, are the cause of the mutational incompatibility and difference in evolvability.

The 3D domain swapping, which occurred in the VIM2 trajectory, is a drastic structural rearrangement, but a few cases have been described in the literature (Y. Liu & Eisenberg 2002; Qu et al. 2000; Bennett et al. 1995; Cameron 1997; Bennett & Eisenberg 2004; Y. Liu et al. 1998; Chirgadze et al. 2004). In most cases, however, a mechanistic and functional reason for 3D domain swapping is still under debate, but effects on protein stability have been discussed (Y. Liu & Eisenberg 2002; Bennett & Eisenberg 2004). We suspect that in the case of VIM2-R10, 3D domain swapping could have improved the enzymes' solubility, which initially decreased from 60% to 40%, but eventually increased by up to 80% over the trajectory (**Figure 5.5 C**). One possible scenario is that initial functional mutations destabilize VIM2 (e.g. D223A) and dimerization through 3D domain swapping is one strategy to regain stability and solubility. Indeed, several examples in the literature support the idea that oligomerization can improve protein stability (Bershtein et al. 2012; Fraser et al. 2016; Qu et al. 2000; Thoma et al. 2000). A recent directed evolution experiment aiming at increasing the thermostability of an α E7 carboxylesterase yielded increased levels of dimeric and tetrameric quaternary structures (Fraser et al. 2016). Similarly, in the case of the capsid proteins of the rice yellow mottle virus, 3D domain swapped variants exhibit higher stability compared to non-swapped capsid proteins of evolutionary related viruses (Qu et al. 2000). Finally, mutational monomerization of a dimeric and thermostable phosphoribosylanthranilate isomerase from *Thermotoga maritima* resulted in a drastic decrease in thermostability, but did not impair its function (Thoma et al. 2000).

Our observations have significant implication for protein design, engineering and evolution in the laboratory. Protein engineers tend to choose a single starting sequence based on the availability of biochemical and structural information (Davids et al. 2013). However, it might be important to explore various starting sequences and examine which sequence is more evolvable (O'Loughlin 2006). Importantly, in our work, an initially less fit variant evolved at a higher rate and ultimately reached a higher fitness plateau: thus

the initial activity may not be a good indicator for evolvability. The molecular causes for contingency can be specific for each protein sequence, and it may be impossible to predict the evolvability of each unique sequence at this point. Together, our results demonstrate that we still need to further investigate and understand the molecular details of evolution in order to decipher evolutionary constraints as well as to develop better algorithms and tools to design and engineer novel enzymes in the laboratory.

Chapter 6: Conclusion and future outlook

Parts of chapter six have been written together with Janine N. Copp in the laboratory of Dr. N. Tokuriki at UBC, Vancouver, Canada and published in “Baier F., Copp J.N., Tokuriki N. (2016): Enzyme superfamilies – new approaches toward systematic mapping of evolutionary sequence-function relationships. *Biochemistry*, 2016, 55 (46), 6375–6388.”

6.1 General summary and conclusion

In summary, this thesis contributed to our understanding of the functional divergence and evolution of enzymes and enzyme superfamilies. In particular, in chapter two, we reveal the evolutionary relationship of functional families of the MBL superfamily as well as show that many MBL enzymes are promiscuous and catalyze the several other MBL functions in addition to their native one. In chapter three, we demonstrate that promiscuous activities of MBL enzymes can stem from a variety of different metal isoforms, which further broadens their function profile, and ultimately could facilitate the functional divergence of metalloenzymes. In chapter four, we visualize how promiscuity leads to significant function connectivity in different enzyme superfamilies and discuss the results in the context of enzyme evolution. In chapter five, we show that cryptic genetic variation can have profound consequences on subsequent evolutionary outcomes through a parallel comparative directed evolution experiment of two orthologs towards a shared promiscuous activity. Effectively, the first chapters highlight that many enzymes promiscuous and functions are highly connected, and thus new functions could readily evolve. In contrast, chapter five emphasizes that not all promiscuous enzymes are (equally) evolvable, due to the distinct structural consequences of mutations in different sequence backgrounds, which means that many of the functional connections are potentially not traversable.

6.2 Future outlook

Despite the advances in the recent years, our understanding of enzyme functions and the evolutionary processes that led to the functional expansion of enzyme superfamilies is far from complete. Many open questions remain and further efforts are required to develop a better understanding of the remarkable functional diversity of enzymes and how it evolved, which we will discuss in the next sections.

6.2.1 Integrating metal ion availability and functional divergence

In chapter three, I showed that different MBL enzymes prefer particular different metal ions for their catalytic function, but the binding of alternative metal ions introduces new promiscuous activities, that are not observed with other metal ions. Thus, only some enzyme/metal ion combinations might offer suitable evolutionary starting points. Beyond this observation, does the availability of metal ions also affect the outcome of evolutionary trajectories? In other words, does the functional effect of a mutation depend on the presence of particular metal ions as an environmental factor (Flynn et al. 2013; Taute et al. 2014)? Furthermore, does the metal ion preference change during a functional transition (Reynolds et al. 2016)? To address these questions we are currently exploring the recent evolution of a pesticide degrading methylparathion hydrolase (MPH) from the MBL superfamily, which most likely evolved from a dihydrocoumarin lactonase (DHCL), and obtained interesting results. Together with the Bornberg-Bauer lab in Germany, we reconstructed the ancestor of MPH, using computational ancestor reconstruction and gene synthesis (Harms & Thornton 2010), which indeed exhibits high levels of DHCL and low levels of MPH activity. Structure and mutational analysis revealed that five active site mutations are responsible for the functional switch, but a particular order of the mutations is required to smoothly transition from DHCL to MPH activity, similar to the previously described mutational analysis of the TEM-1 evolution (Weinreich 2006). Interestingly, when the mutational trajectory is reassessed with different metal ion availability, supplying Zn^{2+} , Mn^{2+} , Co^{2+} , Cd^{2+} , Ca^{2+} , Mg^{2+} , Ni^{2+} or Cu^{2+} in the media and buffer, the accessible mutational paths and overall improvements are strikingly different. This suggests that similar to mutational epistasis, the environment – and in this case, metal ion availability – can severely affect molecular evolution and lead to distinct evolutionary outcomes (Taute et al. 2014). Furthermore, it also suggest that different metal ion requirements, among different evolutionary related enzymes, could have evolved due to such constraints, in addition to catalytic preference (Purg et al. 2016) or extremely metal-depleted environments (Cotruvo & Stubbe 2012; Dudev & C. Lim 2014; Xu et al. 2008; Carter et al. 2011). For this project we are currently analyzing the data and preparing a manuscript together with Dave W. Anderson and Gloria Yang in the Tokuriki lab.

6.2.2 Detailed characterization of the 3D domain swap of VIM2

In chapter five, I described the comparative laboratory evolution of the B1 β -lactamases VIM2 and NDM1 towards improved PMH activity. Structural characterization of the most evolved variants revealed that VIM2 changed from a monomeric protein into a 3D swapped dimer during the evolution, with a complete half of the structure being exchanged symmetrically between two entangled subunits. This potentially helped VIM2 to increase its solubility and consequently activity in the cell lysate, which was our fitness proxy in the directed evolution experiment. A recent study showed that 3D domain swapping is not uncommon and 32 well-defined cases have been identified in the PDB database (Szilágyi et al. 2012). Yet, our case here of VIM2 provides an unprecedented and unique opportunity to study the evolution and molecular basis of 3D domain, because we have all evolutionary intermediates available from the directed evolution experiment for further structural and biochemical analysis. First, however, we need to confirm that the 3D swapped domain is catalytically active and we have to rule out the possibility of a crystallographic artifact, because size exclusion chromatography (SEC) showed that the evolved VIM2-R10 variant still exists as a monomer and a dimer in solution. Therefore, purification, SEC separation of the two states and subsequent measurement of the catalytic activities will confirm if both the dimeric and monomeric fractions of VIM2-R10 are active or only one of them. It could potentially be that VIM-R10 is only active in the monomer form, but adopts the dimer form for stability, and both states coexists in an equilibrium. Subsequently, we aim to reveal the molecular basis and identify the responsible mutations of the 3D domain swap. First, we will perform SEC analysis of all intermediate variants to understand at which point in the evolution the dimerization occurred. This will reveal if the transition happened in a single step or if multiple mutations were responsible. We can then introduce individual mutations and combine them on the background of VIM2-WT and perform SEC to test their effect on dimerization. Structure analysis suggests that two mutations, D223A and N154T, are good candidates, which once introduced into VIM2-WT should provide evidence that the protein adopts a dimeric form, using analytical SEC. Overall, analyzing the 3D domain swapping of VIM2 will provide unprecedented molecular insights into such drastic structural rearrangements and their evolution. In turn, this knowledge could aid efforts to engineer more resilient proteins for biotechnological purposes.

6.2.3 Understanding how completely different catalytic functions evolve

In the four comprehensive enzyme characterization studies, discussed in chapter four, the scope of the functions and substrates that were investigated was relatively narrow compared to the large diversity of functions that can exist in a single superfamily. Furnham *et al.* recently showed that, whereas functional expansion typically occurs *within* each of the six major E.C. classes (*i.e.*, oxidoreductase, transferase, hydrolase, lyase, isomerase and ligase), evolutionary transitions between these classes are also observed, albeit far less common (Furnham *et al.* 2016). For example, MBL enzymes not only perform hydrolytic reactions, but also oxido-reduction, nitric-oxide reduction (Silaghi-Dumitrescu *et al.* 2005) and sulphur-dioxygenation (Holdorf *et al.* 2012), which have completely different catalytic mechanisms and requirements. Hydrolysis requires a hydroxyl ion, which is activated by a divalent metal ion center, whereas the oxidoreductase reaction uses two $\text{Fe}^{2+/3+}$ and FMNH_2 to reduce oxygen to water (Bartlett *et al.* 2003). In the case of the MBL nitric-oxide reductase, a fused structural domain realizes FMN cofactor binding (Silaghi-Dumitrescu *et al.* 2005). How did the structural fusion of the MBL domain and FMN binding domain lead to the evolution of a new enzyme function? Was there already a promiscuous activity nitric-oxide reductase activity without the additional domain and the FMN binding domain fusion enhanced it or was the fusion an evolutionary “hopeful monster” (Toth-Petroczy & Tawfik 2014), which introduced the activity at right time and space and provided a beneficial selective advantage? So far, in most cases, the divergent evolution of such catalytically different functions, *e.g.* oxidoreductases, lyases and hydrolases, remains mainly unexplored (Furnham *et al.* 2012; Furnham *et al.* 2016). Yet, a few anecdotal cases have been described: for example, a reductive dehalogenase of the cytGST superfamily involved in the biodegradation of xenobiotic polychlorinated compounds in *Sphingomonas chlorophenolica*, exhibits high levels of its presumed ancestral isomerase activity (Anandarajah *et al.* 2000). Also, decarboxylases from *E. coli* and *Pseudomonas pavonaceae* have been shown to promiscuously function as synthases (Yew *et al.* 2005) or hydratases, respectively (Poelarends *et al.* 2004). Interestingly, Yew *et al.* also showed that the synthase activity of the *E. coli* decarboxylase could be improved by 260-fold with only four mutations in the active site (Yew *et al.* 2005). Despite these examples, it appears that in some instances large structural and mechanistic reorganizations may be required for the emergence of new catalytic functions, including the insertion of domains with different catalytic properties, that are often observed between functionally distinct enzymes

(Toth-Petroczy & Tawfik 2014). Indeed, laboratory evolution studies of enzymes demonstrated that loop and domain insertions/deletions can cause the emergence of new catalytic functions, which can be subsequently optimized by further adaptive mutations (Park 2006; Afriat-Jurnou et al. 2012). Concomitantly, recent work by Aravind and co-workers highlighted examples of drastic structural divergence of proteins, such as domain incorporation, topological rewiring, loop modifications, and mutations of catalytic residues, that can occur even when their molecular functions remain conserved (Zhang et al. 2014). Such cases of dramatic structural rearrangements that are neutral for the native function lead us to speculate that proteins may, in some cases, be able to explore new regions of sequence and structural space in which they can catalyze new promiscuous activities without necessarily changing the ancestral function or impairing organismal fitness (Zhang et al. 2014).

6.2.4 Exploring and annotating protein functions is far from completion

We generated SSNs of the MBL superfamily to infer the evolutionary relationships between functional families, which we partly annotated using the available database and literature information. What became evident is that only a fraction of available sequences are experimentally characterized and many sequence clusters are completely uncharacterized, but putatively harbor completely new catalytic functions or substrate specificities (Punta et al. 2012; Cvetkovic et al. 2010). To gain further insights into the functional diversity and evolution of enzyme superfamilies, we need to extend our experimental and computational efforts to experimentally explore and uncharacterized sequence clusters. In particular, the integration of high-throughput enzyme assays and metagenomic (environmental DNA) surveys will greatly help the exploration of uncharacterized sequence clusters, potentially leading to the identification of new enzyme functions (Cantarel et al. 2009; Gerlt, Allen, et al. 2011; Akiva et al. 2014; Riaz et al. 2008; Colin et al. 2015). For example, Colin *et al.* recently used a microfluidic picolitre droplet compartments approach to screen over a million sequences in a metagenomic library and identified several enzymes that exhibit promiscuous phosphotriesterase activity (Colin et al. 2015). Similar approaches will hopefully help to further explore and assign enzyme function in a high-throughput manner. Also, careful selection of representative sequences, *e.g.* using SSNs, for experimental characterization might be helpful, as for example performed in the cytGST superfamily (Mashiyama et al. 2014). The development of tools to accurately predict enzyme

function(s) is also essential to support the experimental characterization, because experimental efforts are unlikely to cover the entire sequence space found in nature (Radivojac et al. 2013). Furthermore, there are strong biases of sequence, biochemical and structural information that need to be recognized and accounted for (Schnoes et al. 2013). For example, the focus of many experimental approaches has been generally narrow and traditionally confined to relatively few protein folds and superfamilies, or even isolated to selected functional groups, partly due to biomedical and industrial interests and to the experimental assays that are available (Schnoes et al. 2013). This is illustrated by the family of B1 β -lactamases of the MBL superfamily, for which over 160 protein structures are available, which is roughly one third of all structures solved from the MBL superfamily (Berman et al. 2000). As a result, the extent of the functional repertoire of almost all functional families and superfamilies has yet to be fully explored (Brown & Babbitt 2014). In addition, the misannotation of gene functions is common in public databases that use automated computational annotation predictions based upon homology with existing entries; a study published by Schnoes *et al.* observed that one third of the 37 investigated superfamilies contained misannotation levels greater than 80% (Schnoes et al. 2009). Sequencing bias may also hinder functional annotation and delineation of sequence clusters, as most sequences deposited to genome databases are from relatively few model organisms and isolated microbial strains (Schnoes et al. 2013; Hug et al. 2016). Thus, experimental investigations guided by exhaustive sequence analysis together with high-throughput experimental platforms capable of systematically characterizing a large number of sequences are essential for the effective and comprehensive characterization of enzyme superfamilies (Colin et al. 2015; Davids et al. 2013; Bachovchin et al. 2014).

Bibliography

- Aqvist, J. & Warshel, A., 1989. Calculations of free energy profiles for the staphylococcal nuclease catalyzed reaction. *Biochemistry*, 28(11), pp.4680–4689.
- Abraham, M.J. et al., 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1(2), pp.19–25.
- Acker, M.G. & Auld, D.S., 2014. Considerations for the design and reporting of enzyme assays in high-throughput screening applications. *Perspectives in Science*, 1(1-6), pp.56–73.
- Afriat, L. et al., 2006. The Latent Promiscuity of Newly Identified Microbial Lactonases Is Linked to a Recently Diverged Phosphotriesterase. *Biochemistry*, 45(46), pp.13677–13686.
- Afriat-Jurnou, L., Jackson, C.J. & Tawfik, D.S., 2012. Reconstructing a Missing Link in the Evolution of a Recently Diverged Phosphotriesterase by Active-Site Loop Remodeling. *Biochemistry*, 51(31), pp.6047–6055.
- Aharoni, A. et al., 2005. The “evolvability” of promiscuous protein functions. *Nature Genetics*, 37(1), pp.73–76.
- Ahmed, F.H. et al., 2015. Sequence-Structure-Function Classification of a Catalytically Diverse Oxidoreductase Superfamily in Mycobacteria. *J. Mol. Biol.*, 427(22), pp.3554–3571.
- Akiva, E. et al., 2014. The Structure-Function Linkage Database. *Nuc. Acids Res.*, 42, pp.D521–30.
- Aloy, P. et al., 2002. Structural similarity to link sequence space: new potential superfamilies and implications for structural genomics. *Protein science : a publication of the Protein Society*, 11(5), pp.1101–1116.
- Altschul, S.F. et al., 1990. Basic local alignment search tool. *J. Mol. Biol.*, 215(3), pp.403–410.
- Amitai, G., Gupta, R.D. & Tawfik, D.S., 2007. Latent evolutionary potentials under the neutral mutational drift of an enzyme. *HFSP Journal*, 1(1), pp.67–78.
- Anandarajah, K. et al., 2000. Recruitment of a double bond isomerase to serve as a reductive dehalogenase during biodegradation of pentachlorophenol. *Biochemistry*, 39(18), pp.5303–5311.
- Aravind, L., 1999. An evolutionary classification of the metallo-beta-lactamase fold proteins. *In silico biology*, 1(2), pp.69–91.
- Armougom, F. et al., 2006. Espresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nuc. Acids Res.*, 34(Web Server issue), pp.W604–8.
- Armstrong, R.N., 1997. Structure, catalytic mechanism, and evolution of the glutathione

- transferases. *Chemical research in toxicology*, 10(1), pp.2–18.
- Atkinson, H.J. et al., 2009. Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS ONE*, 4(2), p.e4345.
- Avison, M.B. et al., 2001. Plasmid location and molecular heterogeneity of the L1 and L2 beta-lactamase genes of *Stenotrophomonas maltophilia*. *Antimicrobial Agents and Chemotherapy*, 45(2), pp.413–419.
- Baas, B.-J. et al., 2013. Recent Advances in the Study of Enzyme Promiscuity in the Tautomerase Superfamily. 14(8), pp.917–926.
- Babtie, A., Tokuriki, N. & Hollfelder, F., 2010. What makes an enzyme promiscuous? *Curr. Opin. Chem. Biol.*, 14(2), pp.200–207.
- Bachovchin, D.A. et al., 2014. A high-throughput, multiplexed assay for superfamily-wide profiling of enzyme activity. *Nature Chemical Biology*, 10(8), pp.656–663.
- Badarau, A. & Page, M.I., 2006. The variation of catalytic efficiency of *Bacillus cereus* metallo-beta-lactamase with different active site metal ions. *Biochemistry*, 45(35), pp.10654–10666.
- Baier, F. & Tokuriki, N., 2014. Connectivity between catalytic landscapes of the metallo- β -lactamase superfamily. *J. Mol. Biol.*, 426(13), pp.2442–2456.
- Baier, F. et al., 2015. Distinct Metal Isoforms Underlie Promiscuous Activity Profiles of Metalloenzymes. *ACS chemical biology*, 10(7), pp.1684–1693.
- Bar-Even, A. et al., 2011. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry*, 50(21), pp.4402–4410.
- Barber, A.E. & Babbitt, P.C., 2012. Pythoscape: a framework for generation of large protein similarity networks. *Bioinformatics*, 28(21), pp.2845–2846.
- Barbeyron, T. et al., 1995. Arylsulphatase from *Alteromonas carrageenovora*. *Microbiology*, 141 (Pt 11), pp.2897–2904.
- Bartlett, G.J., Borkakoti, N. & Thornton, J.M., 2003. Catalysing new reactions during evolution: economy of residues and mechanism. *Journal of Molecular Biology*, 331(4), pp.829–860.
- Bastard, K. et al., 2014. Revealing the hidden functional diversity of an enzyme family. *Nature Chemical Biology*, 10(1), pp.42–49.
- Baykov, A.A., Evtushenko, O.A. & Avaeva, S.M., 1988. A malachite green procedure for orthophosphate determination and its use in alkaline phosphatase-based enzyme immunoassay. *Analytical Biochemistry*, 171(2), pp.266–270.
- Bebrone, C., 2007. Metallo- β -lactamases (classification, activity, genetic organization, structure, zinc coordination) and their superfamily. *Biochem. Pharmacol.*, 74(12), pp.1686–1701.

- Bebrone, C. et al., 2001. CENTA as a chromogenic substrate for studying beta-lactamases. *Antimicrob. Agents Chemother.*, 45(6), pp.1868–1871.
- Bellinzoni, M. et al., 2011. 3-Keto-5-aminohexanoate cleavage enzyme: a common fold for an uncommon Claisen-type condensation. *The Journal of biological chemistry*, 286(31), pp.27399–27405.
- Ben-David, M. et al., 2013. Catalytic metal ion rearrangements underline promiscuity and evolvability of a metalloenzyme. *Journal of Molecular Biology*, 425(6), pp.1028–1038.
- Ben-David, M. et al., 2012. Catalytic versatility and backups in enzyme active sites: the case of serum paraoxonase 1. *Journal of Molecular Biology*, 418(3-4), pp.181–196.
- Benkovic, S.J. & Hammes-Schiffer, S., 2003. A perspective on enzyme catalysis. *Science*, 301(5637), pp.1196–1202.
- Bennett, M.J. & Eisenberg, D., 2004. The Evolving Role of 3D Domain Swapping in Proteins. *Structure (London, England : 1993)*, 12(8), pp.1339–1341.
- Bennett, M.J., Schlunegger, M.P. & Eisenberg, D., 1995. 3D domain swapping: A mechanism for oligomer assembly. *Protein Science*, 4(12), pp.2455–2468.
- Berman, H.M. et al., 2000. The Protein Data Bank. *Nucleic Acids Research*, 28(1), pp.235–242.
- Bershtein, S. et al., 2012. Soluble oligomerization provides a beneficial fitness effect on destabilizing mutations. *Proceedings of the National Academy of Sciences of the United States of America*, 109(13), pp.4857–4862.
- Bloom, J.D. et al., 2007. Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution. *Biology Direct*, 2(1), p.17.
- Bloom, J.D. et al., 2006. Protein stability promotes evolvability. *Proceedings of the National Academy of Sciences*, 103(15), pp.5869–5874.
- Bloom, J.D., Gong, L.I. & Baltimore, D., 2010. Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science*, 328(5983), pp.1272–1275.
- Blow, D., 2000. So do we understand how enzymes work? *Structure (London, England : 1993)*, 8(4), pp.R77–81.
- Boucher, J.I. et al., 2014. An atomic-resolution view of neofunctionalization in the evolution of apicomplexan lactate dehydrogenases. *eLife*, 3.
- Bovallius, A. & Zacharias, B., 1971. Variations in the metal content of some commercial media and their effect on microbial growth. *Applied microbiology*, 22(3), pp.260–262.
- Breen, M.S. et al., 2012. Epistasis as the primary factor in molecular evolution. *Nature*, 490(7421), pp.535–538.

- Bridgham, J.T., Ortlund, E.A. & Thornton, J.W., 2009. An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature*, 461(7263), pp.515–519.
- Broder, D.H. & Miller, C.G., 2003. DapE can function as an aspartyl peptidase in the presence of Mn²⁺. *Journal of Bacteriology*, 185(16), pp.4748–4754.
- Brown, S.D. & Babbitt, P.C., 2012. Inference of functional properties from large-scale analysis of enzyme superfamilies. *The Journal of biological chemistry*, 287(1), pp.35–42.
- Brown, S.D. & Babbitt, P.C., 2014. New Insights about Enzyme Evolution from Large Scale Studies of Sequence and Structure Relationships. *The Journal of biological chemistry*, 289(44), pp.30221–30228.
- Brown, S.D. et al., 2006. A gold standard set of mechanistically diverse enzyme superfamilies. *Genome Biology*, 7(1), p.R8.
- Burch, C.L. & Chao, L., 2000. Evolvability of an RNA virus is determined by its mutational neighbourhood. *Nature*, 406(6796), pp.625–628.
- Burroughs, A.M. et al., 2006. Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes. *J. Mol. Biol.*, 361(5), pp.1003–1034.
- Cameron, A.D., 1997. Crystal structure of human glyoxalase I: evidence for gene duplication and 3D domain swapping. *The EMBO journal*, 16(12), pp.3386–3395.
- Campbell, E. et al., 2016. The role of protein dynamics in the evolution of new enzyme function. *Nature Chemical Biology*, 12(11), pp.944–950.
- Campos-Bermudez, V.A. et al., 2007. Biochemical and structural characterization of Salmonella typhimurium glyoxalase II: new insights into metal ion selectivity. *Biochemistry*, 46(39), pp.11069–11079.
- Cantarel, B.L. et al., 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nuc. Acids Res.*, 37(Database issue), pp.D233–8.
- Carny, O. & Gazit, E., 2005. A model for the role of short self-assembled peptides in the very early stages of the origin of life. *The FASEB Journal*, 19(9), pp.1051–1055.
- Carpenter, A.E. & Sabatini, D.M., 2004. Systematic genome-wide screens of gene function. *Nat. Rev. Genet.*, 5(1), pp.11–22.
- Carter, E.L. et al., 2011. Iron-containing urease in a pathogenic bacterium. *Proceedings of the National Academy of Sciences of the United States of America*, 108(32), pp.13095–13099.
- Carvalho, A.T.P. et al., 2014. Challenges in computational studies of enzyme structure, function and dynamics. *Journal of molecular graphics & modelling*, 54, pp.62–79.

- Chirgadze, D.Y. et al., 2004. Snapshot of Protein Structure Evolution Reveals Conservation of Functional Dimerization through Intertwined Folding. *Structure (London, England : 1993)*, 12(8), pp.1489–1494.
- Cieplak, P. et al., 1995. Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *Journal of Computational Chemistry*, 16(11), pp.1357–1377.
- Clugston, S., Yajima, R. & Honek, J., 2004. Investigation of metal binding and activation of Escherichia coli glyoxalase I: kinetic, thermodynamic and mutagenesis studies. *Biochem. J.*, 377, pp.309–316.
- Colin, P.-Y. et al., 2015. Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. *Nature Communications*, 6, p.10008.
- Condon, C. & Gilet, L., 2011. The Metallo- β -Lactamase Family of Ribonucleases. In *Nucleic Acids and Molecular Biology*. Nucleic Acids and Molecular Biology. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 245–267.
- Copley, S.D., 2009. Evolution of efficient pathways for degradation of anthropogenic chemicals. *Nature Chemical Biology*, 5(8), pp.559–566.
- Cotruvo, J.A. & Stubbe, J., 2012. Metallation and mismetallation of iron and manganese proteins in vitro and in vivo: the class I ribonucleotide reductases as a case study. *Metallomics : integrated biometal science*, 4(10), pp.1020–1036.
- Cravatt, B.F., Wright, A.T. & Kozarich, J.W., 2008. Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Annual Review of Biochemistry*, 77, pp.383–414.
- Culotta, V.C., Yang, M. & O'Halloran, T.V., 2006. Activation of superoxide dismutases: putting the metal to the pedal. *Biochimica et biophysica acta*, 1763(7), pp.747–758.
- Currin, A. et al., 2015. Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chemical Society reviews*, 44(5), pp.1172–1239.
- Cvetkovic, A. et al., 2010. Microbial metalloproteomes are largely uncharacterized. *Nature*, 466(7307), pp.779–782.
- Dai, Y., Wensink, P.C. & Abeles, R.H., 1999. One protein, two enzymes. *The Journal of biological chemistry*, 274(3), pp.1193–1195.
- Daiyasu, H. et al., 2001. Expansion of the zinc metallo-hydrolase family of the beta-lactamase fold. *FEBS Letters*, 503(1), pp.1–6.
- Das, S., Dawson, N.L. & Orengo, C.A., 2015. Diversity in protein domain superfamilies. *Current opinion in genetics & development*, 35, pp.40–49.
- Davids, T. et al., 2013. Strategies for the discovery and engineering of enzymes for biocatalysis.

- Curr. Opin. Chem. Biol.*, 17(2), pp.215–220.
- Davies, J. & Davies, D., 2010. Origins and evolution of antibiotic resistance. *Microbiology and molecular biology reviews : MMBR*, 74(3), pp.417–433.
- Dellus-Gur, E. et al., 2015. Negative Epistasis and Evolvability in TEM-1 β -Lactamase--The Thin Line between an Enzyme's Conformational Freedom and Disorder. *Journal of Molecular Biology*, 427(14), pp.2396–2409.
- DePristo, M.A., Weinreich, D.M. & Hartl, D.L., 2005. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat. Rev. Genet.*, 6(9), pp.678–687.
- Dickinson, B.C. et al., 2013. Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 110(22), pp.9007–9012.
- Dimitrov, J.D. & Vassilev, T.L., 2009. Cofactor-mediated protein promiscuity. *Nature Biotechnology*, 27(10), p.892.
- Dong, Y.-J. et al., 2005. Crystal Structure of Methyl Parathion Hydrolase from *Pseudomonas* sp. WBC-3. *Journal of Molecular Biology*, 353(3), pp.655–663.
- Duarte, F. et al., 2014. Force field independent metal parameters using a nonbonded dummy model. *The journal of physical chemistry. B*, 118(16), pp.4351–4362.
- Dudev, T. & Lim, C., 2014. Competition among metal ions for protein binding sites: determinants of metal ion selectivity in proteins. *Chemical reviews*, 114(1), pp.538–556.
- Dunwell, J.M. et al., 2001. Evolution of functional diversity in the cupin superfamily. *Trends Biochem. Sci.*, 26(12), pp.740–746.
- Durrant, J.D. et al., 2014. POVME 2.0: An Enhanced Tool for Determining Pocket Shape and Volume Characteristics. *Journal of chemical theory and computation*, 10(11), pp.5047–5056.
- Durrant, J.D., de Oliveira, C.A.F. & McCammon, J.A., 2011. POVME: an algorithm for measuring binding-pocket volumes. *Journal of molecular graphics & modelling*, 29(5), pp.773–776.
- Dutta, T. & Deutscher, M.P., 2009. Catalytic Properties of RNase BN/RNase Z from *Escherichia coli* RNase BN IS BOTH AN EXO- AND ENDORIBONUCLEASE. *The Journal of biological chemistry*, 284(23), pp.15425–15431.
- Elias, M. & Tawfik, D.S., 2011. Divergence and Convergence in Enzyme Evolution: Parallel Evolution of Paraoxonases from Quorum-quenching Lactonases. *Journal of Biological Chemistry*, 287(1), pp.11–20. Available at: <http://www.jbc.org/cgi/doi/10.1074/jbc.R111.257329>.

- Erecińska, M. & Wilson, D.F., 1978. Homeostatic regulation of cellular energy metabolism. *Trends in Biochemical Sciences*, 3(4), pp.219–223.
- Estell, D.A. et al., 1986. Probing steric and hydrophobic effects on enzyme-substrate interactions by protein engineering. *Science*, 233(4764), pp.659–663.
- Fariás-Rico, J.A., Schmidt, S. & Höcker, B., 2014. Evolutionary relationship of two ancient protein superfolds. *Nature Chemical Biology*, 10(9), pp.710–715.
- Fernández-Gacio, A. et al., 2006. Transforming Carbonic Anhydrase into Epoxide Synthase by Metal Exchange. *Chembiochem : a European journal of chemical biology*, 7(7), pp.1013–1016.
- Field, S.F. & Matz, M.V., 2010. Retracing evolution of red fluorescence in GFP-like proteins from Faviina corals. *Molecular Biology and Evolution*, 27(2), pp.225–233.
- Flynn, K.M. et al., 2013. The environment affects epistatic interactions to alter the topology of an empirical fitness landscape. *PLoS genetics*, 9(4), p.e1003426.
- Foster, A.W., Osman, D. & Robinson, N.J., 2014. Metal Preferences and Metallation. *The Journal of biological chemistry*, 289(41), pp.28095–28103.
- Fraser, N.J. et al., 2016. Evolution of Protein Quaternary Structure in Response to Selective Pressure for Increased Thermostability. *Journal of Molecular Biology*, 428(11), pp.2359–2371.
- Frisch, M.J. et al., 2009. *Gaussian 09, Revision C. 01. Wallingford, CT, USA: Gaussian*,
- Furnham, N. et al., 2012. Exploring the evolution of novel enzyme functions within structurally defined protein superfamilies. *PLoS Computational Biology*, 8(3), p.e1002403.
- Furnham, N. et al., 2016. Large-Scale Analysis Exploring Evolution of Catalytic Machineries and Mechanisms in Enzyme Superfamilies. *Journal of Molecular Biology*, 428(2 Pt A), pp.253–267.
- Gabaldón, T. & Koonin, E.V., 2013. Functional and evolutionary implications of gene orthology. *Nat. Rev. Genet.*, 14(5), pp.360–366.
- Garau, G., Di Guilmi, A.M. & Hall, B.G., 2005. Structure-Based Phylogeny of the Metallo-β-Lactamases. *Antimicrobial Agents and Chemotherapy*, 49(7), pp.2778–2784.
- Garces, F. et al., 2010. Molecular Architecture of the Mn²⁺-dependent Lactonase UlaG Reveals an RNase-like Metallo-β-lactamase Fold and a Novel Quaternary Structure. *Journal of Molecular Biology*, 398(5), pp.715–729.
- Garcia-Saez, I. et al., 2008. The Three-Dimensional Structure of VIM-2, a Zn-β-Lactamase from *Pseudomonas aeruginosa* in Its Reduced and Oxidised Form. *Journal of Molecular Biology*, 375(3), pp.604–611.

- Gerlt, J.A. & Babbitt, P.C., 2001. Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annual Review of Biochemistry*, 70, pp.209–246.
- Gerlt, J.A. et al., 2015. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochimica et biophysica acta*, 1854(8), pp.1019–1037.
- Gerlt, J.A., Allen, K.N., et al., 2011. The Enzyme Function Initiative. *Biochemistry*, 50(46), pp.9950–9962.
- Gerlt, J.A., Babbitt, P.C., et al., 2011. Divergent Evolution in Enolase Superfamily: Strategies for Assigning Functions. *The Journal of biological chemistry*, 287(1), pp.29–34.
- Glasner, M.E., Gerlt, J.A. & Babbitt, P.C., 2006. Evolution of enzyme superfamilies. *Current Opinion in Chemical Biology*, 10(5), pp.492–497.
- Gobeil, S.M.C. et al., 2014. Maintenance of native-like protein dynamics may not be required for engineering functional proteins. *Chemistry & Biology*, 21(10), pp.1330–1340.
- Goddard, J.-P. & Reymond, J.-L., 2004. Recent advances in enzyme assays. *Trends in Biotechnology*, 22(7), pp.363–370.
- Goldman, A.D., Beatty, J.T. & Landweber, L.F., 2016. The TIM Barrel Architecture Facilitated the Early Evolution of Protein-Mediated Metabolism. *Journal of Molecular Evolution*, 82(1), pp.17–26.
- Gong, L.I., Suchard, M.A. & Bloom, J.D., 2013. Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife*, 2, p.e00631.
- Grant, C.L. & Pramer, D., 1962. Minor element composition of yeast extract. *Journal of Bacteriology*, 84(4), p.869.
- Gu, S.-Y., Yan, X.-X. & Liang, D.-C., 2008. Crystal structure of Tflp: A ferredoxin-like metallo- β -lactamase superfamily protein from *Thermoanaerobacter tengcongensis*. *Proteins: Structure, Function, and Bioinformatics*, 72(1), pp.531–536.
- Hagelueken, G. et al., 2006. The crystal structure of SdsA1, an alkylsulfatase from *Pseudomonas aeruginosa*, defines a third class of sulfatases. *Proceedings of the National Academy of Sciences*, 103(20), pp.7631–7636.
- Hammes-Schiffer, S. & Benkovic, S.J., 2006. Relating protein motion to catalysis. *Annual Review of Biochemistry*, 75, pp.519–541.
- Harms, M.J. & Thornton, J.W., 2010. Analyzing protein structure and function using ancestral gene reconstruction. *Curr. Opin. Struct. Biol.*, 20(3), pp.360–366.
- Harms, M.J. & Thornton, J.W., 2013. Evolutionary biochemistry: revealing the historical and

- physical causes of protein properties. *Nat. Rev. Genet.*, 14(8), pp.559–571.
- Harms, M.J. & Thornton, J.W., 2014. Historical contingency and its biophysical basis in glucocorticoid receptor evolution. *Nature*, 512(7513), pp.203–207.
- Hartl, D.L., 2014. What can we learn from fitness landscapes? *Current Microbiology*, 21, pp.51–57.
- Hayden, E.J., Ferrada, E. & Wagner, A., 2011. Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. *Nature*, 474(7349), pp.92–95.
- He, M.M. et al., 2000. Determination of the structure of Escherichia coli glyoxalase I suggests a structural basis for differential metal activation. *Biochemistry*, 39(30), pp.8719–8727.
- Henzler-Wildman, K.A. et al., 2007. Intrinsic motions along an enzymatic reaction trajectory. *Nature*, 450(7171), pp.838–844.
- Holdorf, M.M. et al., 2012. Arabidopsis ETHE1 encodes a sulfur dioxygenase that is essential for embryo and endosperm development. *Plant physiology*, 160(1), pp.226–236.
- Hu, Z. et al., 2009. Structure and Mechanism of Copper- and Nickel-Substituted Analogues of Metallo- β -lactamase L1 †. *Biochemistry*, 48(13), pp.2981–2989.
- Hu, Z., Gunasekera, T.S., et al., 2008. Metal content of metallo-beta-lactamase L1 is determined by the bioavailability of metal ions. *Biochemistry*, 47(30), pp.7947–7953.
- Hu, Z., Periyannan, G., et al., 2008. Role of the Zn1 and Zn2 sites in metallo-beta-lactamase L1. *Journal of the American Chemical Society*, 130(43), pp.14207–14216.
- Huang, H. et al., 2015. Panoramic view of a superfamily of phosphatases through substrate profiling. *Proceedings of the National Academy of Sciences of the United States of America*, 112(16), pp.E1974–83.
- Huang, R. et al., 2012. Enzyme functional evolution through improved catalysis of ancestrally nonpreferred substrates. *Proceedings of the National Academy of Sciences of the United States of America*, 109(8), pp.2966–2971.
- Huang, Y. et al., 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. 26(5), pp.680–682.
- Hug, L.A. et al., 2016. A new view of the tree of life. *Nature*.
- Hult, K. & Berglund, P., 2007. Enzyme promiscuity: mechanism and applications. *Trends Biotechnol.*, 25(5), pp.231–238.
- Humphrey, W., Dalke, A. & Schulten, K., 1996. VMD: visual molecular dynamics. *Journal of molecular graphics ...*, 14(1), pp.33–8–27–8.
- Imlay, J.A., 2014. The mismetallation of enzymes during oxidative stress. *The Journal of*

- biological chemistry*, 289(41), pp.28121–28128.
- Innan, H. & Kondrashov, F., 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nature Reviews Genetics*, 11(2), pp.97–108.
- Jackson, C.J. et al., 2009. Conformational sampling, catalysis, and evolution of the bacterial phosphotriesterase. *Proceedings of the National Academy of Sciences of the United States of America*, 106(51), pp.21631–21636.
- Jackson, C.J. et al., 2008. In Crystallo Capture of a Michaelis Complex and Product-binding Modes of a Bacterial Phosphotriesterase. *Journal of Molecular Biology*, 375(5), pp.1189–1196.
- Jacob, F., 1977. Evolution and tinkering. *Science*, 196(4295), pp.1161–1166.
- Jensen, R.A., 1976. Enzyme recruitment in evolution of new function. *Annual review of microbiology*, 30, pp.409–425.
- Jing, Q., Okrasa, K. & Kazlauskas, R.J., 2009. Stereoselective Hydrogenation of Olefins Using Rhodium-Substituted Carbonic Anhydrase-A New Reductase. *Chemistry - A European Journal*, 15(6), pp.1370–1376.
- Jonas, S. & Hollfelder, F., Mapping catalytic promiscuity in the alkaline phosphatase superfamily. *Pure and Applied Chemistry*, 81(4).
- Jorgensen, W.L. & Chandrasekhar, J., 1983. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2), pp.10.1063–1.445869.
- Jorgensen, W.L. & Maxwell, D.S., 1996. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45), pp.11225–11236.
- Kaltenbach, M. & Tokuriki, N., 2014. Dynamics and constraints of enzyme evolution. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 322(7), pp.468–487.
- Kaltenbach, M. et al., 2016. Functional Trade-Offs in Promiscuous Enzymes Cannot Be Explained by Intrinsic Mutational Robustness of the Native Activity. *PLoS Genet.*
- Kaltenbach, M. et al., 2015. Reverse evolution leads to genotypic incompatibility despite functional and active site convergence. *eLife*, 4.
- Karsisiotis, A.I., Damblon, C.F. & Roberts, G.C.K., 2014. A variety of roles for versatile zinc in metallo- β -lactamases. *Metallomics : integrated biometal science*, 6(7), p.1181.
- Kawabata, T., 2010. Detection of multiscale pockets on protein surfaces using mathematical morphology. *Proteins: Structure, Function, and Bioinformatics*, 78(5), pp.1195–1211.

- Khan, A.I. et al., 2011. Negative epistasis between beneficial mutations in an evolving bacterial population. *Science*, 332(6034), pp.1193–1196.
- Khanal, A. et al., 2015. Differential effects of a mutation on the normal and promiscuous activities of orthologs: implications for natural and directed evolution. *Mol. Biol. Evol.*, 32(1), pp.100–108.
- Khersonsky, O. & Tawfik, D.S., 2010. Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective. *Annual Review of Biochemistry*, 79(1), pp.471–505.
- Khersonsky, O., Roodveldt, C. & Tawfik, D.S., 2006. Enzyme promiscuity: evolutionary and mechanistic aspects. *Current Opinion in Chemical Biology*, 10(5), pp.498–508.
- King, D. & Strynadka, N., 2011. Crystal structure of New Delhi metallo- β -lactamase reveals molecular basis for antibiotic resistance. *Protein Science*, 20(9), pp.1484–1491.
- King, D.T. et al., 2012. New Delhi metallo- β -lactamase: structural insights into β -lactam recognition and inhibition. *Journal of the American Chemical Society*, 134(28), pp.11362–11365.
- King, G. & Warshel, A., 1989. A surface constrained all-atom solvent model for effective simulations of polar solutions. *The Journal of Chemical Physics*, 91(6), pp.10.1063–1.456845.
- Kitagawa, M. et al., 2005. Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research. *DNA research : an international journal for rapid publication of reports on genes and genomes*, 12(5), pp.291–299.
- Klitgord, N. & Segrè, D., 2011. Ecosystems biology of microbial metabolism. *Current Opinion in Biotechnology*, 22(4), pp.541–546.
- Kobayashi, M. & Shimizu, S., 1999. Cobalt proteins. *European journal of biochemistry / FEBS*, 261(1), pp.1–9.
- Koonin, E.V., 2016. Orthologs, paralogs, and evolutionary genomics. *Annual review of genetics*, 39, pp.309–338.
- Kostecky, B. et al., 2006. The Crystal Structure of the Zinc Phosphodiesterase from Escherichia coli Provides Insight into Function and Cooperativity of tRNase Z-Family Proteins. *Journal of Bacteriology*, 188(4), pp.1607–1614.
- Lang, G.I. & Desai, M.M., 2014. The spectrum of adaptive mutations in experimental evolution. *Genomics*, 104(6 Pt A), pp.412–416.
- Lapalikar, G.V. et al., 2012. Cofactor promiscuity among F420-dependent reductases enables them to catalyse both oxidation and reduction of the same substrate. *Catalysis Science & Technology*, 2(8), p.1560.

- Larion, M. et al., 2007. Divergent evolution of function in the ROK sugar kinase superfamily: role of enzyme loops in substrate specificity. *Biochemistry*, 46(47), pp.13564–13572.
- Leemhuis, H., Kelly, R.M. & Dijkhuizen, L., 2009. Directed evolution of enzymes: Library screening strategies. *IUBMB life*, 61(3), pp.222–228.
- Lešćić Ašler, I. et al., 2010. Probing enzyme promiscuity of SGNH hydrolases. *Chembiochem : a European journal of chemical biology*, 11(15), pp.2158–2167.
- Lim, S.A. et al., 2016. Evolutionary trend toward kinetic stability in the folding trajectory of RNases H. *Proceedings of the National Academy of Sciences of the United States of America*, 113(46), pp.13045–13050.
- Limphong, P. et al., 2009. Human Glyoxalase II Contains an Fe(II)Zn(II) Center but Is Active as a Mononuclear Zn(II) Enzyme. *Biochemistry*, 48(23), pp.5426–5434.
- Liu, D. et al., 2008. Mechanism of the quorum-quenching lactonase (AiiA) from *Bacillus thuringiensis*. 1. Product-bound structures. *Biochemistry*, 47(29), pp.7706–7714.
- Liu, Y. & Eisenberg, D., 2002. 3D domain swapping: as domains continue to swap. *Protein science : a publication of the Protein Society*, 11(6), pp.1285–1299.
- Liu, Y. et al., 1998. The crystal structure of a 3D domain-swapped dimer of RNase A at a 2.1-Å resolution. *Proceedings of the National Academy of Sciences*, 95(7), pp.3437–3442.
- Lunzer, M. et al., 2005. The biochemical architecture of an ancient adaptive landscape. *Science*, 310(5747), pp.499–501.
- Lunzer, M., Golding, G.B. & Dean, A.M., 2010. Pervasive cryptic epistasis in molecular evolution. *PLoS genetics*, 6(10), p.e1001162.
- Lutz, S., Lichter, J. & Liu, L., 2007. Exploiting temperature-dependent substrate promiscuity for nucleoside analogue activation by thymidine kinase from *Thermotoga maritima*. *Journal of the American Chemical Society*, 129(28), pp.8714–8715.
- Makarova, K.S. & Grishin, N.V., 1999. The Zn-peptidase superfamily: functional convergence after evolutionary divergence. *J. Mol. Biol.*, 292(1), pp.11–17.
- Marelius, J. et al., 1998. Q: a molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems. *Journal of molecular graphics & modelling*, 16(4-6), pp.213–25– 261.
- Masel, J. & Trotter, M.V., 2010. Robustness and evolvability. *Trends in genetics : TIG*, 26(9), pp.406–414.
- Mashiyama, S.T. et al., 2014. Large-Scale Determination of Sequence, Structure, and Function Relationships in Cytosolic Glutathione Transferases across the Biosphere. *PLoS Biology*, 12(4), p.e1001843.

- McCall, K.A. & Fierke, C.A., 2000. Colorimetric and fluorimetric assays to quantitate micromolar concentrations of transition metals. *Analytical Biochemistry*, 284(2), pp.307–315.
- McGuigan, K. & Sgrò, C.M., 2009. Evolutionary consequences of cryptic genetic variation. *Trends in ecology & evolution*, 24(6), pp.305–311.
- McLoughlin, S.Y. & Copley, S.D., 2008. A compromise required by gene sharing enables survival: Implications for evolution of new enzyme activities. *Proceedings of the National Academy of Sciences of the United States of America*, 105(36), pp.13497–13502.
- Meier, M.M. et al., 2013. Molecular engineering of organophosphate hydrolysis activity from a weak promiscuous lactonase template. *Journal of the American Chemical Society*, 135(31), pp.11670–11677.
- Meng, E.C. & Babbitt, P.C., 2011. Topological variation in the evolution of new reactions in functionally diverse enzyme superfamilies. *Curr. Opin. Struct. Biol.*, 21(3), pp.391–397.
- Meyer, J.R. et al., 2012. Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science*, 335(6067), pp.428–432.
- Miller, B.G. & Wolfenden, R., 2002. Catalytic proficiency: the unusual case of OMP decarboxylase. *Annual Review of Biochemistry*, 71, pp.847–885.
- Mills, D.R., Peterson, R.L. & Spiegelman, S., 1967. An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proceedings of the National Academy of Sciences*, 58(1), pp.217–224.
- Mir-Montazeri, B. et al., 2011. Crystal structure of a dimeric archaeal Cleavage and Polyadenylation Specificity Factor. *Journal of Structural Biology*, 173(1), pp.191–195.
- Miton, C.M. & Tokuriki, N., 2016. How mutational epistasis impairs predictability in protein evolution and design. *Protein science : a publication of the Protein Society*, 25(7), pp.1260–1272.
- Mohamed, M.F. & Hollfelder, F., 2013. Efficient, crosswise catalytic promiscuity among enzymes that catalyze phosphoryl transfer. *Biochimica et biophysica acta*, 1834(1), pp.417–424.
- Ngaki, M.N. et al., 2012. Evolution of the chalcone-isomerase fold from fatty-acid binding to stereospecific catalysis. *Nature*, 485(7399), pp.530–533.
- Nielsen, M.M. et al., 2011. Substrate and metal ion promiscuity in mannosylglycerate synthase. *The Journal of biological chemistry*, 286(17), pp.15155–15164.
- Nobeli, I., Favia, A.D. & Thornton, J.M., 2009. Protein promiscuity and its implications for biotechnology. *Nature Biotechnology*, 27(2), pp.157–167.

- Noor, S. et al., 2012. Intramolecular Epistasis and the Evolution of a New Enzymatic Function R. Dobson, ed. *PLoS ONE*, 7(6), p.e39822.
- O'Brien, P.J. & Herschlag, D., 1999. Catalytic promiscuity and the evolution of new enzymatic activities. *Chemistry & Biology*, 6(4), pp.R91–R105.
- O'Loughlin, T.L., 2006. Natural history as a predictor of protein evolvability. *Protein Engineering Design and Selection*, 19(10), pp.439–442.
- Okrasa, K. & Kazlauskas, R.J., 2006. Manganese-Substituted Carbonic Anhydrase as a New Peroxidase. *Chemistry - A European Journal*, 12(6), pp.1587–1596.
- Olsson, M.H.M. et al., 2011. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *Journal of chemical theory and computation*, 7(2), pp.525–537.
- Paaby, A.B. & Rockman, M.V., 2014. Cryptic genetic variation: evolution's hidden substrate. *Nat. Rev. Genet.*, 15(4), pp.247–258.
- Packer, M.S. & Liu, D.R., 2015. Methods for the directed evolution of proteins. *Nature Reviews Genetics*, 16(7), pp.379–394.
- Pandya, C. et al., 2014. Enzyme Promiscuity: Engine of Evolutionary Innovation. *Journal of Biological Chemistry*, 289(44), pp.30229–30236.
- Parera, M. & Martinez, M.A., 2014. Strong Epistatic Interactions within a Single Protein. *Mol. Biol. Evol.*, 31(6), pp.1546–1553.
- Park, H.S., 2006. Design and Evolution of New Catalytic Activity with an Existing Protein Scaffold. *Science*, 311(5760), pp.535–538.
- Pettinati, I. et al., 2016. The Chemical Biology of Human Metallo- β -Lactamase Fold Proteins. *Trends in Biochemical Sciences*, 41(4), pp.338–355.
- Phillips, G. et al., 2012. Functional promiscuity of the COG0720 family. *ACS chemical biology*, 7(1), pp.197–209.
- Phillips, P.C., 2008. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics*, 9(11), pp.855–867.
- Pieper, U. et al., 2009. Target selection and annotation for the structural genomics of the amidohydrolase and enolase superfamilies. *Journal of structural and functional genomics*, 10(2), pp.107–125.
- Poelarends, G.J. et al., 2004. The hydratase activity of malonate semialdehyde decarboxylase: mechanistic and evolutionary implications. *Journal of the American Chemical Society*, 126(48), pp.15658–15659.

- Poirel, L. et al., 2000. Characterization of VIM-2, a carbapenem-hydrolyzing metallo-beta-lactamase and its plasmid- and integron-borne gene from a *Pseudomonas aeruginosa* clinical isolate in France. *Antimicrobial Agents and Chemotherapy*, 44(4), pp.891–897.
- Pordea, A., 2015. Metal-binding promiscuity in artificial metalloenzyme design. *Curr. Opin. Chem. Biol.*, 25C, pp.124–132.
- Pordea, A. & Ward, T.R., 2009. Artificial metalloenzymes: combining the best features of homogeneous and enzymatic catalysis. *Synlett*, 2009(20), pp.3225–3236.
- Pougach, K. et al., 2014. Duplication of a promiscuous transcription factor drives the emergence of a new regulatory network. *Nature Communications*, 5, p.4868.
- Prosser, G.A., Larrouy-Maumus, G. & de Carvalho, L.P.S., 2014. Metabolomic strategies for the identification of new enzyme functions and metabolic pathways. *EMBO reports*, 15(6), pp.657–669.
- Puehringer, S., Metlitzky, M. & Schwarzenbacher, R., 2008. The pyrroloquinoline quinone biosynthesis pathway revisited: a structural approach. *BMC Biochemistry*, 9, p.8.
- Punta, M. et al., 2012. The Pfam protein families database. *Nucleic Acids Research*, 40(Database issue), pp.D290–301.
- Purg, M. et al., 2016. Probing the mechanisms for the selectivity and promiscuity of methyl parathion hydrolase. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 374(2080).
- Qu, C. et al., 2000. 3D domain swapping modulates the stability of members of an icosahedral virus group. *Structure (London, England : 1993)*, 8(10), pp.1095–1103.
- Radivojac, P. et al., 2013. A large-scale evaluation of computational protein function prediction. *Nature methods*, 10(3), pp.221–227.
- Ragsdale, S.W., 2009. Nickel-based Enzyme Systems. *The Journal of biological chemistry*, 284(28), pp.18571–18575.
- Ranea, J.A.G. et al., 2006. Protein Superfamily Evolution and the Last Universal Common Ancestor (LUCA). *Journal of Molecular Evolution*, 63(4), pp.513–525.
- Rauwerdink, A. et al., 2016. Evolution of a Catalytic Mechanism. *Molecular Biology and Evolution*, 33(4), pp.971–979.
- Razvi, A. & Scholtz, J.M., 2006. Lessons in stability from thermophilic proteins. *Protein science : a publication of the Protein Society*, 15(7), pp.1569–1578.
- Rees, D.C. & Robertson, A.D., 2001. Some thermodynamic implications for the thermostability of proteins. *Protein science : a publication of the Protein Society*, 10(6), pp.1187–1194.

- Reetz, M.T., 2013. The Importance of Additive and Non-Additive Mutational Effects in Protein Engineering. *Angewandte Chemie International Edition*, 52(10), pp.2658–2666.
- Renata, H., Wang, Z.J. & Arnold, F.H., 2015. Expanding the enzyme universe: accessing non-natural reactions by mechanism-guided directed evolution. *Angewandte Chemie (International ed. in English)*, 54(11), pp.3351–3367.
- Reymond, J.-L. & Wahler, D., 2002. Substrate arrays as enzyme fingerprinting tools. *Chembiochem : a European journal of chemical biology*, 3(8), pp.701–708.
- Reynolds, E.W. et al., 2016. An Evolved Orthogonal Enzyme/Cofactor Pair. *Journal of the American Chemical Society*, 138(38), pp.12451–12458.
- Riaz, K. et al., 2008. A metagenomic analysis of soil bacteria extends the diversity of quorum-quenching lactonases. *Environmental Microbiology*, 10(3), pp.560–570.
- Riddles, P.W., Blakeley, R.L. & Zerner, B., 1983. [8] Reassessment of Ellman's reagent. *Methods in enzymology*, 91, pp.49–60.
- Risso, V.A. et al., 2013. Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian β -lactamases. *Journal of the American Chemical Society*, 135(8), pp.2899–2902.
- Romero, P.A. & Arnold, F.H., 2009. Exploring protein fitness landscapes by directed evolution. *Nature Reviews Molecular Cell Biology*, 10(12), pp.866–876.
- Roodveldt, C. & Tawfik, D.S., 2005. Shared Promiscuous Activities and Evolutionary Features in Various Members of the Amidohydrolase Superfamily. *Biochemistry*, 44(38), pp.12728–12736.
- Rufo, C.M. et al., 2014. Short peptides self-assemble to produce catalytic amyloids. *Nature chemistry*, 6(4), pp.303–309.
- Rulísek, L. & Vondrásek, J., 1998. Coordination geometries of selected transition metal ions (Co^{2+} , Ni^{2+} , Cu^{2+} , Zn^{2+} , Cd^{2+} , and Hg^{2+}) in metalloproteins. *Journal of Inorganic Biochemistry*, 71(3-4), pp.115–127.
- Salverda, M.L.M. et al., 2011. Initial Mutations Direct Alternative Pathways of Protein Evolution J. Zhang, ed. *PLoS genetics*, 7(3), p.e1001321.
- Sanchez-Ruiz, J.M., 2010. Protein kinetic stability. *Biophysical chemistry*, 148(1-3), pp.1–15.
- Sandegren, L. & Andersson, D.I., 2009. Bacterial gene amplification: implications for the evolution of antibiotic resistance. *Nature Reviews Microbiology*, 7(8), pp.578–588.
- Sánchez-Moreno, I. et al., 2009. From kinase to cyclase: an unusual example of catalytic promiscuity modulated by metal switching. *Chembiochem : a European journal of chemical biology*, 10(2), pp.225–229.

- Schaper, S., Johnston, I.G. & Louis, A.A., 2012. Epistasis can lead to fragmented neutral spaces and contingency in evolution. *Proceedings. Biological sciences / The Royal Society*, 279(1734), pp.1777–1783.
- Schmidt, D.M.Z. et al., 2003. Evolutionary potential of (beta/alpha)₈-barrels: functional promiscuity produced by single substitutions in the enolase superfamily. *Biochemistry*, 42(28), pp.8387–8393.
- Schnoes, A.M. et al., 2009. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Computational Biology*, 5(12), p.e1000605.
- Schnoes, A.M. et al., 2013. Biases in the Experimental Annotations of Protein Function and Their Effect on Our Understanding of Protein Function Space C. A. Orengo, ed. *PLoS Computational Biology*, 9(5), p.e1003063.
- Schramm, V.L., 2011. Enzymatic Transition States, Transition-State Analogs, Dynamics, Thermodynamics, and Lifetimes. *Annual Review of Biochemistry*, 80(1), pp.703–732.
- Schulenburg, C. et al., 2015. Comparative laboratory evolution of ordered and disordered enzymes. *Journal of Biological Chemistry*, 290(15), pp.9310–9320.
- Seffernick, J.L. et al., 2001. Melamine deaminase and atrazine chlorohydrolase: 98 percent identical but functionally different. *J. Bacteriol.*, 183(8), pp.2405–2410.
- Seibert, C.M. & Raushel, F.M., 2005. Structural and Catalytic Diversity within the Amidohydrolase Superfamily. *Biochemistry*, 44(17), pp.6383–6391.
- Shannon, P. et al., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11), pp.2498–2504.
- Silaghi-Dumitrescu, R. et al., 2005. X-ray Crystal Structures of Moorella thermoacetica FprA. Novel Diiron Site Structure and Mechanistic Insights into a Scavenging Nitric Oxide Reductase. *Biochemistry*, 44(17), pp.6492–6501.
- Sillitoe, I. et al., 2015. CATH: comprehensive structural and functional annotations for genome sequences. *Nuc. Acids Res.*, 43(Database issue), pp.D376–81.
- Singh, B.K., 2009. Organophosphorus-degrading bacteria: ecology and industrial applications. *Nature Reviews Microbiology*, 7(2), pp.156–164.
- Smith, J.M., 1970. Natural selection and the concept of a protein space. *Nature*, 225(5232), pp.563–564.
- Socha, R.D. & Tokuriki, N., 2013. Modulating protein stability - directed evolution strategies for improved protein function. *The FEBS journal*, 280(22), pp.5582–5595.
- Song, N. et al., 2008. Sequence Similarity Network Reveals Common Ancestry of Multidomain Proteins C. Vogel, ed. *PLoS Computational Biology*, 4(5), p.e1000063.

- 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(4), pp.1484–1489.
- Soskine, M. & Tawfik, D.S., 2010. Mutational effects and the evolution of new protein functions. *Nature Reviews Genetics*, 11(8), pp.572–582. Available at: <http://www.nature.com/doi/10.1038/nrg2808>.
- Spencer, J. et al., 2005. Antibiotic recognition by binuclear metallo-beta-lactamases revealed by X-ray crystallography. *Journal of the American Chemical Society*, 127(41), pp.14439–14444.
- Starr, T.N. & Thornton, J.W., 2016. Epistasis in protein evolution. *Protein Science*, 25(7), pp.1204–1218.
- Stefani, M., 2004. Protein misfolding and aggregation: new examples in medicine and biology of the dark side of the protein world. *Biochimica et biophysica acta*, 1739(1), pp.5–25.
- Suttisansanee, U. & Honek, J.F., 2011. Bacterial glyoxalase enzymes. *Seminars in Cell & Developmental Biology*, 22(3), pp.285–292.
- Szilágyi, A., Zhang, Y. & Závodszky, P., 2012. Intra-chain 3D segment swapping spawns the evolution of new multidomain protein architectures. *Journal of Molecular Biology*, 415(1), pp.221–235.
- Søndergaard, C.R. et al., 2011. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *Journal of chemical theory and computation*, 7(7), pp.2284–2295.
- Tamura, K. et al., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, 28(10), pp.2731–2739.
- Taute, K.M. et al., 2014. Evolutionary constraints in variable environments, from proteins to networks. *Trends in genetics : TIG*.
- Teilum, K., Olsen, J.G. & Kragelund, B.B., 2011. Protein stability, flexibility and function. *Biochimica et biophysica acta*, 1814(8), pp.969–976.
- Thoma, R. et al., 2000. Structure and function of mutationally generated monomers of dimeric phosphoribosylanthranilate isomerase from *Thermotoga maritima*. *Structure (London, England : 1993)*, 8(3), pp.265–276.
- Tocchini-Valentini, G.D., Fruscoloni, P. & Tocchini-Valentini, G.P., 2005. Structure, function, and evolution of the tRNA endonucleases of Archaea: an example of subfunctionalization. *Proceedings of the National Academy of Sciences*, 102(25), pp.8933–8938.
- Todd, A.E., Orengo, C.A. & Thornton, J.M., 2001. Evolution of function in protein

- superfamilies, from a structural perspective. *J. Mol. Biol.*, 307(4), pp.1113–1143.
- Tokuriki, N. & Tawfik, D.S., 2009a. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature*, 459(7247), pp.668–673.
- Tokuriki, N. & Tawfik, D.S., 2009b. Protein dynamism and evolvability. *Science*, 324(5924), pp.203–207.
- Tokuriki, N. & Tawfik, D.S., 2009c. Stability effects of mutations and protein evolvability. *Current Opinion in Structural Biology*, 19(5), pp.596–604.
- Tokuriki, N. et al., 2012. Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nature Communications*, 3, p.1257.
- Tokuriki, N. et al., 2008. How protein stability and new functions trade off. *PLoS Computational Biology*, 4(2), p.e1000002.
- Tomatis, P.E. et al., 2008. Adaptive protein evolution grants organismal fitness by improving catalysis and flexibility. *Proceedings of the National Academy of Sciences*, 105(52), pp.20605–20610.
- Toth-Petroczy, A. & Tawfik, D.S., 2014. Hopeful (protein InDel) monsters? *Structure (London, England : 1993)*, 22(6), pp.803–804.
- Tracewell, C.A. & Arnold, F.H., 2009. Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Current Opinion in Chemical Biology*, 13(1), pp.3–9.
- Tufts, D.M. et al., 2015. Epistasis constrains mutational pathways of hemoglobin adaptation in high-altitude pikas. *Molecular Biology and Evolution*, 32(2), pp.287–298.
- UniProt Consortium, 2015. UniProt: a hub for protein information. *Nucleic Acids Research*, 43(Database issue), pp.D204–12.
- Valdez, C.E. et al., 2014. Mysteries of metals in metalloenzymes. *Accounts of chemical research*, 47(10), pp.3110–3117.
- van Loo, B. et al., 2010. An efficient, multiply promiscuous hydrolase in the alkaline phosphatase superfamily. *Proceedings of the National Academy of Sciences of the United States of America*, 107(7), pp.2740–2745.
- Vogel, A., 2002. ElaC Encodes a Novel Binuclear Zinc Phosphodiesterase. *The Journal of biological chemistry*, 277(32), pp.29078–29085.
- Voordeckers, K., Brown, C.A. & Vanneste, K., 2012. Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. *PLoS Biology*, 10(12), p.e1001446.
- Voordeckers, K., Pougach, K. & Verstrepren, K.J., 2015. How do regulatory networks evolve and

- expand throughout evolution? *Curr. Opin. Biotechnol.*, 34, pp.180–188.
- Wagner, A., 2008. Neutralism and selectionism: a network-based reconciliation. *Nat. Rev. Genet.*, 9(12), pp.965–974.
- Waldron, K.J. & Robinson, N.J., 2009. How do bacterial cells ensure that metalloproteins get the correct metal? *Nature Reviews Microbiology*, 7(1), pp.25–35.
- Waldron, K.J. et al., 2009. Metalloproteins and metal sensing. *Nature*, 460(7257), pp.823–830.
- Wang, G. et al., 2010. A novel hydrolytic dehalogenase for the chlorinated aromatic compound chlorothalonil. *Journal of Bacteriology*, 192(11), pp.2737–2745.
- Wang, G. et al., 2011. Recent Advances in the Biodegradation of Chlorothalonil. *Current Microbiology*, 63(5), pp.450–457.
- Wang, J. et al., 2006. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of molecular graphics & modelling*, 25(2), pp.247–260.
- Watson, E., Yilmaz, L.S. & Walhout, A.J.M., 2015. Understanding Metabolic Regulation at a Systems Level: Metabolite Sensing, Mathematical Predictions, and Model Organisms. *Annual review of genetics*, 49, pp.553–575.
- Weinreich, D.M., 2006. Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. *Science*, 312(5770), pp.111–114.
- Weng, J.-K., Philippe, R.N. & Noel, J.P., 2012. The rise of chemodiversity in plants. *Science*, 336(6089), pp.1667–1670.
- Wolfenden, R., 2011. Benchmark Reaction Rates, the Stability of Biological Molecules in Water, and the Evolution of Catalytic Power in Enzymes. *Annual Review of Biochemistry*, 80(1), pp.645–667.
- Wyganowski, K.T., Kaltenbach, M. & Tokuriki, N., 2013. GroEL/ES buffering and compensatory mutations promote protein evolution by stabilizing folding intermediates. *Journal of Molecular Biology*, 425(18), pp.3403–3414.
- Xu, Y. et al., 2008. Structure and metal exchange in the cadmium carbonic anhydrase of marine diatoms. *Nature*, 452(7183), pp.56–61.
- Yang, G. et al., 2016. Conformational Tinkering Drives Evolution of a Promiscuous Activity through Indirect Mutational Effects. *Biochemistry*, 55(32), pp.4583–4593.
- Yang, H. et al., 2014. Spectroscopic and mechanistic studies of heterodimetallic forms of metallo- β -lactamase NDM-1. *Journal of the American Chemical Society*, 136(20), pp.7273–7285.
- Yew, W.S. et al., 2005. Evolution of enzymatic activities in the orotidine 5'-monophosphate

decarboxylase suprafamily: enhancing the promiscuous D-arabino-hex-3-ulose 6-phosphate synthase reaction catalyzed by 3-keto-L-gulonate 6-phosphate decarboxylase. *Biochemistry*, 44(6), pp.1807–1815.

Yokoyama, S. et al., 2014. Epistatic Adaptive Evolution of Human Color Vision J. Zhang, ed. *PLoS genetics*, 10(12), p.e1004884.

Yu, S. et al., 2009. Structure Elucidation and Preliminary Assessment of Hydrolase Activity of PqsE, the PseudomonasQuinolone Signal (PQS) Response Protein. *Biochemistry*, 48(43), pp.10298–10307.

Yuen, C.M. & Liu, D.R., 2007. Dissecting protein structure and function using directed evolution. *Nature methods*, 4(12), pp.995–997.

Zang, T.M., 2000. Arabidopsis Glyoxalase II Contains a Zinc/Iron Binuclear Metal Center That Is Essential for Substrate Binding and Catalysis. *Journal of Biological Chemistry*, 276(7), pp.4788–4795.

Zhang, D. et al., 2014. Resilience of biochemical activity in protein domains in the face of structural divergence. *Current Opinion in Structural Biology*, 26, pp.92–103.

Zhao, H. et al., 1998. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nature Biotechnology*, 16(3), pp.258–261.

Zou, T. et al., 2015. Evolution of conformational dynamics determines the conversion of a promiscuous generalist into a specialist enzyme. *Molecular Biology and Evolution*, 32(1), pp.132–143.

Appendices

Appendix A Supplementary material for chapter two

A.1 Individual kinetic parameters

Table A.1 Kinetic parameters

PDB ID or protein name	Reaction	k_{cat} [s^{-1}]	K_M [μM]	k_{cat}/K_M [$\text{M}^{-1}\text{s}^{-1}$]
1x8g	BLA ^a	120 ± 10	680 ± 110	1.8×10^5
	BLA ^b	0.003 ± 0.0001	150 ± 10	2.2×10^1
2fhx	BLA ^b	50 ± 1	2.9 ± 0.3	1.7×10^7
	PDE	0.004 ± 0.0003	450 ± 110	8.8×10^{-1}
	PPP	0.001 ± 0.0003	3100 ± 160	3.4×10^{-1}
1bc2	BLA ^b	2.4 ± 0.1	42 ± 2	5.8×10^4
	BLA ^b	21 ± 1	14 ± 8	1.5×10^6
1ko3	PDE	0.002 ± 0.0001	430 ± 20	5.2×10^0
	PPP	0.002 ± 0.0001	660 ± 50	2.7×10^0
	PTE ^c	0.0008 ± 0.000008	530 ± 20	1.5×10^0
3spu	BLA ^b	55 ± 5	21 ± 8	2.6×10^6
	PDE	0.005 ± 0.001	2500 ± 230	1.8×10^0
	PTE ^c	0.0007 ± 0.0001	450 ± 30	1.4×10^0
	PPP	0.001 ± 0.0002	4500 ± 1100	2.8×10^{-1}
2cfu	AKS	0.6 ± 0.01	11 ± 7	4.9×10^4
2gcu	BLA ^b	0.5 ± 0.1	2300 ± 510	2.3×10^2
1qh5	SLG	1.5 ± 0.1	370 ± 40	3.9×10^3
	PDE	0.0001 ± 0.00002	6300 ± 1500	2.1×10^{-2}
1xm8	SLG	9.0 ± 0.3	500 ± 40	1.8×10^4
	BLA ^b	0.008 ± 0.0003	290 ± 30	3.1×10^1
1k07	BLA ^b	150 ± 6	420 ± 10	3.5×10^5
	PTE ^c	0.0009 ± 0.0004	5000 ± 3700	1.9×10^{-1}
2aio	BLA ^b	22 ± 3	93 ± 3	2.4×10^5
	PPP	0.0004 ± 0.0001	3800 ± 1100	9.7×10^{-2}
	PDE	10 ± 2	5600 ± 1200	1.8×10^3
2az4	PCE	0.01 ± 0.002	23000 ± 5500	3.4×10^{-1}
	PPP	7 ± 3	40000 ± 19000	1.7×10^2
3h3e	BLA ^b	0.002 ± 0.0002	230 ± 30	1.0×10^1
	PPP	0.04 ± 0.003	4100 ± 600	9.7×10^0
	PDE	0.002 ± 0.0002	11000 ± 1400	1.6×10^{-1}
1zkp	PDE	0.4 ± 0.01	65 ± 10	5.3×10^3
	PPP	0.04 ± 0.001	500 ± 30	8.0×10^1
	PCE	0.02 ± 0.001	22000 ± 2300	7.0×10^{-2}

table continued

PDB ID or protein name	Reaction	k_{cat} [s^{-1}]	K_{M} [μM]	$k_{\text{cat}}/K_{\text{M}}$ [$\text{M}^{-1}\text{s}^{-1}$]
1ztc	HSL	10 ± 2	19000 ± 3900	5.4×10^2
	PPP	0.01 ± 0.001	2200 ± 720	4.7×10^0
	PDE	0.01 ± 0.001	2000 ± 600	4.3×10^0
	PTE ^c	0.01 ± 0.002	19000 ± 4900	6.8×10^{-1}
3aj3	HSL	4 ± 0.1	6300 ± 360	5.8×10^2
	BLA ^b	0.02 ± 0.001	620 ± 60	3.4×10^1
	PTE ^d	0.01 ± 0.002	12000 ± 3500	8.6×10^{-1}
3dhb	HSL	43 ± 1	2200 ± 20	2.0×10^4
	SLG	0.003 ± 0.0002	350 ± 70	8.6×10^0
	PTE ^c	0.002 ± 0.0005	10000 ± 3000	2.0×10^{-2}
1p9e	PTE ^d	12 ± 1	1000 ± 20	1.2×10^4
	PTE ^c	0.03 ± 0.01	480 ± 30	6.6×10^1
1xto	BLA ^b	0.04 ± 0.006	1230 ± 240	3.4×10^1
	SLG	0.01 ± 0.001	710 ± 110	1.7×10^1
2cbn	PDE	10 ± 1	3600 ± 390	2.9×10^3
	BLA ^b	0.001 ± 0.0003	460 ± 30	2.0×10^1
	PPP	4 ± 0.1	6100 ± 1200	5.9×10^0
	PCE	0.02 ± 0.01	66000 ± 24000	3.6×10^{-1}
1wra	PCE	2.3 ± 0.1	1600 ± 50	1.3×10^3
	PPP	n.d. [†]	n.d. [†]	3.4×10^{-1}
1vjn	BLA ^b	0.08 ± 0.001	13 ± 1	5.9×10^3
AtsA	ARS	2.4 ± 0.1	2500 ± 200	1.0×10^3
	PDE	2.0 ± 0.2	12000 ± 1900	1.7×10^2
ChD	TPN	0.1 ± 0.01	2.0 ± 0.4	8.0×10^4
	BLA ^b	0.0004 ± 0.00001	240 ± 80	1.5×10^0

^aSubstrate measured: Imipenem.

^bSubstrate measured: Centa.

^cSubstrate measured: paraoxon.

^dSubstrate measured: parathion-methyl.

[†]n.d means not determined because no saturation kinetics was observed. The data was fitted to pseudo first order kinetics in which the slope directly corresponds to $k_{\text{cat}}/K_{\text{M}}$. The means and standard deviation of the kinetic parameters were calculated from at least three independent measurements.

Appendix B Supplementary information for chapter three

B.1 Kinetic parameters of bla-L1

Table B.1 Kinetic parameters of bla-L1.

Metal ion	Reaction	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($\text{M}^{-1} \text{s}^{-1}$)
Zn	BLA	90 ± 2.3	24 ± 2.2	3.7×10^6
Mn	BLA	6.7 ± 0.2	65 ± 4.8	1.0×10^5
Co	BLA	6.8 ± 0.5	60 ± 12	1.1×10^5
	SLG	0.01 ± 0.001	630 ± 100	2.3×10^1
	PDE	0.02 ± 0.0007	5330 ± 360	2.9×10^0
Cd	PTE	0.01 ± 0.008	> 10 mM	2.7×10^{-1}
	BLA	0.5 ± 0.002	800 ± 45	5.8×10^2
	SLG	0.03 ± 0.003	1180 ± 170	2.2×10^1
Ni	BLA	1.1 ± 0.1	640 ± 80	1.7×10^3
	SLG	0.1 ± 0.005	450 ± 60	2.3×10^2
	PDE	0.2 ± 0.02	4360 ± 740	4.4×10^1
	PTE	0.03 ± 0.003	9980 ± 1060	3.2×10^0
	LAC	0.01 ± 0.001	3460 ± 450	2.5×10^0
Fe	BLA	0.06 ± 0.004	120 ± 23	4.4×10^2
LB purified	BLA	20 ± 0.9	53 ± 4	3.9×10^5

n.d means not determined because no saturation kinetics was observed. The data was fitted to pseudo first order kinetics in which the slope directly corresponds to k_{cat}/K_M . The means and standard deviation of the kinetic parameters were calculated from at least three independent measurements.

B.2 Table of individual kinetic parameters of bla-VIM2

Table B.2 Kinetic parameters of bla-VIM2.

Metal ion	Reaction	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($\text{M}^{-1} \text{s}^{-1}$)
Zn	BLA	22 ± 0.5	11 ± 1	2.0×10^6
	PDE	0.0002 ± 0.00002	590 ± 160	3.9×10^{-1}
	PTE	n.d.	> 10 mM	2.9×10^{-2}
Mn	BLA	3 ± 0.6	8 ± 1	2.9×10^5
Co	BLA	10 ± 0.3	20 ± 3	5.1×10^5
Cd	BLA	1 ± 0.03	6 ± 0.8	2.1×10^5
Ni	BLA	0.3 ± 0.02	75 ± 10	4.6×10^3
Fe	BLA	0.2 ± 0.006	440 ± 30	3.9×10^2
	LAC	n.d.	> 10 mM	3.2×10^{-1}
	EST	0.002 ± 0.0005	5100 ± 1800	4.1×10^{-1}
LB purified	BLA	14 ± 0.4	24 ± 1	5.9×10^5
	PDE	0.007 ± 0.0003	1485 ± 155	4.7×10^0
	PTE	0.0007 ± 0.00004	679 ± 118	1.0×10^0

n.d means not determined because no saturation kinetics was observed. The data was fitted to pseudo first order kinetics in which the slope directly corresponds to k_{cat}/K_M . The means and standard deviation of the kinetic parameters were calculated from at least three independent measurements.

B.3 Table of individual kinetic parameters of mph

Table B.3 Kinetic parameters of mph.

Metal ion	Reaction	k_{cat} (s^{-1})	K_{M} (μM)	$k_{\text{cat}}/K_{\text{M}}$ ($\text{M}^{-1} \text{s}^{-1}$)
Zn	PTE	0.1 ± 0.006	2150 ± 250	5.3×10^1
	EST	0.2 ± 0.01	> 10 mM	1.8×10^1
Mn	PTE	3.2 ± 0.1	1080 ± 80	2.9×10^3
	EST	1.3 ± 0.02	2490 ± 80	5.4×10^2
	LAC	0.002 ± 0.0002	660 ± 90	3.0×10^0
Co	PTE	n.d.	> 10 mM	1.9×10^2
	EST	0.2 ± 0.01	1610 ± 180	1.4×10^2
	LAC	n.d.	> 10 mM	2.5×10^{-1}
Cd	PTE	0.2 ± 0.01	400 ± 60	4.7×10^2
	EST	0.1 ± 0.01	2890 ± 370	4.8×10^1
Ni	PTE	4.8 ± 0.3	1330 ± 170	3.6×10^3
	EST	3.7 ± 0.09	2160 ± 100	1.7×10^3
	PDE	0.1 ± 0.005	1940 ± 140	5.7×10^1
Fe	PTE	n.d.	> 10 mM	1.0×10^0
	EST	0.001 ± 0.0001	870 ± 210	1.3×10^0
LB purified	PTE	0.05 ± 0.002	2063 ± 223	2.1×10^1
	PDE	0.002 ± 0.0002	515 ± 243	3.4×10^0

n.d means not determined because no saturation kinetics was observed. The data was fitted to pseudo first order kinetics in which the slope directly corresponds to $k_{\text{cat}}/K_{\text{M}}$. The means and standard deviation of the kinetic parameters were calculated from at least three independent measurements.

B.4 Table of individual kinetic parameters for atsA

Table B.4 Kinetic parameters of atsA.

Metal ion	Reaction	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($\text{M}^{-1} \text{s}^{-1}$)
Zn	ARS	0.1 ± 0.01	1120 ± 50	9.3×10^1
	PDE	0.3 ± 0.006	1470 ± 60	2.0×10^2
Mn	ARS	5.4 ± 0.2	890 ± 63	6.1×10^3
	PDE	0.7 ± 0.02	25 ± 2	2.9×10^4
	PCE	n.d.	> 10 mM	8.5×10^{-1}
Co	ARS	3.4 ± 0.09	26 ± 2	1.3×10^5
	PDE	0.8 ± 0.02	16 ± 2	4.9×10^4
	PCE	0.6 ± 0.02	> 10 mM	7.3×10^1
Cd	ARS	0.01 ± 0.006	1390 ± 120	9.9×10^0
	PDE	1 ± 0.06	340 ± 40	2.8×10^3
Ni	ARS	0.3 ± 0.02	680 ± 100	4.5×10^2
	PDE	0.005 ± 0.0002	5.8 ± 0.8	8.1×10^2
Fe	ARS	0.005 ± 0.0002	650 ± 80	7.4×10^1
	PDE	0.002 ± 0.0002	350 ± 67	6.6×10^0
LB purified	ARS			
	PDE	0.3 ± 0.02	164 ± 29	2.1×10^3
	PCE	n.d.	> 10 mM	6.0×10^{-3}

n.d means not determined because no saturation kinetics was observed. The data was fitted to pseudo first order kinetics in which the slope directly corresponds to k_{cat}/K_M . The means and standard deviation of the kinetic parameters were calculated from at least three independent measurements.

B.5 Table of individual kinetic parameters for rbn

Table B.5 Kinetic parameters of rbn.

Metal ion	Reaction	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($\text{M}^{-1} \text{s}^{-1}$)
Zn	PDE	n.d.	> 10 mM	1.4×10^1
	EST	0.001 ± 0.00003	1500 ± 90	6.7×10^{-1}
	BLA	0.002 ± 0.0001	110 ± 14	2.0×10^1
Mn	PDE	40 ± 1	100 ± 8	4.0×10^5
	PCE	n.d.	> 10mM	5.1×10^2
	EST	0.001 ± 0.00005	1600 ± 120	8.0×10^{-1}
	BLA	0.002 ± 0.0002	280 ± 56	7.8×10^0
Co	PDE	53 ± 3	560 ± 60	9.4×10^4
	PCE	n.d.	> 10 mM	6.6×10^2
	EST	0.001 ± 0.00007	1300 ± 170	8.0×10^{-1}
	BLA	0.002 ± 0.00009	180 ± 30	8.2×10^0
Cd	PDE	2 ± 0.2	2500 ± 530	7.9×10^2
	PCE	n.d.	> 10 mM	3.9×10^0
	EST	0.002 ± 0.00007	2800 ± 180	7.1×10^{-1}
	BLA	0.004 ± 0.001	560 ± 310	6.4×10^0
Ni	PDE	1.3 ± 0.1	390 ± 120	3.3×10^3
	PCE	0.3 ± 0.006	2000 ± 130	1.5×10^2
	EST	0.001 ± 0.0001	1900 ± 770	5.6×10^{-1}
	BLA	0.002 ± 0.0001	200 ± 27	9.9×10^0
	ARS	0.002 ± 0.0005	> 10 mM	2.6×10^{-2}
Fe	PDE	0.3 ± 0.01	1120 ± 130	2.4×10^2
	EST	0.002 ± 0.0003	2300 ± 830	1.1×10^0
	BLA	0.004 ± 0.0004	420 ± 80	1.0×10^1
LB purified	PDE	24 ± 2	5092	4.7×10^3
	PCE	n.d.	> 10 mM	2.6×10^{-1}

n.d means not determined because no saturation kinetics was observed. The data was fitted to pseudo first order kinetics in which the slope directly corresponds to k_{cat}/K_M . The means and standard deviation of the kinetic parameters were calculated from at least three independent measurements.