EXPLORING THE ROLE OF MUTATIONS IN THE SIGNAL PEPTIDES OF VIM-2, NDM-1, AND IMP-1 IN THE DEVELOPMENT OF ADVANTAGEOUS PHENOTYPES IN *ESCHERICHIA COLI*

by

Hee Jin Jun

B.Sc., Simon Fraser University, 2014

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Biochemistry and Molecular Biology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2021

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, a thesis entitled:

Exploring the role of mutations in the signal peptides of VIM-2, NDM-1, and IMP-1 in the development of advantageous phenotypes in *Escherichia coli*

submitted by Hee Jin Jun in partial fulfillment of the requirements for

the degree of Master of Science

in Biochemistry and Molecular Biology

Examining Committee:

Dr. Nobuhiko Tokuriki, Associate Professor, Biochemistry & Molecular Biology, UBC
Supervisor

Dr. Robert E. W. Hancock, Professor, Microbiology & Immunology, UBC
Supervisory Committee Member

Dr. Natalie Strynadka, Professor, Biochemistry & Molecular Biology, UBC
Supervisory Committee Member

Dr. Elitza Tocheva, Assistant Professor, Microbiology & Immunology, UBC
Additional Examiner

Abstract

Metallo-β-lactamases (MBLs) are powerful enzymes conferring antibiotic resistance to various pathogens. They are actively disseminated in a plethora of pathogenic organisms *via* horizontal gene transfer, and raise clinical concerns. Interestingly, a predominant number of pathogens that favor MBL expression belong to γ-proteobacteria such *as Pseudomonas aeruginosa*, implying biases in the host distribution. A previous study in the Tokuriki lab also revealed restricted MBL gene expression in phylogenetically distanced bacterial species. However, it is still obscure how the enzymes adapt to bacterial hosts. This thesis aims to understand mutations in the signal peptides of MBLs contributing to the advantageous phenotype development in *Escherichia coli*. A series of dose-response curve assays revealed a highly positive correlation between the sequenced individual variants and their respective population where a considerably narrow phenotypic diversity is observed. When the phenotypic diversity is in a limited range at a high ampicillin concentration, the variability in mutational effects associated with phenotypic variations may be confined by fitness costs, resulting in the development of similar phenotypes. This suggests that the sequence profiles of the variants are a reasonable representation of the mutational population. A bioinformatic analysis of the sequence profiles of NDM-1, VIM-2, and IMP-1 variants reveals that mutation-driven changes in translational steps are moderately correlated with minimum inhibitory concentration (MIC), implying that the mutations in the signal peptides are associated with advantageous phenotypes elevating MBL gene expression. In addition, I synthesized a new broad host range vector using three individual plasmid vectors by the Golden Gate assembly technique to analyze mutational effects of MBLs in *Acinetobacter baumannii, Escherichia coli, Pseudomonas aeruginosa,* and *Pseudomonas putida*. The newly synthesized plasmid vector is compatible with the listed bacterial species enabling future

research. Overall, these results in the thesis will help understand how the mutations in the signal peptide of MBLs, in part, promote advantageous phenotypes in *E. coli* to elevate MBL gene expression, and may help develop future therapeutic strategies.

Lay Summary

Metallo-β-lactamases (MBLs) are enzymes that inactivate currently available β-lactam drugs and cause life-threatening events. *Via* horizontal gene transfer, the genetic information of the enzymes is disseminated in various pathogenic organisms worldwide, urging us to develop new and stable antibiotic derivatives. If this gruesome situation continues, an increasing number of deaths by multi-drug resistant infections will be inevitable. Even now, it remains elusive how MBLs develop advantageous phenotypes in a new host and confer high antibiotic resistance to multiple β-lactams. In this study, I sought to determine the connection between the mutations in the signal peptides and the advantageous phenotypes of MBLs that lead to increased gene expression. This study will contribute to a better understanding of the importance of adaptability of MBLs related to antibiotic resistance.

Preface

All work described in this thesis was conducted in the Tokuriki lab under the supervision of Dr. Nobuhiko Tokuriki. The experiments were all performed at the University of British Columbia in Vancouver, British Columbia, between 2018 and 2021 (the publication of this thesis).

Chapters 1 to 5 were written by myself and edited by Dr. Tokuriki. In Chapter 3, the dose-response curve assay that was used to analyze antibiotic resistance of variants' population was conceived and designed through discussion with my supervisor, Dr. Tokuriki. I conducted all experiments. The sequence profiles of VIM-2, NDM-1, and IMP-1 variants were acquired from Dr. Socha, who was a previous Ph.D. student in the Tokuriki lab. I performed all bioinformatic and data analysis, and interpreted results. I wrote the dose-response curve assay analysis in Chapter 3 in collaboration with Dr. Tokuriki, and the rest of the chapter was written entirely by myself. Dr. Tokuriki and John Chen, a Ph.D. student in the same lab, edited the chapter.

The new broad host range vector in Chapter 4 was conceived and designed by Dr. Tokuriki, Sevan Gholipoor, a Master's student in the same lab, and myself. Sevan Gholipoor and I created three sets of primers to arrange the three genes of interest in desired directions. I designed the primers to ligate gentamicin and the vector fragments at the final stage of the vector synthesis. The entire process of vector synthesis was performed solely by myself. I wrote the chapter myself, and Dr. Tokuriki edited the chapter.

Table of Contents

List of Tables

List of Figures

List of Symbols

α	alpha
β	beta
γ	gamma
μ	micro
Π	genetic diversity
%	percent
>	greater than
<	less than
=	equal to
IC_{50}	half maximal inhibitory concentration

List of Abbreviations

AMP	ampicillin
A. baumannii	*Acinetobacter baumannii*
A. baylyi	*Acinetobacter baylyi*
BHR	broad host range
CAI	Codon Adaptation Index
CARD	Comprehensive Antibiotic Resistance Database
CTX	cefotaxime
DNA	deoxynucleic acid
E. coli	*Escherichia coli*
GFP	green fluorescent protein
HGT	horizontal gene transfer
HMM-PBP	high molecular mass penicillin-binding protein
HSD	helical scaffold domain
K. pneumoniae	*Klebsiella pneumoniae*
MBL	metallo-β-lactamase
MBLED	Metallo-β-Lactamase Engineering Database
MEM	meropenem
MIC	minimum inhibitory concentration
mRNA	messenger ribonucleic acid
MRSA	Methicillin-resistant *Staphylococcus aureus*
NAM	N-acetylmuramic acid
NAG	N-acetylglucosamine

OD_{600}	optical density at 600 nm
P. aeruginosa	*Pseudomonas aeruginosa*
PBP	penicillin-binding protein
P. putida	*Pseudomonas putida*
SBL	serine-β-lactamase
SDM	site directed mutagenesis
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
PPXD	peptide cross linking domain
PCR	polymerase chain reaction
UBC	University of British Columbia
UTR	untranslated region

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 Hist)	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 Qln)	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

I would like to thank Dr. Nobuhiko Tokuriki for giving me the opportunity to explore the world of mutations, one of the crucial driving forces of evolution. I appreciate my supervisor for giving me the inspiration and motivation in this research. His guidance helped me realize what I needed to improve on and how to become a better researcher. Experience working in a stimulating and challenging environment has impressed me greatly, which I will cherish in my journey ahead.

I was also most privileged to have wonderful committee members, Dr. Robert E. W. Hancock, Dr. Natalie Strynadka, and Dr. Elitza Tocheva. I appreciate their thoughtful insights and consideration of my thesis. This thesis would not have been possible without their guidance. I especially would like to express my deepest gratitude and appreciation to Dr. Robert E. W. Hancock for reading my thesis critically and giving me invaluable feedback that truly helped me become a better critical thinker. The profound insights of the committee members guided me to develop a keen eye for research that will be a tremendous asset in my future scientific career.

Moreover, I want to express my appreciation to Dr. Devki Nandan, who has always been a great mentor and provided me with wisdom when I had difficult times in research. And, this tough journey wouldn't have been possible without my friends, especially Stella Chen. I would like to thank past and present members of Tokuriki lab and Dr. Charlotte Miton, Dan Kehila, John Chen, and Sevan Gholipoor for their help on thesis writing.

Lastly, I would like to thank my family for their endless love, support, and patience as I pursued my master's study at UBC. My family has always given me the strength to continue whenever I struggled and was discouraged. I genuinely appreciate their love and support from the bottom of my heart. Thank you so much.

To my dear family

Chapter 1: Introduction

1.1 Overview

From the soil to our fingertips, although invisible, microorganisms have adapted and evolved to survive in extreme conditions over the centuries. In the post-antibiotic world, decades of arms race between human beings and pathogens are perhaps inevitable in the pursuit of survival. As an effective defense mechanism, numerous pathogenic organisms have acquired metallo-β-lactamases (MBLs) *via* horizontal gene transfer (HGT) to gain antibiotic resistance (Walsh 2005; González et al 2016; Bello-López et al 2019). It raises clinical concerns owing to a high dissemination rate, continual diversification with mutations, and broad-specificity to the currently available β-lactam drugs (Kohler et al 2020; Walsh 2005; Martínez-García et al 2018; Chiou et al 2014; González et al 2018; Gerald et al 1990). MBLs are composed of the N-terminal signal peptide and the catalytic domain (Makena et al 2015). The mature domain is characterized by the αβ/βα sandwich scaffold containing Zn (II) ions at the active site for efficient β-lactam drug inactivation (Makena et al 2015), whereas the signal peptide gets cleaved during a translocation step (Owji et al 2018). Codon usage proximal to the start codon at the N-terminus of mRNA influences gene expression regulation (Verma et al 2019; Qing et al 2008). In the case of preproteins, signal peptides are located at the N-terminus and may be responsible for the efficiency of protein synthesis. Furthermore, the hydrophobic threshold and hydrophobic residues of signal peptides govern the efficiency of mature peptide exportation *via* the SecYEG system mediated by the SecA ATPases (Tuller et al 2010; Kozak et al 2005; Bouvier et al 2008; Gu et al 2010; Brockmann et al 2007; Vogel et al 2007; Huang et al 2011; Owji et al 2018; Heijne 1984; Fikes et al 1990; Kebir et al 2002; Duffy et al 2010; Auclair et al 2010). These lines of evidence imply that signal peptides of MBLs may contribute to the MBL gene expression and

the efficiency of mature peptide translocation in a new host after being transferred *via* HGT. Given this role of signal peptides, mutations in the signal peptides of MBLs may be responsible for increasing advantageous phenotypes in a new environment at the translational and translocational levels.

To develop advantageous phenotypes in a new bacterial host, rare and beneficial mutations tend to fix in a sequence space and increase their frequencies in the sequence space of MBLs as well as in a finite subdivided population under positive selection. The genotypes of MBL enzymes corresponding to their phenotypes, which are antibiotic resistant, can be analyzed by the directed evolution experiment. While, stochastic mutations tend to prevail in the sequence space of an enzyme under neutral drift evolution, the extent of mutational tolerance is regulated by a selective threshold. All in all, this chapter will provide a comprehensive background for β-lactam antibiotic drugs and MBLs.

1.2 Introduction of β-lactams

β-Lactam production from the mold *Penicillium rubrum* (Kong et al 2010) was first discovered in 1928 by Sir. Alexander Fleming, who noticed a growth inhibition of *Staphylococcus aureus* around the mold. In search of viable derivatives of penicillin, the β-lactam discovered by Fleming, scientists have since identified a number of β-lactam antibiotic classes and further modified them into semi-synthetic therapeutics. The four representative classes of natural β-lactam antibiotics are i) penicillins ii) cephalosporins, iii) carbapenems, and iv) monobactams. β-lactam antibiotics functions by targeting components of bacterial cell walls. Bacterial cell walls comprise alternating N-acetylmuramic acid (NAM) and N-acetylglucosamine (NAG),

known as a peptidoglycan polymer (Mojica et al 2016). To withstand osmotic pressure and maintain cell integrity (Park & Strominger 1957; Tipper & Strominger 1965), NAM and NAG units are cross-linked by penicillin-binding proteins (PBPs). PBPs are divided into two classes based on the molecular weight: low- and high-molecular-weight PBPs (Zervosen et al 2012). According to a previous study, high-molecular-weight PBPs (HMW PBPs) play a significant role in cell survival and are further classified into class A and class B. Class A PBPs are bifunctional enzymes that catalyze both trans-glycosylation and transpeptidation, whereas class B PBPs are transpeptidases (Haenni et al 2006). Both class A and class B PBPs cross-link peptidoglycan stem-peptide (Mojica et al 2016; Zervosen et al 2012), and at least one of the enzymes is employed for cell division and elongation processes in *Escherichia coli, Bacillus subtilis, Caulobacter crescentus, Staphylococcus aureus* and *Streptococcus pneumoniae* (Macheboeuf et al 2006). Class A PBPs particularly catalyze the trans-glycosylation reactions (Lovering et al 2012). Because of the catalytic activity, HMW-PBPs become a target of β-lactams (Zervosen et al 2012), which inhibit bacterial peptidoglycan cell wall synthesis utilizing the universally shared β-lactam ring.

Naturally derived penicillin G exhibits a basic conformation composed of a thiazolidine ring and an acyl side chain connected to a β-lactam ring (Soares et al 2012). Although penicillin G is used regularly in clinical settings, it is easily hydrolyzed by protonation of the nitrogen of the β-lactam functional group (Kong et al 2010; Lima et al 2020) and, consequently, recognized for developing more stable derivatives compared to the first kind. Penicillin V, the first semi-synthetic derivative, is more stable than penicillin G; however, it also is easily inactivated by Gram-negative bacteria (Soares et al 2012). To date, while subsequent generations of semi-

synthetic penicillin, such as ampicillin and methicillin, have been generated with additional modifications and the acyl side chain substitutions, their activity is still limited to the specific bacterial species of Gram positive and negative bacteria (Soares et al 2012).

Cephalosporin, initially isolated from *Cephalosporium acremonium* cultures, contains the dihydrothiazine ring fused to the β-lactam ring and is relatively less reactive to the hydrolysis (Lima et al 2020). The compound's efficacy was first demonstrated by inhibition of *Salmonella typhi*, which causes typhoid fever (Lima et al 2020). Because of its stability, bactericidal efficacy, and broad anti-bacterial spectrum, cephalosporins are actively used in clinical practices to treat pneumonia, respiratory and urinary tract, as well as bloodstream-related infections (Das et al 2019). At the moment, five generations of cephalosporin are available. Cefotaxime and cefoperazone, which belong to the third generation, are commonly used in pre-and post-chemotherapy for treating multiple surgery infections such as cardiac and vascular surgery infections (Gerald et al 1990; Das et al 2019). The efficacy of cephalosporins is more potent against Gram-negative bacteria compared to Gram-positive bacteria (Das et al 2019). Fortunately, subsequent semi-synthetic cephalosporins synthesized by substituting new molecules to the penam and cephem nuclei are shown to be effective against Gram-negative and Gram-positive bacteria (García-Estrada et al 2011).

The carbapenem backbone is originated from olivanic acids, which are naturally produced by *Streptomyces clavuligerus*. Insufficient effectiveness in bacterial cells prompted for further investigations that led to the discovery of two different compounds from *S. clavuligerus* and *Streptomyces cattleya: a* clavulanic acid and thienamycin, respectively. The currently available

carbapenems stem from thienamycin, which serves as the parent compound (Papp-Wallace et at 2011). Carbapenem shares the identical β-lactam ring with penicillin except for a double bond between C2 and C3 with the substitution of carbon at C1. The unsaturated C-2 and C-3 bonds confer a hybrid chemical reactivity in comparison to penicillin and cephalosporin (Dalhoff et al 2003; Lima et al 2020). Previous studies suggest three major components that strengthen carbapenems' chemical stability against β-lactamases. First, the carbon atom at C1 enhances the potency as well as spectrum of carbapenems and increases the structural stability of the compound against β-lactamases. Second, the hydroxyethyl side chain at R2 further increases resistance to hydrolysis by β-lactamases (Pradel et al 2008). Third, the trans-configuration of the β-lactam ring at C5 and C6 provides additional stability against β-lactamases (Fabiane et al 1998). However, the compound's sensitivity to mild base hydrolysis urged demand for stable analogous derivatives such as imipenem and meropenem (Nesmeyanova et al 1997; Gilbert et al 2007; Brewer et al 1991). An elevating pace of resistance to currently available carbapenems is identified for *Pseudomonas spp., Acinetobacter spp*., and *Enterobacteriaceae,* which belong to nonfermenting Gram-negative bacteria (Codjoe et al 2018). To date, gentamicin and aminoglycoside antibiotics are widely used to cure antibiotic drug-resistant Gram-negative bacteria infections (Brewer et al 1991) that specifically depend on MBLs to gain antibiotic resistance.

Monobactams are derivatives of naturally produced antibiotics characterized by their monocyclic β-lactam ring (Bonner et al 1984). They share a structural resemblance to ceftazidime, and the scope of action is comparable to that of the third generation cephalosporines (Raimondi et al 1989). The efficacy of monobactams is potent against Gram-negative bacteria such

as *Enterobacteriaceae* and *Pseudomonas aeruginosa* owing to the stable resistance to MBLs (Dean et al 2018; Raimondi et al 1989), whereas they are insufficient to treat Gram-positive bacterial or anaerobe infections. Previous studies determined a lack of molecular stability to some serine β-lactamases (SBLs), which emerge together with MBLs to confer antibiotic resistance (Dean et al 2018; Raimondi et al 1989; Sykes et al 1985). To date, aztreonam is considered as an adequate alternative to aminoglycosides to treat aerobic Gram-negative infections (Hellinger et al 1999).

1.3 Defense mechanisms against β-lactams

PBPs play a crucial role in cross-linking the glycan strands in peptidoglycan by catalyzing covalent bond formations between bridging pentapeptides. β-Lactams such as penicillins, cephalosporins, carbapenems and monobactams (Meini et al 2015) contain the β-lactam ring that mimics the terminal D-Ala-D-Ala, subunits of the peptidoglycan layer. Because of the structural resemblance to the D-Ala-D-Ala moiety, β-lactam irreversibly acylates the Ser403 residue of the enzyme's active site. The irreversibly inhibited PBPs are incapable of catalyzing transpeptidation of the nascent peptidoglycan and induce cell wall disruption, rarely cell division and distorted cell shape (Fisher et al 2005; Meini et al 2015; Zervosen et al 2012; Macheboeuf et al 2006; Varma et al 2004), leaving bacteria vulnerable to osmotic pressure (Mojica et al 2016). There are three different mechanisms through which antibiotic drugs can penetrate pathogens: diffusion *via* porins, diffusion *via* the bilayer, and self-absorption (Kapoor et al 2017).

Because of their small in size and hydrophilic property, β-lactams can cross the outer membrane, mediated only by porins (Kapoor et al 2017). As a defense mechanism, pathogens reduce the

number of porins to lower the overall amount of β-lactam antibiotics coming into the cell and ultimately become resistant to β-lactams (Kapoor et al 2017; Lucidi et al 2018). Efflux pumps are another antibiotic resistance mechanism by which cells efficiently remove antimicrobials from the cytoplasm (Džidic et al 2008; Wise et al 1999; Kapoor et al 2017). Although most pumps can be specialized to distinct antibiotics, the majority of them can export multiple unrelated antibiotics such as macrolides and tetracyclines (Džidic et al 2008).

Spontaneous mutations can also manipulate the chromosomal genes of antimicrobial targeted regions. The altered binding interactions may shift the downstream reactions for effective antimicrobial activity. One of the well-described examples is an alteration in PBPs. Kapoor et al (2017) showed that mutagenized PBPs revealed a weak binding affinity to β-lactam antibiotics, which enabled *Enterococcus faecium* and *Streptococcus pneumoniae* to be resistant to ampicillin and penicillin, respectively. Moreover, in addition to four native PBPs, MRSA (Methicillin-resistant *Staphylococcus aureus*) produces PBP2a that is a unique transpeptidase. PBP2a is encoded by the gene *mecA* and is not effectively inactivated by currently available β-lactams (Fishovitz et al 2014). As a result, PBP2a enables MRSA to maintain cell-wall integrity and synthesis. In contrast, Gram-negative bacteria mainly rely on β-lactamase to combat antibiotics (Kapoor et al 2017). Together with the active dissemination rate among pathogens, β-lactamases are also recognized by a considerable level of stochastic mutations accumulated within the sequences of enzymes, potentially contributing to developing sub-variants. All in all, β-lactamase is another crucial source of antibiotic resistance for numerous pathogenic Gram-negative organisms.

1.4 Metallo-β-lactamases

1.4.1 General descriptions of β-lactamases

To date, there are two different classes of well-studied β-lactamases: serine-β-lactamases (SBLs) and metallo-β-lactamases (MBLs). Both SBLs and MBLs hydrolyze the β-lactam ring by utilizing the nucleophilic hydroxy group to attack carbonyl carbon and cleave the β-lactam ring, inactivating the antibiotics (Tooke et al 2019). A significant distinction between the two is that class A, C, and D β-lactamases belong to SBLs and hydrolyze a β-lactam ring using the conserved serine residue (Brem et al 2016; Tooke et al 2019). The deprotonated active site serine residue attacks the amide bond and forms a tetrahedral intermediate before the β-lactam ring is cleaved. Then, the transient acyl-enzyme intermediate is produced from the tetrahedral intermediate. At this stage, an activated water molecule induces the release of the product from the active site by attacking the acyl-enzyme intermediate. On the other hand, MBLs, class B β-lactamases, hydrolyze a β-lactam ring using a water molecule instead of a serine residue. As a result, Zn (II) ions at the catalytic site are essential for the effective catalytic activity of MBLs (Garcia-Saez et al 2008). Both SBL and MBL enzymes are located mainly in the periplasm (Pradel et al 2008), or the outer membrane vesicles of bacteria in the case of NDM-1 (González et al 2016). MBLs, which employ the SecA ATPases and the SecYEG system, may interact with chaperones such as SecB to maintain the unfolded state of MBL preproteins until they are delivered to the SecA ATPases (Mao et al 2009). In the periplasm, the enzymes harness available Zn (II) ions in the environment and properly fold into a functional structure (González et al 2012). In addition, it is determined that the unusual stability of NDM-1 is linked to lipidation and anchorage to the outer membrane of Gram-negative bacteria (González et al 2016).

MBLs are classified into B1, B2, and B3 subfamilies (Makena et al 2015). B1 MBLs such as VIM, NDM, and IMP are particularly raising clinical concerns because of active dissemination *via* HGT among diverse genera of microorganisms and broad specificity to diverse antibiotics (Gonzalez et al 2012). In previous studies, IMP- and VIM-variants are frequently identified from mobile gene cassettes inserted into class 1 integrons **(**Arakawa et al 1995; Lauretti et al 1999; Poirel et al 2000), which enable integration by a site-specific recombination mechanism (Fluit et al 1999). With β-lactamases variants on the rise, Martínez-García et al. established the maximum-likelihood (ML) tree that is focused specifically on nucleotide sequences of VIM β-lactamases. The researchers grouped VIM β-lactamases into four different layers, and VIM-2, the fourth group, exhibits highly diversified sub-variants 22 members (Figure 1.1A). Using the Network program by Fluxus Technology, they built the reconstruction displaying diversified VIM-1, VIM-4, and VIM-2 nodes that are created by single and non-synonymous mutations. As the VIM-2 phylogroup shows in Figure 1.1B, 20 out of 27 variants are derived from VIM-2. At present, the Comprehensive Antibiotic Resistance Database (CARD) reports 46, 29, and 54 sub-variants identified for VIM, NDM, and IMP orthologs, respectively. Based on CARD and Metallo-β-Lactamase Engineering Database (MBLED), the B1 MBL genes are predominantly disseminated among α-, β-, and especially γ-proteobacteria. For instance, VIM β-lactamase is most commonly identified in *Pseudomonas aeruginosa* followed by *Klebsiella pneumoniae* (Figure 1.2B), whereas NDM β-lactamase is mainly identified in *Escherichia coli* followed by *Klebsiella pneumoniae* (Figure 1.2A). IMP β-lactamase is also found most frequently in *Pseudomonas aeruginosa* followed by *Acinetobacter baumannii* (Figure 1.2C).

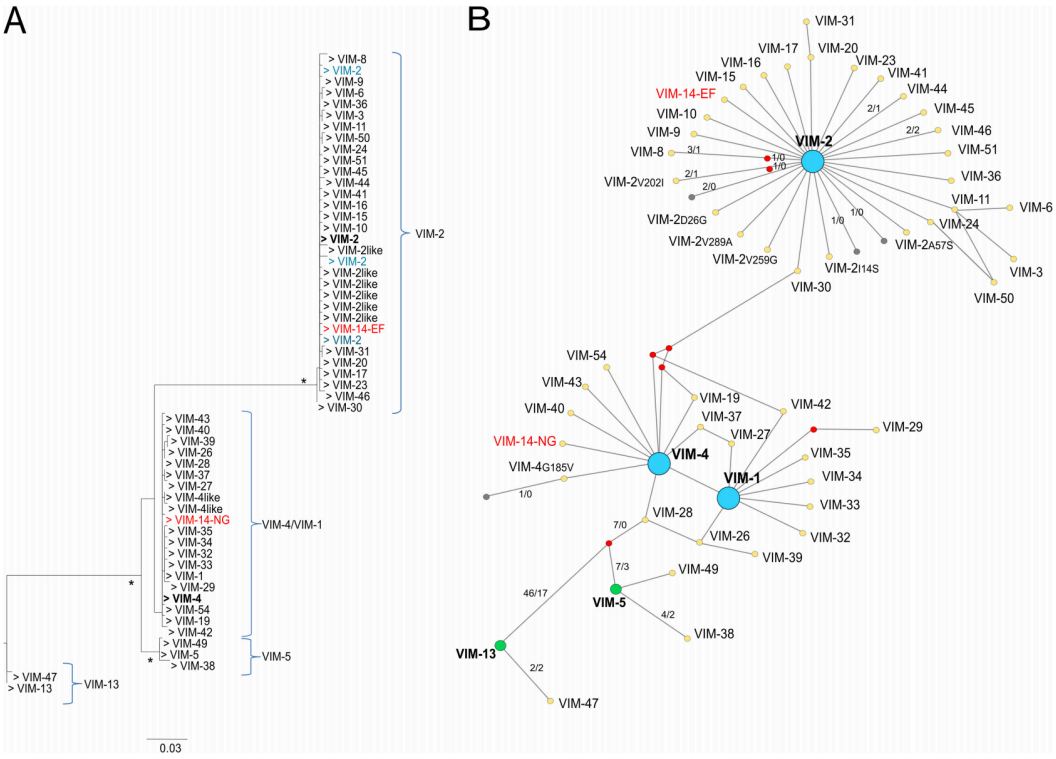


Figure 1.1 Maximum- likelihood (ML) and phylogenetic tree of blaVIM sequences analyzed by Martínez-García et al.

(A) The researchers inferred the ML phylogenetic tree using the PhyML 3.0 program. The blaVIM sequences without gaps and recombinant ones are selected for this network. They estimated the relevant nodes by 1,000 replicates of bootstrapping, and considered them statistically significant only when support values are over 90%, which is marked by asterisk. Among many, VIM-2, VIM-4/VIM-1, VIM-5, and VIM-13 are identified by assigning sequences that are found more than once. Reference sequences of VIM-2 and VIM-4 are represented in boldface. VIM variants highlighted in blue indicates sequences with single-nucleotide polymorphisms (SNPs) compared to the respective reference sequence. blaVIM-14 (EF055455 and NG050341) are considered as incongruent finding which is highlighted in red. These sequences are grouped to VIM-2 and VIM-4, respectively. (B) This graph is created using the Network 5.0.0.1 (Fluxus Technology) software. The four different colors represent i) the three main diversifications (blue dots), ii) other events of diversification (green dots), iii) hypothetical ancestors (red dots), and iv) variant with the number of nucleotide change/number of amino acid changes (gray dots). For those that are 1/1 change, no numbers are assigned.

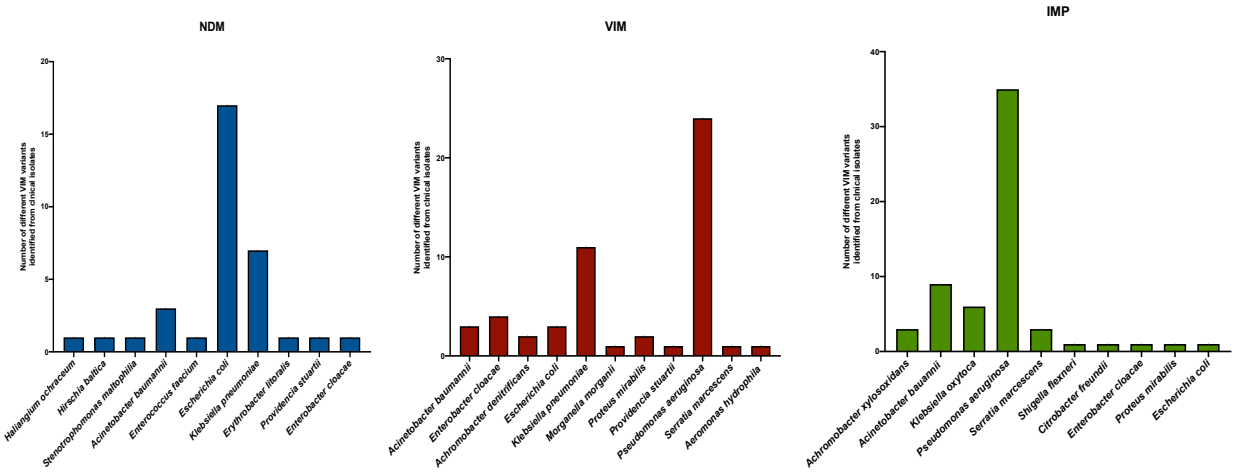


Figure 1.2 List of pathogenic organisms harboring VIM-, NDM-, and IMP-variants.
The histograms visualize data from Comprehensive Antibiotic Resistance Database (CARD) and Metallo-Beta-Lactamase Engineering Database (MBLED). Each MBL shows a bias in dissemination among different species of bacterial hosts. (*A*) Clinical isolates harboring NDM variants (Blue). The list of identified 10 pathogens harbors various types of NDM variants as shown on the y-axis. It is identified that *Escherichia coli* harbors 17 different types of NDM variants. (*B*) Clinical isolates harboring VIM variants (Red). List of 11 pathogens is displayed with bar graphs representing various VIM variants which is identified from the respective pathogens. *Pseudomonas aeruginosa* harbors the highest number of VIM variants, which is 24. (*C*) Clinical isolates harboring IMP variants (Green). List of 10 pathogenic organisms is presented in the x-axis. IMP variants are profoundly identified from Pseudomonas aeruginosa, harboring 35 different variants.

1.4.2 Metallo-β-lactamases' structure and mechanism

1.4.2.1 Catalytic domain and mechanism

Despite a low sequence identity, which can be less than 25%, (Mojica et al 2017), the 220 to 250 aa long catalytic domains of MBLs share the αβ/βα sandwich scaffold containing dinuclear or seldomly mononuclear Zn (II) ion at the active site (Makena et al 2015). The Zn (II) ion at the M1 site is in tetrahedral coordination stabilized by His116, His118, and His196, and another Zn (II) at the M2 site is coordinated by Asp120, Cys221, and His263 in a trigonal bipyramidal sphere (Meini et al 2015; Fabiane et al 1998; Concha et al 2000; Garcia-Saez et al 2008; Concha et al 1996; Nauton et al 2008; Zhang et al 2011). Bridging-residue-lacking MBLs substitute a water molecule as the bridging protein residue in the presence of Cys at position 221 (Bebrone et al 2007). Since metal ions like Zn (II) are electrophiles, the acid dissociation constant, pKa of the

bridging water molecule can be effectively lowered from the regular pKa. As a result, the hydroxy group of the water molecule is induced as an effective nucleophile. In addition, Zn (II) mediates β-lactam binding *via* electrostatic stabilization of opened ring anionic intermediate species of invariant β-lactam carboxylate moiety (Rasia et al 2004; Zhang et al 2011; Spencer et al 2005; Garau et al 2005) and activates the proton donor to protonate at less hindered α- or at C2 of β-lactam in the case of carbapenem (Lisa et al 2017). A highly conserved charged residue at position 224 interacts with the carboxylate moiety as well (Zhang et al 2011; Spencer et al 2005; Garau et al 2005). In general, β-lactam hydrolysis is mediated by hydroxide as a nucleophile to attack the carbonyl group of the β-lactam ring, resulting in C-N bond cleavage and the protonation of the nitrogen (Sally et al 2018). Among the two Zn (II) ions at the active site, Zn1 is assumed to play a significant role in lowering the pKa of water molecule because the distance between the Zn (II) ion and a water molecule is typically 1.9-2.0 Å (Figure 1.3A). According to a recent paper, carbapenem hydrolysis is carried out as the bridging water molecule is detached from the Zn (II) upon β-lactam binding (Tooke et al 2019). As illustrated in Figure 1.3B, the terminal hydroxide eventually becomes an effective nucleophile and attacks the carbonyl group of the β-lactam ring, while the negative charge intermediates are stabilized by interacting with dinuclear Zn (II) ions (Page et al 2008).

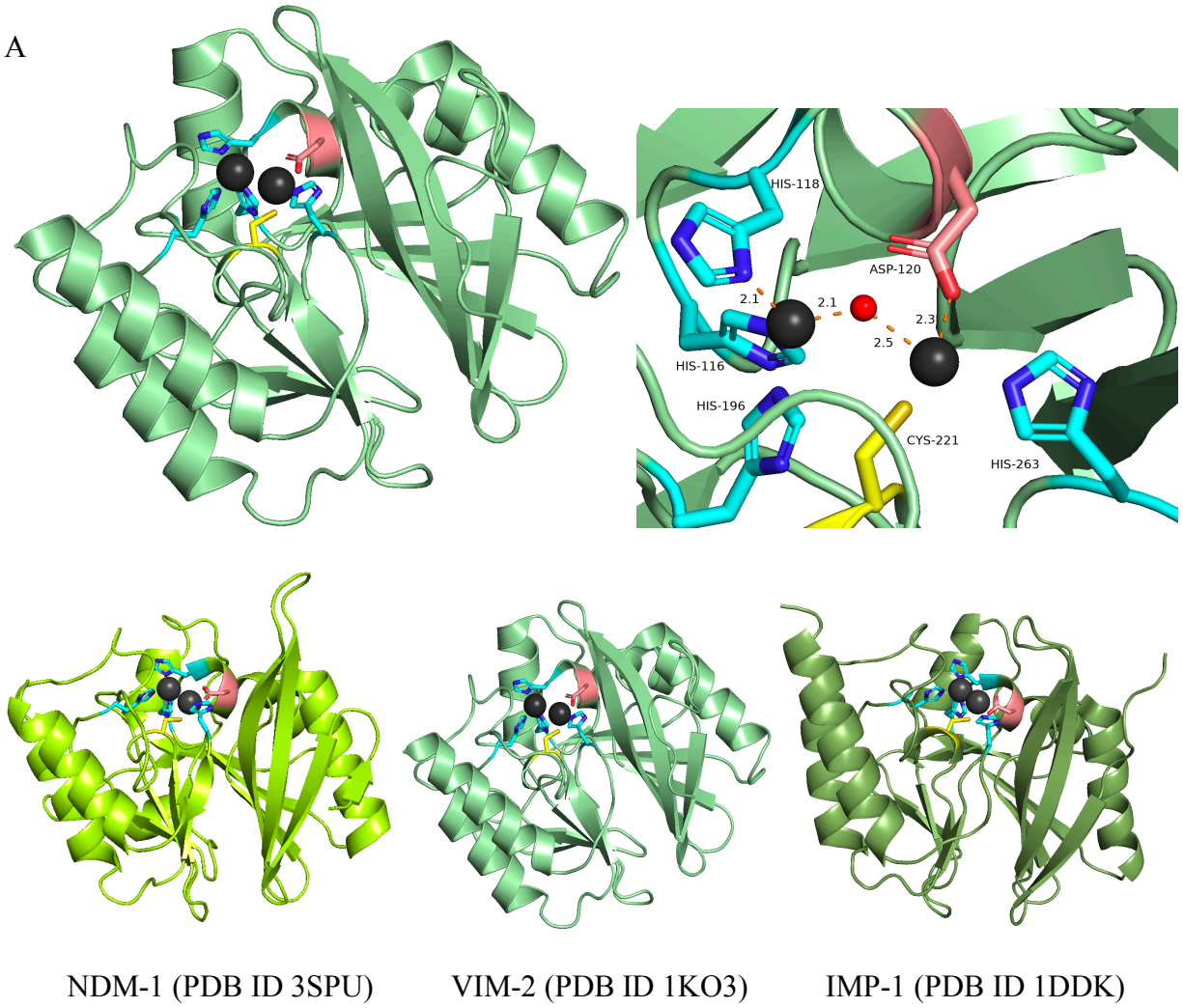


Figure 1.3. Structure of VIM-2 and reaction mechanism for penicillin hydrolysis by B1 enzymes.
(A) Structure of NDM-1 (PDB ID 3SPU), VIM-2 (PDB ID 1KO3), and IMP-1 (PBD ID 1DDK) showing dinuclear Zn (II) ions stabilized by H-H-H and D-C-H at the catalytic domain. General structure of MBLs is shown with dinuclear Zn (II) ions (dark grey) at the catalytic site. Histidine, cysteine, and aspartic acid are colored in blue, yellow, and salmon, respectively. The catalytic domain is enlarged to show electrostatic interactions formed between dinuclear Zn (II) ions as well as His 118, water (red), and Asp 120. The orange dashed lines depict distances

between the molecules. (B) Mechanism of hydrolysis of β-lactam substrate by dinuclear Zn (II) ions at the active site of CcrA. The bridging hydroxide ion is induced as a nucleophile by Zn (II) ions to attack the partially positive carbonyl carbon (Page et al 2008). A negatively charged intermediate is stabilized by the oxyanion hole of the enzyme. The water molecule bound to Zn (II) is oriented to donate a proton to the leaving nitrogen, and the newly formed hydroxide ion occupies the site again as the product dissociates from the active site (Page et al 2008).

1.4.2.2 Signal peptide

Signal hypothesis was first proposed in 1971 by Günter Blobel together with David Sabatini (Anderson et al 1999). Since then, many signal peptides of pre-proteins have been identified and studied in both eukaryotes and prokaryotes. A fundamental role of signal peptide in microorganisms is to sort and localize essential proteins to a periplasmic space or outer membranes *via* translocases, such as the SecYEG or the Tat machinery (Owji et al 2018; Natale et al 2008; Pradel et al 2008). Despite the common functionality, they share a low extent of sequence homology (Nesmeyanova et al 1997). In general, signal peptides can be characterized by i) a positively charged n-region, ii) a central hydrophobic h-region, and iii) a polar c-region as shown in Figure 1.4 and Table 1.1 (Owji et al 2018; Nesmeyanova et al 1997). Based on the signal and the hydrophobicity of the signal peptide (Nilsson et al., 2015), either the Sec- or Tat-secretory pathway is mediated to translocate mature peptides and to sustain the rate of growth (Owji et al 2018; Natale et al 2008; Pradel et al 2008). In the case of Gram-negative bacteria, the SecYEG pathway mainly conducted post-translational translocation of pre-proteins in their unfolded state in a SecA ATPases-dependent manner (Owji et al 2018; Nesmeyanova et al 1997). On the other hand, the Tat secretory pathway was largely observed in Gram-positive bacteria carrying mature and rapid folding proteins (Natale et al 2008).

Figure 1.4. Overview of signal peptide.
A schematic view of signal peptide showing three sub-regions: positively charged N-terminus (N- region, red), hydrophobic residues (H-region, yellow), and signal peptide cleavage site (C-regions, cyan). The N-region contains positively charged residues such as lysine and is normally one to five amino acid residues. The H-region consists of long hydrophobic residues such as leucine. The C-region contains polar and small amino acid residues which may make it easier for signal peptide peptidases (SPPs) to cleave the signal peptide after successful translocation. SPP: Signal peptide peptidase; Pro: Pro-region; mature peptide: protein segment exported to the periplasm or outer membrane in prokaryotic cells.

Table 1.1. General characteristics of signal peptides.

Signal Sequence Regions	*NH_2*	*Hydrophobic*	*COOH*
Charge	+	No charge	Polar
Length	1-5	7-15	3-7
Frequently observed amino acids	Arg, Lys	Leu	Ala, Val

1.4.2.2.1 N-region

In general, positively charged basic residues such as Lys were frequently identified in the n-region of signal peptides. Although the peptide length varies depending on the type of facilitated secretory pathways and organisms, the n-region consisted of five residues on average (Owji et al 2018). The functional roles of this region were i) to interact with negatively charged phosphate groups of the lipid bilayer (Owji et al 2018; Nesmeyanova et al 1997; Van et al 2000), ii) to interact with chaperones such as SecB or DnaK (Owji et al 2018; Batey et al 2000; Randall et al 2004), and iii) to form a complex with SecA ATPases (Karamanou et al 2007; Sianidis 2001). The charge of the n-region played another substantial role in promoting efficient mature peptide

translocation (Wang et al 2000). A consistent charge difference between the n- and c-region served to fix the n-region toward the cytosol and to make the c-region available for the active site of signal peptide peptidases (SPases) (Amaya et al 2015; Hartmann et al 1989; Paetzel et al 2002).

1.4.2.2.2 H-region

The hydrophobic region is normally composed of 7-15 residues, predominantly hydrophobic residues such as Leu (Rusch et al 1994; Chen et al 1996). Even though the global amino acid distribution seems random, the h-region was characterized by its conserved sets of amino acid arrays unique to species (Duffy et al 2010). In fact, hydrophobicity plays a significant role in determining the α-helical conformation of a signal peptide, orienting signal peptide cleavage residues at their position, defining a specific secretory pathway, and processing proteins (Owji et al 2018; Engelman et al 1981). Within the SecYEG channel, the hydrophobic-rich region adopts an α-helical secondary structure (Li et al 2016; Gelis et al 2007). The conformation helps reduce unfavorable free energy associated with polar or charged residues which induce a membrane partitioning by the hydrophobic effect (Engelman et al 1981). Moreover, the intensity of hydrophobicity of a signal peptide is an essential component that governs interactions with different types of chaperones such as SecB or SRP (Morán-Barrio et al 2009; Draycheva et al 2018; Petriman et al 2018). As polypeptide emerges from a ribosome, highly hydrophobic residues in a signal peptide interact with SRP, inhibiting the translation processes until the complex interacts with the SecYEG system (Draycheva et al 2018). A signal peptide with less hydrophobicity such as MBLs were translated in the cytoplasm and may interact with chaperones

such as SecB to maintain the unfolded structure of polypeptide until the peptide was properly directed to the SecYEG system (Mao et al 2009).

1.4.2.2.3 C-region

The c-region plays an important role in signal peptide cleavage. Normally, this region consists of 3 to 7 polar or neutral amino acid residues (Heijne 1990). According to a previous study, a length of five amino acids is considered optimum for the cleavage efficiency, and increasing the length of the c-region to more than nine amino acids would impair or abolish secretion (Suominem et al 1995). Most significantly, the cleavage efficiency at the c-region determined the protein secretion level, suggesting that it is a rate-limiting step (Geukens et al 2004). Although the canonical -3 and -1 rule or AxA motif is commonly accepted as a universal cleavage site for a signal peptide in prokaryotes (Heijne 1984; Paetzel et al., 2002). Payne et al. suggested that the novel SPaseI cleavage motifs were unique to each taxonomic group and correlate with a proteome-wide reduction in alanine use (Payne et al 2012). In previous studies, Gly, Ser, and Cys are highly tolerated in both -3 and -1 positions (Fikes et al 1990; Shen et al 1991; Karamyshev et al 1998), while Val, Leu, and Ile can also substitute Ala at position -3 due to a larger pocket size at position -3 (Choo et al 2005). However, big polar, aromatic, and charged amino acids are not commonly preferred at position -3 (Choo et al 2005)**.**

1.5 Mutation drives evolution

1.5.1 Directed evolution

Positive selection tends to increase allele frequency with advantageous phenotypes within a finite population, fostering adaptive evolution (Schaffner 2008). In this regime, the selective

advantage, s, characterizes the diverse reproductive success (fitness) or the frequency among alleles. The fixation frequency of beneficial mutants in a population relies on two important and yet subtle parameters: the population size and the mutation rate (Desai 2007; Kosheleva et al 2013; Silander et al 2007). The larger the population size and the stronger the mutation rate, the higher the probability that rare beneficial mutations will emerge, compete, and fix (Desai 2007; Kosheleva et al 2013). Fixation of non-synonymous mutations is one of the characteristics of positive selection and is linked to advantageous fitness effects (Desai 2007). Eventually, they will get edited by selection instead of random genetic drift (Desai 2007). If the fixation frequency of rare beneficial mutations is low in a given generation time in the distribution of fitness effects, a traveling wave-like selective sweep may occur (Desai 2007). While a phenotypically advantageous clone with beneficial mutations promotes its allele frequency, neutral and deleterious mutations may hitchhike in genomic linkages which are commonly detected in microbial organisms (Desai 2007). The extent of hitchhike effect is dependent on the effects of beneficial mutations and the quality of genetic background (Desai 2007).

In nature, MBLs are actively shared among phylogenetically diverged bacterial hosts. As part of an adaptive mechanism, MBLs acquire mutations, evolve, and adapt to a new host under a positive selection. The mutations may have a connection to improve the level of gene expression within a new host. It is also possible that the mutations may be associated with re-arranging an electrostatic network among altered amino acids due to non-synonymous mutations and influence the structural stability of the enzyme, which eventually impacts the enzymatic activity. The MBL mutants with an advantageous phenotype may confer antibiotic resistance to improve the fitness of its host organism.

1.5.2 Directed evolution experiment

In 1993, the first directed evolution experiment was conducted using subtilisin E to tune the catalytic activity (Chen et al 1993; Cobb et al 2013). Since then, the directed evolution experiment is largely practiced in laboratory settings to study enzymes conferring antibiotic resistance because mutational patterns associated with respective phenotypes can be addressed under controlled conditions. Indeed, the TEM-1 β-lactamase strains with improved antibiotic resistance were identified from the directed evolution (Backus & Stauffer 1955; Vournakis & Elander 1983). The enzyme gained antibiotic resistance by 32,000-fold using the DNA shuffling method (Stemmer 1994) and 16-fold using error-prone PCR and cassette mutagenesis (Stemmer 1994; Palzkill et al 1992). Surprisingly, the TEM-1 β-lactamase that evolved by the directed evolution technique with cefotaxime revealed an E104K/M182T/G238S combination equivalent to the clinical isolate TEM-52 (Orencia et al 2001). These discoveries revealed the potential evolvability of TEM-1 β-lactamases and the effectiveness of the technique.

Previously in the Tokuriki lab, the clinically isolated VIM-2, NDM-1, and IMP-1 sequences were used as the wild-type and were subjected to 18 rounds of directed evolution by a previous Ph.D. student, Dr. Socha. Figure 1.5 illustrates six major steps required to complete a single round of the experiment. Wild-type MBL gene is mutagenized by the error-prone PCR technique to introduce random mutations. Then, the mutagenized genes are amplified and digested by respective restriction enzymes. The digested target DNA is sub-cloned into a new vector to reduce possible noise from any mutations encoded in the plasmid DNA during the error-prone PCR step. Therefore, at this point, it is assumed that antibiotic resistance increases are solely dependent on the mutational effects of the MBL genes. The plasmid DNA is transformed into

Ecloni10G cells, and the recovered cells are spread on a LB agar media supplemented by a desired concentration of ampicillin, a selection pressure, combined with a respective selective marker. The ampicillin concentration is increased by two-fold at every round. The plates are incubated at 37 °C overnight. The next day, colonies that can survive at the new higher concentration of ampicillin are collected to extract plasmid DNAs from the *E. coli* cells. Then, another round of directed evolution experiment is performed.

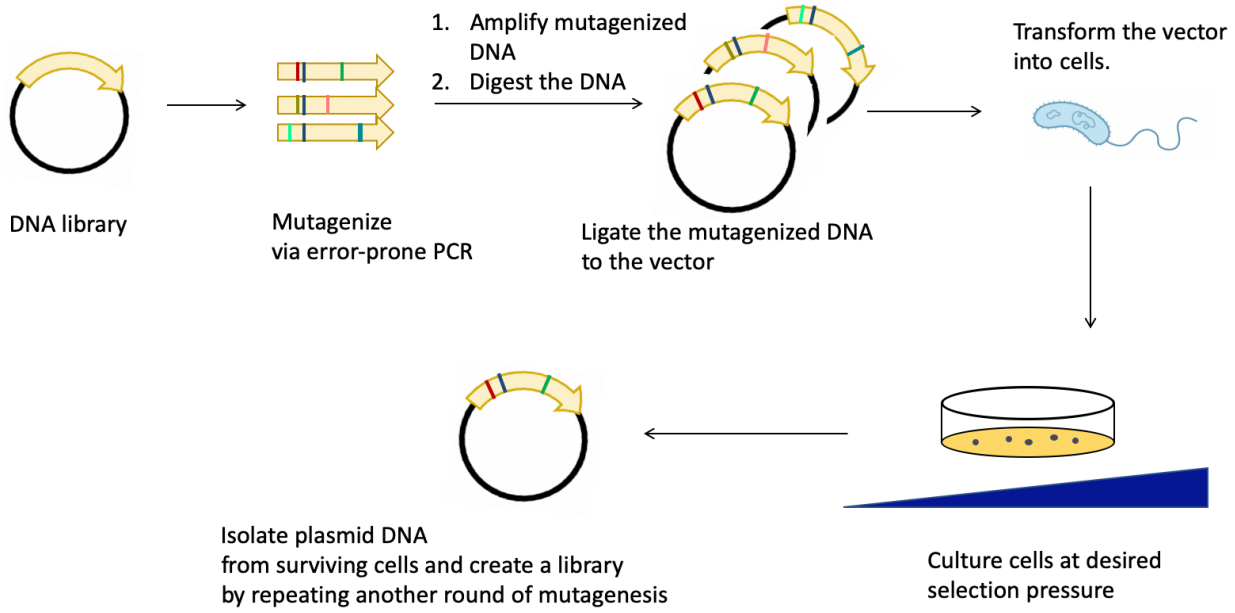


Figure 1.5. Overview of directed evolution experiment.
A directed evolution experiment starts with a known sequence of interest (Top left). The mutagenized MBL directed evolution libraries are created by performing these steps.

1.6 Preliminary data description

Previously, Dr. Socha performed 18 rounds of directed evolution experiment on VIM-2, NDM-1, and IMP-1 using ampicillin. He selected round 3, 6, 9, and 18 of each MBL trajectory and measured the non-synonymous frequency per residue per round for the signal peptide and the catalytic domain separately. He discovered that the non-synonymous mutation frequency per

residue per round was 3- to 7-fold higher in the signal peptides than the mutation frequency observed in the catalytic domains at round 3 from the orthologs (Figure 1.6).

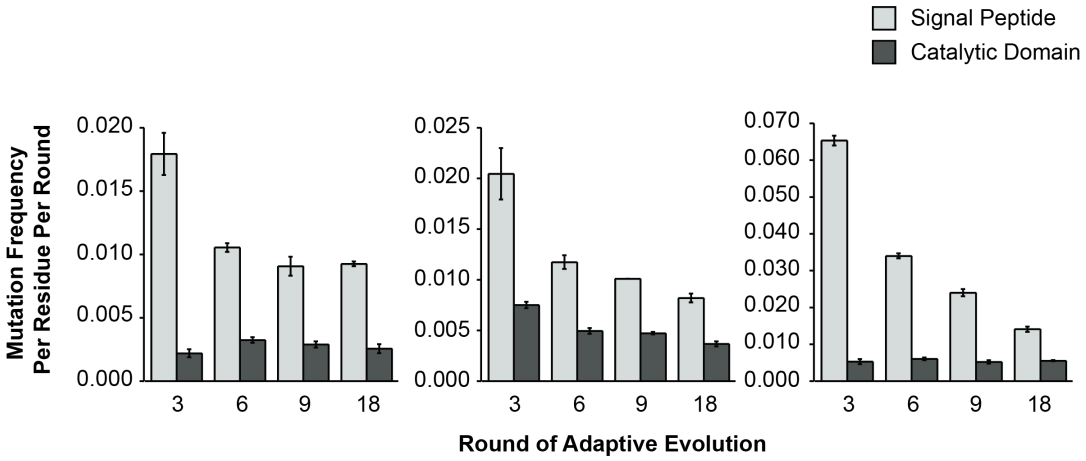


Figure 1.6. The frequency of nonsynonymous mutations per residue, per round, in the signal peptide and catalytic domain regions of the NDM-1, VIM-2, and IMP-1 ampicillin trajectories.
Each bar represents the average mutation frequency of 6 to 10 sequenced variants from round 3, 6, 9, and 18 of the ampicillin trajectories. The signal peptide was defined as the region before the first secondary structure element in the crystal structure of each enzyme, whereas the catalytic domain is defined as the remainder (Socha et al 2019).

The difference in the amount of increase in frequency between the signal peptide and the catalytic domain gradually becomes comparable toward the round 18 trajectory. In contrast, the non-synonymous mutation frequency per residue per round remained almost constant in the catalytic domain at the four rounds for all orthologs. After the mutation acquisition, NDM-1 and VIM-2 exhibited a modest improvement in catalytic efficiency. Both enzymes enhanced the turnover rate and reduced the binding affinity between the enzymes and substrates, whereas IMP-1 showed the opposite trend of decreasing turnover rate and increasing binding affinity between the enzyme and substrates. However, the catalytic efficiency alone is insufficient to explain the dramatic improvement in minimum inhibitory concentration (MIC) (Figure 1.7).

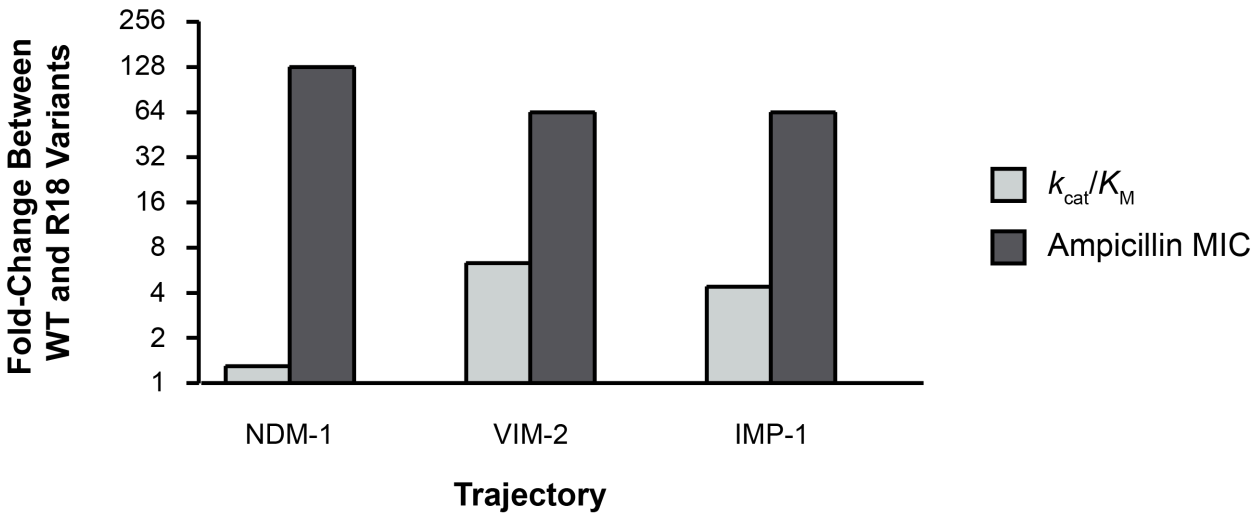


Figure 1.7. A comparison between the improvement in catalytic efficiency and ampicillin resistance over the NDM-1, VIM-2, and IMP-1 ampicillin trajectories.
NDM-1, NA, VIM-2, VA, IMP-1, and IA were used to determine the fold-change in catalytic efficiency and ampicillin MIC over the 18 rounds of their respective trajectories (Socha et al 2019).

He determined a positive correlation between the elevated amount of functional periplasmic enzyme and MIC improvement (Figure 1.8). In addition, the functional periplasmic enzyme activity was shown to increase with the enzyme activity in the cytoplasm (Figure 1.9), implicitly indicating the significant contribution of gene expression elevation to MIC. These lines of evidence led him to conclude that improved antibiotic resistance is through a combination of the catalytic activity of each enzyme and increased functional periplasmic enzyme expression. This discovery showed that, the adaptation of NDM-1 and IMP-1 extensively depends on the functional periplasmic enzyme production, while VIM-2 is the combination of both enzyme production and catalytic activity.

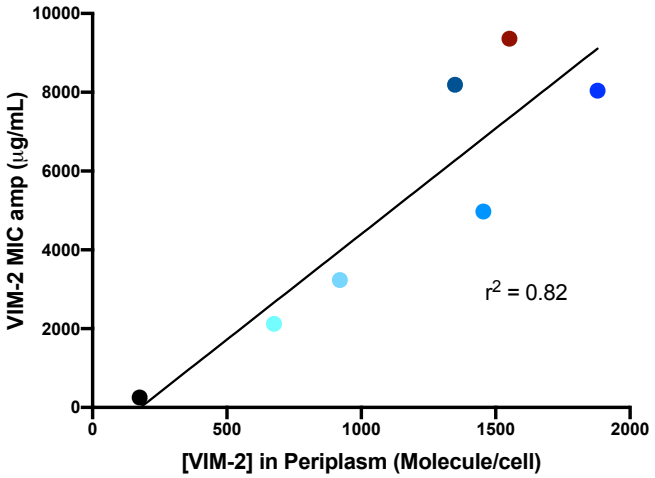


Figure 1.8. Scatter plot showing a correlation between VIM-2 expression in the periplasmic space and MIC. The amount of VIM-2 in the periplasmic space is obtained from round 3 (light blue), 6 (sky blue), 9 (aqua blue), 12 (blue), 15 (dark blue), and 18 (red) in the previous experiment by Dr. Socha. The determined amount of VIM-2 is averaged and compared with MIC (μg/mL). This plot suggests a considerable connection between VIM-2 expression in the periplasmic space and antibiotic resistance. The black dot represents wild-type VIM-2. The positive correlation is estimated by Pearson correlation coefficient.

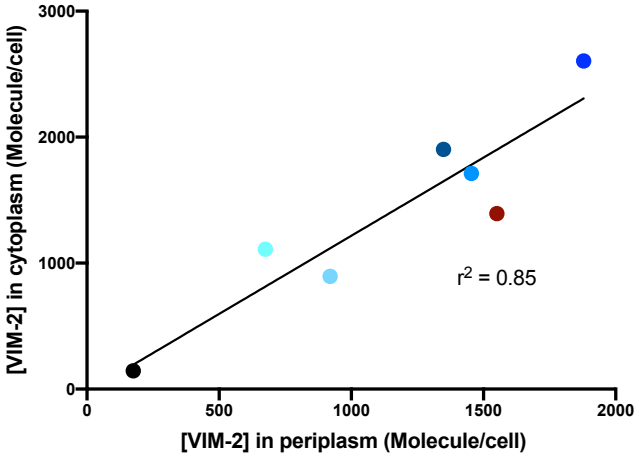


Figure 1.9. Scatter plot showing a correlation between VIM-2 expression in the cytoplasm and the periplasmic space.
As described in Figure 6, the amount of VIM-2 in the cytoplasm and periplasmic space is obtained from round 3 (light blue), 6 (sky blue), 9 (aqua blue), 12 (blue), 15 (dark blue), and 18 (red) in the previous experiment by Dr. Socha. The black dot represents wild-type VIM-2. This plot suggests that increased VIM-2 expression in the periplasmic space is partly dependent on the level of VIM-2 expression in the cytoplasm. The positive correlation is estimated by Pearson correlation coefficient.

Furthermore, the replacement of native signal peptides of round 18 VIM-2, IMP-1, and NDM-1 with PelB, the 22 N-terminal signal peptide of pectate lyase B of *Erwinia carotovora* CE,

resulted in a two-fold reduction in antibiotic resistance of VIM-2, and complete abolishment of antibiotic resistance for NDM-1 and IMP-1 in *Escherichia coli* (Figure 1.10). Overall, the lines of evidence support the hypothesis that mutations in signal peptides of MBLs contribute to optimizing enzyme expression, potentially by overcoming functional constraints derived from phylogenetic distances across bacterial strains to enhance antibiotic resistance.

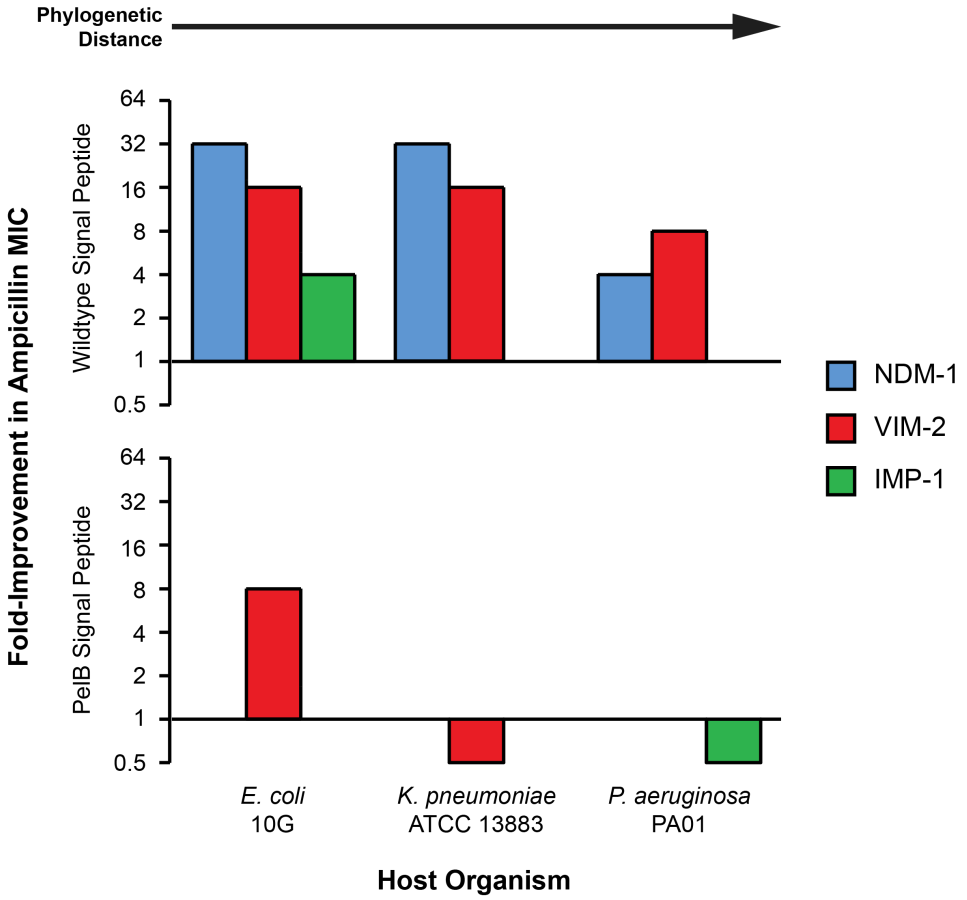


Figure 1.10. The fold-improvement between the evolved round-18 variants and wild type enzymes for each trajectory in three organisms.
(A) The improvement in ampicillin MIC for the three MBL ampicillin trajectories was determined in E. coli 10G, K. pneumoniae ATCC 13383, and P. aeruginosa PA01. An arrow displays the order in phylogenetic distance from E. coli. (B) The improvement in ampicillin MIC for the three trajectories with their native signal peptides replaced by the PelB leader sequence was determined in the three organisms (Socha et al 2019).

1.7 Thesis objective and outline

Global dissemination of the metallo-β-lactamases (MBLs) enables numerous pathogens to gain β-lactam antibiotic resistance and threatens the world with the rise of multi-drug resistant infections. Many of clinical isolates with MBLs belong to proteobacteria, showing a sign of biases in the host distribution. In the midst of antibiotic crisis, it is crucial to have a better understanding of how MBLs adapt in a new host and drive active dissemination in phylogenetically related pathogens. The aim of this thesis is to explore the role of mutations in the signal peptides of MBLs in developing adaptive (advantageous) phenotypes in *E. coli.* As a means of investigation, this thesis also aims to synthesize a new broad host range vector by ligating three individual pieces of plasmid DNA.

Chapter 2 explains the methodology of the dose-response curve assay. The method provides two parameters: the Hill coefficient and the IC_{50} values. Based on these parameters, the average fitness of mutant variants is estimated at the population level.

Chapter 3 focuses on mutations in the signal peptides of NDM-1, VIM-2, and IMP-1 variants evolved in *E. coli* during 18 rounds of directed evolution against ampicillin using the dose-response curve assay and bioinformatics analysis.

Chapter 4 describes how a new broad host range vector is synthesized using three separate plasmid DNA fragments. I identify that the newly synthesized vector is compatible with *Acinetobacter baumannii, Escherichia coli, Pseudomonas aeruginosa,* and *Pseudomonas putida*, enabling future research.

Chapter 2: Dose-response curve assay methodology

2.1 Survey of available methods for antibiotic analysis

To date, several standardized approaches are widely practiced to validate the susceptibility of bacteria to the currently available antibiotic drugs, such as disk diffusion and the dilution method. The disk diffusion technique, also known as Bauer and Kirby's experiment (Bauer et al 1966), uses standardized suspension plated on the agar plate with an antibiotic-containing paper. During the overnight incubation, the antibiotics diffuse through the agar and inhibit bacterial growth. The effective antibiotic concentrations can be estimated based on the formation of inhibition zone (Reller et al 2009; Guardino et al 2005). To estimate the zone of inhibition in diameter, the clear edge to edge distance across the center of the antibiotic disk is measured by a ruler. Then, the result is analyzed based on the bacterial species, the type of antibiotic used, and the zone of inhibition in millimeters using given charts (Hudzicki 2009). The measurements of Bauer and Kirby's experiment are reported only as one of three labels: susceptible, intermediate, or resistant.

The dilution method is another microbiological approach that helps identify the bacterial population in suspension (Guardino et al 2005). This method can be performed by two basic types: macro-dilution and micro-dilution, using either broth or agar as medium. In this method, antibiotics are consecutively diluted two-fold into micro-centrifuge tubes or 96-well microtiter plates. A dose-response curve assay is a type of micro-dilution method in which standardized pre-inoculated culture is mixed with a series of exponentially diluted antibiotics in 96-well microtiter plates. Then, the antibiotic resistance of the bacterial population is measured by checking the turbidity after incubation. The same technique can be conducted using an agar

medium which also provides normalized minimum inhibitory concentration (MIC) of the bacteria (Khan et al 2019; Reller et al 2009; Guardino et al 2005; Bauer et al 1966; Wheat 2001; Tang et al 2012). Although both approaches yield standardized MIC results, broth dilution can render antibacterial activity at the population level (Khan et al 2019).

2.2 Dose-response curve assay

MIC was used to perform a directed evolution experiment by Dr. Socha in the Tokuriki lab. He used Luria-Bertani (LB) agar medium to prepare an array of ampicillin dilutions at two-fold increments into plates. Then, the antibiotic resistance of *E. coli* cells harboring mutagenized MBLs is determined based on the corresponding ampicillin concentration. This analysis is a practical method to approximate the resistance of individual strain and to identify a unique genotypic population. However, this approach only allows to check the concentration that will inhibit 90% of cell growth, IC_{90}, and it is difficult to estimate how fitness varies among mutants in the population. A dose-response curve assay resolves the issue by providing two important parameters: the Hill coefficient and IC_{50} value. The Hill coefficient determines the shape of a dose response curve, meaning that it allows estimation of diversified MBL mutants linked to various fitness effects within a population by providing an approximated range of phenotypic diversity. Because the IC_{50} value represents the average fitness of an evolving population, it allows identification of the extent of improved antibiotic resistance at each trajectory. Hence, the fitness of *E. coli* cells harboring MBL mutants can be analyzed at the population level. Furthermore, equilibrated dose-response curves let us compute the IC_{10}, IC_{50}, and IC_{90} values of the clonal population using the equation, $IC_x = \left(\frac{x}{100-x}\right)^{\sqrt{\text{Hill coefficient}}} \cdot IC_{50}$, where x is the

desired level in the distribution. Hence, the heterogeneity of fitness effects related to their distinct mutational trait, a phenotype diversity, can also be estimated.

2.3 Determination of antibiotic resistance of population by dose-response curve analysis

The dose-response curve measured in the experiment is sigmoidal in shape. The slope of the curves can be inferred by fitting the Hill equation (1), which is a non-linear logistic equation with four-parameters. The IC_{50} and the Hill coefficient values are approximated by Prism 9. The minimum and maximum asymptotes are inferred by using the raw data in excel. *Y* is the expected response at dosage *X*, *a* is the baseline response (the minimum asymptote) at 0 dosages, in (1), *b* is the stabilized response at infinite dosage (the maximum asymptote), *c* is the dosage at which 50% of the subjects are expected to show the desired response, and *d* is the Hill coefficient. When the response falls with increasing dosage, the Hill coefficient is a negative value.

$$Y = a + \frac{(b-a)}{\left[1 + \left(\frac{C}{X}\right)^{d}\right]} \quad (1)$$

In this work, the curve shape is uniformly set by the potency of β-lactam drugs and the activity of MBLs. As a result, a shift in IC_{50} values to the respective wild-type helps estimate the average antibiotic resistance of *E. coli* cells harboring mutagenized MBL variants in a population. The Hill coefficient describes the slope of the response curve at IC_{50} and is determined by the fitness of *E. coli* cells with MBL mutants in a given range of β-lactam drug concentrations. If the Hill coefficient is reduced, the growth inhibition induced by β-lactam drug is less potent, and a higher fraction of clones with different fitness effects survive in the population. Because of the highly positive correlation i) between $\log_2$ MIC and $\log_2 IC_{50}$ (Figure 2.1A) and ii) between

variations observed in MIC (Coefficient of variance of MIC) and the Hill coefficient (1/|Hill coefficient|) (Figure 2.1B), the dose-response curve assay can roughly capture the antibiotic resistance of diverse resistance strains in the population. This approach reveals that the Hill coefficient represents the variability of antibiotic resistance in the variant's population.

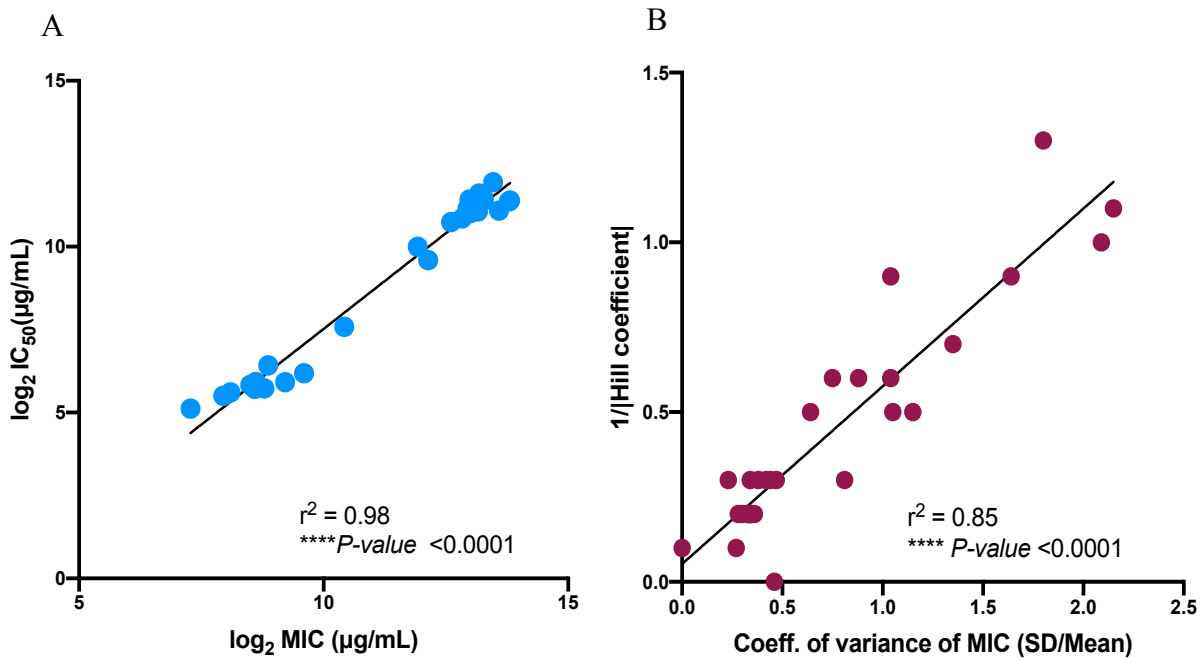


Figure 2.1. Relationship between log_2 MIC and log_2 IC_{50} of VIM-2 variants.
(A) The scatter plot shows a strong positive correlation between antibiotic resistance of individual variants determined by log_2 MIC and the antibiotic resistance of the variants' respective population by log_2 IC_{50}. The data points were collected from round 1, 2, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 of T10 and T1000 neutral drift trajectories together with round 1, 2, 3, 6, 9, 12, 15, 18 of directed evolution trajectories. log_2 MIC values were acquired from a colleague in the Tokuriki lab, and log_2 IC_{50} values were extracted from the dose-response curve assays that I performed. (B) The scatter plot shows a positive correlation between the coefficient of variance of individual variants measured in MIC (the x-axis) and the reciprocal of the absolute value of Hill coefficient (the y-axis). This suggests that populations with steeper Hill curves have a lower variability in phenotypes. These two results show that phenotypes of individual variants measured in MIC and the mean phenotypes of variants' population represented in IC_{50} are highly interconnected to each other. The Pearson correlation coefficient is used to estimate the association between MIC and IC_{50} values of the VIM-2 variants collected from neutral drift and directed evolution experiments.

Chapter 3: Mutations in the signal peptide of VIM-2, NDM-1, and IMP-1 may induce advantageous phenotypes in *Escherichia coli*

3.1 Introduction

MBLs are disseminated in diverse pathogenic bacteria *via* HGT and potentially encounter multiple challenges due to the changed environment of their new bacterial hosts, such as variations in codon usage among different species (López et al 2019; Iriarte et al 2013; Klumpp et al 2012) leading to lower overall resistance. Over time, MBLs continuously acquire naturally occurring stochastic mutations, with the majority of mutations having deleterious effects and being purged by purifying selection (Chen et al 2020). However, some mutations can be neutral or beneficial, and can eventually be fixed within the MBL sequence. Beneficial, and often non-synonymous mutations, with advantageous phenotypes may increase their frequencies and ultimately confer high antibiotic resistance to a new host, owing to the effect of a positive selection. Positive selection is an essential evolutionary process that improves the fitness of a population by increasing the frequency of alleles with advantageous phenotypes (Schaffner et al 2008; Desai et al 2007). Elevation in adaptive allele frequency during adaptive evolution may reveal mutation-driven sequence convergence due to a biased subgroup of beneficial mutations (Storz 2016). Patterns of converged sequence may represent inherited biases in creating mutational variations because a tendency of mutational development can promote a chance of similar features to evolve in different lineages (Storz 2016). In the preliminary study, a subset of mutations identified from signal peptides of VIM-2, NDM-1, and IMP-1 were non-synonymous mutations repeatedly fixed and propagated at identical positions through descendent populations. In addition, a sign of mutation-driven sequence convergence was observed in signal peptides of

the orthologs, as the enzymes enhanced antibiotic resistance to ampicillin during the directed evolution experiment, implying a potential association with advantageous phenotypes in *E. coli*. Hence, the uninvestigated role of mutations in signal peptides of the three MBLs is analyzed in this chapter.

Previously in the Tokuriki lab, the clinically identified VIM-2, NDM-1, and IMP-1 sequences were subjected to the directed (adaptive) evolution experiment by challenging the orthologs with two-fold increasing concentrations of ampicillin by MIC assay using *E. coli.* Interestingly, a subset of mutations identified from the signal peptides of VIM-2, NDM-1, and IMP-1 during the directed evolution experiment are non-synonymous mutations repeatedly fixed and propagated at identical positions through descendent populations. The fixation of these signal peptide mutations also occurs in tandem with the enzymes' enhanced resistance to ampicillin, suggesting these mutations may be responsible for advantageous phenotypes in *E. coli*. A number of observations support the importance of signal peptide mutations in host adaptation.

First, when the adapted MBL evolved in round 18 variants of *E. coli* was inoculated in *P. aeruginosa* and *K. pneumoniae*, the increase in MIC from that of wild-type was generally lower than what was measured in *E. coli* (Figure 1.10, Top). Second, the replacement of native signal peptide with PelB, the 22 N-terminal signal peptide of pectate lyase B of *Erwinia carotovora* CE, abolished enhanced antibiotic resistance of NDM-1 and IMP-1 in *E. coli* and reduced antibiotic resistance to a lower level for VIM-2 which showed an increase in catalytic activity unlike the others (Figure 1.10, Bottom). Third, the substitution of the PelB signal peptide for eight MBL orthologs resulted in antibiotic resistance of *E. coli* cells proportional to their

catalytic activity of enzymes, which is a noticeable difference from the eight orthologs' original antibiotic resistance to ampicillin. (Socha et al 2018). Fourth, improvement in catalytic efficiency alone was insufficient to explain the dramatic MIC increase, suggesting other factors such as expression may also play a role (Figure 1.7). Thus, there is a variety of evidence to support that a portion of MBL function, and subsequently adaptation, is dependent on the signal peptide. However, at present, the functionality of these mutations in the signal peptide is still obscure.

Although sequence profiles acquired from Dr. Socha help identify mutational patterns in the signal peptide, it may be a challenge to make a logical connection to the clonal population due to the low sample size. To confirm that these mutations are reflective of the variant population, I performed a series of dose-response curve assays using the entire population of the identical trajectories evolved and analyzed by MIC. Subsequently, I conducted a bioinformatic analysis to unveil a potential mechanism that improved MBLs' gene expression. Bioinformatic analysis suggests that the mutations are responsible for the altered mRNA folding energy, translation rate, translation initiation rate, and ribosome density, showing a moderate positive correlation with the enhanced MIC for NDM-1, VIM-2, and IMP-1. Overall, based on the previous discoveries and the positive correlation determined from the dose-response curve assay, these mutations may be attributed to codon usage of MBL genes assimilating to the codon usage of endogenous *E. coli* genes which reduces potential functional constraints derived by phylogenetic distances. To that end, the mutations in the signal peptide may be moderately accountable for the considerable improvement in antibiotic resistance by elevating MBL gene expression in *E. coli*.

3.2 Materials and methods

3.2.1 Creation of mutagenized libraries for directed evolution trajectory (Preliminary experiment performed by Dr. Socha)

Randomly mutagenized libraries of NDM-1, VIM-2, and IMP-1 were created by error-prone PCR with the nucleotide analogues, 8-oxo-2'-deoxyguanosine-5'-triphosphate (8-oxo-dGTP) and 2'-deoxy-P-nucleoside-5'-triphosphate (dPTP) (TriLink). For each library, two independent PCRs were used to ensure a balanced mutation rate with only one of the analogues featured in each reaction. Each 25 μL PCR consisted of 1 x GoTaq Buffer (Promega), 3 μM MgCl2, 0.1 μM of each primer, 0.2 mM of dNTPs, 1.25 U of GoTaq DNA polymerase (Promega), 1 ng of template plasmid, and either 100 μM of 8-oxo-dGTP or 1 μM of dPTP. The 'mutagenesis' PCR was set-up and ran as follows: an initial denaturation (95°C for 2 minutes), followed by 20 cycles of denaturation (95°C for 30 seconds), annealing (58°C for 60 seconds), and extension (72°C for 60 seconds), before a final extension step (72°C for 3 minutes). The PCR products were subsequently purified with the EZNA®Cycle Pure Kit (Omega Bio-tek), quantified, and used as template in a combined second 'amplification' PCR. Each 50 μL 'amplification' PCR consisted of 1 x GoTaq Buffer (Promega), 3 μM MgCl2, 0.1 μM of each primer, 0.25 mM of dNTPs, 1.25 U of GoTaq DNA polymerase (Promega), and 5 ng of each PCR product from the two previousseparate reactions. The reaction was run as before with the exception of an increase in the number of cycles from 20 to 35. Each product was purified again with the EZNA®Cycle Pure Kit, treated with *DpnI* (Fermentas) to remove the original supercoiled plasmid template, *NcoI* (Fermentas), and either and *HindIII* (Fermentas), *XhoI* (Fermentas), or *EcoRI* (Fermentas) for NDM-1, VIM-2, and IMP-1 respectively, in an 1 hour incubation at 37°C to prepare the PCR products for cloning.

Three versions of the pIDR-5.1 plasmid were prepared for use as the vector in this experiment. The selective resistance marker in the plasmid was replaced with kanamycin resistance for use with NDM-1, chloramphenicol resistance for VIM-2, and tetracycline resistance for IMP-1. Each plasmid was then prepared with the restriction enzymes flanking each of their corresponding genes (*NcoI* and *HindIII* for NDM-1, *NcoI* and *XhoI* for VIM-2, and *NcoI* and *EcoRI* for IMP-1). Each pIDR-5.1 plasmid was concurrently prepared by incubation with their set of restriction enzymes for 3 hours at 37°C for use as the vector in this experiment. While the digested plasmid was subsequently isolated via gel electrophoresis, the digested PCR products were again purified with the EZNA®Cycle Pure Kit before ligation with the prepared vector. Each 10 µL ligation consisted of 1 Å~ T4 DNA ligase buffer (ThermoFisher), 5 U of T4 DNA Ligase (ThermoFisher), 10 ng of prepared vector, and 30-40 ng of prepared mutagenized insert, before incubation at room temperature for an 1 hour. The ligations were then purified with a micro Elute Kit (ENZA) and eluted in 20 µL of water.

3.2.2 Application of selection (Performed by Dr. Socha)

The mutagenized libraries were electroporated with E. cloni® 10G E. coli cells (Lucigen), suspended in 1 mL of LB media, and allowed to recover for an 1 hour at 37°C. They were then grown overnight at 30°C in 10 mL of LB media supplemented with each trajectory's selective antibiotic (40 µg/mL of kanamycin for NDM-1, 34 µg/mL of chloramphenicol for VIM-2, and 15 µg/mL of tetracycline for IMP-1). The next day, a 1:100 dilution was made of each overnight culture in LB media and 100 µL was plated on each selection plate. Each trajectory was plated onto a series of LB agar plates containing 2-fold increases in the concentration of ampicillin from 2 to 8192 µg/mL, in addition to the constant selective antibiotic. The plate with the highest

concentration of ampicillin that had between 100 and 1000 colonies was collected. The plasmids were then extracted and used as the template for the next round.

3.2.3 Measuring the minimum inhibitory concentration of selected variants (Performed by Dr. Socha)

To assess the fitness of individual variants within the trajectories, single colonies were obtained from selected rounds and grown in a deep-96-well plate overnight at 30°C in 500 μL of LB media with 40 μg/mL of kanamycin for the NDM-1 trajectory, 34 μg/mL of chloramphenicol for the VIM-2 trajectory, and 15 μg/mL of tetracycline for IMP-1 trajectory. The next day, 5 μL of overnight culture was used to inoculate 195 μL of LB media with each trajectory's corresponding selective resistance in quadruplicate in a 96-well plate and grown for 3 hours at 37°C. The cultures were then plated with 96-well replicator pins on a series of LB agar plates with increasing levels of antibiotics (two-fold increases in cefotaxime, meropenem and ampicillin from 0.032 to 4096 μg/mL, 0.016 to 64 μg/mL, and 2 to 32768 μg/mL respectively). The agar plates were subsequently incubated overnight at 37°C. The next day, the minimum inhibitory concentration (MIC) was determined by identifying the concentration of antibiotics by which no growth was observed in at least three of the four replicates for each variant.

3.2.4 Sanger sequencing of selected variants (Performed by Dr. Socha)

Between 5-10 variants from selected rounds of each trajectory were isolated and sent for Sanger sequencing (Genewiz). The sequence results were visually inspected in Geneious® (8.1.9) and the mutations were identified.

3.2.5 Stock preparation for dose-response curve assay (Performed in this study)

The eight mutagenized libraries, the negative control, and the respective wild-type MBL genes used in the dose-response curve assay are created by Dr. Socha as described in "Creation of mutagenized libraries for directed evolution trajectory". The mutagenized libraries from the rounds 1, 2, 3, 6, 9, 12, 15, and 18 trajectories of VIM-2, NDM-1, and IMP-1 were individually electroporated with E. cloni® 10G *E. coli* cells (Lucigen). The *E. coli* cells were suspended in 1mL of Luria-Bertani (LB) media and recovered for an 1 hour at 37°C. Negative control (the inactive wild-type gene) and wild-type MBL genes were prepared by the same method. After the recovery, the mutagenized trajectories, the wild-type, and the negative control were mixed with glycerol stock (25% final glycerol concentration) and stored at -80°C.

3.2.6 Performing dose-response curve assay (Performed in this study)

The frozen stock of i) the negative control, ii) the wild-type, and iii) the mutagenized libraries of VIM-2, NDM-1, and IMP-1 were inoculated to 3 mL of LB media supplemented with a respective selective antibiotic (40 μg/mL of kanamycin for NDM-1, 34 μg/mL of chloramphenicol for VIM-2, and 15 μg/mL of tetracycline for IMP-1) and grown 16 hours at 30°C. The next day, the overnight cultures were normalized and diluted by 1000-fold with LB media. The subsequent cultures were incubated at 37°C for an 1 hour and 20 min. While the cultures were under incubation, two-fold increasing concentrations of ampicillin, cefotaxime, and meropenem were prepared in reagent reservoirs (Fisher Scientific). The mutagenized VIM-2 variants were challenged by $0 – 2^{14.6}$ (25,000) μg/mL of ampicillin, $0 – 2^{10.0}$ (1024) μg/mL of cefotaxime, and $0 – 2^{3.00}$(8) μg/mL of meropenem. The concentration ranges used for NDM-1 is $0 – 2^{15.0}$ (32,770) μg/mL for ampicillin, $0 – 2^{13.0}$ (8,200) μg/mL for cefotaxime, and $0 – 2^{4.00}$ (16)

μg/mL for meropenem. The concentration ranges used for IMP-1 were $0 – 2^{12.3}$ (5,000) μg/mL for ampicillin, $0 – 2^{11.0}$ (2,048) μg/mL for cefotaxime, and $0 – 2^{3.00}$ (8) μg/mL for meropenem. The resistance of IMP-1 was lower than the resistance of VIM-2 and NDM-1. To accurately capture the antibiotic resistance of IMP-1, the concentrations of the three representative antibiotic drugs were partitioned into 24 instead of 12 wells. 20 μL of the diluted antibiotic drug stocks were dispensed into separate polystyrene 96-well plates (VWR). Once the incubation was over, the turbidity of the samples was measured. When the absorbance of the culture at 600 nm, OD_{600}, of the subsequent cultures was in a range between 0.01-0.03, 180 μL of the cultures were mixed with 20 μL of respective antibiotic drugs (10-fold dilution). Then, the plates were covered by adhesive films (VWR) and incubated for 6 hours at 37 °C. After the 6 hours of incubation, OD_{600} of the cultures with antibiotic drugs were measured by Gen5 Data Collection and Analysis Software using a BioTekTM EpochTM Microplate spectrophotometer. The data was processed in excel to plot the dose-response curves using Prism 9.0.

3.2.7 Using Transim and the four translational aspects (In this study)

Transim is a bioinformatics tool developed by Gilad Shaham and Tamir Tuller (Shaham et al 2017). The rationale behind selecting Transim is as follows: i) the model's performance is examined and analyzed based on *Escherichia coli* genes, ii) the model investigates transcription sequences to estimate mRNA translation dynamics, and iii) the model takes the relative effect of mutations into account to predict the translation initiation and elongation. The analysis provides a list of components regarding translational events: i) ΔG_{mRNA}, ii) the normalized initiation rate, iii) the translation rate, and iv) the ribosome density. The ΔG_{mRNA} approximates the folding energy of the mRNA subsequences before the interaction with the 30S ribosome complex. A

poor correlation between protein abundance and the mRNA level implies a gene expression regulation at multiple levels (Huang et al 2011). Consequently, I considered the above four aspects as important to examine for the following reasons. First, mRNA local secondary structure was under the influence of the evolutionary process (Peeri et al 2020) and regulated translation efficiency in *E. coli* (Gu et al 2010). Second, translation initiation rate was determined as a rate-limiting step in the protein synthesis from mRNA transcripts and relied on the start of a coding sequence together with the 5'-untranslated region (UTR) (Reeve et al 2014). Third, translation rate was one of the substantial components that influence protein concentration together with the mRNA degradation rate (Lu et al 2007). Fourth, ribosome density was used as an approximation of the translation efficiency, which is defined as the number of proteins produced per mRNA per unit of time (Li et al 2014).

3.2.8 Data preparation and processing for Transim (Performed in this study)

The nucleotide sequences of NDM-1, VIM-2, and IMP-1 were acquired from Dr. Socha's data. The entire enzyme sequence of each variant at rounds 3, 6, 9, and 18 trajectories was individually converted into a fasta format and analyzed by Transim (Sharam et al 2017). The entire P_{bla} sequences were used in the analysis as the 5' untranslated region (UTR).

3.2.9 Signal peptide prediction by SingalP 5.0 (Performed in this study)

The putative signal peptide of NDM-1, VIM-2, and IMP-1 was estimated using the bioinformatics tool called SignalP 5.0 (Nielsen et al 2019). The entire amino acid sequences were arranged in Multiple Sequence Comparison by Log-Expectation (Muscle) fasta format, and

the NDM-1, VIM-2, and IMP-1 files were submitted directly to the SignalP-5.0 server. The Gram-negative was selected as an organism group.

3.2.10 Genetic diversity analysis using signal peptide sequences of the enzymes (Performed in this study)

The genetic diversity (Π) estimates the extent of diversity at the nucleotide level between lines of sequences within an intrapopulation of a specific trajectory using equation (2). *N* represents the total number of sequences analyzed for a given trajectory, x_i and x_j mean the frequency of *i*th and *j*th sequences from the overall sequences, and π_{ij} means the fraction of nucleotide differences between x_i and x_j. Only the signal peptide sequences of NDM-1, VIM-2, and IMP-1 were analyzed by the method. The total numbers of analyzed sequences were 50, 43, and 46 for NDM-1, VIM-2, and IMP-1, respectively, from the round 3, 6, 9, 12, 15, and 18 trajectories.

$$\Pi = \frac{N}{(N-1)} \sum_{i<j} x_i \, x_j \, \pi_{ij} \qquad (2)$$

3.2.11 The *dN/dS* ratio approximation (Performed in this study)

The *dN/dS* ratio of signal peptides is the ratio of the number of non-synonymous (changes in the amino acid sequence) and synonymous (no changes in the amino acid sequence) mutations substituted in the sequence space of the analyzed variants from the round 3, 6, 9, 12, 15, and 18 trajectories. For the estimation, I excluded the sequence(s) that did not have any synonymous mutations, meaning that the approximation includes some degrees of biases. The trajectories with the biased approximation were marked by an asterisk in Table 3.2.

3.2.12 Heatmap construction (Performed in this study)

The frequency of mutational substitutions in the signal peptide regions was estimated by normalizing the number of altered codons by the number of wild-type codons at the respective codon positions. The heatmaps depict the frequency obtained from the signal peptide of the enzymes in percentage. The codon positions where there weren't any mutational substitutions were represented as 0, whereas the codon positions that are completely dominated by mutations are denoted with 100.

3.3 Results

3.3.1 The MIC of sampled variants are representative of the resistance in their respective population

In a previous study, wild-type NDM-1, VIM-2, and IMP-1 were individually subjected to the directed evolution experiment for 18 rounds by Dr. Socha, as stated in the methodology; the evolution experiment over 18 rounds for each enzyme is a "trajectory." Then, he determined the sequence and resistance level of representative steps in each trajectory by selecting random variants from rounds 1, 2, 3, 6, 9, 12, 15, and 18 and conducting Sanger sequencing and MIC assays. The sequence profiles revealed the mutational patterns and allele frequencies linked to the resistance at each trajectory. Owing to a low sample size of sequenced variants, I conducted a series of dose-response curve assays using the entire population at each of the eight trajectories to verify a correlation between the sequenced individual variants and their corresponding populations. Each respective wild-type MBL orthologs was used as a positive control, while the negative control is a wild-type MBL with the majority of the gene fragment truncated, losing the original activity. As shown in the dose-response curves in Figure 3.1, the three orthologs

exhibited a uniform trend. The collective IC_{50} value in the population of variants at each round of each trajectory steadily propagated to the higher concentrations of ampicillin as the variants at their respective trajectory gained random mutations (Table 3.1 and Figure 3.1). There was a 30-, 15-, and 107-fold increase in ampicillin resistance in the round 18 variants of each enzyme population compared to the wild-type. As Figure 3.2 shows, the population-level IC_{50} values of the three orthologs strongly correlated with the MIC values of individual variants, with R^2 values of 0.93, 0.99, and 0.96 for the NDM-1, VIM-2, and IMP-1 trajectory, respectively. The Hill coefficients of the dose-response curves (Table 3.1) also supported the interconnection between the sequenced individuals and the respective population. The Hill coefficient can be used to approximate the diversity of mutants and their fitness effects. If the mutants are linked to phenotypes that are distributed in a relatively narrow fitness distribution, the fitness of the mutants will collapse abruptly at a certain concentration due to the lack of fitness variations in a population. This phenomenon was manifested as a high absolute value of Hill coefficient (>1) of wild-type which is in a homogeneous state at the nucleotide level (only a single resistance level). As Table 3.1 shows, the absolute values of Hill coefficients of the evolved orthologs' population at each trajectory were nearly identical to the monoclonal positive and negative controls, suggesting that the mutational substitutions are a reasonable representation of the entire population with low variability in phenotypes. It is worth noting that a gradual antibiotic resistance enhancement was also consistently identified from the dose-response curve assays performed using cefotaxime (CTX) and meropenem (MEM) (Appendix), indicating the MBLs have also maintained broad-specificity to various β-lactam drugs during the evolution.

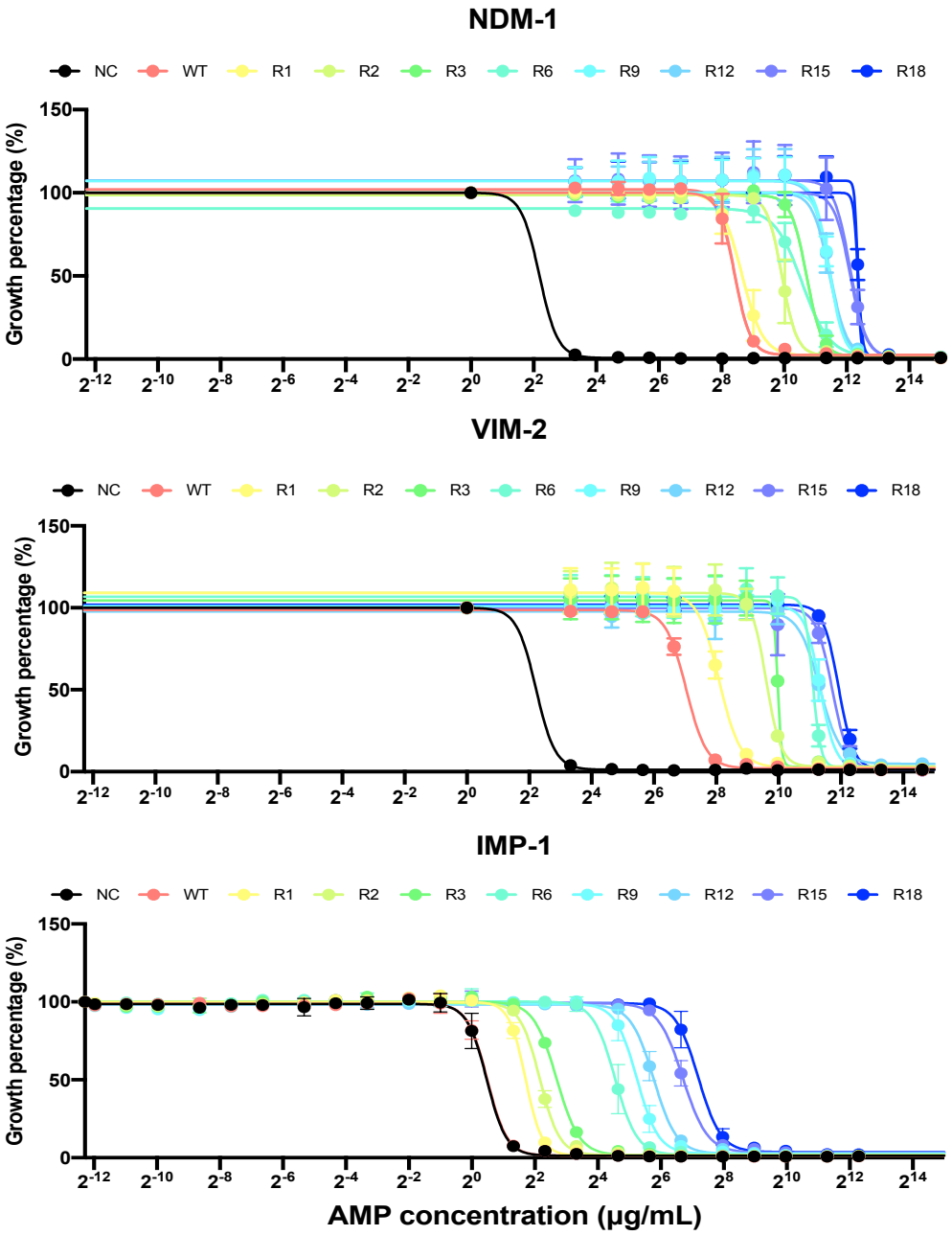


Figure 3.1. Dose-response curves for eight selected rounds of the NDM-1, VIM-2, and IMP-1 directed evolution trajectories when challenged with ampicillin (AMP).
Of the 18 rounds in each trajectory, round (R) 1 (yellow), 2 (light green), 3 (green), 6 (spindrift), 9 (light blue), 12 (sky blue), 15 (orchid), and 18 (dark blue) trajectories of the three MBL orthologs are challenged by ampicillin (AMP). The black represents negative control (NC), and the pink reflects the wild-type used as a positive control. Each dot represents the data point that is obtained from the assay. The determined cell growth is normalized and converted to a percentage (%), shown on the y axis. AMP concentrations are displayed on the x axis in a power of two and capture the ranges of AMP concentration that are used for each MBL. The dose-response curves of NC for NDM-1 and VIM-2 lack data points near IC_{50} and therefore may not be a reliable fit.

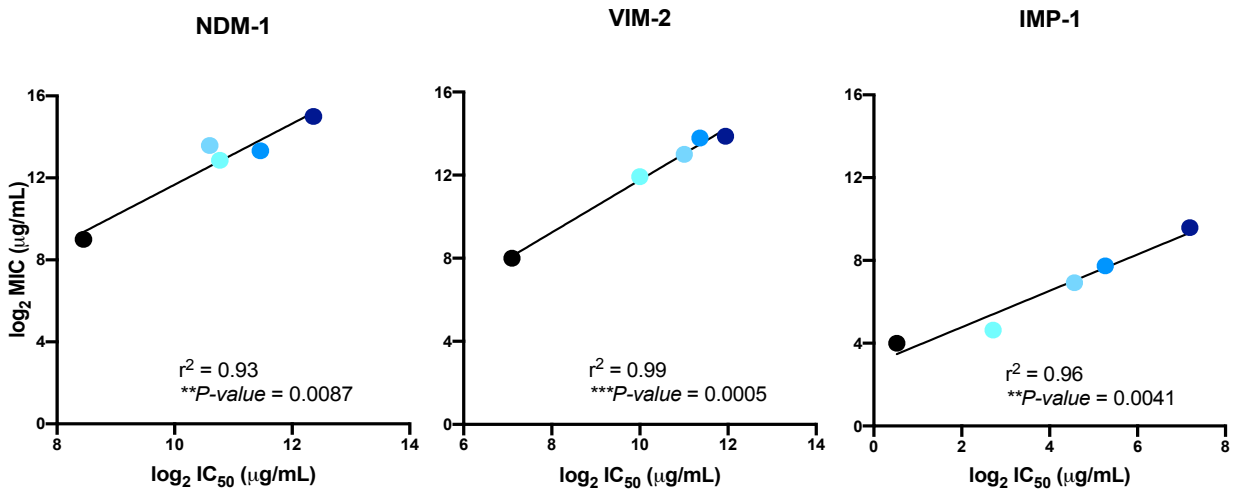


Figure 3.2. Scatter plots showing a correlation between $\log_2$ IC_{50} and $\log_2$ MIC of NDM-1, VIM-2, and IMP-1.
The plots universally show the correlation between the antibiotic resistance of individual variants (MIC) and the antibiotic resistance of the variants' respective populations (IC_{50}) at round 3 (light blue), 6 (sky blue), 9 (blue), and 18 (dark blue) of the enzymes. The antibiotic resistance of individual variants was determined by MIC using ampicillin (μg/mL). In contrast, dose-response curve assays tested the antibiotic resistance of the respective populations using ampicillin (μg/mL). IC_{50} represents the average antibiotic resistance of the respective populations shown on the x axis. The black dot represents the respective wild-type of the three enzymes. The ampicillin concentrations for MIC and IC_{50} are converted in $\log_2$. A strong positive correlation observed from each of the enzymes reveals that individual variants and their respective population gained antibiotic resistance progressively. This result suggests that the sequenced individuals are reasonable representations of their respective population.

Table 3.1 IC_{50} values and the respective slopes for eight trajectories of NDM-1, VIM-2, and IMP-1 challenged by ampicillin (AMP). The IC_{50} values and Hill coefficients were obtained by performing dose-response curve assays on directed evolution trajectories. NC and WT represent negative control and wild type, respectively. The R1, R2, R3, R6, R9, R12, R15, and R18 represent each trajectory of directed evolution. Asterisks indicate that there weren't sufficient data for accurate estimation.

	Ampicillin (AMP)					
	NDM-1		VIM-2		IMP-1	
	IC_{50} (μg/ml)	Hill coefficient	IC_{50} (μg/ml)	Hill coefficient	IC_{50} (μg/ml)	Hill coefficient
NC	3.3 ± 0.1*	-1.2 ± 0.0*	3.3 ± 0.0*	-1.2 ± 0.0*	1.5 ± 0.2	-5.3 ± 1.4
WT	$3.3x10^2$±4.7x10	-5.8 ± 0.6	$1.3x10^2$ ± 6.5	-4.4 ± 0.4	1.4 ± 0.1	-4.7 ± 0.8
R1	$4.0x10^2$±7.8x10	-4.8 ± 0.4	$2.7x10^2$ ± 5.4	-3.8 ± 0.1	2.5 ± 0.7	-3.5 ± 2.0
R2	$9.6x10^2$±$1.2x10^2$	-1.0x10±7.0	$7.7x10^2$±6.1x10	-6.1 ± 1.0	4.0 ± 0.1	-3.7 ± 1.4
R3	$1.8x10^3$±$3.9x10^2$	-8.5 ± 4.8	$1.0x10^3$±4.9x10	-9.9 ± 8.1	6.4 ± 0.2	-3.9 ± 0.3
R6	$1.6x10^3$±$3.0x10^2$	-3.6 ± 0.3	$2.0x10^3$±$2.5x10^2$	-7.7 ± 2.9	2.5x10 ± 0.8	-4.5 ± 0.4
R9	$2.8x10^3$±6.9x10	-5.9 ± 0.2	$2.6x10^3$±$2.3x10^2$	-7.1 ± 2.2	3.8x10 ± 4.2	-4.5 ± 0.5
R12	$2.8x10^3$±8.7x10	-5.1 ± 0.3	$2.6x10^3$±8.0x10	-4.4 ± 0.9	5.5x10 ± 4.7	-4.1 ± 0.2
R15	$4.3x10^3$±$3.1x10^2$	-5.6 ± 0.7	$3.3x10^3$±$4.4x10^2$	-4.9 ± 2.3	$1.0x10^2$ ± 8.8	-3.9 ± 0.3
R18	$5.2x10^3$±$1.4x10^2$	-1.1x10±8.2	$3.9x10^3$±$4.3x10^2$	-6.8 ± 2.0	$1.5x10^2$±1.4x10	-4.5 ± 0.1

3.3.2 Non-synonymous mutations are consistently identified at identical codon positions in MBL signal peptides

The sequence profiles of evolved variants for the three orthologs revealed the fixation of multiple mutations in the signal peptide of the MBL orthologs. To determine whether the mutations are neutral or adaptive, the *dN/dS* ratios of the signal peptide regions of 23, 6, and 6 sequence samples of VIM-2, NDM-1, and IMP-1, respectively, were analyzed. The *dN/dS* ratio is a widely accepted approach that measures the ratio of non-synonymous (*dN)* and synonymous (*dS)* mutations. A *dN/dS* ratio of > 1 represents the biased fixation of non-synonymous mutations that is often interpreted as a sign of positive selection (Gilchrist et al 2015; Kryazhimskiy et al 2008). The three MBL enzymes consistently acquired non-synonymous mutations in their signal peptides, exhibiting *dN/dS* ratios greater than one for all trajectories. In fact, several rounds of VIM-2 and NDM-1 as well as all rounds of IMP-1 had sequences where non-synonymous mutations completely dominated. As Table 3.2 shows, the round 18 populations of VIM-2, NDM-1, and IMP-1 had approximately 2-, 1.1-, and 6-fold more non-synonymous mutations than the number of synonymous mutations on average. Interestingly, these non-synonymous mutations were consistently fixed at certain codon positions and propagated through the sub-divided populations. In the case of VIM-2, the 5^{th} and 8^{th} codon positions were absolutely dominated by non-synonymous mutations starting at round three, and the same trend was observed for the 10^{th} position after round 12 (Figure 3.3B). Similarly, NDM-1 had a conspicuous amount of substitutions at codon positions 5, 7, 8, and 11, only with a mutation at the 7^{th} position spread from round 1 (Figure 3.3A). IMP-1 also displayed such tendency at positions 5, 9, 10, and 11 (Figure 3.3C). It is worth noting that the orthologs consistently fixed a non-synonymous mutation at the 5^{th} position, potentially implying the significance of this codon.

Table 3.2. *dN/dS* ratio determined from the signal peptide of NDM-1, VIM-2, and IMP-1

Avg. of dN/dS of Signal peptide (Ampicillin)	*R3*	*R6*	*R9*	*R12*	*R15*	*R18*
NDM-1	*2.00 ± 0.33	2.00 ± 0.0	1.82 ± 0.65	3.00 ± 0.82	2.50 ± 0.0	1.09 ± 0.15
VIM-2	*1.39 ± 0.49	1.92 ± 0.71	*3.08 ± 0.51	1.67 ± 0.33	1.88 ± 0.44	2.02 ± 0.53
IMP-1	*2.00 ± 0.0	*2.60 ± 0.60	Only non-synonymous mutation	*4.00 ± 0.0	*4.00 ± 0.0	*6.0 ± 0.0

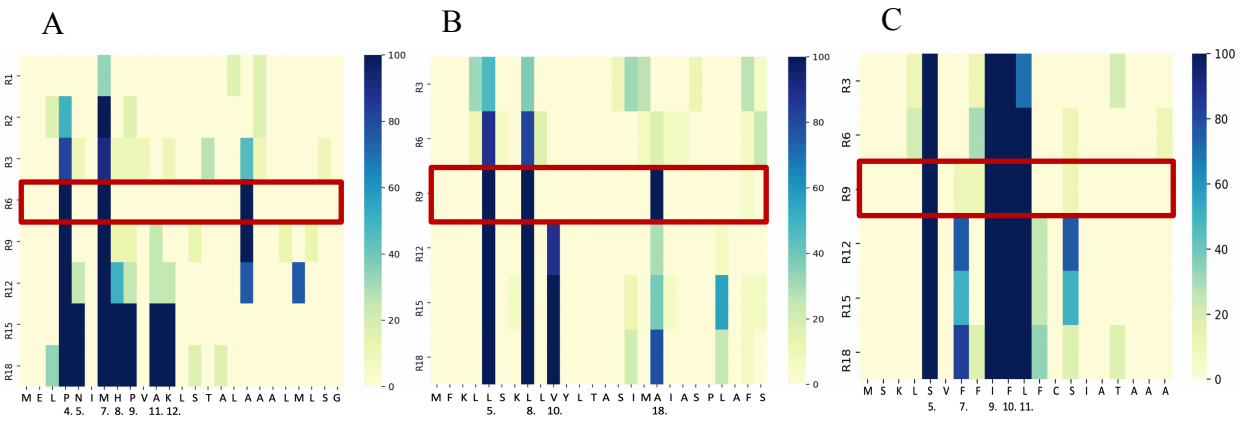


Figure 3.3. Mutational substitutions from directed (adaptive) evolution trajectories of NDM-1, VIM-2, and IMP-1.
The heatmaps show codon positions with mutational substitutions in the signal peptide of (*A*) NDM-1, (*B*) VIM-2, and (*C*) IMP-1. The frequency of mutational substitution is described based on the color bar on the right. The codon position which is completely dominated by mutation is colored in dark blue. The codon position in the absence of mutational substitutions is colored in yellow. The substitution involves both synonymous and non-synonymous mutations. For each ortholog, the putative signal peptide is estimated from the wild-type sequence by SignalP5.0. Each trajectory where sequence profiles were collected is presented on the y axis. The amino acid residues corresponding to the codon positions are listed on the x axis along with codon positions that obtained predominant mutational substitutions. The red boxes highlight the trajectories where the least sequence diversity was detected. Inherent mutational bias and the natural selection for preferred codon usage in *E. coli* may result in a sign of sequence convergence.

3.3.3 A sign of sequence convergence is observed from the MBL orthologs

In adaptive evolution, variants with similar functional traits tend to arise and evolve through sub-divided populations. The mutations that drive the adaptive evolution may be a limited portion of all possible beneficial mutations because of the inherently stochastic nature in the creation and fixation of mutational variations - only a subset of all beneficial mutations may be sampled, and

those that are fixed will propagate to future lineages. As a result, patterns of converged sequences may appear in evolved lineages (Storz 2016). To estimate whether there is a sign of mutation-driven sequence convergence, the genetic diversity (Π) was computed using the sequence profiles of the three orthologs by equation (2) described in the method. The Π measures the nucleotide diversity in the sequenced variants within an intrapopulation of a specific trajectory. Therefore, it simply indicates that the lower the Π is, the more similar their sequences are to each other.

Of the trajectories, round 6 and round 9 trajectories of NDM-1 and VIM-2, respectively, revealed a noticeable extent of sequence convergence (Table 3.3 and Figure 3.3A and B). A relatively mild degree of sequence convergence was detected at round 9 for IMP-1, as well (Table 3.3 and Figure 3.3C). The genetic diversity analysis clearly coincided with the highlighted regions in Figure 3.3 exhibiting fixed mutations only at positions 5, 8, 10, and 18 in the case of VIM-2 and at positions 4, 7, and 18 in the case of NDM-1. As the highlighted regions indicate, IMP-1 also showed a similar trend of having mutations at positions 5, 9, 10, and 11 with a moderate level of additional mutations at positions 7, 8, and 14 (Figure 3.3C). Overall, the non-synonymous mutations fixed and prevailed at the highlighted positions through the sub-divided populations. The tendency of developing a functionally similar phenotype connected with these mutations in the signal peptides of the MBLs may have caused sequence convergence observed in a distinct trajectory during the adaptive evolution.

Table 3.3. Genetic diversity values determined from signal peptide of NDM-1, VIM-2, and IMP-1

	Genetic diversity (Π) of signal peptide sequences of adaptive evolution trajectories					
	R3	*R6*	*R9*	*R12*	*R15*	*R18*
NDM-1	0.028	0.008	0.020	0.039	0.029	0.026
VIM-2	0.031	0.027	0.008	0.012	0.021	0.020
IMP-1	0.051	0.052	0.045	0.093	0.093	0.083

3.3.4 A rare codon is consistently detected in proximity to the start codon at the N-terminals of the three enzymes

At the translational level, codon usage is one of the substantial elements, influencing the efficiency of protein synthesis in prokaryotes (Gingold et al 2011; Quax et al 2015; Verma et al 2019; Zhou et al 2016). Owing to the mutational substitutions, the sequence profiles of the round 18 variants showed multiple altered codon usages. A rare codon usage seemed to be maintained selectively at codon positions 4 and 5 for all enzymes. It is interesting to note that the 5th codon position, uniformly substituted by a non-synonymous mutation, consistently assigned codons with low usage for all MBL enzymes in *E. coli*. For instance, the round 18 variants possessed a rare codon at the 5th codon position, unlike the wild-type VIM-2. NDM-1 and IMP-1 already assembled a low-usage codon at these regions and exhibited no signs of a recognizable shift. This trend shared among the orthologs suggests a potential regulatory mechanism, such as the translational ramp (Figure 3.4), associated with the MBL gene expression. However, these non-synonymous substitutions did not show a particular preference for a specific amino acid across all orthologs (L5S for VIM-2, N5K for NDM-1, and S5L/S5R for IMP-1).

Next, the rare codon distribution in the signal peptides was analyzed using the list of rare codons provided by Chen et al 2006. The rare codons were identified at codon positions 4, 5, 6, 8, 12,

15, 21, and 26 in VIM-2. NDM-1 possessed rare codons at positions 4, 7, 8, 9, 11,17, and 25, whereas IMP-1 contained rare codons at positions 4, 5, 9, 10, and 11. Surprisingly, a rare codon remained persistently at the 4th position for the wild-type and round 18 variants of VIM-2 and NDM-1, whereas IMP-1 consistently placed two rare codons consecutively at the 4th and 5th positions for wild-type and round 18 variants (Table 3.4). Only VIM-2 gained another rare codon at the 5th position at round 18 trajectory. In the case of NDM-1, a type of rare codon changed from Group I to Group II in round 18 trajectory (Table 3.4). According to Chen et al. 2006, rare codons in Group II (Table 3.5) were used at a frequency of more than 0.5% in *E. coli*. None of the orthologs gained mutations at the 3rd position that can be important for ribosome processivity. Interestingly, NDM-1 gained high-scoring codons, AAG and ATT (Figure 3.4), at the 5th and 6th positions respectively owing to the non-synonymous mutation at position 5, as shown in Table 3.4 (Top). The high-scoring codons (Figure 3.4) indicate sets of mRNA codons associated with a high level of eGFP expression. According to Verma et al. 2019, the sequences encoded with AAD and UAU, where D can be either A, G, or U at the 3rd and 4th positions, as well as AAV and AUU, where V can be either A, C, or G at the 4th and 5th codon positions, could strongly express eGFP.

Table 3.4. Codons identified from position from three to six in three orthologs.
The codons written in bold and red letters are rare codons used in MBLs. The italicised letters are putative high-score codons, as shown in Figure 3.4, that may increase protein synthesis.

	NDM-1	
	WT	R18
Codon 3	TTG	TTG
Codon 4	**CCC**	**CCT**
Codon 5	AAT	*AAG*
Codon 6	ATT	*ATT*
	VIM-2	
	WT	R18
Codon 3	*AAA*	*AAA*
Codon 4	**CTT**	**CTT**
Codon 5	TTG	**TCG**
Codon 6	AGT	AGT
	IMP-1	
	WT	R18
Codon 3	*AAG*	*AAG*
Codon 4	**TTA**	**TTA**
Codon 5	**TCT**	**CTC** (5) /CGC
Codon 6	GTA	GTA

Table 3.5. Rare codons in *E. coli* suggested by Chen et al. 2006.

Rare codons (20 amino acids in *E. coli*)	
Group I (Rare codon used at a freq. of < 0.5%)	AGG, AGA, CGA, CUA, AUA, **CCC**, and CGG
Group II (Rare codon used at a freq. of > 0.5%)	ACA, **CCU**, UCA, GGA, AGU, **UCG**, CCA, UCC, GGG, **CUC**, **CUU**, **UCU**, and **UUA**

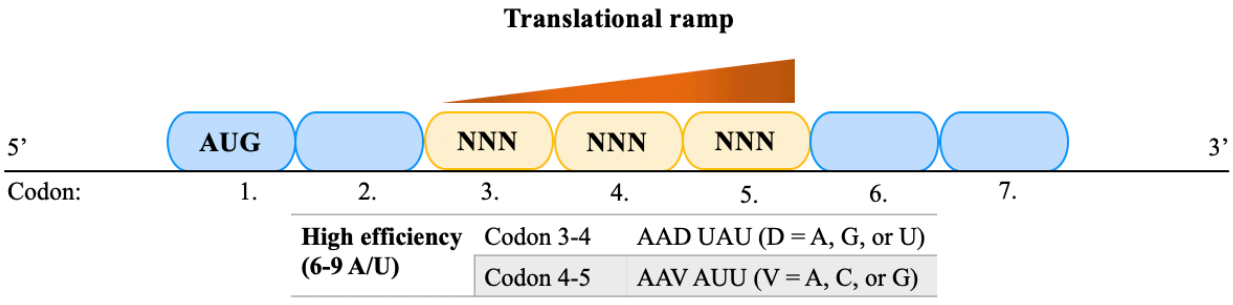


Figure 3.4. Overview of the translational ramp.
The sketch of the short translational ramp in *E. coli* suggested by Verma et al 2019. The solid black line represents an mRNA transcript with 5' and 3' at the left end and right end of the line, respectively. The codon positions are

denoted in numbers below the line. The orange units represent the short translational ramp determined in *E. coli*. The NNN represents potential codons that can be placed. The high translational efficiency is suggested to contain six to nine numbers of AU in 3-5 codon positions. The efficiency of protein synthesis is affected by combinations of AAD UAU at 3-5 codon positions and AAV AUU at 4-5 codon positions.

3.3.5 The ribosome density and the translational rate are correlated with MIC of NDM-1, VIM-2, and IMP-1

Gene expression takes two important steps, transcription and translation. The translation is a crucial step for protein synthesis and is influenced by multiple factors such as the tRNA abundance and the number of ribosomes on an mRNA transcript (Huang et al 2011; Lu et al 2007). This implies that the amount of cellular protein production can vary based on the efficiency of translation rather than the total amount of mRNA. In a previous study in Tokuriki lab, the periplasmic enzyme amount of eight wild-type B1 MBL orthologs was measured by Dr. Socha (Socha et al 2018), and a 210-fold disparity between the two orthologs with the lowest and the highest enzyme production was detected in *E. coli* at the translational stage. To understand the underlying mechanisms, I analyzed the entire nucleotide sequences of the round 3, 6, 9, and 18 trajectories of NDM-1, VIM-2, and IMP-1 using a bioinformatic tool called Transim (Shaham et al 2017). Transim approximates mRNA translation dynamics solely based on the coding sequences of a given DNA. The analysis predicted that the ribosome density increased slightly in all orthologs compared to the respective wild-type. However, the Spearman rank correlation between ribosome density and the respective MIC unveiled a distinctive trend in three orthologs. The ribosome density for NDM-1 was inversely correlated with MIC ($r = -0.42$, *P-value* = 0.0089, Figure 3.5B). On the other hand, VIM-2 and IMP-1 had an increased number of ribosomes translating their mRNA strands. In the case of IMP-1, the ribosome density increase may be one of many substantial factors that enhanced MIC ($r = 0.38$, *P-value* = 0.0212, Figure

3.5F). Although VIM-2 showed a marginal increase in ribosome density, its effect was rather negligible to explain the enhanced MIC (r = 0.27, *P-value* = 0.0507, Figure 3.5D). For the translation rate, only VIM-2 exhibited a positive correlation with MIC. Interestingly, VIM-2 revealed a moderate level of correlation at a significant level (r = 0.53, *P-value* < 0.0001, Figure 3.5C). This result implicitly indicates that the increase in translation rate for VIM-2 may be one of the crucial elements that lead to antibiotic resistance increase. However, the inverse correlation was consistently observed in NDM-1 (r = -0.45, *P-value* = 0.0047, Figure 3.5A). Moreover, enhanced antibiotic resistance could not be explained by the changes in translation rate for IMP-1 (r = -0.22, *P-value* = 0.1924, Figure 3.5E).

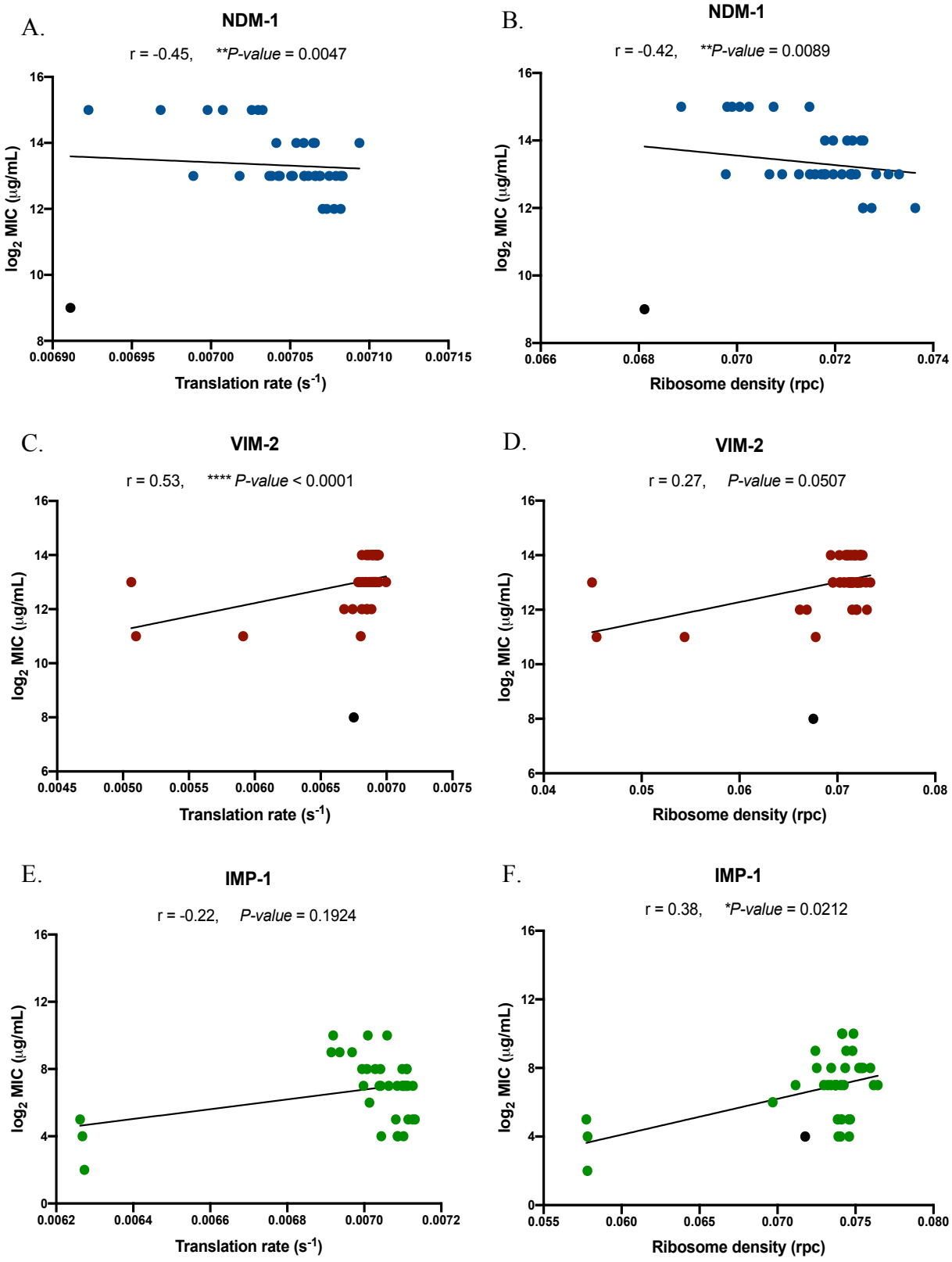

A.
NDM-1
r = -0.45, **P-value = 0.0047
log2 MIC (μg/mL)
Translation rate (s-1)
B.
NDM-1
r = -0.42, **P-value = 0.0089
log2 MIC (μg/mL)
Ribosome density (rpc)
C.
VIM-2
r = 0.53, **** P-value < 0.0001
log2 MIC (μg/mL)
Translation rate (s-1)
D.
VIM-2
r = 0.27, P-value = 0.0507
log2 MIC (μg/mL)
Ribosome density (rpc)
E.
IMP-1
r = -0.22, P-value = 0.1924
log2 MIC (μg/mL)
Translation rate (s-1)
F.
IMP-1
r = 0.38, *P-value = 0.0212
log2 MIC (μg/mL)
Ribosome density (rpc)

Figure 3.5. Scatter plots showing a correlation between log_2 MIC and translation rate as well as between log_2 MIC and ribosome density.
The bioinformatic analysis is performed using the 37 sequenced variants of NDM-1 (Top, Blue) and IMP-1 (Bottom, Green) as well as the 51 sequenced variants of VIM- 2 (Middle, Red). These variants are collected from rounds 3, 6, 9, and 18 trajectories of directed evolution. Antibiotic resistance of each variant was determined by log_2 MIC (μg/mL) using ampicillin from the previous experiment in the Tokuriki lab (y axis). MIC of the variants is compared with the translation rate (s^{-1}) shown in the first column and with the ribosome density (ribosomes per codons (rpc)) in the second column. The black dot represents the wild-type of the respective orthologs. Spearman's rank correlation coefficient is used to estimate their relationship. (*A, B*) In the case of NDM-1, both translation rate and the ribosome density are inversely correlated with MIC. (*C*) MIC of the VIM-2 variants is compared with the translation rate. The three widely scattered dots on the left side of the plot belong to round 3 trajectories. MIC of each variant is positively connected with its translation rate at the moderate level. (*D*) The ribosome density of the VIM-2 variants does not reveal a strong sign of correlation with MIC. Again, the most scattered dots belong to the round 3 trajectories. (*E*) The translation rate of IMP-1 variants are inversely correlated with MIC without statistical significance. The two variants that are located at the far left belong to the round 3 trajectories. Wild-type was overlaid with variants. (*F*) MIC of IMP-1 variants are moderately correlated with the ribosome density, implying a slight improvement in translation efficiency.

3.3.6 The predicted shifts in mRNA folding energy and translation initiation rate were positively correlated with the enhanced antibiotic resistance determined in MIC

As determined in multiple studies, the secondary structure of the mRNA 5' untranslated region (UTR) is highly conserved in the bacterial kingdom, and regulates translation initiation and overall protein production (Tuller et al 2010; Peeri et al 2020; Gu et al 2012; Kudla et al 2009; Bentele et al 2013; Chursov et al 2013). One of its fundamental roles is to accommodate ribosome complexes onto the N-terminal mRNA transcripts, where the reduced thermodynamic stability is selectively favored by allocating AU-rich codons (Reeve et al 2014; Duval et al 2013; Goodman et al 2013).

To determine if the fixed mutations had any effect on the secondary structure of mRNA, the following mRNA sequences were analyzed: the 35 upstream nucleotides from the start codon in UTR and 35 downstream nucleotides from the start codon in coding sequences for all mRNAs of variants from rounds 3, 6, 9, and 18 of the orthologs. The coding sequences up to approximately

the 12th codon were analyzed in the bioinformatics; the non-synonymous mutations identified in the signal peptides were all included.

As the mutations accumulated in the signal peptides, the folding energies of the first 35 coding sequences of the mRNAs, where the 30S complex binds, were increased by a different extent. The analysis predicted the most conspicuous mRNA conformational shift in NDM-1; over the course of the trajectories from the wild-type to round 18, the value changed from -7.0 to -3.2 kcal/mol (Figure 3.6A, Blue and Table 3.6). In fact, the local folding energy of the round 18 variants of NDM-1 was predicted to be 1.6-fold less stable than the mean folding energy of each mRNA of 4,226 *E. coli* genes, which is -5 kcal/mol, for the first 40 nucleotides (Tuller et al 2010). However, the local folding energy of the wild-type NDM-1, -7.0 kcal/mol, is comparable to the mean folding energy, -7.95 kcal/mol, of the same *E. coli* genes between 41 and 80 nucleotides. Such folding energy almost consistently remained through the entire downstream transcripts of up to 500 nucleotides (Tuller et al 2010), where ribosomes unravel the folded transcripts with the helicase activity as they translate (Takyar et al 2005). Hence, the mutation-mediated shift in the secondary structure of NDM-1 mRNAs may provide a favorable condition for ribosome complexes to bind at the translation initiation region. VIM-2 also exhibited an mRNA folding energy shift to a moderate degree, from -7.4 to -6.3 kcal/mol over the course of the trajectories (Figure 3.6A, Red and Table 3.6). IMP-1 showed a relatively marginal change from -5.3 to -5.2 kcal/mol in mRNA folding energy (Figure 3.6A, Green and Table 3.6). Taking the mean folding energy of the first 40 nucleotides of *E. coli* into consideration, the folding energy of the wild-type IMP-1 was already similar to endogenous *E. coli* genes and had no significant folding energy changes in the round 18 IMP-1 variants after the adaptive evolution.

To determine an association between MIC and the shifted mRNA folding energy, variants from the round 3, 6, 9, and 18 trajectories of the three orthologs were compared with their respective MIC using the Spearman rank correlation. mRNA folding energy shift in NDM-1 was moderately correlated with MIC with a statistical significance (r = 0.56, *P-value* = 0.0004, Figure 3.7A). Although the magnitude of correlation was weaker than that of NDM-1, mRNA folding energy shift in VIM-2 also exhibited a minor positive correlation with MIC at a statistically significant level (r = 0.35, *P-value* = 0.0125, Figure 3.7B). However, the folding energy of IMP-1 variants revealed no sign of correlation, presumably owing to the limited mRNA conformational change (r = 0.08, *P-value* = 0.6429, Figure 3.7C).

The bioinformatic analysis revealed that the MBL mutants increased the translation initiation rate to elevate the MBL gene expression, as shown in Figure 3.6B. Hence, I estimated a correlation between the translation initiation rate and MIC. The trend between the translation initiation rate and MIC for the three orthologs was identical to what was detected from the mRNA folding energy shift to MIC. The considerable similarity between translation initiation rate and mRNA folding energy may have resulted because the analysis depends mainly on the thermodynamic energy of mRNA secondary structure and derives translation initiation rate from it. A coding sequence of mRNA is one of the significant regulators in translation initiation rate (Kudla et al, 2009; Pop C et al, 2014; Tuller et al, 2010; Simmons et al, 1995). Therefore, further investigations will help in better understanding the functional role of the mutations in the signal peptides of MBLs.

Despite the inconsistency among the orthologs, they share two common trends. First, the round 3 trajectory showed the biggest range of mRNA folding energy and the translation initiation rate across all trajectories. The broadest range for mRNA folding energy and the translation initiation rate belonged to IMP-1. The variants were distributed the most widely between -4.6 and -8 kcal/mol in the case of the mRNA folding energy (Figure 3.7C) and between 0.09 and 0.02 (s^{-1}) for the translation initiation rate. The mRNA folding energies and translation initiation rates of individual VIM-2 variants were also the most spread out in the round 3 trajectory. The estimated values were within a range from -6.6 to -9.4 kcal/mol (Figure 3.7B) and from 0.04 to 0.01 (s^{-1}), respectively. While the range was the largest in round 3 for NDM-1 as well, the variance was the smallest amongst the three enzymes. They were distributed in a range between -4.2 and -6 kcal/mol (Figure 3.7A) and between 0.11 and 0.05 (s^{-1}), respectively. Second, the amount of variance progressively tapered down to a confined range as the directed evolution continues. To estimate the difference, a single variant with the highest and the lowest values of mRNA folding energy and translation initiation rate was selected from the round 3 and 18 trajectories, respectively. In the case of IMP-1, the range for mRNA folding energy at round 3 was between -4.6 and -8.0 kcal/mol, then the range changed from -3.9 to -5.5 kcal/mol at round 18 trajectory. VIM-2, which kept the mRNA folding energy range between -6.6 to -9.4 kcal/mol at round 3, showed a unified mRNA folding energy at -6.3 kcal/mol for round 18. NDM-1 also exhibited a similar pattern as VIM-2 with mRNA folding energy distributed in a range between -4.2 to -6.0 kcal/mol at round 3 and had mRNA folding energy only at -3.2 kcal/mol at round 18 trajectory. The translation initiation rate shared the similar pattern with the mRNA folding energy.

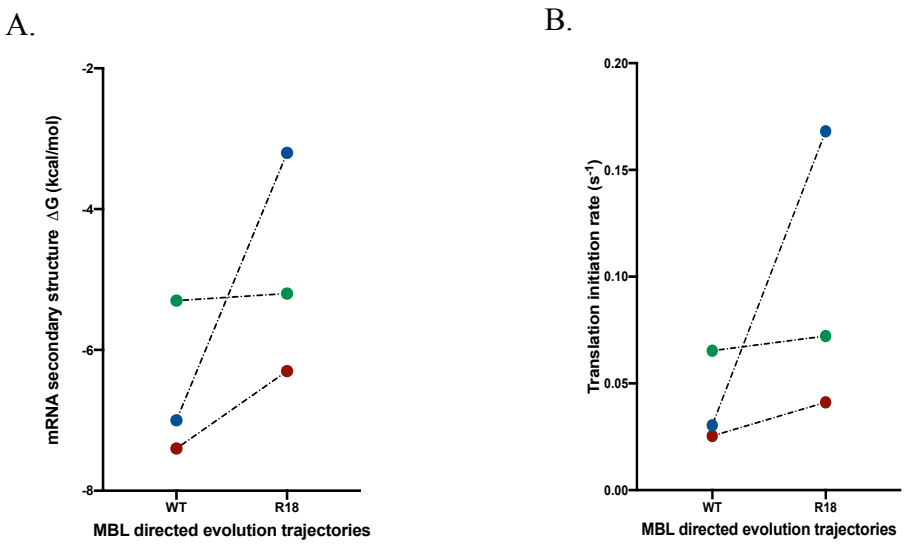


Figure 3.6. mRNA folding energy and translation initiation rate shift in three enzymes.
(A) The mRNA folding energy (kcal/mol) of wild-type (WT) and round 18 (R18) variants are shown on the left. Thermodynamically less stable conformation of mRNA is detected in the R18 variants as free folding energy shifted from -7.4 to -6.3 kcal/mol in VIM-2 (Red). NDM-1 is with the most dramatic mRNA conformational change by shifting from -7.0 to -3.2 kcal/mol (Blue). IMP-1 shows the least transformation with a change from -5.3 to -5.2 kcal/mol (Green). (B) Translation initiation rate (s^{-1}) for the WT and the R18 variants of VIM-2 (Red), NDM-1 (Blue), and IMP-1 (Green) are shown on the right. The most considerable change is a 5.5-fold increase in NDM-1 (from 3.0 x 10-2 to 1.7 x 10-1 s-1). VIM-2 also exhibits moderate improvement with a 1.6-fold increase (from 2.5 x 10-2 to 4.1 x 10-2 s^{-1}). Again, IMP-1 is with a marginal change in translation initiation rate by elevating only 1.1-fold (from 6.5 x 10-2 to 7.2 x 10-2 s^{-1}).

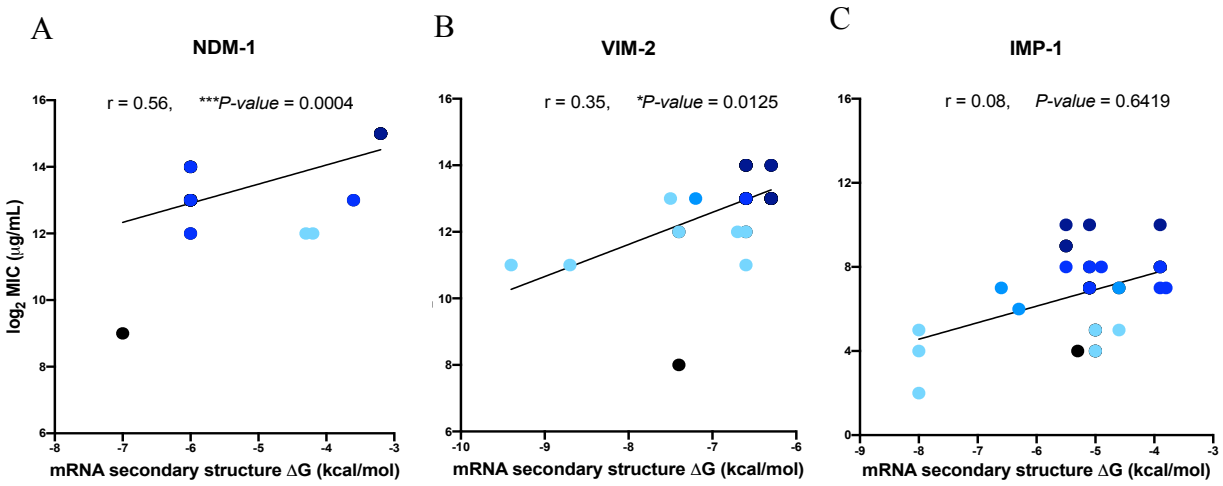


Figure 3.7. Scatter plots showing a correlation between log_2 MIC and mRNA folding energy shift in three enzymes.
The individual dots represent the sequenced variants analyzed by Transim. Same as in the previous analysis, these variants are collected from round 3 (light blue), 6 (aqua blue), 9 (blue), and 18 (dark blue) trajectories of the directed evolution. MIC (μg/mL) of the individual variants is shown in log_2 on the y-axis. The mRNA folding energy is shown on the x-axis (kcal/mol). Spearman rank correlation coefficient is used to estimate potential connections between the two elements with statistical significance. (A) NDM-1 shows a moderate level of a correlation between log_2 MIC and mRNA folding energy. The majority of the individual variants shared similar characteristics, resulting in many of the dots overlapping one another, especially at -6 kcal/mol. It is possible that NDM-1 variants may have a strong tendency to develop an advantageous phenotype. (B) VIM-2 shows the minor level of a correlation between

$\log_2$ MIC and mRNA folding energy. Again, the majority of the dots overlapped each other, showing a trend of similarity. As observed in previous findings, the variants at the round 3 trajectory are scattered in a wide range. Then, the range of scattered dots narrows gradually towards the round 18 trajectories. (C) IMP-1 shows a lack of correlation between $\log_2$ MIC and mRNA folding energy.

Table 3.6. The average values of mRNA folding energy, translation initiation rate, translation rate, and ribosome density of the orthologs.

	Avg. mRNA ΔG (kcal/mol)	Avg. translation initiation rate (s^{-1})	Avg. translation rate (s^{-1})	Avg. ribosome density (rpc)
		NDM-1		
Wild-type	-7.0	3.0×10^{-2}	6.9×10^{-3}	6.8×10^{-2}
Round 3	-5.7 ± 0.7	$4.8 \times 10^{-2} \pm 2.3 \times 10^{-2}$	$7.1 \times 10^{-3} \pm 2.9 \times 10^{-5}$	$7.2 \times 10^{-2} \pm 1.0 \times 10^{-3}$
Round6	-6.0 ± 0.0	$7.1 \times 10^{-2} \pm 7.3 \times 10^{-18}$	$7.1 \times 10^{-3} \pm 1.8 \times 10^{-5}$	$7.2 \times 10^{-2} \pm 2.6 \times 10^{-4}$
Round 9	-5.4 ± 1.1	$1.7 \times 10^{-1} \pm 4.3 \times 10^{-2}$	$7.1 \times 10^{-3} \pm 1.4 \times 10^{-5}$	$7.3 \times 10^{-2} \pm 5.0 \times 10^{-4}$
Round 18	$-3.2 \pm 4.8 \times 10^{-16}$	$1.7 \times 10^{-1} \pm 3.0 \times 10^{-17}$	$7.0 \times 10^{-3} \pm 4.0 \times 10^{-5}$	$7.0 \times 10^{-2} \pm 8.1 \times 10^{-4}$
		VIM-2		
Wild-type	-7.4	2.5×10^{-2}	6.8×10^{-3}	6.8×10^{-2}
Round 3	-7.6 ± 1.1	$2.5 \times 10^{-2} \pm 1.1 \times 10^{-2}$	$6.4 \times 10^{-3} \pm 7.4 \times 10^{-4}$	$6.3 \times 10^{-2} \pm 1.1 \times 10^{-2}$
Round6	-6.7 ± 0.2	$3.6 \times 10^{-2} \pm 2.7 \times 10^{-3}$	$6.9 \times 10^{-3} \pm 4.2 \times 10^{-5}$	$7.2 \times 10^{-2} \pm 9.8 \times 10^{-4}$
Round 9	-6.6 ± 0.0	$3.6 \times 10^{-2} \pm 7.3 \times 10^{-18}$	$6.9 \times 10^{-3} \pm 2.5 \times 10^{-5}$	$7.2 \times 10^{-2} \pm 5.7 \times 10^{-4}$
Round 18	-6.3 ± 0.1	$4.1 \times 10^{-2} \pm 1.7 \times 10^{-3}$	$6.9 \times 10^{-3} \pm 4.9 \times 10^{-5}$	$7.1 \times 10^{-2} \pm 9.9 \times 10^{-4}$
		IMP-1		
Wild-type	-5.3	6.5×10^{-2}	7.1×10^{-3}	7.2×10^{-2}
Round 3	-5.9 ± 1.5	$6.0 \times 10^{-2} \pm 2.8 \times 10^{-2}$	$6.8 \times 10^{-3} \pm 4.0 \times 10^{-4}$	$6.9 \times 10^{-2} \pm 8.0 \times 10^{-3}$
Round6	-5.3 ± 0.7	$6.9 \times 10^{-2} \pm 1.7 \times 10^{-2}$	$7.1 \times 10^{-3} \pm 3.7 \times 10^{-5}$	$7.3 \times 10^{-2} \pm 1.5 \times 10^{-3}$
Round 9	-4.5 ± 0.7	$9.7 \times 10^{-2} \pm 2.9 \times 10^{-2}$	$7.1 \times 10^{-3} \pm 4.9 \times 10^{-5}$	$7.5 \times 10^{-2} \pm 1.4 \times 10^{-3}$
Round 18	-5.2 ± 0.6	$7.2 \times 10^{-2} \pm 2.5 \times 10^{-2}$	$7.0 \times 10^{-3} \pm 5.7 \times 10^{-5}$	$7.4 \times 10^{-2} \pm 8.9 \times 10^{-4}$

3.4 Discussion

Because signal peptides do not contribute to the enzymatic activities of MBLs, the role of mutations in the signal peptides has remained elusive for years. In this study, the unknown functionality of the mutations in the signal peptide is analyzed by various approaches. First, I performed a series of dose-response curve assays to confirm a correlation between the phenotype of sequenced variants and the phenotype of the variants' respective populations. In Figure 3.2, the three enzymes universally showed a strong positive correlation (> 90%) between IC_{50} and MIC. In addition, the Hill coefficients of the variants exhibited a highly steep slope (|Hill coefficient| > 1), implying that the MBL mutants are linked to a profoundly narrow range of phenotypic diversity. Hence, these results support that the individual variants' mutational

patterns can be representative of the respective populations that gradually gain antibiotic resistance over the experiment.

The sequence profiles of the three enzymes showed the non-synonymous mutations that were repeatedly fixed at multiple positions and propagated through sub-divided populations. As the mutations accumulated, the mutation-driven sequence convergence was identified at round 6 for NDM-1 and round 9 for VIM-2 and IMP-1. What could be the driving forces behind the sequence convergence when almost any mutations are tolerable in signal peptides of MBLs? This phenomenon may be explained by two elements suggested by Iriarte et al., 2013: mutational bias and natural selection. Mutational bias may cause such convergence during the adaptive evolution. Convergence is potentially due to the inherent, limited availability of beneficial mutation for each species. As a result, this bias can create a pattern. The identified mutational substitutions in the signal peptides of MBLs are the outcome of adaptive evolution in *E. coli* to confer enhanced antibiotic resistance.

Natural selection may be another factor that can lead to sequence convergence. To respond to environmental conditions, gene expression needs to be regulated, and translation is a crucial step that impacts the overall protein synthesis. At the translational stage, translation speed and accuracy are largely regulated by the use of optimal (preferred) codons, especially for highly expressed genes showing an adaptation to the translation machinery and tRNA abundance (Klumpp et al, 2012, Plotkin et al, 2011; Stoletzki et al, 2007, Quax et al, 2015; Zhou et al, 2016; López et al, 2019; Gilchrist et al, 2015). To that end, the adapted codon usage gets reinforced for gene expression regulation by natural selection acting at the level of translation speed and

accuracy. Overall, only a handful of mutations from an inherent range may be allowed for fixation in the signal peptides of MBLs to distribute more adaptive codon usage in *E. coli* to elevate gene expression. This hypothesis provides an implication of how the translation stage of the phylogenetically distanced microorganisms can cause functional constraints, such as codon usage, inducing variations in MBL gene expression.

In this study, i) the translation rate, ii) the translation initiation rate, iii) the mRNA folding energy prior to 30S binding subsequences, and iv) the ribosome density were analyzed using bioinformatics, and compared with MIC. The bioinformatic analysis showed that mutations can manifest with different phenotypes during the adaptive evolution in VIM-2, NDM-1, and IMP-1. After acquiring mutations in the signal peptides, the round 18 VIM-2 variants were predicted to have an increased translation rate, translation initiation rate, and reduced local mRNA folding energy. All these mutation-driven changes showed a positive correlation with the antibiotic resistance measured by MIC. These results imply that the mutations influence not only the translation initiation, which is a rate-limiting step of translation (Laursen et al, 2005), but also the elongation step of translation. The round 18 VIM-2 variants may have gained optimal codons, particularly abundant in the *E. coli* system, in their transcripts because translation rate and accuracy are highly correlated with tRNA abundance (Klumpp et al, 2012, Plotkin et al, 2011; Stoletzki et al, 2007, Quax et al, 2015; Zhou et al, 2016; López et al, 2019; Gilchrist et al, 2015). In a previous study, López et al. 2019 observed that the chromosomal genes that were more comparable to the ancestral gene revealed codon usage better adapted to *Sinorhizobium meliloti* tRNA pool, leading to high gene expression. In the case of NDM-1, mutational substitutions were predicted to induce an elevation in translation initiation rate and local mRNA

folding energy. These two changes increased with MIC as well. What was unique about NDM-1 was that both the translation rate and the ribosome density showed inverse correlations with MIC. According to the bioinformatics data, the translation rate and ribosome density remained the same as those of the wild-type, while the antibiotic resistance of the evolved variants continuously increased. This result may indicate that the mutations for the NDM-1 variants, in particular, optimized the translation initiation step to reduce the overall time it takes for the rate-limiting step to increase the protein production. IMP-1, which was predicted to possess a local mRNA folding energy already ideal to *E. coli*, revealed a mutation-mediated ribosome density change correlated with MIC. This result may indicate how important the coding sequence can be at the translation initiation region to elevate the translation initiation efficiency. Because ribosome density can be an implication of how frequently translation initiation occurs for translation, the predicted elevation in ribosome density implies that the mutations in the signal peptide improved the translation initiation efficiency to a moderate level resulting in increased protein production. It is worth noting that the variants at round 3 trajectory of the three orthologs showed the most variability in mRNA folding energy, translation rate, translation initiation rate, and ribosome density as shown in Figure 3.5 and Figure 3.7. It may be because of a noticeable number of stochastic mutations identified from the round 3 variants. The seemingly moderate level of correlation may be, in part, due to the diverse stochastic mutations at the early stage of the directed evolution trajectories.

Owing to the most notable change in the 5th codon position, I conjectured possible hypotheses about how mutations at the codon position, together with the 3rd, 4th, and 6th codon positions, contribute to the increase in the efficiency of MBL enzyme synthesis. First, the nucleotide at the

5th codon position may influence the dynamics of ribosome interaction that can impact on both early elongation and the processivity of protein synthesis (Verma et al, 2019). In a recent study, Verma et al. 2019 identified that the protein synthesis efficiency of eGFP variants could be escalated by altering codon positions from 3 to 5. This result showed a positive correlation between a level of protein expression and a codon arrangement at N-terminal sequences. According to the sequence analysis of MBLs in Table 3.4, the three enzymes universally placed rare codons, which all belong to Group II, at the 4th codon position (Table 3.5). In fact, another rare codon was identified at the 5th codon position for round 18 variants of VIM-2 and IMP-1. In the case of round 18 variants of NDM-1, a low-usage codon was placed at the 5th codon position. These results imply a potential translation arrest that may lead to ribosome drop off from the enzymes' mRNA transcripts for a subsequent recycling. Fascinatingly, as Table 3.4 shows, VIM-2 and IMP-1 possessed one of the high-scoring codons at the 3rd and partially at the 4th codon position which was conserved throughout the entire trajectories. Unlike the other two enzymes, wild-type NDM-1 was devoid of a codon with a high-scoring mRNA at the region. After being adapted in *E. coli*, round 18 variant of NDM-1 gained the non-synonymous mutation, AAG, at the 5th codon position and captured high-scoring codons, AAG and ATT at the 5th and 6th positions (Figure 3.4). Although the high-scoring codon sequences were not arranged exactly at the 4th and 5th position as Verma et al. 2019 and Zahurancik et al. 2020 suggested, the mutation-mediated high-scoring codon arrangement may promote gene expression elevation presumably by discouraging a ribosome drop off and improving a ribosome processivity. According to Zahurancik et al. 2020, even with identical translation initiation complex formation, ribosomes engaged in a high-scoring mRNA as listed in Figure 3.4, were able to complete translation by 84%, whereas the ribosomes with a low-scoring mRNA ended up completing the translation by

27%, implying the significance of mRNA nucleotide compositions and the ribosome processivity (Verma et al 2019; Zahurancik et al 2020). All in all, this study has a significant implication that the mutation at the 5th codon of the enzymes may help ribosomes avoid translation arrest and encourage an elongation commitment that are essential steps in early elongation processes. Ultimately, the improved ribosome processivity may have promoted the efficiency of protein synthesis, resulting in increased gene expression, especially for NDM-1.

Second, mutational substitution at the 5th codon position together with the 4th codon position may act as a translational ramp that regulate ribosome allocation on mRNA transcripts and that improves the efficiency of protein synthesis. The importance of the translational ramp is that it delays ribosome speed at a specific position of a mRNA transcript and minimizes a chance of ribosome collisions during a later stage of elongation (Tuller et al, 2011; Verma et al, 2019). Ultimately, it helps avoid activation of the ribosome quality control mechanism that is stimulated by collided ribosomes. In other words, the translational ramp is a significant regulatory mechanism that governs protein levels based on the interplay among tRNA abundance, codon usages in mRNAs, ribosomes, and nascent polypeptide chain reactions (López et al 2015; Verma et al 2019). In fact, a recent study suggested that codons 3-5 and their amino acids form a translational ramp in a sequence-dependent manner in *E. coli* (Zahurancik et al 2020). Although it is possible that unfavorable amino acids may also be preferred together with the rare codon usage (Zahurancik et al 2020; Verma et al 2019), the inconsistent non-synonymous substitution, N5K for NDM-1, L5S for VIM-2, and S5L/S5R for IMP-1, implies that the translational ramp formation for MBLs may be nucleotide sequence dependent.

A mild level of correlation was also detected from a notable shift in the folding energy of the mRNA sequence prior to the 30S complex interaction in the case of NDM-1 and VIM-2. The mutations in the signal peptide of IMP-1 did not alter the local mRNA folding energy in the evolved variants probably because of the already suitable mRNA secondary structure of the wild-type in *E. coli*. The unfolded secondary structure of 5' mRNA was widely recognized, especially in prokaryotes, to regulate ribosome accessibility (Klumpp et al 2012; Zhang et al 2009; Kudla et al 2009). For an mRNA without a Shine-Dalgarno (SD) sequence, called a leaderless mRNA, a higher degree of unfoldedness than that of mRNA with an SD sequence was strongly necessary (Scharff et al 2011). A noticeable unfolded mRNA structure was identified from a region around -35 to +5 nucleotides of mRNA, where the 30S ribosome subunit binds, and the trend was also shown in α- and γ-proteobacteria with an SD sequence (Scharff et al 2011). Previous studies suggested that the codon adaptation index (CAI) and the folding free energy (ΔG) of the upstream region of mRNA were positively correlated with mRNA half-life as well (Dressaire et al 2012; Tuller et al 2010; Kozak 2005). In this investigation, i) the entire P_{bla} region, which possesses partial SD sequence, AGG which is nine nucleotides upstream from the start codon, and ii) the 35 downstream coding sequences from the start codon were analyzed for the mRNA folding energy shift. Because the non-synonymous mutations in the signal peptides were distributed up to the 11^{th} codon position for all enzymes, mRNA folding energy shift was predicted to be fully mediated by the mutational substitutions. The result implies a crucial role of the non-synonymous mutations that were predicted to reducing mRNA folding energy. In fact, three out of the four non-synonymous mutations detected in the round 18 variants of NDM-1 belonged to rare codons, whereas two-thirds of the non-synonymous mutations were rare codons for the round 18 VIM-2 variants. The variants of IMP-1 revealed four rare codons out of five

non-synonymous mutations in the round 18 trajectory. As determined in previous studies, the AU enriched rare codons are widely identified at the N-terminus and are attributed to reducing local mRNA folding energy, resulting in unfolded mRNA secondary structure (Goodman et al 2013; Scharff et al 2011, Pop C et al 2014; Duval et al 2013; Chursov et al 2013; Gu et al 2010). Overall, the non-synonymous mutations may induce mRNA folding energy shift close to the translation initiation region to accommodate ribosome complexes effectively.

Obviously, the mRNA secondary structure change and the translation initiation rate increase are only a part of complicated biological pathways that a signal peptide is involved in to improve MIC. In the process, a signal peptide interacts not only with residues in 2b and 7 transmembrane helices (TMHs) of the lateral gate to unlock a plug helix from the channel but also with SecA ATPases to initiate mature peptide translocation (Jelger et al 2012). This implies that these mutations may participate in multiple essential steps to induce efficient mature peptide translocation. Previous studies indicate the importance of a hydrophobicity threshold and the order of amino acids in the h-region of a signal peptide influencing the SecA ATPase activity, which governs the efficiency of a mature peptide transport (Zimmer et al 2008; Duffy et al 2010; Auclair et al 2010; Kebir et al 2002). It is possible that the increased polar residue(s) and the altered amino acid arrangement at the h-region by the non-synonymous substitutions may contribute to a dramatic increase in enzyme expression in the periplasm.

According to Moreira et al. 2019, *Klebsiella oxytoca* and *Enterobacter asburiae* also exhibited high-scoring sequences at 3-5 codon positions together with *Escherichia coli*, implying a positive selection applied to these bacterial species by a similar extent (Moreira et al 2019). In

addition, LaRue et al. 1976 presented the divergence of the isoaccepting tRNA sequences among different species (LaRue et al 1976). In line with LaRue's finding, the most distantly related species revealed noticeable differences in the optimal codon usage and the greatest frequency of synonymous and non-synonymous mutations within the sequences (Klumpp et al 2012). These lines of evidence may be a subtle and yet substantial piece of information to understand how phylogenetic distance of bacterial species raises functional constraints and impacts on MBL ortholog gene expression, as shown in the preliminary data.

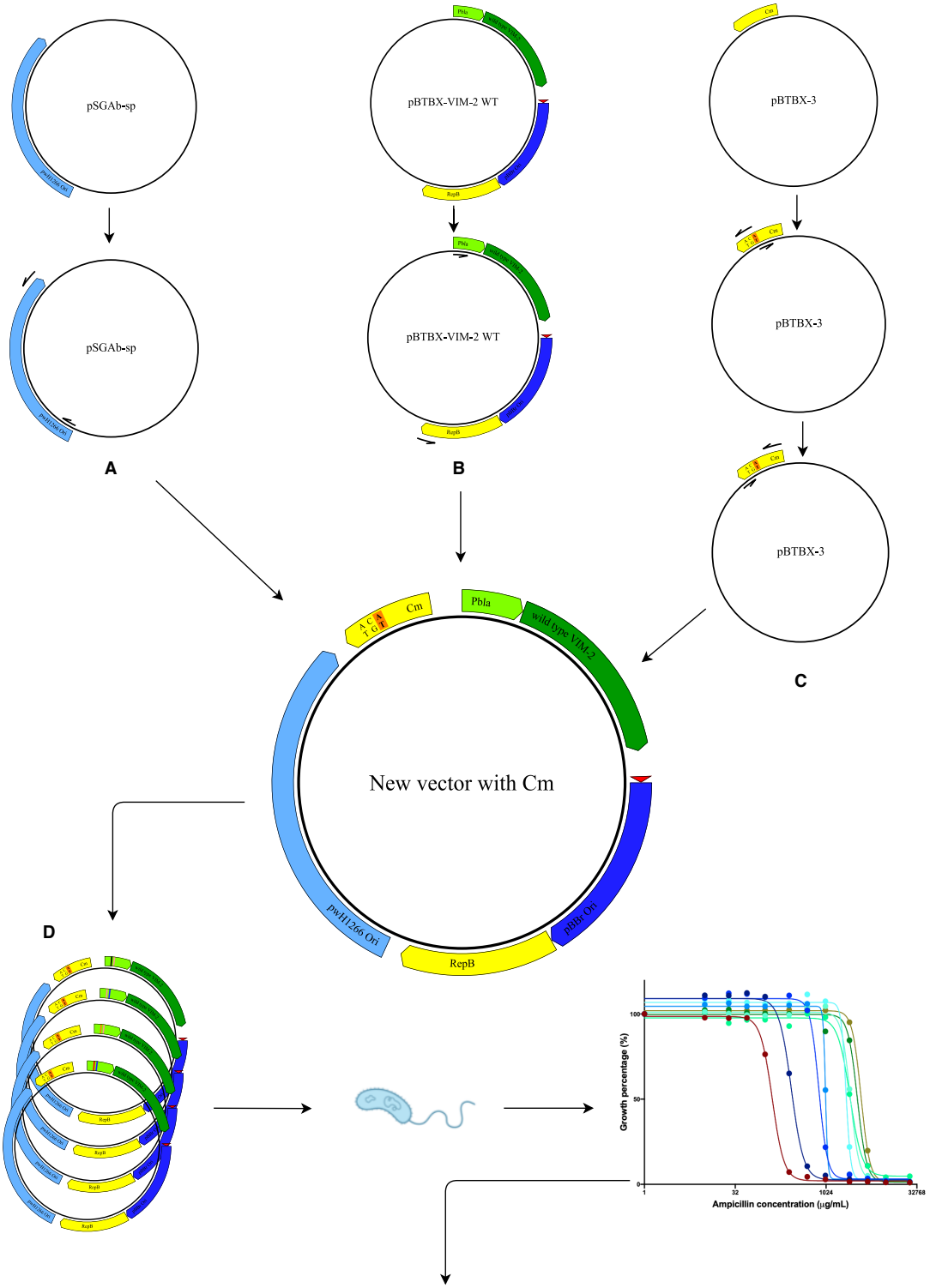
The bioinformatic analysis focuses only on the nucleotide sequences of MBL orthologs that gained multiple synonymous and non-synonymous mutations to develop advantageous phenotypes in *E. coli* and, ultimately, to confer enhanced antibiotic resistance. Consequently, the analysis of the sequence profiles may be a piece of meaningful evidence that represents the functional role of the mutations in the signal peptides. To the best of my knowledge, this is the first time that these mutations in the signal peptides of VIM-2, NDM-1, and IMP-1 are related to multiple factors involved in translation events, which may correlate with the enzymes' gene expression elevation. This result implies that mutations in the signal peptides could be a significant part of the adaptive mechanisms of MBLs in different host organisms to induce protein expression level. This phenomenon is perhaps to compensate the physicochemical limit of enzymes for their activities. This is an interesting perspective of evolution in action to optimize the fitness effect in a dynamic environment.

Chapter 4: New vector construction for transformation into *Acinetobacter baumannii, Escherichia coli, Pseudomonas aeruginosa,* and *Pseudomonas putida*

4.1 Introduction

In nature, VIM-2, NDM-1, and IMP-1 genes are widespread among pathogens due to transfer *via* HGT. To transform the MBL genes to *Acinetobacter baumannii, Escherichia coli,* and *Pseudomonas aeruginosa,* a broad host range (BHR) vector plays a crucial role. The previously available BHR vector, pBTBX-2, is perfectly compatible with *Escherichia coli*, *Pseudomonas aeruginosa*, and *Pseudomonas putida.* Unfortunately, the vector is not compatible with *Acinetobacter baumannii* and *Acinetobacter baylyi*. To circumvent the problem, a new BHR vector is synthesized by ligating three different pieces of genes of interest. The first piece of DNA is combined with three distinct gene segments: the pBBR1 ori, repB, and wild-type VIM-2. The second piece of DNA is the pWH1266 ori, another origin of replication. The last segment of DNA encodes gentamicin (Gen), which is a selective marker. At first, chloramphenicol (Cm) was used as a selective marker for the new vector synthesis. However, owing to the high Cm resistance, it was later replaced by Gen. The pWH1266 ori is integrated into the plasmid along with other pieces of DNA to enhance the new vector's compatibility, especially with *Acinetobacter spp*. To synthesize the vector, the individually amplified fragments of DNA are cleaved and ligated together by the Golden Gate assembly technique. During the new vector synthesis, the poorly expressive TEM-1 promoter in P_{bla} is modified by specifically designed primers. The TEM-1 modified BHR plasmid is compatible with *Acinetobacter*

baumannii, *Acinetobacter baylyi, Escherichia coli, Pseudomonas aeruginosa*, and *Pseudomonas putida* and hold promise for future antibiotic research.



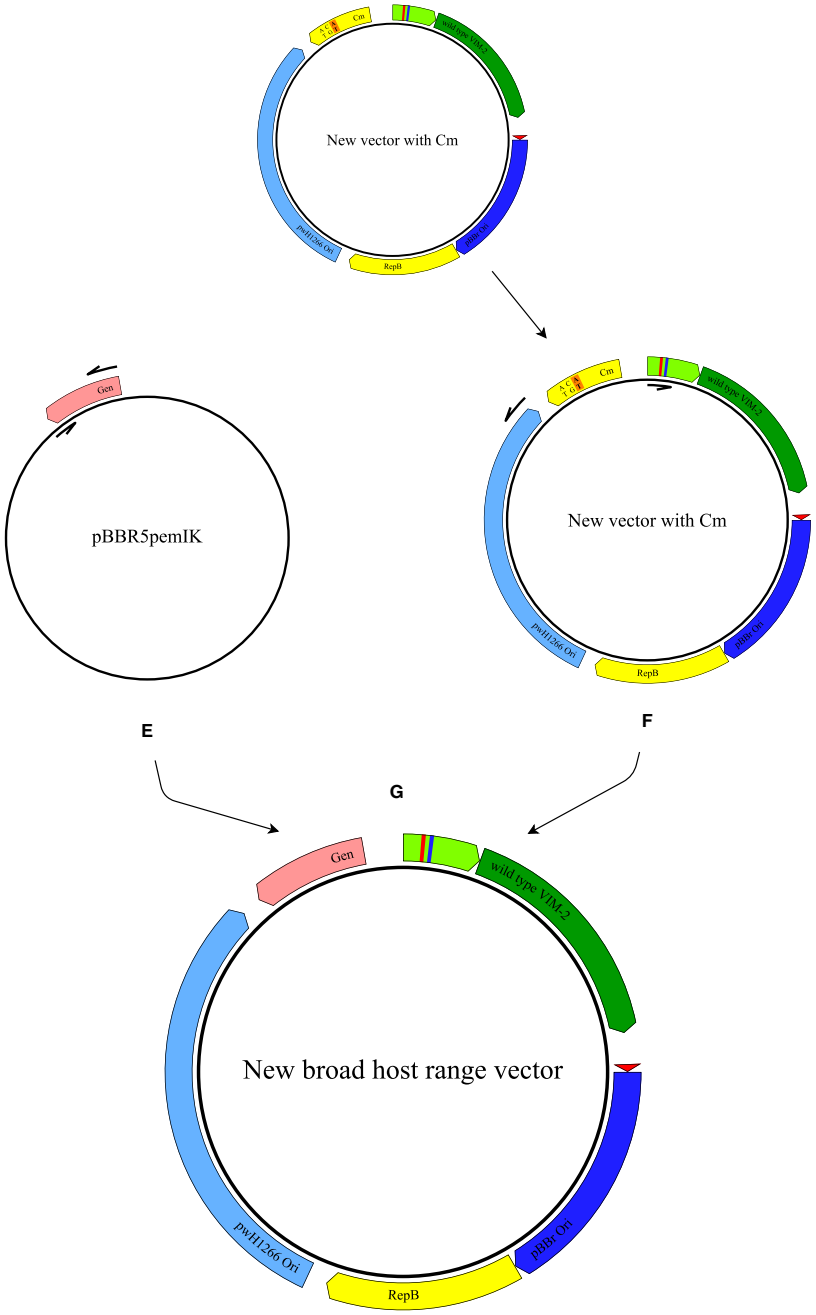


Figure 4.1. Overview of new broad host range vector synthesis.
(A) pSGAb-sp vector is to amplify pWH1266 ori. The black arrows around the vectors represent primers amplifying the genes of interest. (B) pBTBX-VIM-2 WT is to amplify pBBR1, repB, and wild-type VIM-2 gene. (C) pBTBX-3 is to amplify chloramphenicol (Cm). Nucleotides highlighted in orange indicate the region where *NcoI* is removed by the site-directed mutagenesis (SDM) technique. (D) An array of new broad host range vectors with modified TEM-1 promoter are shown. Red and blue vertical lines in the P_{bla} region indicate the modified regions. (E) pBBR5pemIK is to amplify gentamicin (Gen). F) The newly synthesized vector segment except the Cm fragment is amplified to construct the final plasmid construct shown in G. (G) Final construct of the newly synthesized broad host range vector is shown at the bottom.

4.2 Materials and methods

4.2.1 *NcoI* restriction site removal from Cm of pBTBX-3

The *NcoI* restriction site from chloramphenicol (Cm) selective marker, originating from the pBTBX-3 vector, was removed by using the forward, 5' TTTTTGGTCTCTCAC**A**ATGGGCAAATATTATACGCAAG, and the reverse primer, 5' TTTTTGGTCTCT**T**GTGAAAACGGGGGCGAA 3'. The bold letters indicate a nucleotide substitution causing a synonymous mutation at Thr (ACC to ACA). For the PCR reaction, 20 ng of template DNA was used along with i) 5x Q5 polymerase reaction buffer (New England Biology), ii) 10mM Deoxynucleotide (dNTP) Solution Mix (NEB), iii) 1 µM forward and reverse primers, iv) Q5 High-Fidelity DNA Polymerase (New England Biology), and v) Milli-Q water. The PCR condition was set up for 25 cycles of reaction to avoid potential mutations in the wild-type VIM-2 insert. The PCR reaction was carried out based on the following condition: an initial denaturation (98 °C for 10s), followed by 25 cycles of denaturation (98°C for 20s), annealing (62 °C for 20s), and extension (72 °C for 2min). Then, the PCR product was incubated with *DpnI* (10U/µL 2500U Fisher Scientific) at 37 °C for an 1 hour to remove the original plasmid DNA. After the *DpnI* treatment, the PCR products were subsequently purified with the EZNA®Cycle Pure Kit (Omega Bio-Tek). Then, the PCR product was cleaved by the Type IIS restriction enzyme, *BsaI* 5,000 units (New England Biology) and ligated by T4 DNA Ligase 5U/UL 1000U (Fisher Scientific) concurrently, which is known as the Golden Gate assembly technique. After an 1 hour of incubation at room temperature, the gene of interest was purified by micro Elute DNA Cleanup Kit (VWR). The purified DNA product was transformed with *E. cloni* 10G *E. coli* cells and incubated overnight at 37 °C. The next day, the individual colonies

were selected for colony PCR to verify gene insert using the primers stated above. After confirmation, the plasmid DNA was extracted by QIAprep Spin Miniprep Kit (Qiagen).

4.2.2 Restriction enzyme digest diagnosis using *NcoI*

The original pBTBX-3 and *NcoI* restriction site removed plasmid DNA were individually digested by the *NcoI* (Fisher Scientific). The reaction mixture contained 500 ng of either original pBTBX-3 or modified pBTBX-3 plasmid DNA that was combined with Fast digest buffer (Fisher Scientific), *NcoI*, and Milli-Q water. The reaction was carried out for 20 min at 37 °C. The diagnosis reaction was stopped by adding 6x loading dye buffer and analyzed by DNA electrophoresis.

4.2.3 Amplification of pBTBX-2, pWH1266 ori, and modified Cm regions by PCR reaction

By amplifying each piece of the gene inserts with designed primers, the new broad host range vector could orient the genes of interest in the desired direction. Each primer contained a *BsaI* restriction enzyme site followed by a cut site, four nucleotide ligation sequences, and a plasmid DNA binding region. Based on the ligation sequences, the individual pieces of gene insert were oriented and ligated in a direction shown in Figure 4.1A, B, and C by the Golden Gate assembly technique. The primers for pBTBX-VIM-2 were 5' TTTTTGGTCTCTTCTGTTCTGCTATGGAGGTCAGGTA 3' for the forward and 5' TTTTTGGTCTCTTTCCGGAACCTCTTACGTACTACC 3' for the reverse. The pBTBX-VIM-2 forward and reverse primers were coupled with Cm forward and pWH1266 ori forward primer, respectively. The forward primer for pWH1266 ori was 5'

TTTTTGGTCTCTGGAAGCGGCCGC GGATTTTAACATTTTGCGTTGTTC 3', and the reverse primer was 5'TTTTTGGTCTCTGACTGCGGCCGCGATCGTAGAAATATCTATGATTATCTTGAAG 3'. These primers contained a *NotI* restriction site in the case of vector modification. The pWH1266 ori forward and reverse primers were coupled with pBTBX-VIM-2 reverse and Cm reverse primer, respectively. The primers for the *NcoI* restriction site removed Cm fragment were 5' TTTTTGGTCTCTCAGAGGATCCTACGTAAGAGGTTCCAACTTTC 3' for the forward and 5' TTTTTGGTCTCTAGTCCCTGACCTCCATAGCAGAAAG 3' for the reverse. *BamHI* restriction enzyme site was encoded in both primer sequences. Cm forward and reverse primers were coupled with pBTBX-2 forward and pWH1266 ori reverse primers, respectively. The PCR reaction was run under the following condition described here. In the case of pBTBX-VIM-2, the 2,526 bp insert was initially denatured at 98 °C for 10s. Then, the PCR reaction ran for 25 cycles of denaturation (at 98 °C for 20s), annealing (at 59 °C for 20s), and extension (at 72 °C for 75s). The same PCR condition were applied to pWH1266 ori and Cm segments except for the annealing temperature and extension time. The pWH1266 ori, which is 1,337 bp, was annealed at 58 °C for 20s and extended at 72 °C for 40s. The Cm fragment, 825 bp, was annealed at 59 °C for 20s and extended at 72 °C for 30s.

4.2.4 Ligation of three individual pieces by the Golden Gate assembly and restriction enzyme digest diagnosis using four restriction enzymes

The amplified PCR products were treated with the *DpnI* over an 1 hour at 37 °C. Then, the products were purified by EZNA®Cycle Pure Kit. The purified gene fragments were mixed in a reaction solution containing 10x T4 DNA ligase buffer and *BsaI* reaction buffer in a 1:1 ratio,

BsaI, T4 DNA ligase, and Milli-Q water and then incubated at room temperature for an 1 hour. The ligated plasmid sample was purified by micro Elute DNA Cleanup Kit. Then, the synthesized vector was eluted in 10μL of Milli-Q water. The plasmid DNA was examined by diagnostic restriction enzyme digest method that was applied in the previous step using *NcoI*, *HindIII* (Fisher Scientific), *XhoI* (Fisher Scientific), and *BamHI* (Fisher Scientific).

4.2.5 TEM-1 promoter modification

The sequence of the TEM-1 promoter was altered by introducing random nucleotides. The bold letters labeled "N" within the primer sequences indicate positions where random nucleotides were introduced. The changes in nucleotide sequence affected the binding interactions with transcription factors, which shifted the strength of the TEM-1 promoter for wild-type VIM-2 expression. Colonies harboring plasmids expressing a high level of wild-type VIM-2 were selected owing to the enhanced antibiotic resistance. For this experiment, the forward primer, 5' AAAAAGGTCTCCATGA GCGGATACATAT**N**TT**N**AATGTATTTAGAA 3', and the reverse primer, 5' TTTTTGGTCTCCTCATGA**N**AA**N**AATAACCCTGATAAATGCTTCAATA 3', were designed. Then, the primer sequences were synthesized by IDT. The purified new broad host range plasmid vector was amplified by a Q5 High-Fidelity DNA polymerase using the primers listed above (New England Biology) in the Q5 PCR reaction. The reaction was carried out for 30 cycles. The plasmid DNA was denatured initially at 98 °C for 10s. Then, during the reaction cycle, the plasmid DNA was denatured at 98 °C for 20s, annealed at 58 °C for 20s, and extended at 72 °C for 1 min. The original plasmid was removed by being treated with *DpnI* enzyme and purified by micro Elute DNA Cleanup Kit. The TEM-1 modified new broad host range vector was re-circularized by the Golden Gate assembly technique.

4.2.6 Gentamicin replacement of the new vector

To replace chloramphenicol (Cm) with gentamicin (Gen), the Gen fragment from pBBR5pemIK was amplified by the forward primer 5' TTTGGTCTCAGAGTCGTGGAAACGGATGAAGGC A 3' and the reverse primer 5'AAAGGTCTCATCAGAGCCGATCTCGGCTTGAACGAATTG 3'. The new broad host range vector was also amplified except the Cm fragment as shown in Figure 4.1F using the forward primer 5' TTTGGTCTCAACTCCTCTTACGTAGGATCCTTCTGCTATGG 3' and the reverse primer 5' AAAGGTCTCACTGACAGTACTGCGATGAGTGGCAGGG 3'.

Gen and the newly synthesized vector plasmid were individually amplified using Q5 High-Fidelity DNA polymerase in the PCR reaction. Both Gen and the vector DNA fragments were initially denatured at 98 °C for 10s, then introduced to 25 cycles of Q5 PCR reaction. In the case of Gen fragment, the pBBR5pemIK plasmid was denatured at 98 °C for 10s, annealed at 65 °C for 20s, and elongated at 72 °C for 1min. The new broad host range vector was denatured at 98 °C for 10s, annealed at 63 °C for 20s, and extended at 72 °C for 2min. As stated above, the amplified PCR product was treated with *DpnI* for an 1 hour at 37 °C to remove the original plasmid. Next, they were purified by EZNA®Cycle Pure Kit. The purified DNA fragments of interest were combined with the Golden Gate assembly reaction mixture. The mixture contained 10x T4 DNA ligase buffer and *BsaI* reaction buffer with the same ratio combined with T4 ligase, BsaI restriction enzymes, and Milli-Q water. The Golden Gate assembly reaction was carried out at room temperature for an 1 hour. After the incubation, the new host range vector was purified using micro Elute DNA Cleanup Kit and eluted in 10 μL of Milli-Q water.

4.3 Results

4.3.1 Additional *NcoI* restriction site from chloramphenicol region is removed by site directed mutagenesis

ATG is one of the transcription initiation codons for mRNA synthesis. Because of the naturally encoded ATG within the restriction site, *NcoI* was employed prior to the MBL gene inserts for cloning. To achieve an efficient gene expression regulation, the additional *NcoI* restriction site in the chloramphenicol (Cm) fragment was modified by site-directed mutagenesis (SDM), causing a synonymous mutation at Thr (ACC to ACA). Then, the manipulated Cm fragment was cleaved and ligated by Golden Gate assembly. The vector plasmid was transformed into Ecloni10G cells and extracted from five random variants shown in lanes from four to eight in Figure 4.2. The black arrows indicated the modified plasmid variants selected for amplification. Then, the modified region of Cm fragment was confirmed by performing a diagnostic restriction enzyme digest experiment using *NcoI.* The original and modified Cm fragments were amplified by primers containing *BsaI* sites and ligation regions for the pBTBX-2 and pWH1266 ori. According to the gel image in Figure 4.3A, the amplified original (lane 3) and modified Cm fragments (lanes 4, 5, and 6) were aligned at an identical region. However, after the *NcoI* treatment, the modified Cm fragments (lanes from 4 to 6 in Figure 4.3B) remained intact, whereas the original Cm fragment was clearly digested by *NcoI*, resulting in two separate bands in lane 3 (Figure 4.3B). This result support that the *NcoI* site was abolished from the Cm fragment of the pBTBX-3 vector.

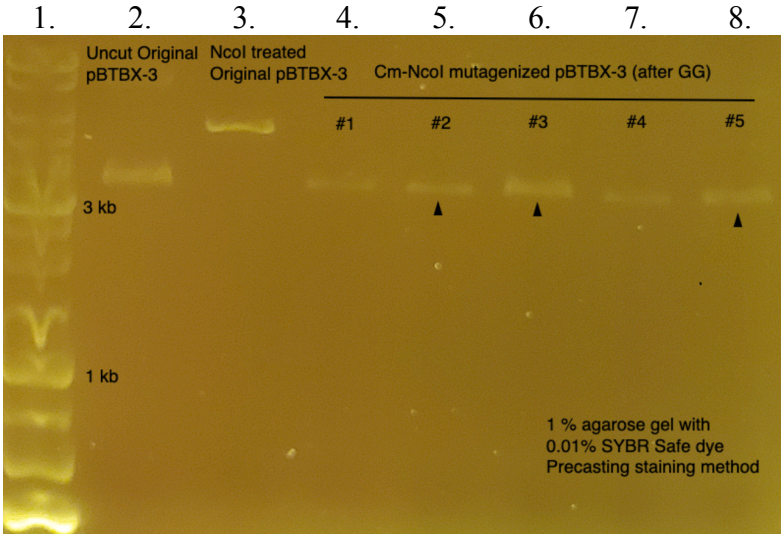


Figure 4.2. Electrophoresis gel image of modified chloramphenicol (Cm) segment of pBTBX vector.
NcoI restriction site at Cm segment is first removed by the site directed mutagenesis (SDM) method targetting Thr residue. To accurately estimate the sizes of DNAs of interest, a precasting staining method is utilized for the verification of the NcoI site deletion. For this experiment, 1% agarose gel is stained with 0.01% SYBR Safe dye. The original vector treated without and with *NcoI* is shown in lane 2 and 3, respectively. The modified ones are shown in lanes from 4 to 8. The DNA bands that are marked by small arrows indicate selected variants for the further experiment. This result implies that the plasmid vector is no longer reactive to the *NcoI* restriction enzyme due to mutagenesis.

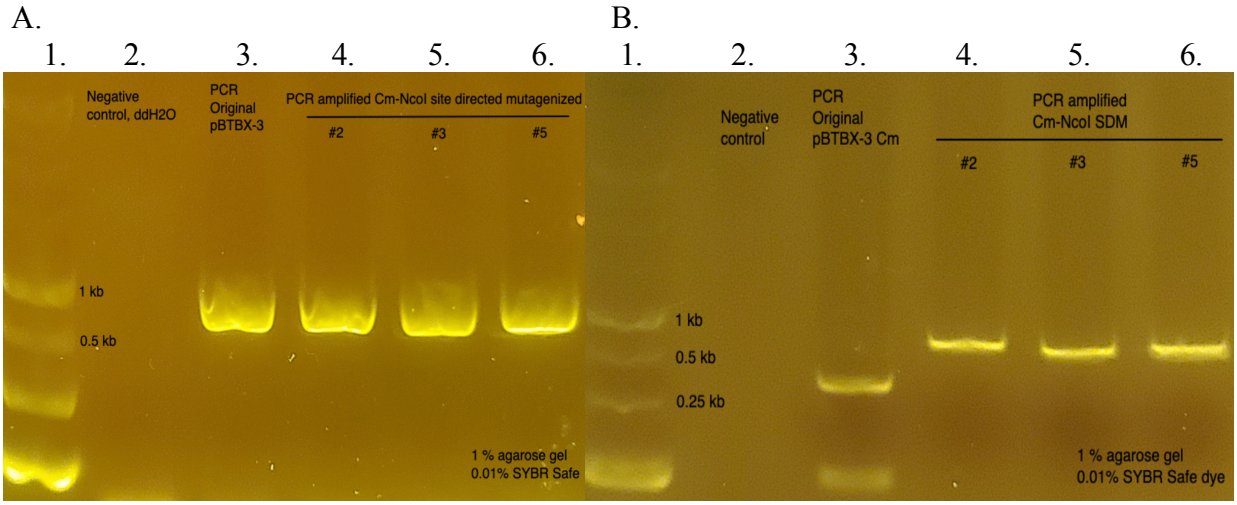


Figure 4.3. Amplification of NcoI site removed Cm segment, and NcoI diagnosis performed on the modified Cm segment.
Electrophoresis experiment is performed using 1% agarose gel mixed with 0.01% SYBR Safe dye. (A) Negative control in lane 2 contains Milli-Q water. The original pBTBX-3 vector is shown in lane 3. In lanes 4 to 6, the *NcoI* site removed Cm fragments are individually amplified and analyzed by electrophoresis. Each band originates from the plasmid DNA that was selected in the previous experiment. The size of the vector is near 1kb which is close to the actual band size, 825 bp. (B) Negative control in lane 2 contains Milli-Q water. As shown in lane 3, the original pBTBX-3 vector is divided into two pieces after treatment. The amplified genes of interest shown in lanes 4 to 6 are treated with *NcoI* to confirm that the restriction site is completely gone. The modified DNA fragments maintain their original band size even after the *NcoI* treatment.

4.3.2 Three pieces of DNA of interest are successfully amplified for the BHR vector synthesis

Modified Cm, pWH1266 ori, and pBTBX-VIM-2 fragments were amplified in triplicates. In Figure 4.4, the three bands on the right, positioned at around 2 kb, represented the gene of interest containing i) the pBBR1 ori, ii) the wild-type VIM-2, and iii) the RepB. The bands in the middle at around 1kb represented the pWH1266 ori. Lastly, the bands on the left, located slightly below 1kb, represented the Cm fragment. Both the pWH1266 ori and the Cm fragments were positioned at the sites that correspond to their expected band sizes, 1,337bp and 825 bp, respectively. However, the fragments on the right were located at the region approximately 500 bp lower than what was expected for the actual band size, which is 2,526 bp. The discrepancy may have occurred because of the amount of amplified gene fragments. In DNA electrophoresis, a gene migration depends on the charge and the size of DNA. It means that the amount of DNA can influence the migration speed. Based on the intensity of the bands, densely amplified fragments may have migrated faster than their expected speed, resulting in such discrepancy.

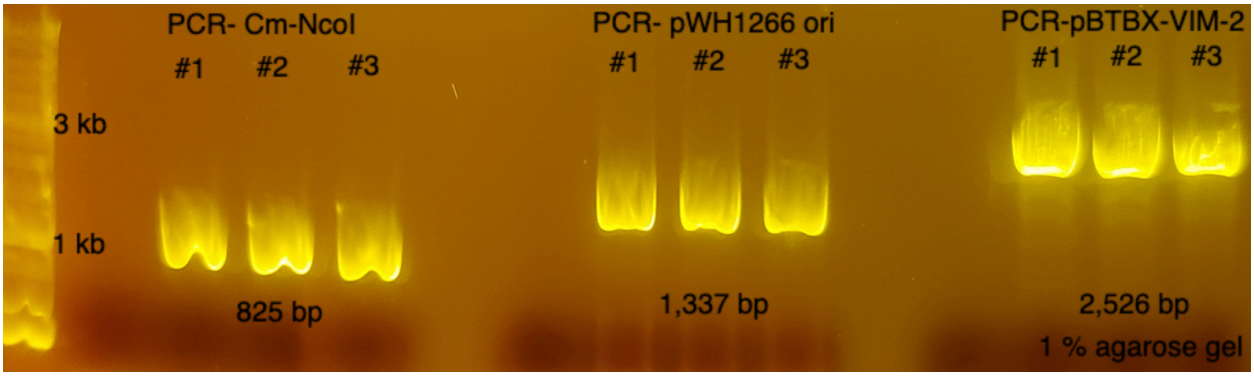


Figure 4.4. Amplified Cm fragment, pWH1266 ori, and pBTBX-VIM-2 WT.
Individually amplified DNA fragments (in triplicates) are analyzed using 1% agarose gel. The actual length of each gene insert is displayed at the bottom. PCR-Cm-*NcoI* represents the Cm fragments without *NcoI* site (left). PCR-pWH1266 ori represents the origin of replication that is compatible with *Acinetobacter baumannii* (middle). PCR-pBTBX-VIM-2 represents plasmid DNA with the wild-type VIM-2 gene (right).

4.3.3 New broad host range (BHR) vector is successfully synthesized

The amplified pieces of DNA were cleaved and ligated together by the Golden Gate assembly technique. Then, the ligated vector was run through DNA electrophoresis to estimate their sizes. As Figure 4.5 shows, the newly synthesized vector was presented in lane 3, indicated by red arrows. Two bands from the top aligned at around 4 and 5 kb of the DNA ladder. The size of the two bands matched closely to the expected size of the new vector plasmid, assuming that the BHR vector was in two different conformations. Even though the control pBTBX-VIM-2 fragment was undetectable in lane 2, the third band from the top appeared at around 2.5 kb in lane 3, matching the expected band size of pBTBX-VIM-2. The bottom band aligned with the Cm fragment, a control in lane 2, although it was difficult to verify its band size. It is possible that the two bands at the bottom were the remnants of pBTBX-VIM-2 and Cm fragments after the Golden Gate assembly reaction.

Then, the vector in lane 3 of Figure 4.5 was examined using two restriction enzymes in two different combinations: i) *NotI* only (left) and ii) *NotI* plus *BamHI* (right). Two *NotI* sites were integrated around the pWH1266 ori fragment. Therefore, the *NotI* treatment could result in two separate bands, one band representing the pWH1266 ori at 1,337bp, and another representing the rest at 3.311 bp. Out of the five variants, variants 4 and 5 displayed two visible bands (left) at around 1 kb and 3 kb, highlighted by red arrows in lanes 4 and 5 as shown in Figure 4.6. The diagnostic restriction enzyme digest was also performed using *NotI* and *BamHI* restriction enzymes. Because the *BamHI* site was integrated around the Cm fragment, the two-restriction enzyme treatment should produce three separate bands: i) the pWH1266 ori, ii) the Cm, and iii) the pBTBX-VIM-2. Again, variants 4 and 5 consistently showed three separate bands that

correspond to pBTBX-VIM-2 (2,526 bp), pWH1266 ori (1,337 bp) and Cm (825 bp) as marked by three red arrows (Figure 4.6 right). Overall, these results suggest that the new BHR vector is synthesized with the three pieces of gene inserts.

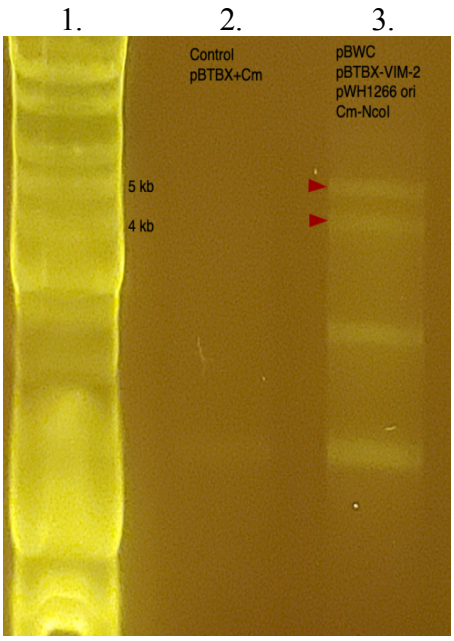


Figure 4.5. Image of a newly synthesized vector after Golden Gate assembly technique.
After treatment by Golden Gate assembly, the plasmid DNA is loaded to a 1% agarose gel mixed with SYBR Safe dye. The two bands indicated by two red arrows at around 4 and 5kb are the newly synthesized vector plasmid in two different conformations. The actual band size is 4,688 bp.

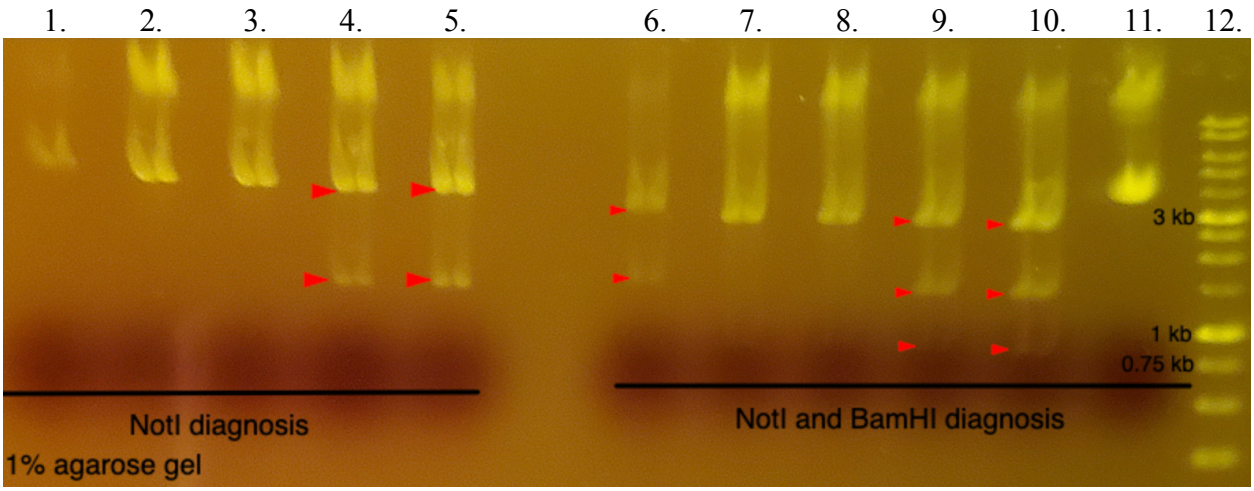


Figure 4.6. Diagnostic restriction enzyme digest test for the newly synthesized vector.
To verify that the new broad host range (BHR) vector is truly ligated by three gene fragments, the vector is digested by *NotI* and *BamHI* in two combinations: i) *NotI* only and ii) *NotI* and *BamHI*. A pair of *NotI* restriction enzyme sites is located prior to and after pWH1266 ori. Consequently, *NotI* treatment should generate two separate bands.

As shown in lanes 4 and 5, when BHR vector is digested by *NotI*, pWH1266 ori is cleaved off and appears at around 1.5 kb, indicated by red arrows at the bottom. The *BamHI* works in the same fashion and this restriction enzyme can isolate Cm fragment after cleavage. When the same vector is cleaved by *NotI* and *BamHI*, three bands are clearly generated in lanes 9 and 10, respectively. The three bands correspond to Cm (bottom), pWH1266 ori (middle), and pBTBX-VIM-2WT (top) fragments. They are all indicated by the red arrows in lanes 9 and 10. These lines of evidence support that the new vector is synthesized by the three pieces of gene fragments.

4.3.4 The wild-type VIM-2 gene lacks unnecessary mutations that might affect antibiotic resistance

A single colony harboring plasmid variant 4 or 5 was selected and inoculated into a liquid LB media supplemented with i) 34 μg/mL of Cm only and ii) 34 μg/mL of Cm with 12.5 μg/mL of ampicillin to prepare overnight cultures for a growth curve assay. Concurrently, the same variants were inoculated into LB liquid media supplemented with 0, 20, 50, and 100 μg/mL of ampicillin combined with 34 μg/mL of Cm. After six hours of incubation at 37 °C, visible growth was not observed from the samples with high ampicillin concentrations. However, after 24h incubation under the identical condition, OD_{600} of *E. coli* cells harboring each of the two plasmid variants became close to one even at 100 μg/mL of ampicillin combined with 34 μg/mL of Cm. This result implies that the new vector expresses wild-type VIM-2 at a slow rate. To confirm that the VIM-2 gene is devoid of unexpected mutations, the wild-type VIM-2 gene insert from variants 4 and 5 was amplified (lane 5 and 6) with i) the original pBTBX-VIM-2 vector (lane 3, a positive control), and ii) pBTBX-Cm, which was another vector control (lane 4, a previously prepared positive control) as shown in Figure 4.7. As panel A presents, the wild-type VIM-2 genes from lanes 3 to 6 were aligned at around 1kb which match the expected band size of 1,111bp. Then, the amplified wild-type VIM-2 from plasmid variants 4 and 5 were analyzed by the Sanger sequencing. The sequence analysis confirmed the absence of additional mutation.

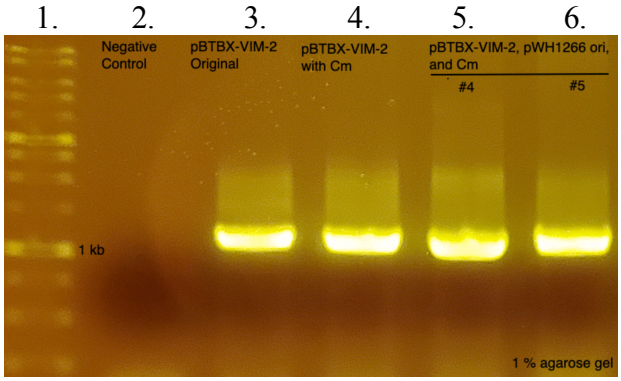
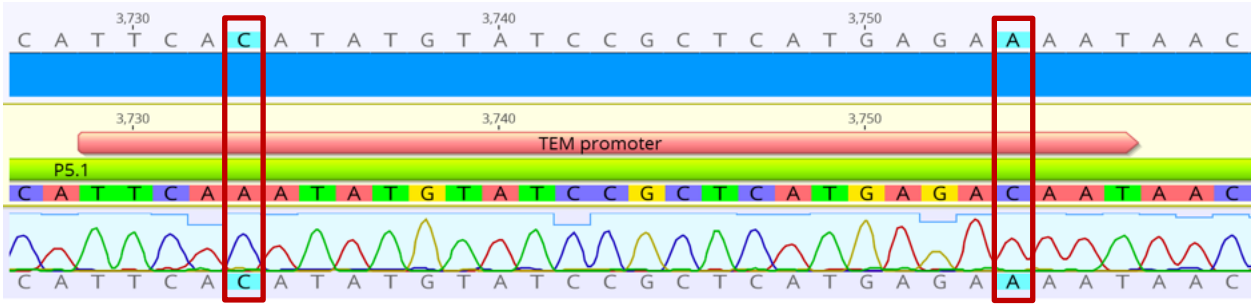


Figure 4.7. Amplification of VIM-2 gene to confirm its intact sequence.
The vector variant 4 and 5 are selected to specifically amplify wild-type VIM-2 sequence (1,111 bp) to analyze by Sanger sequencing. In lane 2, negative control includes milli-Q water instead of the DNA of interest and shows the absence of contamination. Two positive controls are used in the reaction: i) pBTBX-VIM-2 original vector in lane 3 and ii) pBTBX-VIM-2 with Cm, which is two pieces ligated vector in lane 4. As shown in lanes 5 and 6, the amplified gene of interest aligns well with controls in lanes 3 and 4 at 1kb.

4.3.5 TEM-1 promoter sequences are modified to have enhanced wild-type VIM-2 expression

When a dose-response curve assay using two-fold increasing ampicillin concentrations with 34 μg/mL of Cm was conducted, the plasmid variants 4 and 5 expressed wild-type VIM-2 to a level similar to the original pBTBX-VIM-2. It means that wild-type VIM-2 was expressed approximately 1.5-fold weaker than that of the previous vector system used in MIC by Dr. Socha. To induce a VIM-2 expression and to separate from the background resistance, the TEM-1 promoter sequences in the P_{bla} were modified. The promoter sequences were altered by primers that introduced random mutations at -35 and -10 regions, where the transcription factors bind and initiate transcription. After the modification, the diverse sequence library of TEM-1 was first treated with *DpnI*, and then cleaved and ligated by the Golden Gate assembly (Figure 4.1D). Out of 17 variants, six variants from both variants 4 and 5 showed ampicillin resistance, either identical to or 2.2-fold higher than the previous vector system. We selected a variant that showed a 1.2-fold more resistance in ampicillin resistance. The Sanger sequence analysis determined that

the original promoter consensus was modified to TTC ACA at -35 and GAA AAT at -10 as illustrated in the Figure 4.8.



Original consensus of *E. coli* promoter sequence: TTCAAA (-35) GACAAT (-10)
Modified consensus of *E. coli* promoter sequence: TTCA**C**A (-35) GA**A**AAT (-10)

Figure 4.8. Altered TEM-1 promoter sequences to enhance the strength of promoter.
To modify the TEM-1 promoter, primers containing four undefined nucleotides are used. The modified sequences contain A → C at -35 and C → A at -10 as indicated in red. The mutagenized sequences of the promoter are denoted by the two red boxes.

4.3.6 Chloramphenicol (Cm) fragment is replaced by Gentamicin (Gen) by Golden Gate Assembly technique

The new broad host range (BHR) vector was transformed into *Acinetobacter baylyi, Pseudomonas aeruginosa*, and *Pseudomonas putida*. Although it was not shown here, the bacterial hosts with and without the new vector were challenged by ampicillin, ceftazidime, meropenem, piperacillin in a series of dose-response curve assays. The purpose of the experiment was to estimate the background resistance and the level of resistance conferred by *P. aeruginosa* and *P. putida* with the new vector conferred antibiotic resistance against ceftazidime and piperacillin. However, the opposite trend was observed against ampicillin and meropenem. In the case of *Acinetobacter baylyi*, prominent antibiotic resistance against ampicillin was manifested by the new vector. *Pseudomonas putida* was unexpectedly resistant to Cm – up to 200 µg/mL – leading us to replace Cm by Gen as a selective marker.

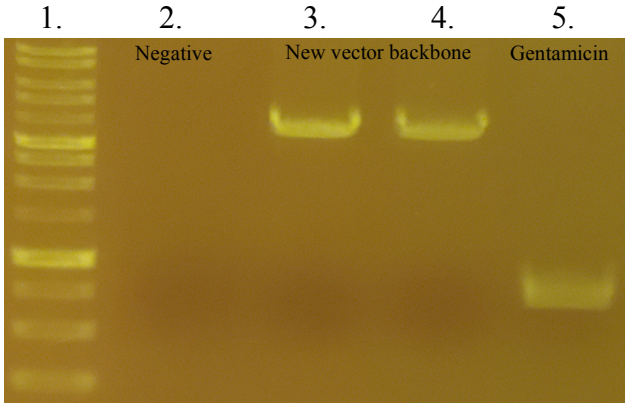


Figure 4.9. Amplified pBTBX-VIM-2 plus pWH1266 ori backbone DNA and gentamicin (Gen) fragment.
The actual band sizes for BHR and Gen are 3,976 bp and 804 bp, respectively. The amplified vector DNA segments are observed in lanes 3 and 4, aligned at around 3kb. The Gen is displayed in lane 5. In lane 2, the negative control contains milli-Q water.

The Gen selective marker and the plasmid DNA without the Cm fragment (lanes 3 and 4) were individually amplified to replace the Cm fragment (Figure 4.1E and F). According to Figure 4.9, vector DNA and Gen closely aligned to around the expected band size, 3,976 bp and 804 bp, respectively. Again, the Golden Gate assembly approach was applied to ligate the two pieces of gene fragments (Figure 4.1G and Figure 4.10). A diagnostic restriction enzyme digest experiment confirmed the new BHR vector synthesis.

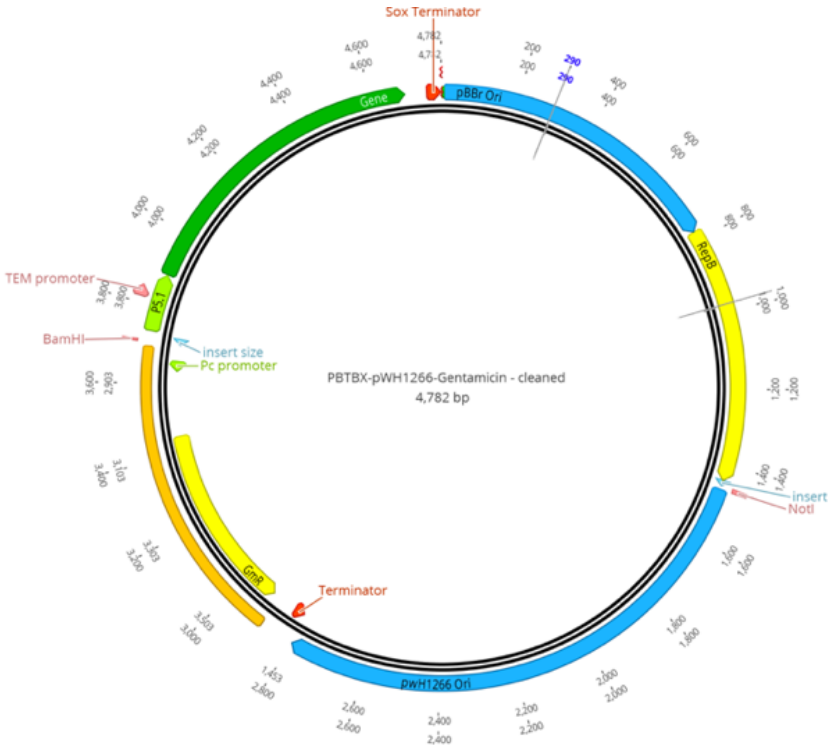


Figure 4.10. The map of a new broad host range (BHR) vector plasmid.
The overview of a newly synthesized vector shows integrated gene inserts: pBBR ori, RepB, wild-type VIM-2, pWH1266 ori, and Gen. The length of BHR is 4,782 bp. Two red marks indicate terminators and the salmon mark represents TEM-1 promoter from P_{bla} (P5.1).

4.4 Discussion

The essence of the experiment is to synthesize a new broad host range (BHR) vector that is compatible with *Acinetobacter baumannii, Acinetobacter baylyi, Escherchia coli, Pseudomonas aeruginosa,* and *Pseudomonas putida*. Because the level of MBL gene expression is tightly connected with the extent of antibiotic resistance, a consistent replication rate and a standardized protein expression level of the enzyme are significant in the investigation of mutation-driven MBL expression level changes, which ultimately affect the antibiotic resistance in different host organisms. The new BHR vector arranges two separate origins of replication: the pBBR1 ori and the pWH1266 ori. The pBBR1 ori exhibited a 4.3-4.4 copy number after an 8 and 24 h incubation in *E. coli*, respectively (Jahn et al 2016), but is not compatible with

the *Acinetobacter* spp. Although its copy number or stability mechanisms are obscure, the pWH1266 ori has been widely used for *Acinetobacter baumannii* (Lucidi et al 2018). In fact, López et al. successfully expressed metallo-β-lactamase proteins in *A. baumannii* using the pWH1266 ori, supporting its compatibility in *Acinetobacter* spp. (López et al 2019). Hence, the new vector system resolves a problem caused by plasmid vector and host incompatibility. Furthermore, the TEM-1 promoter sequences of P_{bla} were modified to enhance interaction with transcription factors. This modification brought the level of MBL gene expression to a similar point as that in an already established system in the previous experiment. In conclusion, the newly synthesized BHR vector can be utilized for various host organisms, enabling broader future research in mutational effects.

Chapter 5: Conclusion and future outlook

5.1.1 General summary

The goal of this thesis is to understand functionalities of mutations fixed in the signal peptide of NDM-1, VIM-2, and IMP-1 which are evolved and adapted in *E. coli* via the directed evolution experiment.

In Chapter 2, I explained the methodology of dose-response curve assay that is used to analyze average fitness of variants at a population level. The dose-response curve assay allows a systematic analysis of the fitness of a population with two parameters: the Hill coefficient and the IC_{50} value. The Hill coefficient is the slope at IC_{50} of the dose-response curves and estimates a fraction of mutations associated with various fitness effects. The IC_{50} value is indicates the average fitness of an entire population. One of the advantages of the assay is that the IC_{10} and IC_{90} can also be approximated based on the two parameters using the Hill equation. In short, a shifted IC_{50} value that is consistently observed in the dose-response curve clearly demonstrates enhanced antibiotic resistance, as the mutations are fixed in the sequence spaces of the three enzymes.

In Chapter 3, I performed a series of dose-response curve assays using eight directed evolution trajectories that Dr. Socha created and measured MIC. The rationale of conducting the assay is to verify a correlation between the phenotypes of the few sequenced variants and the phenotypes of the variants' respective population. The strong positive correlation between the MIC of individual variants and the IC_{50} values of the variants' population suggests that a limited number of sequence samples used in this work are a reasonable representation of the population. This is

further corroborated by the universally high absolute values of Hill coefficients which indicates that the phenotypic diversity of the directed evolution variants is confined in an extensively narrow range. I investigated the potential role of these mutations by conducting bioinformatic analysis. The analysis is focused on the translational stage. The result shows that VIM-2 exhibits a moderate level of correlation between mutation-mediated translation rate change and MIC. Moreover, IMP-1 shares a similar trend by increasing ribosome density. Additionally, a moderate and minor degree of folding energy shift in mRNA sequences prior to 30S complex binding is identified in NDM-1 and VIM-2, respectively. Lastly, NDM-1 and VIM-2 exhibit an elevated translation initiation rate that is crucial for efficient gene expression. These lines of evidence suggest the role of mutations in MBL's signal peptide that may enhance enzyme's gene expression in *E. coli* as an adaptive mechanism. Eventually, the elevated gene expression leads to the high antibiotic resistance to ampicillin.

In Chapter 4, I described a procedure for synthesizing a new broad host range vector (BHR) by the Golden Gate assembly technique. The BHR vector is synthesized using three distinct DNA pieces originating from three separate vector plasmids. The first segment, 2,526 bp, comprises three sub-regions: pBBR1 ori, repB, and wild-type VIM-2. pBBR1 ori is the origin of replication where the transcription initiates. The repB gene produces RepB proteins which are required for the plasmid to replicate at pBBR1 ori. Finally, the last segment belongs to wild-type VIM-2. To circumvent incompatibility with the *Acinetobacter baumannii*, pWH1266 ori is also ligated into the new vector along with other gene fragments. As a selective marker, the chloramphenicol (Cm) resistance gene is used for the vector synthesis. The additional NcoI restriction enzyme site is removed by site directed mutagenesis. To bring antibiotic resistance of a new vector

comparable to the resistance level of a previous vector, the TEM-1 promoter from P_{bla} region is manipulated and verified by a series of dose response curve assays using ampicillin. Lastly, because of the strong Cm resistance from *P. putida*, Cm is replaced with gentamicin (Gen) in the final step of vector construction. In conclusion, the new BHR vector is synthesized to be compatible with the *A. baumannii, E. coli, P. aeruginosa,* and *P. putida* for enabling future research of antibiotic resistance of these organisms.

5.1.2 Future outlook

In Chapter 3, I performed dose-response curve assays using variants of three enzymes created by the directed evolution experiment. In addition, I analyzed the nucleotide sequences of the orthologs using bioinformatics analysis. I was able to identify an implication of the mutations in the signal peptide for all enzymes. It is highly possible that the mutation-driven changes influenced the essential translational steps in *E. coli*, resulting in MBL gene expression elevation. The moderate level of correlation between MIC and the mutation-mediated changes further supports the role of mutation in the signal peptides with advantageous phenotype development. However, I was unable to identify how the mutations in the signal peptide influence MBL translocation via the SecYEG system. Signal peptides bind to a multidomain interface containing mainly the polypeptide cross-linking domain (PPXD), the nucleotide-binding domain (NBD-1), and the helical scaffold domain (HSD) (Auclair et al 2010). This region of the SecA ATPases may involve binding to a nucleotide and the SecYEG channel (Das et al 2012; Auclair et al 2010; Wang et al 2000). An altered amino acid arrangement due to the non-synonymous substitution may influence the degree of the binding interaction between the SecA ATPases and the SecYEG, resulting in a potential change in the exportation rate.

This question may be addressed by adding a green fluorescent protein (GFP) tag at the C-terminus of MBLs and using the SecYEG system whose activity is dependent on temperature. It needs to be verified that i) the addition of the GFP tag does not disturb or hinder the mature peptide translocation and that ii) the temperature-dependent SecYEG is inhibited until the time for translocation initiation. The non-GFP tagged wild-type MBL can be considered as a positive control, whereas a plain GFP tag can be used as a negative control. Then, the periplasmic amount of the GFP-tagged wild-type and the round 18 variants are measured and compared at different time points before the GFP fluorescence intensity reaches a saturation point. The result may help elucidate the mutational impact on translocation efficiency, whereas in this study, I have investigated the potential role of the mutations in the signal peptide of the enzymes at the translational stage.

5.1.3 Conclusion

Mutations in the signal peptides of MBLs haven't been investigated despite their potential role in increasing the concentration of MBL in *Escherichia coli*. To investigate this phenomenon, I performed a series of dose-response assays and bioinformatic analysis using the sequence profiles of MBL variants evolved in directed evolution. The results show a correlation between the mutation-driven changes in the translational steps and MIC, implying that mutations in signal peptides alter MBL gene expression. This study provides a significant implication for the importance of MBL's adaptability in various pathogenic organisms, which needs to be considered to combat antibiotic resistance. Furthermore, to determine mutational effects in diverse bacterial hosts, I synthesized a new broad host range vector. The plasmid vector's compatibility with various hosts such as *Acinetobacter baumannii, Escherichia coli,*

Pseudomonas aeruginosa, and *Pseudomonas putida* will enable future research in studying mutational effects in multiple pathogenic organisms using a single plasmid vector. These findings will help contribute to developing new therapeutic strategies to combat multi-drug resistant pathogens.

Bibliography

Amaya, Y. et al., 2015. Evolutionary well-conserved region in the signal peptide of parathyroid hormone-related protein is critical for its dual localization through the regulation of the ER translocation. *J. Biochem*., 159(4), p.393-406.

Anderson, D. & Walter, P. 1999. Blobel's Nobel. *Cell*. 99(6). p.557-558.

Arakawa, Y. et al., 1995. A novel integron-like element carrying the metallo-β-lactamase gene *bla*$_{IMP}$. *Antimicrob. Agents Chemother.*, 39, p1612-1615.

Auclair, S. M. et al., 2010. Mapping of the Signal Peptide-Binding Domain of Escherichia coli SecA Using Förster Resonance Energy Transfer. *Biochemistry*, 49(4), p.782-792.

Batey, R. T. et al., 2000. Crystal structure of the ribonucleoprotein core of the signal recognition particle. *Science*, 287(5456), p.1232-1239.

Bauer A.W. et al., 1966. Antibiotic susceptibility testing by a standardized single disk method. *Am. J. Clin. Pathol*, 45, p.493–496.

Bebrone, C., 2007. Metallo-beta-lactamases (classification, activity, genetic organization, structure, zinc coordination) and their superfamily. *Biochem Pharmacol*, 74, p.1686–1701.

Bello-López, J. M. et al., 2019. Horizontal Gene Transfer and Its Association with Antibiotic Resistance in the Genus *Aeromonas* spp. *Microorganisms*, *7*(9), p.363.

Bentele, K. et al., 2013. Efficient translation initiation dictates codon usage at gene start. Mol. Syst Biol., 9, p.675.

Boer, D. R. et al., 2009. Plasmid replication initiator RepB forms a hexamer reminiscent of ring helicases and has mobile nuclease domains. *The EMBO journal*, *28*(11), p.1666–1678.

Bouvier, M. et al., 2008. Small RNA Binding to 5' mRNA Coding Region Inhibits Translational Initiation. *Cell*, 32, 827-837.

Brem, J. et al., 2016. Structural basis of metallo-β-lactamase, serine-β-lactamase and penicillin-binding protein inhibition by cyclic boronates. *Nature communications*, 7, p.12406.

Brewer, N. S. & Hellinger, W. C., 1991. The monobactams. *Mayo Clinic Proceedings*, 66(11), p.1152-1157

Brockmann, R. et al., 2007. Posttranscriptional expression regulation: what determines translation rates? *PLoS Comput Biol.,* 3, e57.

Chen D. et al., 2006. Low-usage codons and rare codons of *Escherichia coli*. *Gene Ther Mol Biol* 10, p.1-12.

Chen, H. et al., 1996. Competition between functional signal peptides demonstrates variation in affinity for the secretion pathway. *J. Bacteriol*., 178(23), p.6658-6664.

Chen, J. Z. et al., 2020. Comprehensive exploration of the translocation, stability and substrate recognition requirements in VIM-2 lactamase. *eLife*. 9, e56707.

Chen, K. & Arnold, F.H., 1993. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc Natl Acad Sci USA*, 90, p.5618-5622.

Chiou, J. et al., 2014. Molecular mechanisms of substrate recognition and specificity of New Delhi metallo-β-lactamase. *Antimicrobial agents and chemotherapy*, 58(9), p.5372–5378.

Choo, K. H. et al., 2005. SPdb-a signal peptide database. *BMC Bioinform*., 6(1), p.249.

Chursov, A. et al., 2013. Conservation of mRNA secondary structures may filter out mutations in Escherichia coli evolution. *Nucleic Acids Research*, 41(16). P7854-7860.

Cobb, R. E. et al., 2013. Directed Evolution: Past, Present and Future. *AIChE journal. American Institute of Chemical Engineers*, *59*(5), p.1432–1440.

Codjoe, F. S., & Donkor, E. S., 2017. Carbapenem Resistance: A Review. *Medical sciences (Basel, Switzerland)*, *6*(1), p.1.

Concha, N.O. et al., 1996. Crystal structure of the wide-spectrum binuclear zinc beta-lactamase from Bacteroides fragilis. *Structure*, 4, p.823–836.

Concha, N.O. et al., 2000. Crystal structure of the IMP-1 metallo betalactamase from Pseudomonas aeruginosa and its complex with a mercaptocarboxylate inhibitor: binding determinants of a potent, broad-spectrum inhibitor. *Biochemistry*, 39, p.4288–4298.

Dalhoff. A. & Thomson, C.J., 2003. The art of fusion: from penams and cephems to penems. *Chemotherapy,* 49, 105e120.

Das, N. et al., 2019. An overview of cephalosporin antibiotics as emerging contaminants: a serious environmental concern. *3 Biotech*, 9(6), p.231.

Das, S. et al., 2012. The Variable Subdomain of *Escherichia coli* SecA Functions To Regulate SecA ATPase Activity and ADP Release. *Journal of Bacteriology*, 194 (9), p.2205-2213.

Dean, C. R. et al., 2018. Mode of Action of the Monobactam LYS228 and Mechanisms Decreasing *in Vitro* Susceptibility in *Escherichia coli* and *Klebsiella pneumoniae*. *Antimicrobial Agents and Chemotherapy*, 62(10), e01200-18

Desai, M. M. et al., 2007. Beneficial Mutation-Selection Balance and the Effect of Linkage on Positive Selection. *Genetics,* 176, p.1759-1798.

Draycheva, A. et al., 2018. Co-translational protein targeting to the membrane: Nascent-chain transfer in a quaternary complex formed at the translocon. *Sci Rep,* 8, p.9922.

Dressaire, C. et al., 2012. Role of mRNA stability during Bacterial Adaptation. *PLoS ONE*. 8(3), e59059.

Duffy, J. et al., 2010. Discovery of functional motifs in h-regions of trypanosome signal sequences. *Biobhem. J.,* 426(2), p.135-145.

Duffy, J. et al., 2010. Discovery of functional motifs in h-regions of trypanosome signal sequences. *Biochem J*., 426, p.135-145.

Duval, M. et al., 2013. *Escherichia coli* Ribosomal Protein S1 Unfolds Structured mRNAs Onto the Ribosome for Active Translation Initiation. PLoS Biol., 11(12), e1001731.

Engelman, D. M. et al., 1981. The spontaneous insertion of proteins into and across membranes: the helical hairpin hypothesis. *Cell.* 23, p.411-422.

Fabiane, S.M. et al., 1998. Crystal structure of the zinc-dependent beta-lactamase from Bacillus cereus at 1.9 Å resolution: binuclear active site with features of a mononuclear enzyme. *Biochemistry*, 37, p.12404–12411.

Fikes, J. et al., 1990. Maturation of *Escherichia coli* maltose-binding protein by signal peptidase I in vivo. Sequence requirements for efficient processing and demonstration of an alternate cleavage site. *J. Biol. Chem*., 256(6), p.3417-3423.

Fishovitz, J. et al., 2014. Penicillin-binding protein 2a of methicillin-resistant *Staphylococcus aureus*. *IUBMB life*, 66(8), p.572-577.

Fluit, A. C. & Schmitz, F. J., 1999. Class 1 integrons, gene cassettes, mobility, and epidemiology. *Eur. J. Clin. Microbiol. Infect. Dis.,* 18, p.761-770.

Garau, G. et al., 2005. A metallo-beta-lactamase enzyme in action: crystal structures of the monozinc carbapenemase CphA and its complex with biapenem. *J Mol Biol*, 345, p.785–795.

García-Estrada, C. & Martin J., 2011, Penicillins and cephalosporins. *Compr Biotechnol.,* 3, p.255–268

Garcia-Saez, I. et al., 2008. The three-dimensional structure of VIM-2, a Zn-beta-lactamase from Pseudomonas aeruginosa in its reduced and oxidised form. *J Mol Biol*, 375, p.604–611.

Gerald, L.M. et al., 1990. Penicillins, cephalosporin and other β-lactam antibiotics. Goodman and Gilmans, the pharmacological basis of therapeutics. *New York: Pargamon Press,* pp. 1065–1097.

Geukens, N. et al., 2004. Analysis of type I signal peptidase affinity and specificity for preprotein substrates. *Biochem. Biophys. Res. Commu*., 314(2), p.459-467.

Gilchrist, M. A. et al., 2015. Estimating Gene Expression and Codon-Specific Translational Efficiencies, Mutation Biases, and Selection Coefficients from Genomic Data Alone. Genome. Biol. Evol. 7(6), p.1559-1579.

Gingold, H. & Pilpel, Yitzhak. 2011. Determinants of translation efficiency and accuracy. *Molecular Systems Bilogy*. 7, p. 481.

González, J. M. et al., 2012. Metallo-β-lactamases withstand low Zn(II) conditions by tuning metal-ligand interactions. *Nature chemical biology*, 8(8), p.698–700.

Gonzalez, L. J. et al., 2012. Carbepenem Resistance in *Elizabethkingia meningoseptica* Is Mediated by Metallo-β-Lactamase BlaB. *American Society for Microbiology*, p.1686-1692.

González, L. J. et al., 2016. Membrane anchoring stabilizes and favors secretion of New Delhi metallo-β-lactamase. *Nature chemical biology*, *12*(7), p.516–522.

González, L.J. et al., 2018. Shaping Substrate Selectivity in a Broad-Spectrum Metallo-β-Lactamase. *Antimicrob Agents Chemother*, 62(4), e02079-17.

Goodman, D. B. et al., 2013. Causes and Effects of N-Terminal Codon Bias in Bacterial Genes. *Science.* 342. p.475-478.

Gu, W. et al., 2010. A Universal Trend of Reduced mRNA Stability near the Translation-Initiation Site in Prokaryotes and Eukaryotes. *PLoS Comput Biol.* 6(2), e1000664.

Guardino, R.F., 2005. Early History of Microbiology and Microbiological Methods. *Parenteral Drug Association*, *Wilmington, DE, USA.*

Haenni, M. et al., 2006. Mutational analysis of class A and class B penicillin-binding proteins in *Streptococcus gordonii*. *Antimicrobial agents and chemotherapy*. 50(12). p.4062-4069

Hartmann, E. et al., 1989. Predicting the orientation of eukaryotic membrane-spanning proteins. *Proc. Natl. Acad. Sci*., 86(15), p.5786-5797.

Heijne, G. V., 1984. Analysis of the distribution of charged residues in the N-terminal region of signal sequences: implications for protein export in prokaryotic and eukaryotic cells. *EMBO J*. 3(10), p.2315.

Heijne, G. V., 1990. The signal peptide. *J. Membr. Biol*., 115(3), p.195-201.

Huang, T. et al., 2011. Analysis and prediction of translation rate based on sequence and functional features of the mRNA. *PloS one*, 6(1), e16036.

Iriarte, A. et al., 2013. Evolution of optimal codon choices in the family Enterobacteriaceae. *Microbiology*. 159. p.555-564.

Jelger, A. et al., 2012. The bacterial Sec-translocase: structure and mechanism. *Phil. Trans. R. Soc. B.*, 367, p.1016-1028.

Kapoor, G. et al., 2017. Action and resistance mechanisms of antibiotics: A guide for clinicians. *Journal of anaesthesiology, clinical pharmacology*, *33*(3), p.300–305.

Karamanou, S. et al., 2007. Preprotein-controlled catalysis in the helicase motor of SecA. *EMBO J*. 26, p.2904-2914.

Karamyshev, A. L. et al., 1998. Processing of *Escherichia coli* alkaline phosphatase: role of the primary structure of the signal peptide cleavage region. *J. Mol. Biol*., 277(4), p.859-870.

Kebir, M. O. & Kendall, D. A., 2002. SecA Specificity for Different Signal Peptides. *American Chemical Society*, 41, p.5573-5580.

Khan, Z. A. et al., 2019. Current and Emerging Methods of Antibiotic Susceptibility Testing. *Diagnostics*, *9*(2), p.49.

King, D.T. et al., 2012. New Delhi metallo-β-lactamase: structural insights into β-lactam recognition and inhibition. *J. Am Chem Soc.,* 134, p.11362-11365.

Klumpp, S. et al., 2012. On Ribosome Load, Codon Bias and Protein Abundance. *PLoS ONE,* 7(11), e48542.

Kohler, P. et al., 2020. Dissemination of Verona Integron-encoded Metallo-β-lactamase among clinical and environmental Enterobacteriaceae isolates in Ontario, Canada. *Nature research*, 10, p.18580.

Kong, K. F. et al., 2010. Beta-lactam antibiotics: from antibiosis to resistance and bacteriology. *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica*, 118(1), p.1–36.

Kosheleva, K. & Desai, M. M., 2013. The Dynamics of Genetic Draft in Rapidly Adapting Populations. *Genetics*, 195, p.1007-1025.

Kozak, M., 2005. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*, 361, p.13-37.

Kryazhimskiy, S. & Plotkin, J.B. 2008. The Population Genetics of dN/dS. *PLoS Genet* 4(12), e1000304.

Kudla, G. et al., 2009. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science*, 324(5924), p.255-258.

LaRue, B. et al., 1979. Evolution of methionine initiator and phenylalanine transfer RNAs. *J Mol Evol* 14, p.287–300

Lauretti, L. et al., 1999. Cloning and characterization of bla_{VIM}, a new integron-borne metallo-β-lactamase gene from a *Pseudomonas aeruginosa* clinical isolate. *Antimicrob. Agents Chemother*, 43, p.1584-1590.

Laursen, B. S. et al., 2005. Initiation of Protein Synthesis in Bacteria. *Microbiology and Molecular Biology Reviews*. 69(1), p101-123.

Lima, L. M. et al., 2020. β-lactam antibiotics: An overview from a medicinal chemistry perspective. *European Journal of Medicinal Chemistry*. 15, p.208.

Lisa, M. et al., 2017. A general reaction mechanism for carbapenem hydrolysis by mononuclear and bionuclear metallo-β-lactamases. *Nature communications*, 8, p.538.

López, C. et al*.,* 2019. Protein determinants of dissemination and host specificity of metallo-β-lactamases. *Nat Commun,* 10, p.3617

López, J. L. et al., 2019. Codon usage heterogeneity in the multipartite prokaryote genome: selection-based coding bias associated with gene location, expression level, and ancestry. *mBio,* 10, e00505-19.

Lucidi, M. et al., 2018. New Shuttle Vectors for Gene Cloning and Expression in Multidrug-Resistant Acinetobacter Species. *Antimicrobial agents and chemotherapy*, *62*(4), e02480-17.

Macheboeuf, P. et al., 2006. Penicillin Binding Proteins: key players in bacterial cell cycle and drug resistance processes. *FEMS Microbiology Reviews.* 30(5). p.673-691.

Makena, A. et al., 2015. Comparison of Verona Integron-Borne Metallo-β-Lactamase (VIM) Variants Reveals Differences in Stability and Inhibition Profiles. *Antimicrob. Agents Chemother.* 60, p.1377-1384.

Mao, C. et al., 2009. Maximal efficiency of coupling between ATP hydrolysis and transloation of polypeptides mediated by SecB requires two protomers of SecA. J. *Bacteriol.* 191, p.978-984.

Martínez-García, L. et al., 2018. Ceftazidime Is the Key Diversification and Selection Driver of VIM-Type Carbapenemases, mBio, 9 (3), e02109-17

Meini, M. R., Llarrull, L. I. & Vila, A. J., 2015. Overcoming differences: The catalytic mechanism of metallo-β-lactamases. *FEBS letters*, *589*(22), p.3419–3432.

Mojica, M. F. et al., 2016. B1-Metallo-β-Lactamases: Where Do We Stand?. *Current drug targets*, 17(9), p.1029–1050.

Morán-Barrio, J. et al., 2009. Secretion of GOB metallo-beta-lactamase in Escherichia coli depends strictly on the cooperation between the cytoplasmic DnaK chaperone system and the Sec machinery: completion of folding and Zn(II) ion acquisition occur in the bacterial periplasm. *Antimicrobial agents and chemotherapy*, 53(7), p.2908–2917.

Moreira, M. H. et al., 2019. From reporters to endogenous genes: the impact of the first five codons on translation efficiency in Escherichia coli. *RNA Biol*. 16, p.1806−1816.

Natale, P. et al., 2004. Binding of SecA to the SecYEG complex accelerates the rate of nucleotide exchange on SecA. *J. Biol. Chem*., 279, p.13769-13777.

Natale, P. et al., 2008. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane-Distinct translocases and mechanisms. *Biochimica et Biophysica Acta*, 1778, p.1735-1756.

Nauton, L. et al., 2008. Structural insights into the design of inhibitors for the L1 metallo-beta-lactamase from Stenotrophomonas maltophilia. *J Mol Biol*, 375, p.257–269.

Nesmeyanova, M. A. et al., 1997. Positively charged lysine at the N-terminus of the signal peptide of the Escherichia coli alkaline phosphatase provides the secretion efficiency and is involved in the interaction with anionic phospholipids. *FEBS Letters,* 403(2), p.203-207.

Nilsson, U. et al., 2015. The Code for Directing Proteins for Translocation across ER Membrane: SRP Co-translationally Recognizes Specific Features of a Signal Sequence. *Journal of Molecular Biology*, 426(6), p.1191-1201.

Orencia, M. C. et al., 2001. Predicting the emergence of antibiotic resistance by directed evolution and structural analysis. *Nature structural biology*, 8(3), p.238-242.

Owji, H. et al., 2018. A comprehensive review of signal peptides: Structure, roles, and applications. *European Journal of Cell Biology*, 97(6), p.422-441.

Paetzel, M. et al., 2002. Signal peptidases. *Chem. Rev.*, 102(12), p.4549-4580.

Page, M. & Badarau, A. 2008. The mechanisms of catalysis by metallo beta-lactamases. *Bioinorganic chemistry and applications*. 576297.

Palzkill, T. et al., 1992. Identification of Amino Acid Substitutions That Alter the Substrate Specificity of TEM-1 β-Lactamase. *J. Bacteriol*, 174(16), p.5237-5243.

Payne S.H. et al., 2012. Unexpected diversity of signal peptides in prokaryotes. *mBio*, 3(6), e00339-12.

Peeri, M. & Tuller, T., 2020. High-resolution modeling of the selection on local mRNA folding strength in coding sequences across the tree of life. *Genome Biology*, 21, p.63.

Petriman, N. A. et al., 2018. The interaction network of the YidC insertase with the SecYEG translocon, SRP and the SRP receptor FtsY. *Scientific reports*, *8*(1), p.578.

Plotkin, J. B. & Kudla, G. 2011. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet.* 12, p.32–42.

Poirel, L. et al., 2000. Characterization of VIM-2, a carbapenem-hydrolyzing metallo-β-lactamase and its plasmid- and integron-borne gene from a *Pseudomonas aeruginosa* clinical isolate in France. *Antimicrob. Agents Chemother.*, 44, p891-897.

Pop C, et al., 2014. Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Mol Syst Biol.* 10, p.770.

Pradel, N. et al., 2008. Sec- and Tat-Dependent Translocation of β-Lactamases across the *Escherichia coli* Inner Membrane. *Antimicrobial Agents and Chemotherapy*, 53 (1), p.242-248.

Quax, Tessa E. F. et al., 2015. Codon Bias as a Means to Fine-Tune Gene Expression. Molecular *Cell*. 59, p. 149-161.

Randall, L. et al., 2004. Sites of interaction between SecA and the chaperone SecB, two proteins involved in export. *Protein Sci*., 13, p.1124-1133.

Rasia, R.M. et al., 2004. Structural determinants of substrate binding to Bacillus cereus metallo-betalactamase. *J Biol Chem,* 279, p.26046–26051.

Reeve, B. et al., 2014. Predicting translation initiation rates for designing synthetic biology. *Frontiers in bioengineering and biotechnology*, *2*, p.1.

Reller, L.B. et al., 2009. Antimicrobial Susceptibility Testing: A Review of General Principles and Contemporary Practices. *Clin. Infect. Dis*. 49, p.1749–1755.

Rusch, S. L. et al., 1994. Signal peptide hydrophobicity is finely tailored for function. *J. Cell Biochem*, 55(2), p.209-217.

Sally R Partridge et al., 2018. Mobile Genetic Elements Associated with Antimicrobial Resistance. *Clin Microbiol Rev*, 31(4), p.e00088-17

Scharff, L. B. et al., 2011. Local Absence of Secondary Structure Permits Translation of mRNAs that Lack Ribosome-Binding Sites. *PLoS Genet,* 7(6), e1002155.

Shen, L. M. et al., 1991. Use of site-directed mutagenesis to define the limits of sequence variation tolerated for processing of the M13 procoat protein by the *Escherichia coli* leader peptidase. *Biochemistry*, 30(51), p.11775-11781.

Sianidis, G. et al., 2001. Cross-talk between catalytic and regulatory elements in a DEAD moter domain is essential for SecA function. *EMBO J*, 20, p.961-970.

Silander, O. K. et al., 2007. Understanding the Evolutionary Fate of Finite Populations: The Dynamics of Mutational Effects. *PLoS Biology*, 5(4), e94

Simmons, L. C. & Yansura, D. G. 1995. Translational level is a critical factor for the secretion of heterologous proteins in Escherichia coli. *Nature*. 14. P.629-634.

Soares, G. M. et al., 2012. Mechanisms of action of systemic antibiotics used in periodontal treatment and mechanisms of bacterial resistance to these drugs. *Journal of applied oral science: revista FOB*, *20*(3), p.295–309.

Socha, R. D. et al., 2019. The Molecular Mechanisms Underlying Hidden Phenotypic Variation among Metallo-β-Lactamases. *Elsevier.*

Spencer, J. et al., 2005. Antibiotic recognition by binuclear metallo-beta-lactamases revealed by X-ray crystallography. *J Am Chem Soc*, 127, p.14439–14444.

Stemmer, W. P. C. 1994. DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA*, 91, p.10747-10751.

Stemmer, W.P. 1994. Rapid evolution of a protein in vitro by DNA shuffling, *Nature*, p.370:389–391.

Stoletzki, N. & Eyre-Walker, A. 2007. Synonymous codon usage in *Escherichia coli*: selection for translational accuracy. *Mol Biol Evol*., 24, p.374–381.

Storz. J. F., 2016. Causes of molecular convergence and parallelism in protein evolution. *Nat Rev Genet*, 17(4), p.239-250.

Suominem, I. et al., 1995. Effects of signal peptide mutations on processing of *Bacillus stearothermophilus* α-amylase in *Escherichia coli*. *Microbilogy*, 141(3), p.649-654.

Takyar, S. et al., 2005. mRNA helicase activity of the ribosome. *Cell.* 120(1). p.49-58.

Tang Y.W. & Stratton, C.W., 2012. Advanced Techniques in Diagnostic Microbiology. *Springer, Nashville*, *TN, USA*.

Tuller T, et al., 2010. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell*, 141(2), p.344–354.

Tuller, T. et al., 2010. Translation efficiency is determined by both codon bias and folding energy. *Proc Natl Acad Sci USA*, 107, p.3645-3650.

Van, V. et al., 2000. Role of lipids in the translocation of proteins across membranes. *Biochem. J*., 347(3), p.601-612.

Varma, A. & Young, K. D. 2004. FtsZ Collaborates with Penicillin Binding Proteins To Generate Bacterial Cell Shape in *Escherichia coli. Journal of Bacteriology.* 186(20), p.6768-6774.

Vogel, C. et al., 2007. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol.,* 25(1), p.117-24.

Walsh, T.R., 2005. The emergence and implications of metallo-β-lactamases in Gram-negative bacteria. *Clinical Microbiology and Infection,* 11, Supplement 6. P.2-9.

Wang, L. et al., 2000. Signal Peptide Determinants of SecA Binding and Stimulation of ATPase Activity. *The Journal of Biological Chemistry*, 275(14), p.10154-10159.

Wheat P.F., 2001. History and development of antimicrobial susceptibility testing methodology. *J. Antimicrob. Chemother*., 48, p.1–4.

Zahurancik et al., 2020. Ramping Recombinant Protein Expression in Bacteria. *Biochemistry*. 59(23). p.2122-2124.

Zervosen, A. et al., 2012. Development of New Drugs for an Old Target – The Penicillin Binding Proteins. *Molecules*, 17, p.12478-12505.

Zhang, G. et al., 2009. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat Struct Mol Biol*., 16, p. 274–280.

Zhang, H. et al., 2011. Crystal structure of NDM-1 reveals a common β-lactam hydrolysis mechanism. *FASEB J*, 25, p.2574–2582.

Zhou, Z. et al., 2016. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. PNAS, E6117-E6125.

Zimmer, J. et al., 2008. Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature*, 455, p.936-943.

Appendix: Supplementary Figures and Tables

Appendix contains supplementary figures and tables from Chapter three.

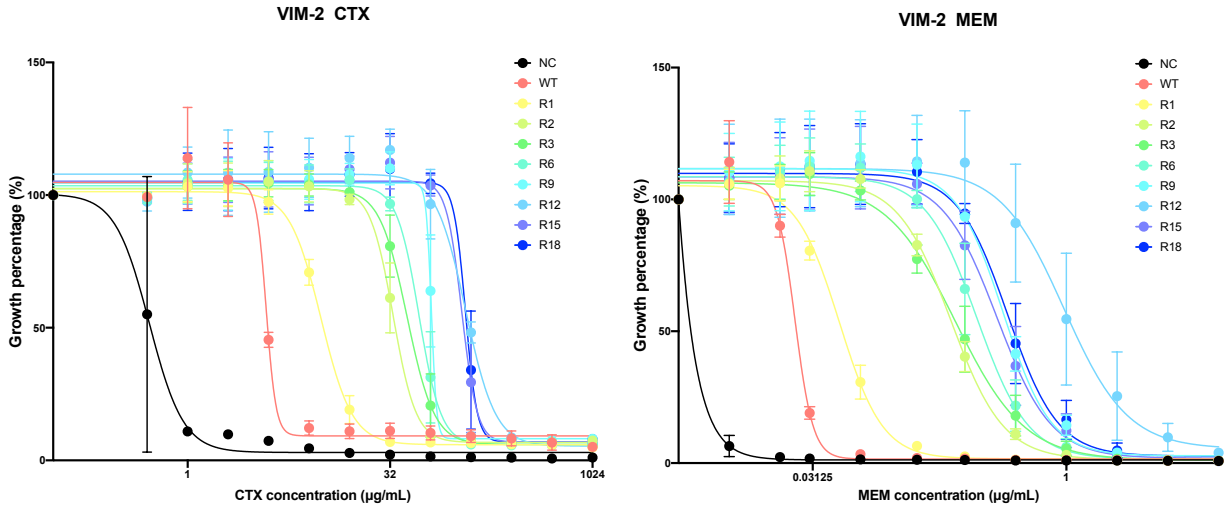


Figure A.1. Dose-response curve assay of VIM-2 challenged with cefotaxime (CTX) and meropenem (MEM). Of the 18 rounds in each trajectory, round (R) 1 (yellow), 2 (light green), 3 (green), 6 (spindrift), 9 (light blue), 12 (sky blue), 15 (orchid), and 18 (dark blue) trajectories of the three MBL orthologs are challenged by cefotaxime (CTX) on the left and meropenem (MEM) on the right. The black represents negative control (NC), and the pink reflects the wild-type used as a positive control. Each dot represents the data point that is obtained from the assay. The determined cell growth is normalized and converted to a percentage (%), shown on the y axis. CTX and MEM concentrations are displayed on the x axis and capture the ranges of the β-lactam concentration that were used for each MBL.

Table A.1. The IC_{50} and Hill coefficients of VIM-2 challenged with cefotaxime (CTX) and meropenem (MEM).

VIM-2	*CTX*		*MEM*	
	IC_{50} (μg/mL)	Hill slope	IC_{50} (μg/mL)	Hill slope
Negative control (NC)	0.52	-3.41	1.38×10^{-3}	-4.21
WT	3.84	-11.9	2.45×10^{-2}	-8.07
R1	9.80	-3.71	4.50×10^{-2}	-3.34
R2	33.7	-5.10	0.21	-2.77
R3	42.8	-4.27	0.22	-2.11
R6	52.5	-5.31	0.29	-2.84
R9	64.9	-22.3	0.42	-3.07
R12	1.17×10^{2}	-3.90	0.97	-2.31
R15	1.09×10^{2}	-7.63	0.39	-2.69
R18	1.16×10^{2}	-10.1	0.45	-2.93

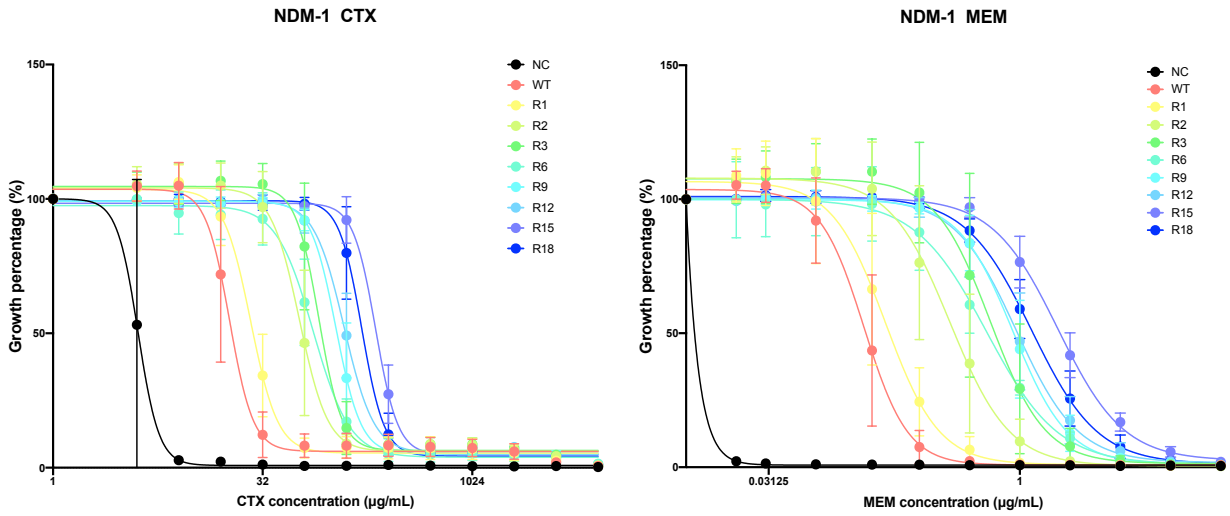


Figure A.2. Dose-response curve assay of NDM-1 challenged with cefotaxime (CTX) and meropenem (MEM). Of the 18 rounds in each trajectory, round (R) 1 (yellow), 2 (light green), 3 (green), 6 (spindrift), 9 (light blue), 12 (sky blue), 15 (orchid), and 18 (dark blue) trajectories of the three MBL orthologs are challenged by cefotaxime (CTX) on the left and meropenem (MEM) on the right. The black represents negative control (NC), and the pink reflects the wild-type used as a positive control. Each dot represents the data point that is obtained from the assay. The determined cell growth is normalized and converted to a percentage (%), shown on the y axis. CTX and MEM concentrations are displayed on the x axis and capture the ranges of the β-lactam concentration that were used for each MBL.

Table A.2. The IC_{50} and Hill coefficients of NDM-1 challenged with cefotaxime (CTX) and meropenem (MEM).

NDM-1	*CTX*		*MEM*	
	IC_{50} (μg/mL)	Hill slope	IC_{50} (μg/mL)	Hill slope
Negative control (NC)	4.08	-5.75	5.28 x10^{-3}	-6.09
WT	18.5	-5.11	0.12	-3.30
R1	26.3	-4.44	0.16	-2.71
R2	58.8	-4.27	0.38	-2.42
R3	81.0	-5.22	0.66	-2.64
R6	73.3	-3.29	0.63	-2.02
R9	1.08 x10^{2}	-4.81	0.91	-2.65
R12	1.23 x10^{2}	-4.03	0.95	-2.35
R15	2.07 x10^{2}	-5.42	1.69	-2.22
R18	1.65 x10^{2}	-5.37	1.18	-2.19

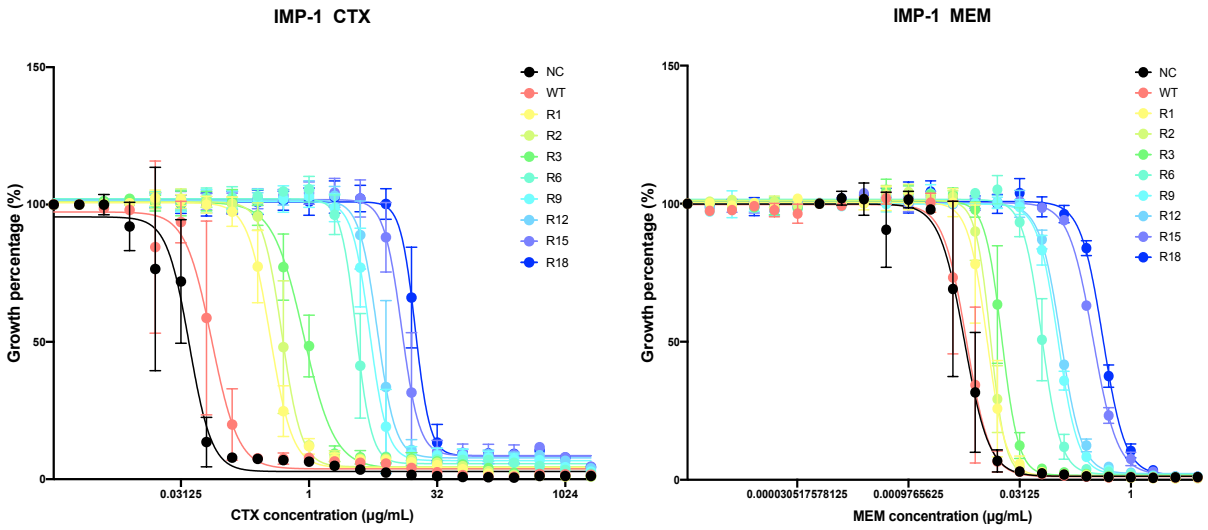


Figure A.3. Dose-response curve assay of IMP-1 challenged with cefotaxime (CTX) and meropenem (MEM). Of the 18 rounds in each trajectory, round (R) 1 (yellow), 2 (light green), 3 (green), 6 (spindrift), 9 (light blue), 12 (sky blue), 15 (orchid), and 18 (dark blue) trajectories of the three MBL orthologs are challenged by cefotaxime (CTX) on the left and meropenem (MEM) on the right. The black represents negative control (NC), and the pink reflects the wild-type used as a positive control. Each dot represents the data point that is obtained from the assay. The determined cell growth is normalized and converted to a percentage (%), shown on the y axis. CTX and MEM concentrations are displayed on the x axis and capture the ranges of the β-lactam concentration that were used for each MBL.

Table A.3. The IC_{50} and Hill coefficients of IMP-1 challenged with cefotaxime (CTX) and meropenem (MEM).

IMP-1	*CTX*		*MEM*	
	IC_{50} (μg/mL)	Hill slope	IC_{50} (μg/mL)	Hill slope
Negative control (NC)	3.92×10^{-2}	-3.28	5.56×10^{-3}	-2.60
WT	7.20×10^{-2}	-2.80	5.95×10^{-3}	-2.76
R1	0.35	-3.31	1.13×10^{-2}	-3.42
R2	0.48	-3.78	1.26×10^{-2}	-4.33
R3	0.87	-2.40	1.79×10^{-2}	-3.82
R6	3.58	-4.76	6.23×10^{-2}	-3.39
R9	5.14	-4.22	0.10	-3.15
R12	6.32	-4.09	0.11	-3.05
R15	12.2	-4.19	0.31	-2.64
R18	17.7	-5.02	0.42	-2.89