이학박사 학위논문

Exploring the Evolutionary Landscape of Artificial β-lactamases

인공 베타-락탐 가수분해효소의 진화적 다양성 연구

2023 년 2월

서울대학교 대학원
화학부 무기화학 전공
유 재 승

Ph. D. Dissertation

Exploring the Evolutionary Landscape of Artificial β-lactamases

Supervisor: Professor Woon Ju Song

Major: Inorganic Chemistry

February 2023

by Jaeseung Yu

Department of Chemistry

The Graduate School

Seoul National University

Exploring the Evolutionary Landscape of Artificial β-lactamases

지도 교수 : 송 윤 주

이 논문을 이학박사 학위논문으로 제출함
2023 년 2 월

서울대학교 대학원
화학부 무기화학 전공
유 재 승

유재승의 이학박사 학위논문을 인준함
2023 년 2 월

위 원 장 이 동 환 (인)

부위원장 송 윤 주 (인)

위 원 김 석 희 (인)

위 원 이 형 호 (인)

위 원 홍 승 우 (인)

Abstract

Exploring the Evolutionary Landscape of Artificial β-lactamases

Jaeseung Yu
Department of Chemistry
The Graduate School
Seoul National University

Numerous enzymes have evolved from ancestor genes to more specialized modern descendants for billions of years. These natural catalysts carry out a variety of chemical transformations with high catalytic efficiency and substrate specificity. The chemical power of enzymes has attracted and inspired numerous researchers to make *de novo* biocatalysts. Along this line, directed evolution has yielded artificial enzymes that enhanced the native functions or catalyzed abiological chemical reactions. However, the evolutionary process still remains unclear about how the target-of-interest has evolved to emerge novel structures and functions. To gain insights into understanding enzyme evolution and to develop more efficient methods for protein evolution, this dissertation explored the evolutionary landscapes of *de novo* artificial β-lactamases as versatile and representative model systems.

Chapter 1 introduces two representative experimental methods and a model system used for this dissertation. First, iterative-saturation-mutagenesis (ISM) was primarily used to target relatively well-studied regions near the active site. Secondly, random mutagenesis was widely used to maximize the potential evolvability of target proteins. The advantages and the limit of these two representative methods will be discussed in this chapter.

Additionally, the advantage of using β-lactamases as a model system will also be discussed. In particular, I adapted two β-lactamases that operate via discrete mechanisms, serine-β-lactamases (SβLs) and metallo-β-lactamases (MβLs).

Chapter 2 demonstrates the importance of balancing the flexibility and rigidity of protein structure in enzyme evolution. I employed *de novo* oligomeric β-lactamase AB5, which forms a tetramer upon adding 2 equivalent Zn^{2+} ion to the monomer, as a model system. Whereas the site-saturation mutagenesis of the residues near the active site showed low evolvability, the single mutation of the 96th residue (C96T) gave rise to drastic changes in protein structure and function. As a result, AB5 acquired structural flexibility and exhibited substantially elevated catalytic activities upon sequence optimization.

Chapter 3 demonstrates a novel guideline to propose hotspots for enzyme mutations. The increasing results in directed evolution suggest that distant mutations via allosteric interactions play significant roles in inducing catalytic activities of interest. However, it is unclear where we should start for sequence optimization. Herein, I targeted symmetry-related residues of oligomeric enzymes as mutational hotspots for their evolution. This hypothesis was based on the results in **Chapter 2** that C96T mutation, which played a significant role in altering the protein structure and function, was located on one of the C_2 rotational axes of AB5. In this chapter, I report the experimental results of directed evolution of AB5 and ^{C96T}AB5 when targeting residues near three different C_2 rotational axes. I also conducted the biochemical characterization and structural analysis.

Chapter 4 explores the bifurcated evolutionary landscapes of β-lactamases, which comprises two main subclasses: serine-β-lactamases (SβLs) and metallo-β-lactamases (MβLs). MβLs and SβLs exhibit the identical functions of hydrolyzing β-lactams. However, they show dissimilar

sequences and protein structures, suggesting that they are derived from distinct origins (as known as convergent evolution) and catalytic motifs. In this chapter, I demonstrated that the chemical mechanism and active site structures could be interchangeable between MβLs and SβLs in altered chemical environments, such as the presence and absence of metal ions. I have applied either iterative saturation mutagenesis or continuous evolution, resulting in substantially increased β-lactamases activities via interchanged catalytic motifs. Thus, this study demonstrated that metal ions could be a key player in dictating the landscape of enzyme evolution.

Keyword: enzyme evolution, β-lactamase, protein symmetry, biocatalysts, artificial metalloenzymes

Student Number: 2016-27405

Table of Contents

List of Figures

List of Tables

List of Schemes

Chapter 1.
Tracing an Evolutionary Landscape Using β-lactamase as a Model System

1.1. Introduction

Nature has very useful catalysts called enzymes carrying out numerous chemical transformations in biological systems. The power of enzymes, such as high reactivity and substrate specificity, has attracted many researchers to utilize them in two representative ways. Firstly, numerous synthetic catalysts have been yielded by mimicking the active site of a particular enzyme that catalyzes the reaction-of-interest. Secondly, the enzyme scaffold itself was engineered to enhance the pre-existing reactivity and/or perform new non-canonical chemical reactions. Since *in vitro* evolutions of RNA molecules for the first examples of artificial evolution in the late 1960s,[1-3] various strategies were adopted for the artificial evolution under *in vitro* evolutions.[4-6] Then, Arnold and co-workers substantially opened a new era of directed evolution in 1993.

Directed evolution is an in-lab mimic of Darwinian evolution with iterative screening and selection of genetically diversified DNA libraries (Figure 1.1). This strategy has become one of the most powerful methods to create artificial enzymes by expediting the long-term evolution in nature fast and mimicking the evolutionary process easily in the laboratory. Although many approaches by directed evolution have yielded novel biocatalysts, elucidating the effect of selected mutations on the target sites remains challenging in many cases. This might be due to the dependence on the randomness of constructed libraries without a *rationale*. In addition, elucidation of the mutation on a specific position of target protein became more important since random mutagenesis was widely used, even if it helped us open new eyes to observe unexpected changes at distal regions from the active site where directed evolutionists targeted in most conventional approaches.

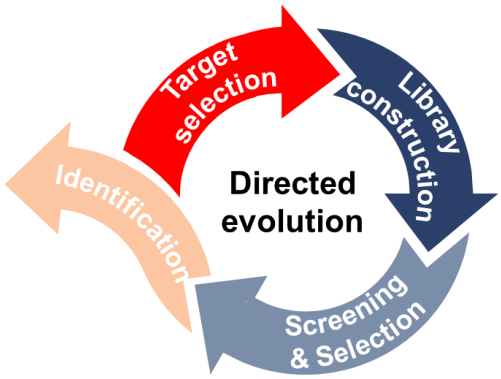


Figure 1.1. Principle of directed evolution

For efficient evolution study, selection of protein scaffolds and/or target reaction is the most considerable.[7] Structural stability and ease in the heterologous expression of selected scaffold can be considered. Evolution studies for *de novo* biocatalysts inevitably require *in vitro* characterization to examine whether the readouts from screenings under *in vivo* or *in vitro* conditions work consistently when isolated. Additionally, conditions for target reactions must be considered. Reactions that are physiologically not available for organisms and/or proteins might have been ruled out.

As a result, it is important to balance between the target scaffold and reaction. Applications of directed evolution started with evolving individual target proteins to improve their stability, activity, selectivity, and compatibility to facilitate for industry.[8-10] Nowadays, the engineering of promiscuous enzymes for abiotic transformations[11-14] is also performed by directed evolution. In this chapter, I will introduce a brief background of representative methodologies for directed evolution. And then, the advantages of adopting β-lactamases as model systems for the evolution study will be explained as well.

1.2. General strategies to construct mutant library for the directed evolution

Directed evolution has become a critical technique available for protein engineering since the 1990s.[15-17] In spite of advances in the understanding of protein sequence–structure–function relationships, prediction of protein function still remains challenging, which makes the statistical and combinatorial approaches mainstream for protein evolution. Those approaches substantially require two main steps: library construction and the screening/selection of the constructed mutant library. The success of directed evolution depends heavily on the quality of mutant libraries prepared and the ease of the designed procedure for screening/selection with a high throughput. However, methodologies for the screening or selection will not be discussed in this dissertation because relatively straightforward screening platforms, which are based on the host organism survival system, were employed in the entire studies here. In addition, qualitatively/quantitively effective and adequate library constructions are considered as more significant steps, since the mutant libraries must be prepared before the screening and selection steps. Thus, general strategies applied to library construction for directed evolution will be discussed instead.

Complete-random genes and semi-random genes targeting specific regions are usually constructed respectively by random mutagenesis (RM) using error-prone PCR (epPCR) and site-saturation mutagenesis (SSM) using degenerate primers. Since SSM is substantially based on the protein structures and sequences, this method can prepare a random mutant library that targets particular positions engaged in the protein function, while random mutagenesis cannot. Both SSM and RM are based on *in vitro* mutagenesis, which was a common and widely used strategy for controlled and efficient gene library construction. Thanks to recent advances, however, *in vivo* mutagenesis is also emerging, which possesses high potential to be performed

continuously in an organism independently.

1.2.1 Gene diversification under *in vitro* condition

Biochemical methods for gene diversifications have been immensely developed because of numerous researches for directed evolution.[17, 18] To prepare mutant libraries under *in vitro* conditions, site-saturation mutagenesis (SSM) and error-prone polymerase chain reaction (epPCR) are the most widely used methods (Figure 1.2). Using degenerated primers like NNK (N = A, G, T, and C / K = G and T), SSM can be performed on specific residues which are related to enzyme reactivity, substrate selectivity, and protein stability. This method also contributed to the construction of the smaller, and more efficient library size.[19] Reetz group reported the novel application of site-saturation mutagenesis by conducting SSM iteratively, which is called as iterative saturation mutagenesis (ISM).[20] Using this method, multiple amino acids, rationally chosen, at a specific location can be optimized iteratively.

On the other hand, SSM was used only when the structure of the target protein was already revealed. This limitation led random mutagenesis to be employed to construct the randomized DNA library for directed evolution again. Mimicking the natural error rates during DNA replication processes in nature (10^{-10} bp^{-1}), RM exploits the DNA polymerase to cause errors with a mutation rate (10^{-4} bp^{-1}) while the whole gene is replicated during the PCR process. Interestingly, it enabled rapid engineering of a target gene without the knowledge of target protein structures. Furthermore, it yielded the evolved products that cannot be easily expected with saturation mutagenesis using our rationales based on the understanding of the protein functions and structures.

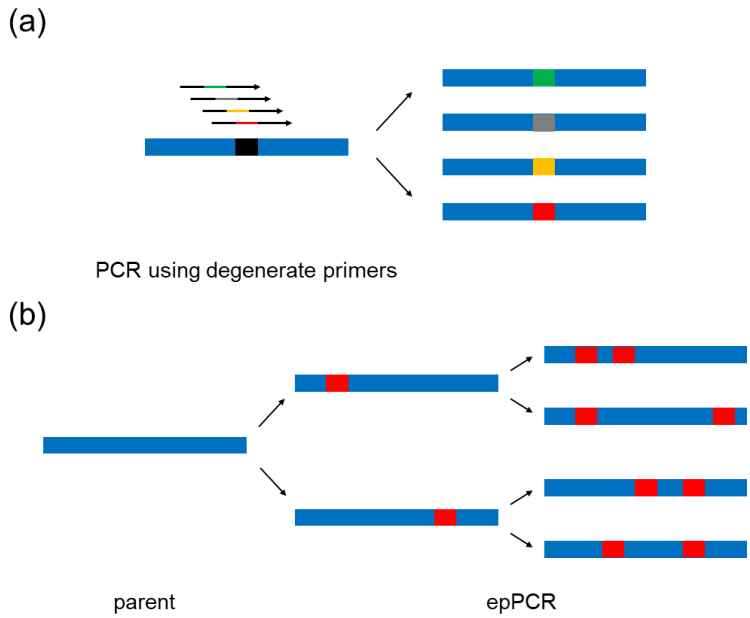


Figure 1.2. *In vitro* gene diversification. (a) site-saturation mutagenesis. Each arrow represents the primer encoding distinct amino acids. (b) random mutagenesis via error-prone PCR (epPCR).

1.2.2 Gene diversification under *in vivo* condition

Despite the power of methodologies described above, *in vitro* mutagenesis is time-consuming and/or requires repetitive cycles of gene diversification, followed by screening/selection and isolation for identification. To cope with this problem, Cell-free systems were also designed and applied but not generally used because of the technical hardness. Recently, *in vivo* mutagenesis methods have been gradually introduced when the desired phenotypes are engaged in the cell growth to avoid repetitive cloning and transformation steps and to alter multiple sequences of the target residues simultaneously. For these reasons, *in vivo* mutagenesis is occasionally preferred since library construction, expression, and even their selection upon the give chemical pressures can be performed in a continuous process, which achieves much faster directed evolution.

The orthogonal system for DNA replication has been widely used, *inter*

alia, to generate mutant libraries in living cells. Moore group employed a chimeric fusion protein between rAPOBEC1, which carries out deamination reaction of cytosine base, and T7 RNA polymerase (Muta T7) that introduces mutations within a recombinant DNA plasmid downstream of T7 promoter selectively in *E. coli*. In this system, the chimeric cytidine deaminase which was fused to T7 RNA polymerase randomly introduced C > T (or G > A) mutations.[21, 22] This system enabled the introduction of mutations randomly over relatively longer sequences without unexpected mutations.

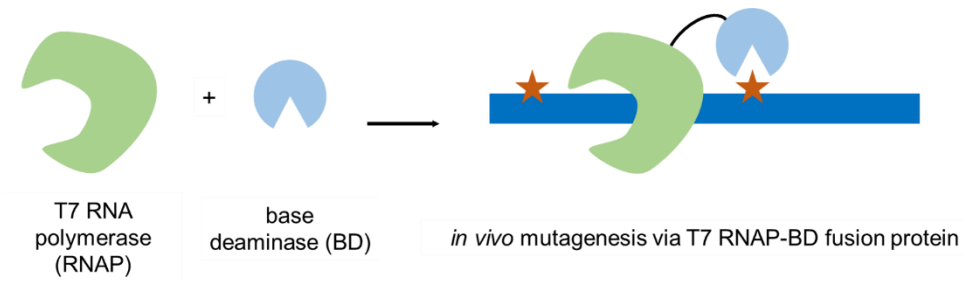


Figure 1.3. Representative example of *in vivo* mutagenesis.

1.3. β-lactamases, versatile model systems for evolution study

1.3.1 A brief history of β-lactam antibiotics

The discovery of penicillin in 1929[23] and the modern era of antibiotics started with its first medicinal use in the 1940s.[24] After penicillin, a variety of β-lactam antibiotics were successively discovered/developed and saved our lives for decades by treating serious bacterial infections (Figure 1.4–5). But unfortunately, human beings are still searching for new antibiotics. This is because of the rapid emergence of antibiotic resistance. Bacterial organisms resistant to penicillin were already reported even before its first prescription to treat bacterial infections. Most β-lactam antibiotics face resistance as soon as they are newly discovered and/or developed. The rapid emergence of resistance against β-lactam antibiotics has become a great obstacle to using them nowadays. Moreover, this is not only applied to β-lactam but also to all

antibiotics. And this is called the 'antibiotics resistance crisis.'[25, 26]

Bacteria become resistant against β-lactam antibiotics with the enzymes called β-lactamases. These enzymes hydrolyze β-lactam antibiotics, causing them to lose antimicrobial effects. It took only a decade for the β-lactamase, which can hydrolyze the first β-lactam antibiotic penicillin, to be discovered. Resistance was accelerated, albeit new β-lactam antibiotics like methicillin and cephalosporin were discovered and developed.

However, there are still benefits of antibiotics. Antibiotics have not only saved many lives but also played a significant role in understanding how they were working and prevented. From another chemical point of view, those enzymes that underwent rapid evolution to break through their archenemy, antibiotics, can give us an opportunity to understand how such a resistance could have emerged and what did happen during the evolution of those enzymes, in other words. This would allow us to peek at the principle of evolution in nature and apply the extended insights to the directed evolution of other enzymes.

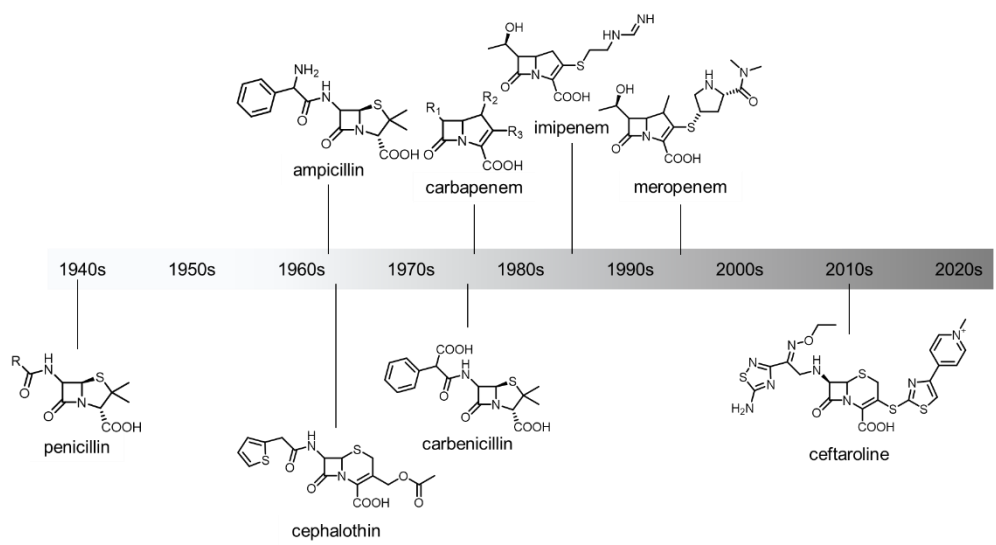


Figure 1.4. Timeline of β-lactam antibiotics.

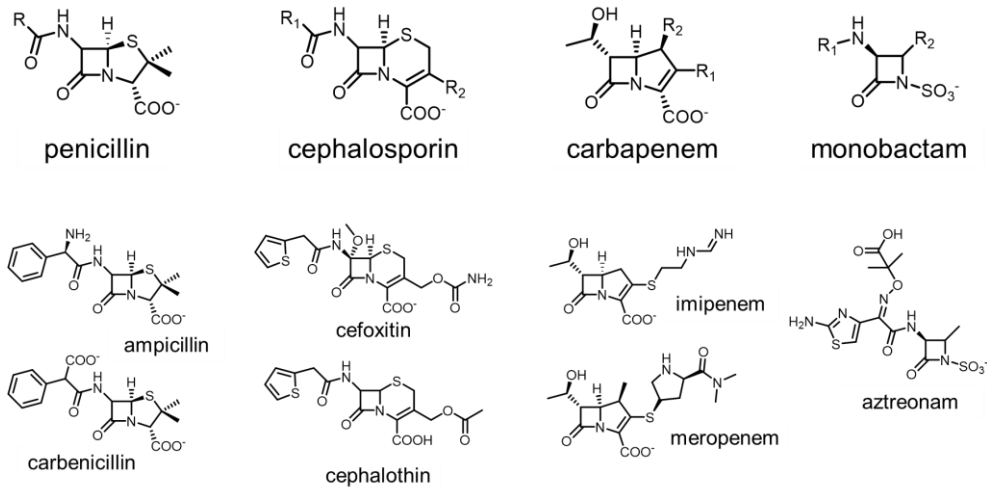
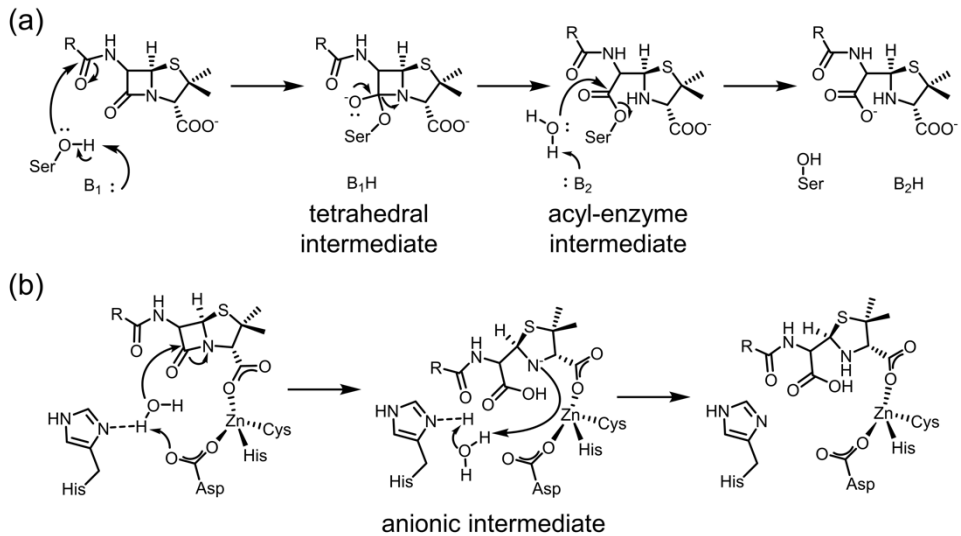


Figure 1.5. (top) Structures of four main β-lactam antibiotic classes. (bottom) Structures of representative examples belonging to each class.

1.3.2 β-lactamases as good model system

β-lactamases would be a very suitable candidate as an attractive model system for the following reasons. First, hydrolysis reactions of β-lactam antibiotics are relatively simple to explore due to the redox-inactive reaction mechanisms. β-lactamases are classified mainly in two representative types by catalytic moieties which carry out catalysis; one is the activated serine and the other is Zn cofactor which serves as a Lewis acid or substrate anchor. Generally, the breakdown of β-lactam antibiotics results from the cleavage of the β-lactam ring by the nucleophilic attack of the activated serine, Zn–OH, and/or probably the water molecule activated by adjacent base amino acid. These nucleophiles attack the carbonyl carbon at β-lactam ring of penicillin-like substrates.[27] More complexed catalysis for cepham substrates and carbapenem substrates was revealed as well. (Scheme 1.1)



Scheme 1.1. Representative proposed reaction mechanisms of (a) serine-β-lactamases and (b) metallo-β-lactamases.

Second, the rapid emergence of resistance can be used by itself. As mentioned above, they have developed the resistance extremely fast for several decades probably due to the fast evolution process with the short generational succession of bacterial organisms.[28] The first resistance against penicillin was discovered even before the first prescription of penicillin. Since then, resistances have emerged almost as soon as new β-lactam antibiotics were discovered and synthesized. This causes difficulty in the clinical usage of β-lactam antibiotics. On the other hand, the rapid evolution of β-lactamases also makes them more attractive to employ as a good guide for exploring the evolutionary landscape of enzymes in the laboratory.

Last, the antimicrobial reaction makes the design for the selection step easier by monitoring survival of cells exposed to the antibiotic substrate. Directed evolution studies are highly dependent on the throughput of the screening process. To make a laborious step efficient, phenotypes of the target reaction must be clear and easy to detect and/or combine with the other process. Antibiotics are literally toxic to cell survival but advantageous for

direct detection of the readouts.[29-31]

Hence, β-lactamases have been widely used as a model system to explore the evolutionary landscapes of enzymes. Vila *et al.*[32] successfully demonstrated the evolution of enzymatic catalysis can take place by remote mutations which tune reactivity. They employed the metallo-β-lactamase from *B. cereus* (BcII) as a model system. The evolved BcII variant acquired the power to catalyze cephalothin, which was a poorly hydrolyzable substrate for the parent enzyme. Ranganathan *et al.*[33] adopted TEM-1, which is the representative serine-β-lactamase, as a robust model system to explore the evolvability of enzymes using a huge size of mutant library constructed via saturation mutagenesis. Furthermore, I will discuss the evolutionary landscape of *de novo* metallo-β-lactamases as promiscuous model enzymes in this dissertation (Figure 1.6).

In particular, *de novo* metallo-β-lactamases, AB5, and evolved variants are homo-oligomeric enzymes that the active sites were located at the protein-protein interfaces. In general, protein oligomerization is strongly related to the protein function.[34-37] there are some advantages for oligomeric proteins to be a target for the evolution. First, about 20% of oligomeric enzymes have an active site located at the domain–domain interfaces (or protein–protein interfaces). Because mutations in oligomeric proteins are inevitably replicated, the mutational effect can be amplified when the interfaces are targeted. Surprisingly, numerous oligomeric proteins have metal ions at the protein–protein interfaces.[38] This implies that transition metal ions play significant roles in biology from the catalysis to the structural stabilization. In **chapter 2–4**, I will discuss how oligomerization of symmetric proteins affects the evolution of enzymes as well. (Figure 1.6c)

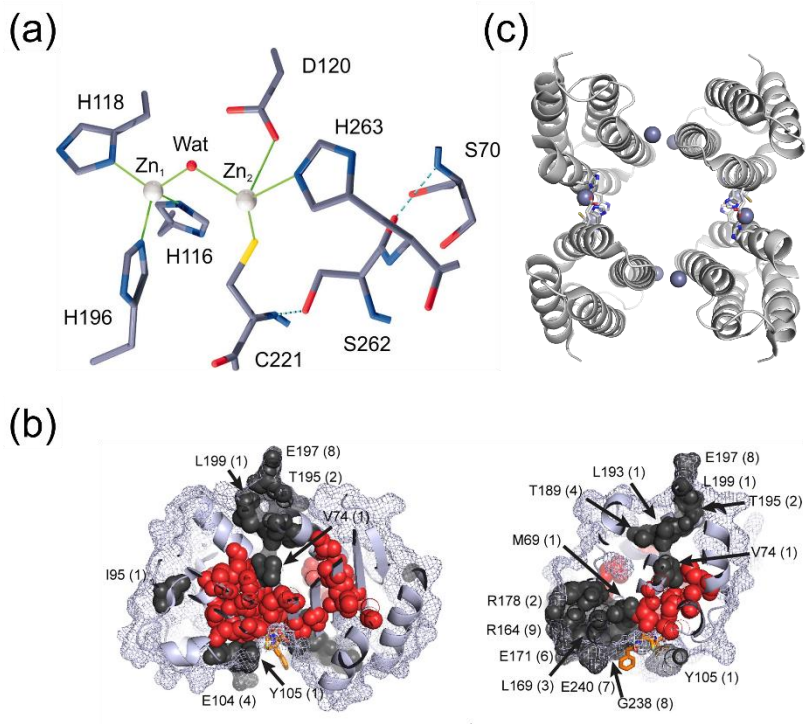


Figure 1.6. Representative β-lactamases used as a model system. (a) Model structure for the active site of metallo-β-lactamase BcII from B. cereus. Reprinted with a permission from ref. 33. Copyright (2005) National Academy of Sciences. (b) serine-β-lactamase TEM-1 adopted to explore the fitness effect by mutations. Reprinted with permission from ref.34. Copyright © 2015 Elsevier Inc. and (c) *De novo* metallo- β-lactamase AB5 (PDB 5XZI)

1.3.3 Convergent evolution of β-lactamases

Two subclasses comprised of serine-β-lactamases (SβLs) and metallo-β-lactamases (MβLs) do not share any common features in structures and sequences; they are products of convergent evolution. This feature could make the study of the evolution of β-lactamases more interesting by examining the dynamics in evolutionary landscapes left behind. Distinctly but convergently evolved two β-lactamases adopted different reaction mechanisms to deal with the same threat. However, the chemical basis of each mechanism was still similar in that activated serine and Zn–OH serve as

nucleophiles in the hydrolysis reactions. (Scheme 1.1) Nevertheless, there was no report about their common ancestors, and/or examples of the conversion between SβLs and MβLs (*vice versa*).

1.4. Conclusion

To evolutionists, evolution strategies are the primary thought for food to engineer proteins to make useful biocatalysts. Directed evolution methods and the application of random mutagenesis might be the product of those demands. In light of the introduction and development of those powerful and useful techniques, plenty of biocatalysts carry out enhanced pre-existing and canonical reactions, or non-canonical chemical transformations. Nonetheless, I still have a lot of challenges to work on for understanding and analyzing the readouts obtained from the designed evolution platform.

As described above, β-lactamases have been our practical threat with antibiotic resistance, despite the benefits of β-lactam antibiotics in our lives for many decades. But at the same time, they could be good model systems for studying evolution by mimicking their rapid evolution in the laboratory. Furthermore, their distinct reaction mechanisms with activated serine and Zn^{2+} ion cofactors were also interesting. In this study, a *de novo* metallo-β-lactamase and a native serine-β-lactamase are employed to explore the evolutionary landscapes of each enzyme, respectively.

Chapter 2 will discuss the importance of balancing the rigidity and the flexibility of proteins for the design and evolution of metalloenzymes. Then, **Chapter 3** will demonstrate how sequence alterations obtained during directed evolution affect the structures and functions and suggest the novel guideline where I target for efficient evolution in the case of oligomeric enzymes using protein symmetry.

Chapter 4 will display the first case of crossover between SβLs and MβLs (and *vice versa*) using metal ion-dependent evolution with a library

constructed by site-saturation mutagenesis and continuous evolution method in which sequences encoding target protein are randomized under in-situ condition to expand the evolutionary landscapes of each type of β-lactamases and to understand the sequence–structure–function relationship.

1.5. References

(1) Kacian, D. L.; Mills, D. R.; Kramer, F. R.; Spiegelman, S., *Proceedings of the National Academy of Sciences* **1972,** *69* (10), 3038-3042.

(2) Levisohn, R.; Spiegelman, S., *Proceedings of the National Academy of Sciences* **1969,** *63* (3), 805-811.

(3) Mills, D. R.; Peterson, R. L.; Spiegelman, S., *Proceedings of the National Academy of Sciences* **1967,** *58* (1), 217-224.

(4) Smith, G. P., *Science* **1985,** *228* (4705), 1315-1317.

(5) Ellington, A. D.; Szostak, J. W., *Nature* **1990,** *346* (6287), 818-822.

(6) Liao, H.; McKenzie, T.; Hageman, R., *Proceedings of the National Academy of Sciences* **1986,** *83* (3), 576-580.

(7) Jeong, W. J.; Yu, J.; Song, W. J., *Chem. Commun.* **2020,** *56* (67), 9586-9599.

(8) Zhao, H.; Chockalingam, K.; Chen, Z., *Current Opinion in Biotechnology* **2002,** *13* (2), 104-110.

(9) Arnold, F. H.; Volkov, A. A., *Current Opinion in Chemical Biology* **1999,** *3* (1), 54-59.

(10) Arnold, F. H., *Nature* **2001,** *409* (6817), 253-257.

(11) Renata, H.; Wang, Z. J.; Arnold, F. H., *Angewandte Chemie International Edition* **2015,** *54* (11), 3351-3367.

(12) Hyster, T. K.; Ward, T. R., *Angewandte Chemie International Edition* **2016,** *55* (26), 7344-7357.

(13) Arnold, F. H., *Angewandte Chemie International Edition* **2018,** *57* (16), 4143-4148.

(14) Reetz, M. T.; Zonta, A.; Schimossek, K.; Jaeger, K.-E.; Liebeton, K., *Angewandte Chemie International Edition in English* **1997,** *36* (24), 2830-2832.

(15) Farinas, E. T.; Bulter, T.; Arnold, F. H., *Current Opinion in Biotechnology* **2001,** *12* (6), 545-551.

(16) Reetz, M. T.; Rentzsch, M.; Pletsch, A.; Maywald, M., *CHIMIA* **2002,** *56* (12), 721.

(17) Taylor, S. V.; Kast, P.; Hilvert, D., *Angewandte Chemie International Edition* **2001,** *40* (18), 3310-3335.

(18) Arnold, F. H.; Georgiou, G. In *Directed enzyme evolution : screening and selection methods*, 2003.

(19) Chica, R. A.; Doucet, N.; Pelletier, J. N., *Current Opinion in Biotechnology* **2005,** *16* (4), 378-384.

(20) Reetz, M. T.; Carballeira, J. D., *Nat. Protoc.* **2007,** *2* (4), 891-903.

(21) Chen, H.; Liu, S.; Padula, S.; Lesman, D.; Griswold, K.; Lin, A.; Zhao, T.; Marshall, J. L.; Chen, F., *Nature Biotechnology* **2020,** *38* (2), 165-168.

(22) Moore, C. L.; Papa, L. J.; Shoulders, M. D., *J. Am. Chem. Soc.* **2018,** *140* (37), 11560-11564.

(23) Fleming, A., *Br J Exp Pathol* **1929,** *10* (3), 226-236.

(24) Sengupta, S.; Chattopadhyay, M.; Grossart, H.-P., *Frontiers in Microbiology* **2013,** *4*.

(25) Ventola, C. L., *P T* **2015,** *40* (4), 277-283.

(26) Piddock, L. J. V., *The Lancet Infectious Diseases* **2012,** *12* (3), 249-253.

(27) Salahuddin, P.; Kumar, A.; Khan, U. A., *Current Protein & Peptide Science* **2018,** *19* (2), 130-144.

(28) Salverda, M. L. M.; De Visser, J. A. G. M.; Barlow, M., *FEMS Microbiology Reviews* **2010,** *34* (6), 1015-1036.

(29) Song, W. J.; Tezcan, F. A., *Science* **2014,** *346* (6216), 1525.

(30) Song, W. J.; Yu, J.; Tezcan, F. A., *J. Am. Chem. Soc.* **2017,** *139* (46), 16772-16779.

(31) Yu, J.; Yang, J.; Seok, C.; Song, W. J., *Chemical Science* **2021,** *12* (14), 5091-5101.

(32) Tomatis, P. E.; Rasia, R. M.; Segovia, L.; Vila, A. J., *Proceedings of*

the National Academy of Sciences **2005,** *102* (39), 13761-13766.

(33) Stiffler, Michael A.; Hekstra, Doeke R.; Ranganathan, R., *Cell* **2015,** *160* (5), 882-892.

(34) Griffin, M. D. W.; Gerrard, J. A., n *Protein Dimerization and Oligomerization in Biology*, Matthews, J. M., Ed. Springer New York: New York, NY, 2012; pp 74-90.

(35) Marianayagam, N. J.; Sunde, M.; Matthews, J. M., *Trends Biochem. Sci.* **2004,** *29* (11), 618-625.

(36) Ali, M. H.; Imperiali, B., *Bioorg. Med. Chem.* **2005,** *13* (17), 5013-5020.

(37) André, I.; Strauss, C. E. M.; Kaplan, D. B.; Bradley, P.; Baker, D., *Proceedings of the National Academy of Sciences* **2008,** *105* (42), 16148-16152.

(38) Song, W. J.; Sontz, P. A.; Ambroggio, X. I.; Tezcan, F. A., *Annu. Rev. Biophys* **2014,** *43*, 409-431.

Chapter 2.
Importance of Scaffold Flexibility/Rigidity in the Design and Directed Evolution of Artificial Metallo-β-lactamases

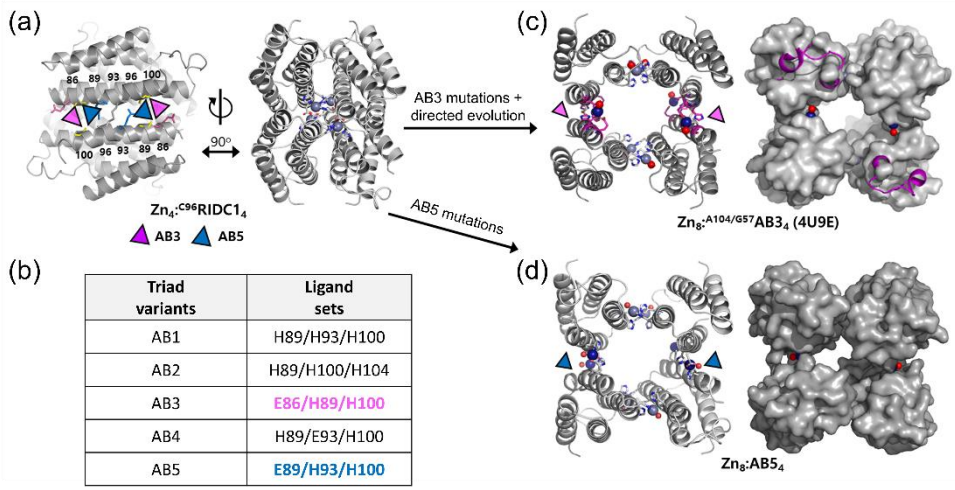
*The Part of this chapter was reprinted with permission from W. J. Song; **J. Yu**; F. A. Tezcan, *J. Am. Chem. Soc.* 2017, 139, 16772, copyright 2017, American Chemical Society.

2.1. Introduction

Owing to their remarkable catalytic efficiency and chemical specificity, natural enzymes have long captured the imagination of scientists.[1-6] Metalloenzymes, in particular, have been a favorite target of protein design and biomimetic synthesis, as they are able to execute challenging chemical transformations in addition to combining the intricacies of inorganic reactivity with the complexity of protein structure and dynamics.[7-11] Metalloenzyme design and engineering have been primarily pursued on two major fronts. One has involved the *de novo* design of small to midsize (typically fewer than 100 residues) peptide/protein scaffolds that house one or more reactive metal centers or metallocofactors.[12-17] On the other front, preexisting proteins have been reengineered with alternative metal sites and/or active-site pockets,[18-23] in a process that is akin to divergent evolutionary processes to generate functional diversity. Because such redesign strategies bypass the enormous complexity of the (reverse) protein folding problem, they have provided rapid access to new metal-based functionalities. Protein scaffolds, such as myoglobin,[19, 24-27] cytochrome P450,[22, 28-30] cytochrome c,[31] TIM barrels,[32-34] (strept)avidin,[21, 35-37] and others,[38, 39] possess well-defined metal/metallocofactor/substrate binding pockets and stable tertiary structures that can withstand extensive mutations and modifications, and have proven to be particularly valuable in protein design and engineering efforts. These features have facilitated the achievement of novel catalytic functions and impressive efficiencies via a combination of rational/computational design and directed evolution in the laboratory.[18, 20-22, 35, 40]

An alternative route for protein evolution and design involves the self-assembly of multiple protein domains into larger architectures.[41-52] This route allows the combination of different protein structures and functionalities in a modular fashion to engender novel catalytic processes, produces new

interfaces for the construction of active sites, and lends itself well to the generation of allostery and regulation through the flexibility of the supramolecular architecture.[41-43] To synthetically mimic this evolutionary pathway, Tezcan *et al.* have devised a design approach (Metal-Templated Interface Redesign, MeTIR), whereby a monomeric protein is directed through metal coordination into oligomeric assemblies.[44, 53-59] Crystal structures of these metal-templated assemblies are subsequently used as a guide for the rational or computational design of the monomeric protein surfaces, which associate to produce stable, self-standing quaternary architectures with "evolutionarily naïve" interfaces. These interfaces are then tailored for desired structural and functional properties. By primarily using a small monomeric heme protein (cytochrome cb_{562} or cyt cb_{562}) as a model system, Tezcan *et al.* constructed, for example, stable quaternary architectures that could form in vitro and *in vivo* and possess controllable flexibilities and rigidities as well as tunable metal coordination sites that allow for high metal selectivity and allosteric coupling.[44, 54-59]



(b)

Triad variants	Ligand sets
AB1	H89/H93/H100
AB2	H89/H100/H104
AB3	E86/H89/H100
AB4	H89/E93/H100
AB5	E89/H93/H100

Figure 2.1 Design and X-ray crystal structures of *de novo* tetrameric metallohydrolases. (a) Cartoon representation of the Zn_4:^{C96}RIDC1$_4$ structure (PDB 3IQ6), which guided the design of AB3 and AB5 variants. Locations of the Zn-binding triad motifs are indicated with magenta and navy triangles for the AB3 and AB5 variants, respectively. Residues 89 and 100, which are used for Zn coordination in both AB3 and AB5, are shown in yellow. Residues 86 and 93, which are used for either AB3 or AB5, respectively, are colored in magenta and blue, respectively. Residues 96 that form a disulfide bond are highlighted in black. (b) List of variants containing the triads of Zn-binding residues examined in previous and current work. (c) X-ray crystal structure of Zn_8:$^{A104/G57}$AB3$_4$ (PDB 4U9E). The flexible loop containing the G57 residue is highlighted in magenta. (d) X-ray crystal structure of Zn_8:AB5$_4$. Core (structural), peripheral (catalytic) Zn ions, and the Zn-bound water molecules are shown as light blue, navy, and red spheres, respectively.

One of the cyt cb$_{562}$ variants, ^{C96}RIDC1, has proven to be a particularly versatile platform because it self-assembles into a stable, yet flexible D_2 symmetric tetramer (^{C96}RIDC1$_4$, $K_{d,\ tetramer} < 100$ nM) via a combination of metal coordination, disulfide and computationally prescribed noncovalent interactions (Figure 2.1a).[55] The ^{C96}RIDC1$_4$ tetramer and its metal-bound

form Zn_4:^{C96}RIDC1$_4$ feature three pairs of C_2 symmetric interfaces that can be independently tailored. In the previous study, Tezcan *et al.* reported that four additional, coordinatively unsaturated Zn coordination sites could be installed into the C96−C96 disulfide-cross-linked interfaces of Zn_4:^{C96}RIDC1$_4$ (Figure 2.1a). The resulting variant Zn_8:^{A104}AB3$_4$ (containing four structural and four catalytic Zn centers) was shown to properly self-assemble in the periplasm of *E. coli* cells and perform *in vivo* catalytic hydrolysis of the β-lactam antibiotic ampicillin.[58] The resulting antibiotic resistance allowed the directed evolution of Zn_8:^{A104}AB3$_4$ through an *in vivo* selection strategy, leading to the discovery of a mutant Zn_8:$^{A104/G57}$AB3$_4$ (Figure 2.1c), which displayed Michaelis−Menten behavior for ampicillin hydrolysis with a catalytic proficiency ($(k_{cat}/K_M)/k_{uncat}$) of 2,300,000. The studies of AB3 thus demonstrated that a protein, which is evolutionarily, functionally, and structurally unrelated to any hydrolase enzyme, could be engineered and evolved into a supramolecular β-lactamase through metal-templated self-assembly.

In this design and evolution process, the simultaneous stability and global/local flexibility of the Zn_4:^{C96}RIDC1$_4$ scaffold emerged as an important element that allowed the structure to adapt to variations in its interfaces. Arguably, a criterion for the evolutionary fitness or designability of a new protein structure is its ability to accommodate different active sites or functions,[60] which underlies the functional versatility of enzyme superfamilies like TIM barrels,[32-34, 60] globins,[19, 24-27] and α/β hydrolases.[61, 62] With this in mind, I asked in the current study the following questions: Can the tetrameric, Zn_4:^{C96}RIDC1$_4$ assembly be engineered to harbor an alternative set of catalytic sites (other than those in Zn_8:$^{A104/G57}$AB3$_4$) in its interfaces? If so, would such a site display different extent of evolvability due to its positioning and local environment? Toward addressing these questions, I report here the structural and functional analysis of a new artificial β-lactamase assembly, Zn_8:AB5$_4$ and its variants.

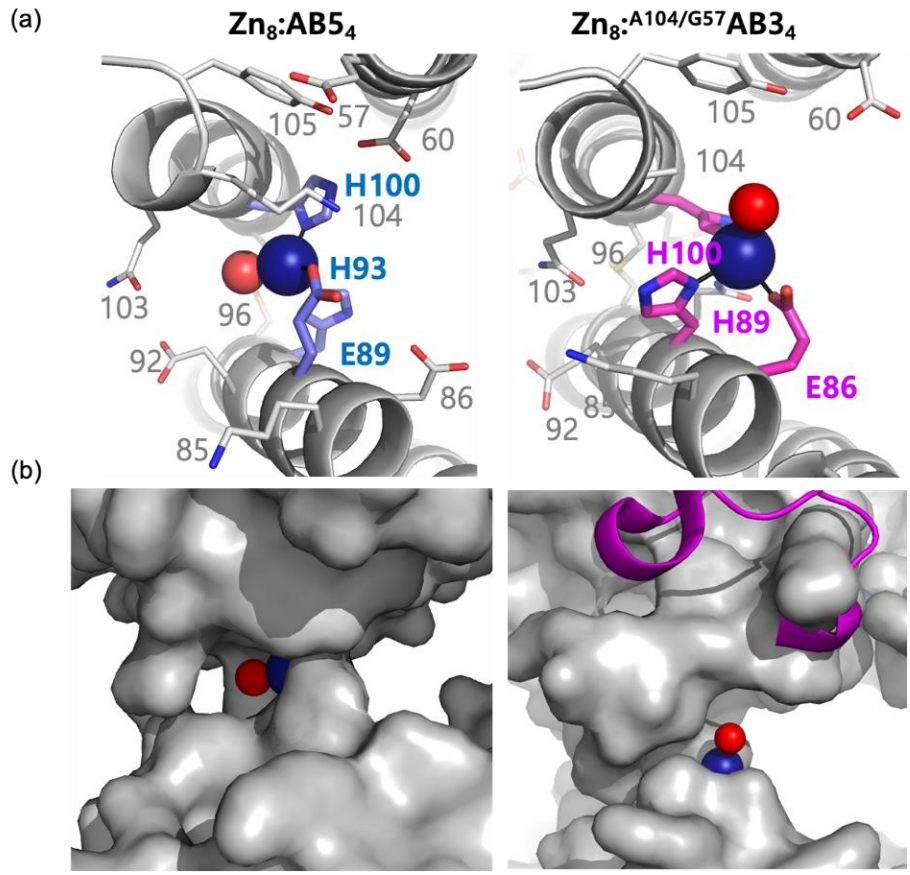


Figure 2.2 Close-up views of the peripheral/catalytic Zn sites in Zn_8:$AB5_4$ (left) and Zn_8:$^{A104/G57}AB3_4$ (right) in (a) cartoon and (b) surface representations. Zn-coordinating and proximal residues are shown as sticks and lines, respectively. Zn centers and associated water molecules are shown as navy and red spheres, respectively. The flexible loop containing the G57 residue in Zn_8:$^{A104/G57}AB3_4$ is colored in magenta. Black arrows indicate the orientation of the Zn−OH_2 moiety in the interfaces of Zn_8:$AB5_4$ and Zn_8:$^{A104/G57}AB3_4$.

2.2. Results and discussion

2.2.1 Design and initial characterization of AB5.

The Zn_8:^{C96}RIDC1$_4$ scaffold contains C96−C96 disulfide bonds at its interfaces that are formed through the antiparallel alignment of two monomers, where the disulfide bond is bisected by a C_2 symmetry axis (Figure 2.1a). This interfacial arrangement gives rise to several triads of amino acid positions that can conceivably be substituted with His and Glu residues to form tripodal Zn^{2+} anchoring sites that resemble the active sites of archetypal mononuclear Zn-hydrolases like carbonic anhydrase[63, 64] (3-His coordination) and carboxypeptidase[65] (2-His/1-Glu coordination) (Figure 2.1b). While these triads abut one another, they can be viewed as completely independent sites differing in terms of their local microenvironments, potential metal coordination geometries, secondary coordination spheres and proximities to the C96−C96 disulfide bond. Several triad variants (AB1, AB2, AB3 and AB4; Figure 2.1b) were prepared in the previous study.[58] Of these four variants, only AB3 (E86/H89/H100) self-assembled into the targeted tetrameric species with sufficiently high abundance (> 90%) upon Zn^{2+} coordination, and therefore was chosen for further study. At the same time, Tezcan group observed that two 89−93−100 permutants AB1 (H89/H93/H100) and AB4 (H89/E93/H100) also bound two equivalents of Zn^{2+} per protein monomer as desired, but they mostly formed heterogeneous oligomeric species.[58]

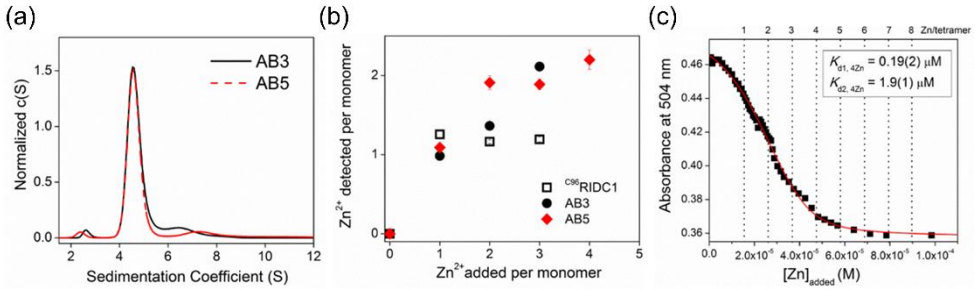


Figure 2.3. Characterization of AB5 variant. (a) AUC spectrum of the Zn-complexed AB5 variant, Zn_8:$AB5_4$. (b) ICP-OES data of the AB5 variant reacted with various ratios of Zn ions per monomeric protein. (c) Determination of Zn-binding affinity of the AB5 protein. A representative spectrum exhibiting the changes from the reaction of pre-mixed Newport green with AB5 protein upon the addition of $ZnCl_2$ is depicted.

Based on the previous observations, I decided to examine in this study if a third permutant of this triad, AB5 (E89/H93/H100), could self-assemble into a tetramer while still binding two Zn ions per monomer. Indeed, analytical ultracentrifugation (AUC) analysis of the protein in the presence of 2-fold Zn^{2+} relative to the monomer, revealed the predominant formation of a tetrameric species with a sedimentation coefficient of 4.2 S (Figure 2.3a). Inductively coupled plasma-optical emission spectroscopy (ICP-OES) measurements indicated that the Zn-AB5 complex contained two Zn^{2+} ions per AB5 monomer (Figure 2.3b), thus confirming the formation of the desired Zn_8:$AB5_4$ complex. Competitive titration experiments with a Zn-responsive fluorophore (Newport Green DCF diacetate) revealed two sets of four-Zn binding equilibria (K_{d1} = 190 ± 20 nM and K_{d2} = 1900 ± 100 nM) (Figure 2.3c), suggesting that the Zn_8:$AB5_4$ complex possesses two distinct sets of Zn-binding sites as planned.

2.2.2 Structural characterization of Zn_8:$AB5_4$.

The crystal structure of Zn_8:$AB5_4$ was determined at 2.65 Å resolution (Figures 2.1d, 2.2, and 2.4); data collection and refinement statistics are listed in Table 2.1. The asymmetric unit of the F_{222}-space group crystals contains a single AB5 monomer, with the crystallographic symmetry generating the perfectly D_2 symmetric tetramer. Interestingly, the tetramer possesses a considerably more open topology compared to that of the parent Zn_4: ^{C96}RIDC1_4 complex, and resembles the open architecture of the catalytically active Zn_8: $^{A104/G57}$AB3_4 variant. In line with analytical measurements, the crystal structure of Zn_8:$AB5_4$ reveals two sets of Zn-binding sites, termed structural and peripheral (or catalytic). The four identical structural sites are composed of three His residues (H63, H73, and H77) with a solvent-derived molecule in a tetrahedral geometry (Figures 2.1d and 2.4), and are nearly identical to those observed in the Zn_8: $^{A104/G57}$AB3_4 complex (Figure 2.1c). Although these Zn-binding sites are coordinatively unsaturated, they are considerably less exposed to solvent compared to the peripheral sites due to occlusion by W41, W66 and R62 side chains, and were previously shown not to be involved in catalysis.[58] As designed, the peripheral Zn sites in Zn_8:$AB5_4$ are coordinated by the newly installed triad motif composed of E89 (monodentate), H93 (N_ε), and H100 (N_δ) at an average coordination distance of 2.2 Å and a solvent molecule (H_2O or OH^-) as the fourth ligand at a coordination distance of 2.1 Å (Figure 2.2a and Table 2.2). To accommodate this Zn binding mode, the protein monomers undergo a significant rotational/ bending motion around the C96−C96 hinges with respect to the Zn_4:^{C96}RIDC1_4 structure (Figure 2.1a), which enforces the tetrameric assembly to adopt an open conformation and distorts the structural Zn centers into coordinative unsaturation.

Protein	Zn_8:AB5$_4$	Zn_8:^{C96T}AB5$_4$
Crystallization Condition	31.5% 2-methyl-2,4-pentanediol 400, 0.1M HEPES (pH 7.5), 0.2M $MgCl_2$	45% 2-methyl-2,4-pentanediol 400, 0.1M HEPES (pH 6.5)
Data collection location	SSRL 12-2	SSRL 12-2
Unit cell dimensions	50.29×98.00×99.33 α = β = γ = 90°	60.75×3.48×127.79 α = β = γ = 90°
Symmetry group	F222	$P2_12_12_1$
Resolution (Å)	26.61–2.65	36.18–1.98
X-ray wavelength (Å)	0.9795	1.0000
Number of unique reflections	3383	35176
Redundancy*	3.8 (3.7)	5.8 (6.0)
Completeness* (%)	91.5 (82.4)	99.9 (100.0)
$< I /\sigma >$*	25.6 (16.5)	17.8 (3.4)
R_{symm}* (%)	4.5 (6.4)	6.4 (55.9)
R_{work}/R_{free} (%)	20.5/24.5	17.1 (21.4)
Number of atoms		
Protein	827	3312
Ligands/ions	43/5	172/15
Water	66	288
B-factors (Å^2)	45.0	31.2
R.m.s deviations		
Bond lengths (Å)	0.0111	0.0200
Bond angles (°)	1.4150	1.8423

Table 2.1. Crystallographic data collection and refinement statistics for Zn_8:AB5$_4$ and Zn_8:^{C96T}AB5$_4$ structures. *denotes values from the highest resolution shell.

Proteins/ residues	**Zn–Nδ in H63**	**Zn–Nε in H73**	**Zn–O in D74**	**Zn–Nε in H77**	**ions/water**
Zn_8:AB5$_4$	2.0	1.8	nd	2.0	2.2 (Cl^-)
Zn_8:^{C96T}AB5$_4$	2.1 2.1 2.0 2.1	2.1 2.1 2.1 2.1	2.0 1.9 2.1 2.1	2.1 2.1 2.1 2.1	nd nd nd nd
Proteins/ residues	Zn–O in E89	Zn–Nε in H93	Zn–Nδ in H77		ions/water
Zn_8:AB5$_4$	2.0	1.8	2.0		2.1 (H_2O)
Zn_8:^{C96T}B5$_4$	2.1 2.1 2.0 2.1	2.1 2.1 2.1 2.1	2.1 2.1 2.1 2.1		2.1 (H_2O) nd 2.2 (Cl^-) 2.3 (H_2O)

Table 2.2. Distances (Å) of Zn and coordinating residues in (top) core/structural and (bottom) peripheral/catalytic Zn-binding sites.

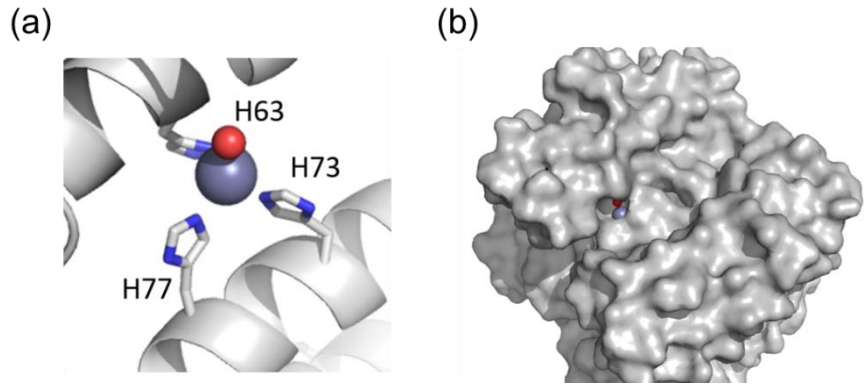
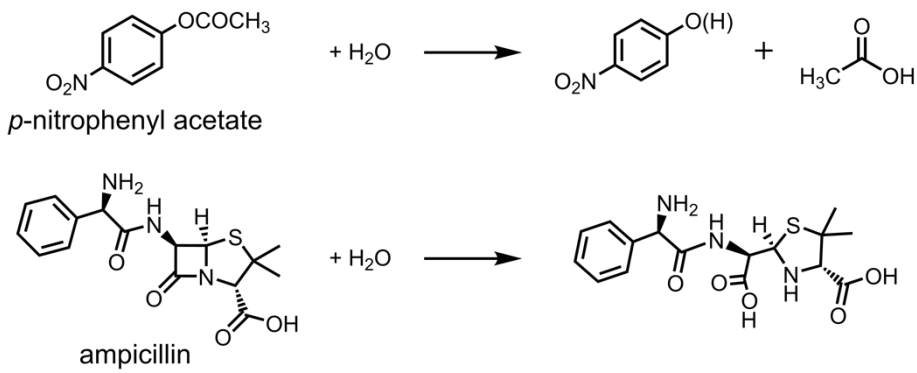


Figure 2.4. Views of core/structural Zn-binding sites in Zn_8:$AB5_4$. (a) Close-up view in cartoon representation and (b) surface representation.



Scheme 2.1. Hydrolytic reactions monitored in this study

Despite the fact that the triads in AB3 and AB5 share two common positions (89 and 100), the peripheral Zn coordination microenvironments in Zn_8:$AB5_4$ are entirely different from those in Zn_8:$^{A104/G57}AB3_4$ (Figure 2.2). Notably, the Zn−OH_2 moiety in Zn_8:$AB5_4$ is directed toward the exterior of the tetramer rather than to the interior as observed in Zn_8:$^{A104/G57}AB3_4$. In this orientation, the putative active site of Zn_8:$AB5_4$ is wedged between two interfacial α-helices and surrounded by a different set of side chains (K85, E92, Q103 and K104) compared to that in Zn_8:$^{A104/G57}AB3_4$ (G57, D60, A104, Y105). Additionally, the Zn−OH_2 moiety in Zn_8:$AB5_4$ is directed away from the hydrophobic 43−57 loop (shown in magenta in Figure 2.2b, right), which had emerged as a key element in the directed evolution of Zn_8: $^{A104/G57}AB3_4$

for enabling ampicillin binding interactions.[58]

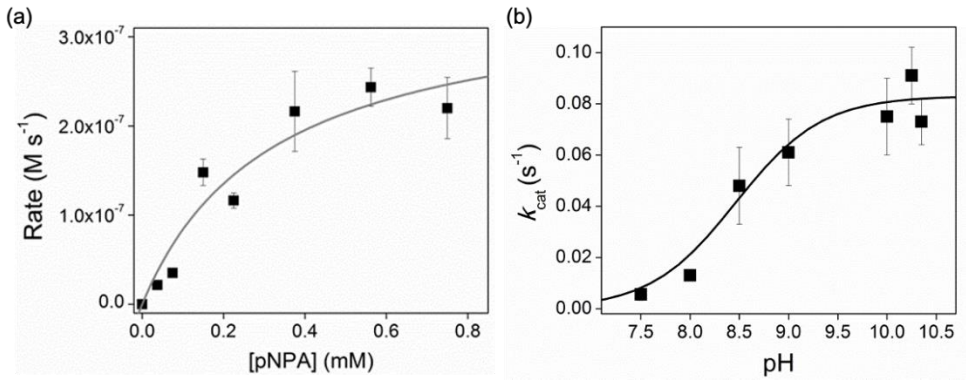


Figure 2.5. Esterase activity of Zn_8:$AB5_4$, with *p*-nitrophenyl acetate as a substrate. (a) Michaelis−Menten kinetics. (b) pH-dependent esterase activity. Data points and error bars are obtained from average and propagated standard deviations, respectively for at least three independent measurements after background correction.

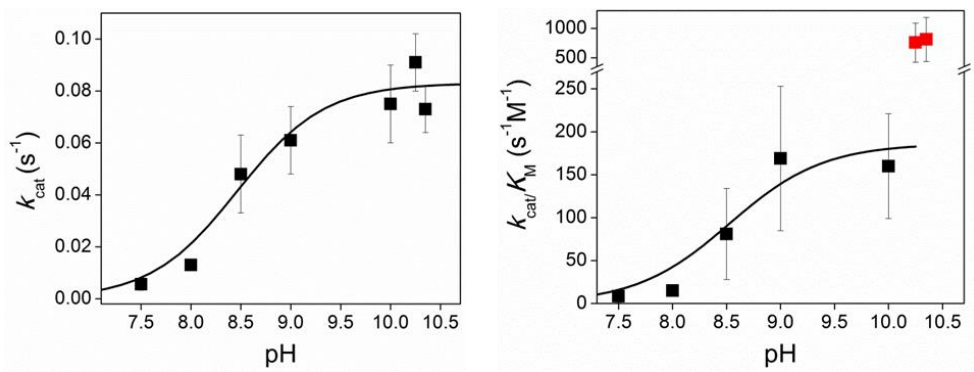


Figure 2.6. Kinetic analysis of Zn_8:$AB5_4$ for the pH-dependent esterase activity with *p*-nitrophenyl acetate as the substrate. In the plot of k_{cat}/K_M of Zn_8:$AB5_4$ versus pH, data points at pH 10.25 and 10.35 are masked (red squares) to estimate the first pK_a value related to the catalysis.

2.2.3 *In vitro* esterase and β-lactamase activity of Zn_8:$AB5_4$.

Having established that Zn_8:$AB5_4$ presents a new set of coordinatively unsaturated Zn coordination sites within distinct microenvironments, I first investigated its catalytic activity toward the hydrolysis of *p*-nitrophenyl

acetate (*p*NPA) as a chromogenic substrate (Scheme 2.1). Time-dependent spectral changes at 400 nm, stemming from the formation of the product p-nitrophenolate, were monitored, and the net hydrolytic activity of Zn_8:$AB5_4$ was determined by measuring the difference between the product formation rates in the presence and absence of Zn^{2+}. The background reaction due to free Zn^{2+} was also taken into account. Zn_8:$AB5_4$ was found to catalyze pNPA hydrolysis with a $k_{cat} = 0.06 \pm 0.01$ s^{-1} and a $k_{cat}/K_M = 180 \pm 90$ s^{-1} M^{-1} (at pH 9) (Figure 2.5a), and these values are several-fold higher than those for Zn_8: $^{A104/G57}AB3_4$ ($k_{cat} = 0.019$ s^{-1} , $k_{cat}/K_M = 29$ s^{-1} M^{-1}),[58] and comparable to the most efficient artificial esterases.[15, 16, 66-68] Through the determination of the Michaelis−Menten parameters for *p*NPA hydrolysis at different pHs, I determined the pK_a value of the Zn–OH_2 species to be 8.4 ± 0.1 (Figures 2.5b and 2.6). This value is considerably lower than that for Zn_8: $^{A104/G57}AB3_4$ (8.9 ± 0.2) and likely accounts for the higher catalytic rate of Zn_8:$AB5_4$ at pH 9.[58] It is important to note that the primary coordination spheres of Zn_8:$AB5_4$ and Zn_8: $^{A104/G57}AB3_4$ are identical to one another in terms of composition (His N_ε/His N_δ/Glu O) and the lack of H bonding to these ligands by protein side chains or solvent. These similarities in the coordination environments suggest that the large difference in the Lewis acidities of the Zn−OH_2 moieties must stem from their specific orientation within the protein−protein interface and corresponding variations in the surrounding microenvironment as observed in other systems.[68]

Next, I investigated the hydrolytic activity of Zn_8:$AB5_4$ using ampicillin as a substrate (Scheme 2.1) and monitoring the reaction by HPLC analysis of the reaction mixtures. The results demonstrate that Zn_8:$AB5_4$ is a β-lactamase ($k_{cat} = 1.03 \pm 0.06$ min^{-1} and $k_{cat}/K_M = 120 \pm 10$ min^{-1} M^{-1}) and displays a substrate saturation behavior at increasing ampicillin concentrations (Figure 2.7a and Table 2.3), implying that the protein forms a Michaelis complex with the substrate. This contrasts with the Zn_8:$^{A104}AB3_4$ construct, which did not show Michaelis−Menten behavior until it was subjected to directed evolution

via saturation mutagenesis of the surrounding residues to yield the improved variant Zn_8:$^{A104/G57}AB3_4$.[58] Docking simulations indicate that ampicillin can be situated in a small niche on the protein surface in a catalytically viable conformation in which the carbonyl group of the ampicillin lactam ring is oriented toward the Zn−OH_2 site (Figure 2.7b).

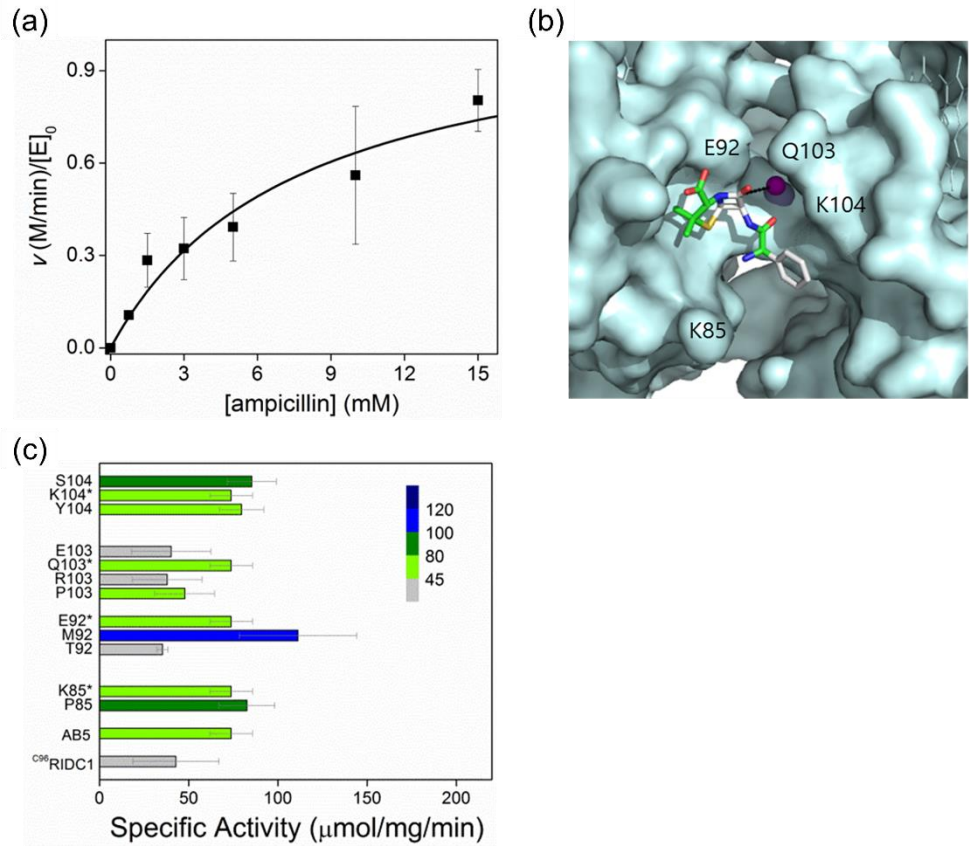


Figure 2.7. β-lactamase activity of Zn_8:$AB5_4$ with ampicillin as substrate. (a) Michaelis−Menten kinetics. Data points and error bars were obtained from at least two independently prepared set of samples and measurements. (b) Docking simulation of Zn_8:$AB5_4$ with ampicillin, which is depicted as sticks. Zn center and associated water molecule are shown with as gray and purple spheres, respectively. Residues nearby the Zn center are labeled on the surface representation. The distance between the Zn-bound atom and the carbon in the lactam-ring is represented as a dotted black line. (c) *In vitro* specific activity of AB5 single variants with ampicillin as the substrate.

Variant	pK_a	k_{cat} (min^{-1})	K_M (mM)	k_{cat}/K_M or k_2 (min^{-1} M^{-1})[a]
AB5	8.4(1)	1.03(6)	9(1)	120(10)
^{P85}AB5	8.5(1)	0.29(7)[a]	nd[a]	nd[a]
^{M92}AB5	9.0(2)	1.1(3)	8(4)	140(40)
^{S104}AB5	8.9(1)	nd[a]	nd[a]	67(15)[a]
^{C96T}AB5	9.1(2)	1.4(2)	11(3)	120(20)

Table 2.3. Net β-lactamase activities of *de novo* metallohydrolases at pH 9.0
[a]This variant exhibits no substrate saturation behavior or displays too low of an activity for the determination of an accurate K_M value

2.2.4 *In vivo* assembly, activity, and directed evolution of Zn_8:$AB5_4$.

Encouraged by the nascent β-lactamase activity of Zn_8:$AB5_4$, I examined whether this activity could also be realized in *E. coli* cells and used for the directed evolution of the substrate binding pocket as described previously.[58] Toward this end, AB5 was expressed in kan^+ /amp^- *E. coli* cells which display negligible background levels of β-lactamase activity. To enable proper maturation of its c-type heme cofactor, AB5 contains an N-terminal leader sequence for translocation to the periplasm. The periplasmic space is a generally favorable cellular compartment for directed evolution of artificial enzymes as the chemical environment in this compartment can be more easily manipulated than the tightly regulated cytoplasm.[35] Specifically, in this case, the periplasmic expression promotes disulfide bond formation and allows access to sufficiently high ambient Zn^{2+} concentrations (through the adjustment of Zn^{2+} concentration in the growth medium), which should favor the proper self-assembly of Zn_8:$AB5_4$. Indeed, the analysis of the periplasmic cell extracts by size-exclusion chromatography (SEC) followed by the ICP-OES analysis of the fractions indicated that AB5 primarily existed as a tetramer that contained two Zn ions per protein monomer (Figure 2.8),

confirming the proper *in vivo* self-assembly of Zn_8:$AB5_4$.

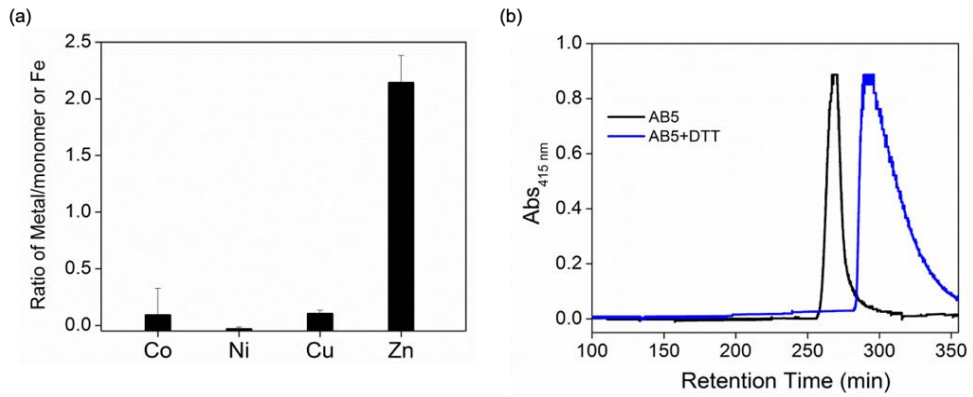


Figure 2.8. Characterization of AB5 extracted from the periplasm of *E. coli*. (a) Metal contents of AB5 protein detected by ICP-OES. (b) Size-exclusion chromatograms (Sephadex S75) of AB5 protein before and after the treatment with 1,4-dithiothreitol (DTT).

For the directed evolution of Zn_8:$AB5_4$, I selected four residues, K85, E92, Q103, and K104, that encircle the putative active site pockets for saturation mutagenesis.[69] Upon creation of the single-site libraries, *in vivo* selection was carried out using LB agar plates containing ≥ 1 mg/L ampicillin. Surviving colonies were isolated as positive hits and sequenced, and the frequency of colonies expressing specific AB5 variants was determined (Figures 2.9 and 2.6). Surprisingly, there was no variant that displayed substantially higher survival frequency than the parent AB5 for any of the four positions. This finding contrasts with the observations on AB3 whose single-site variants (E57G and Y105T) showed more than 10-fold higher survival frequencies over the parent species as a result of significantly improved catalytic performance.[58]

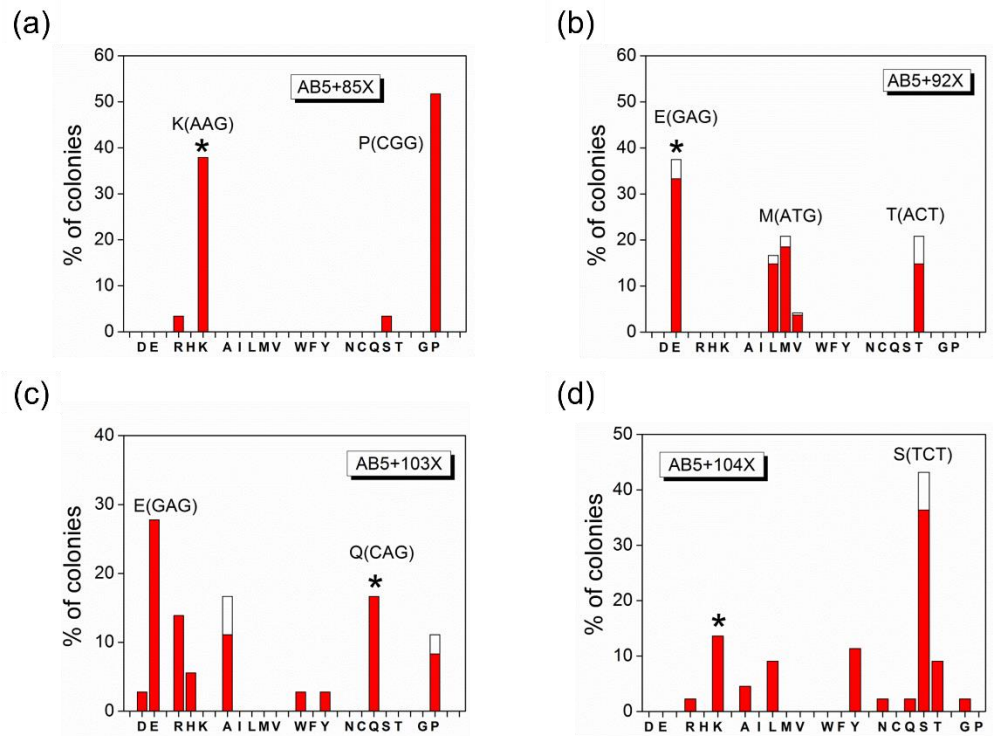


Figure 2.9. Frequencies of ampicillin surviving colonies in saturated AB5 single mutant libraries. To obtain more accurate statistics of the screening results, each codon, rather than amino acid, were counted so that there was no bias for the amino acids that are encoded by more than one codon. The dominant and minor outputs for each amino acid were shown with red and white bars, respectively. The original sequence of AB5 protein is shown with an asterisk. (a) AB5+K85X, (b) AB5+E92X, (c) AB5+Q103X and (d) AB5+K104X.

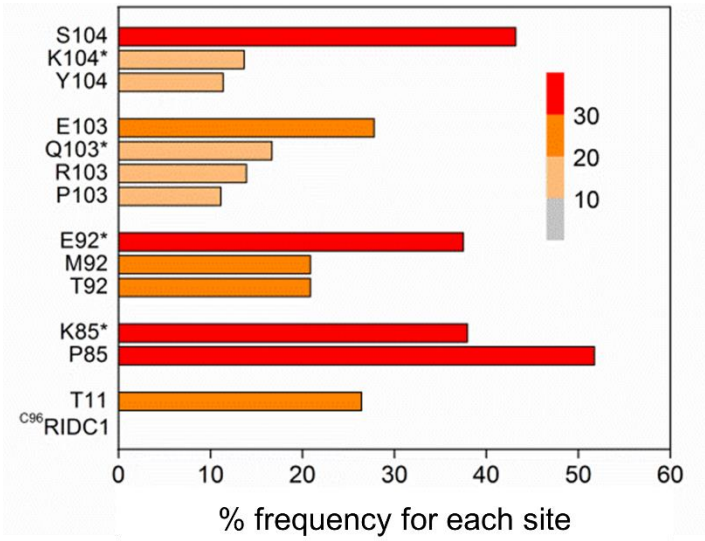


Figure 2.10. *In vivo* survival frequencies of AB5 variants (best ones for each position) adapted from the statistics in Figure 2.9

I expressed and isolated several single-site AB5 variants that showed the highest survival frequency (e.g., K85P, E92M and K104S) at each position (Figure 2.2a) and measured their *in vitro* β-lactamase activities upon forming the Zn-assembly (Figures 2.7c and 2.11, and Table 2.1). These analyses confirmed that these variants indeed could catalyze ampicillin hydrolysis with appreciable improvement over k_{uncat} and that the M92 variant displayed enzyme-like Michaelis−Menten (i.e., substrate saturation) behavior. Nevertheless, the overall β-lactamase efficiencies of these single-site variants were not substantially better than that of AB5 (Table 2.3), in accordance with the observation that they did engender significantly different survival frequencies. Only the M92 variant displayed a comparable k_{cat}/K_M value to AB5. I note, however, that this improvement is likely an underestimate given that the pK_a of the Zn−OH_2 species in the M92 variant is 0.6 pH units higher than that of the parent AB5 complex and thus contains considerably less of the active, deprotonated species (～51% vs ～75% as deduced from the pH-dependent activity data) at pH 9, where the enzymatic assays were conducted. Given the modest changes in catalytic activity, it is difficult to pinpoint why

single-site mutations like E92M give rise to higher *in vivo* survival frequencies or improved kinetic parameters, particularly in the absence of information on a) the *in vivo* expression levels and self-assembly yields of the mutants and b) structural data on protein−substrate complexes. I also prepared several multisite variants (P85/M92, P85/S104, M92/S104, P85/ M92/S104) combining the most promising single-site mutations, but these did not display any additive effects in mutations (Figure 2.12).[70] These findings suggest that the limit of evolvability through single-site saturation may have been reached for the Zn_8:$AB5_4$ scaffold as in its current structural state.

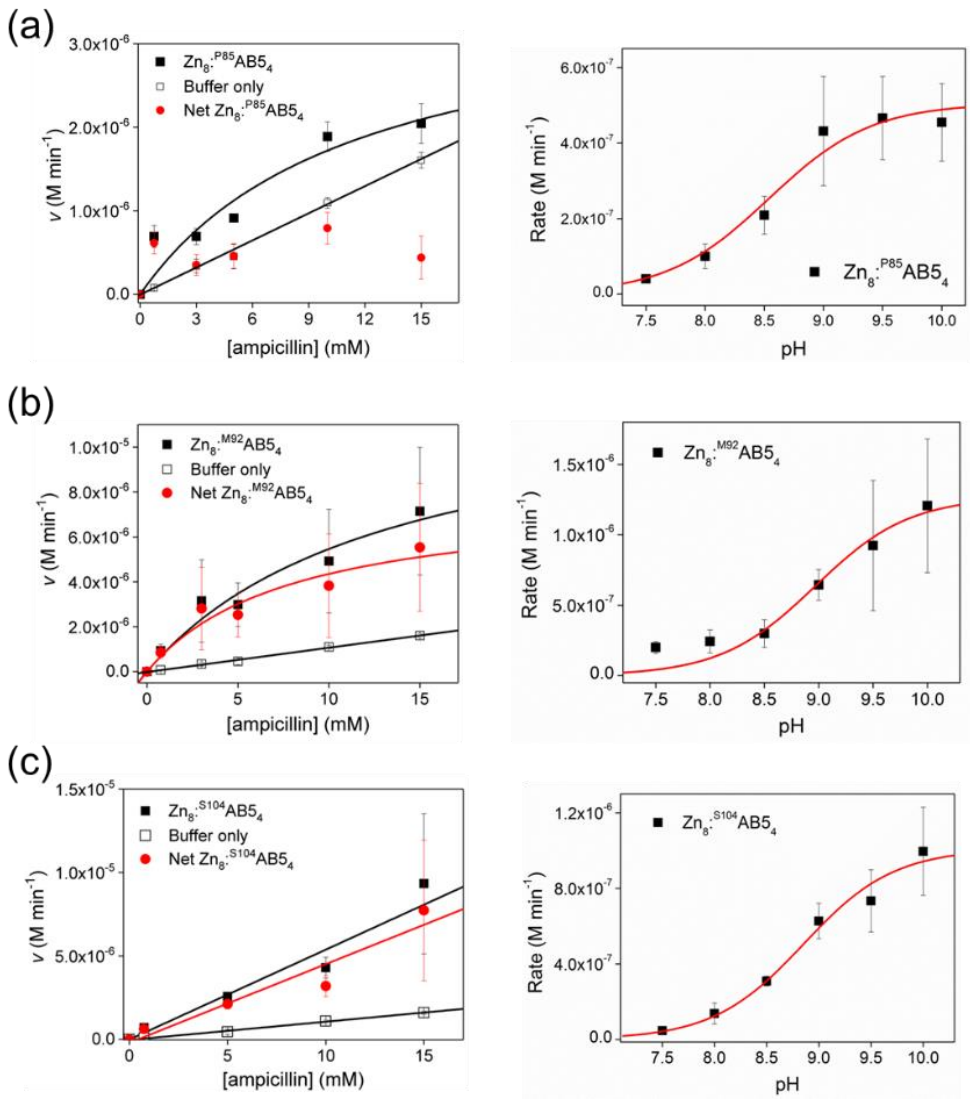


Figure 2.11. Michaelis–Menten kinetics (left) and pH-dependence (right) of β-lactamase activities. (A) Zn_8:$^{P85}AB5_4$ (B) Zn_8:$^{M92}AB5_4$ (C) Zn_8:$^{S104}AB5_4$

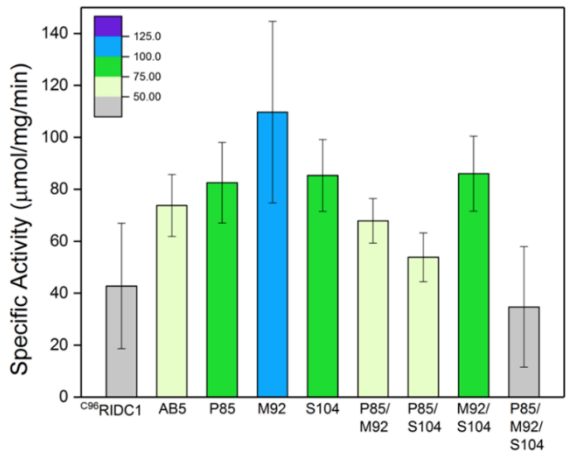


Figure 2.12. Specific ampicillin hydrolytic activities of single and multiple mutants of AB5.

2.2.5 Structural and functional characterization of Zn_8:^{C96T}AB5_4, a variant lacking peripheral disulfide bonds

The lack of improvement in the β-lactamase activity of Zn_8:$AB5_4$ in the course of directed evolution was surprising at first because this complex appeared to be a better primordial scaffold than Zn_8:$^{A104}AB3_4$ due to its higher starting catalytic efficiency and its apparently well-defined substrate binding site. Yet, I had observed that Zn_8:$^{A104}AB3_4$ benefitted significantly, in the context of directed evolution and the discovery of the high-performing E57G variant, from structural flexibility of the supramolecular scaffold, having metal-coordinating residues near the monomer C-terminus and a flexible loop near the catalytic Zn sites. I therefore speculated that the modest changes in the catalytic performance of the Zn_8:$AB5_4$ scaffold may be linked to the proximity of catalytic Zn sites to the C96−C96 disulfide linkages (Figure 2.1a), which may restrict local and global flexibility. These disulfide bonds were previously introduced to stabilized the tetrameric scaffold at low protein concentrations and/or in the absence of Zn ions.[55] Now, given the additional stability afforded by the presence of the catalytic Zn sites in the same protein−protein interfaces, I reasoned that the C96−C96 linkages may not be

necessary for tetrameric assembly. Thus, I reverted C96 back to the wild-type threonine, and isolated the resulting ^{C96T}AB5 construct for structural and functional examination.

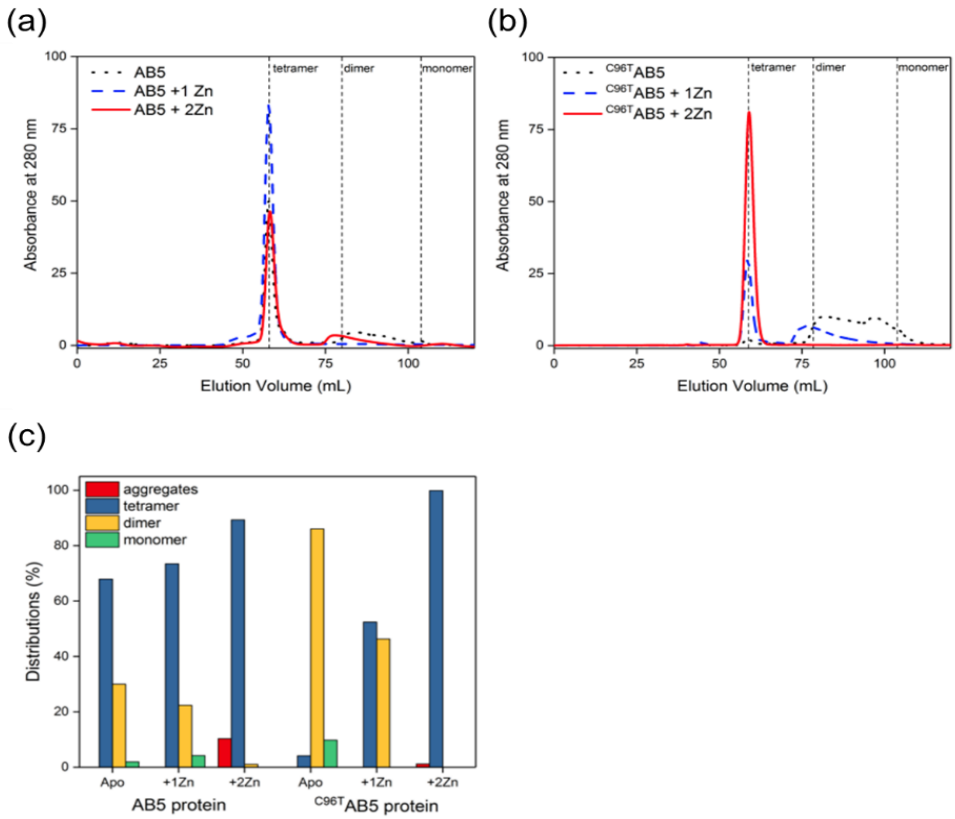


Figure 2.13. Size-exclusion chromatograms of AB5 and ^{C96T}AB5 proteins in the absence of Zn (black dotted lines), in the presence of 1 Zn/monomer (blue dashed lines), and in the presence of 2 Zn/monomer (red solid lines). (a) AB5 (b) ^{C96T}AB5 (c) Distribution of tetrameric, dimeric, and monomeric states from the data in (a) and (b). Small fractions of protein eluted earlier than tetramer were tentatively assigned as aggregates.

The ^{C96T}AB5 protein is found primarily in a dimeric state in the absence of metal ions, but forms a stable tetramer upon addition of 1 or 2 equiv of Zn^{2+} (Figure 2.13). The X-ray crystal structure of Zn_8:^{C96T}AB5$_4$ was determined at 1.98 Å resolution (Table 2.1), confirming the tetrameric assembly state and the stoichiometry of two Zn ions per protein monomer, with the expected sets of core/structural and peripheral/catalytic sites (Figure 2.14a). Interestingly, Zn_8:^{C96T}AB5$_4$ possesses a compact architecture very similar to the "grandparent" Zn_4:^{C96}RIDC1$_4$ scaffold, rather than the open, parent Zn_8:AB5$_4$ assembly. Apparently, the simultaneous presence of Zn-coordination sites and C96−C96 disulfide bonds in the same interfaces imposes a structural strain on the supramolecular assembly to favor an open state; this strain is released upon removal of either the Zn coordination or the C96−C96 linkages (Figure 2.14b).

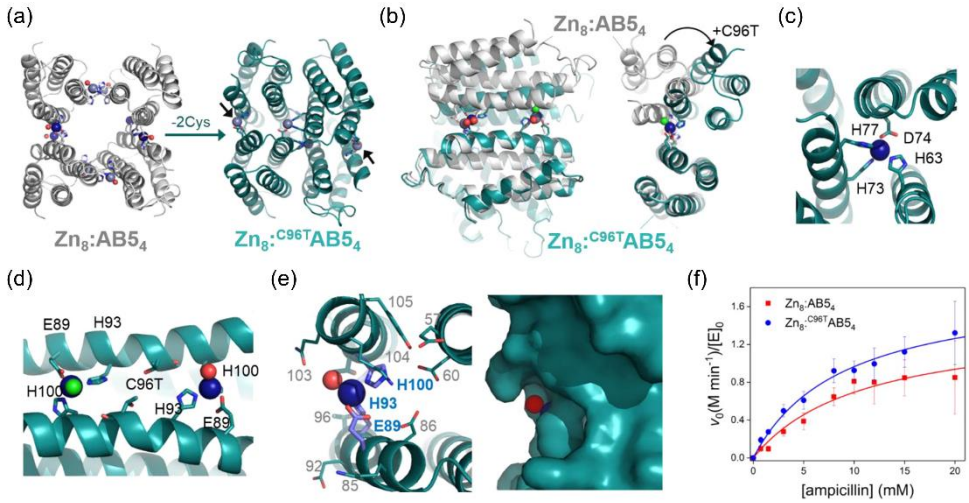


Figure 2.14. Characterization of AB5 and ^{C96T}AB5 variants. (a–d) X-ray crystal structure of Zn_8:^{C96T}AB5_4. Catalytic Zn-binding sites are indicated with black arrows. (b) Overlay of Zn_8:^{C96T}AB54 (cyan) and Zn_8:AB5_4 (gray) structures. Zn centers (navy) and Zn-coordinating water (red) or chloride ions (green) are shown as spheres. (c) Close-up view of the core/structural Zn-binding sites. (d) Close-up view of the peripheral/catalytic Zn-binding sites adjacent to the C96T mutation site. Zn-bound water and chloride ion are shown as red and green spheres, respectively. (e) Enlarged peripheral/catalytic Zn-binding sites Zn-coordinating and adjacent residues are shown as sticks and lines, respectively. (f) *In vitro* β-lactamase activity of Zn_8:AB5_4 and Zn_8:^{C96T}AB5_4 at pH 9.5. Data points and error bars are obtained from at least two independently prepared set of samples and measurements.

As a result of this compaction, the structural Zn sites assume the original, coordinately saturated conformation as found in Zn_4:^{C96}RIDC1$_4$ (Figures 2.1a and 2.9c), whereas the catalytic sites sustain the unsaturated, three-coordinate geometry as in the Zn_8:AB5$_4$ structure (Figure 2.14d, e). The geometric changes in the structural sites are also manifested in the Zn binding constants determined by competitive titration experiments (Table 2.4) one set of Zn-binding affinities increase by > 6-fold from a K_d = 190 nM to K_d = 31 nM (consistent with the increase in coordination number from three to four), whereas the other set (K_d = 2 μM) remains the same. In the course of the structural compaction, the intermonomer interfaces that previously contained the C96−C96 linkages undergo a drastic conformational change (Figure 2.14b), and the peripheral Zn sites become more exposed to solvent and no longer surrounded by the side chains of position 92 and 103 residues (Figure 2.14e).

Protein	Zn-binding constants (μM)
AB5	K_{d1} = 0.19(2) K_{d2} = 1.9(1)
^{C96T}AB5	K_{d1} = 0.031(4) K_{d2} = 2.04(6)

Table 2.4. Zn-binding affinity of AB5 variants estimated by the competition assay with Zn-responsive fluorophores.

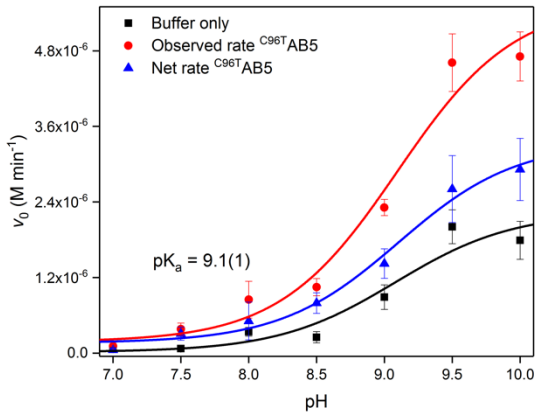


Figure 2.15. pH-dependent *in vitro* β-lactamase activity of Zn_8:$^{C96T}AB5_4$. The net activity from the Zn_8:$^{C96T}AB5_4$ protein was measured by monitoring the background reaction with the identical buffer used for the assay.

These changes are reflected in catalytic properties of the protein assembly. Although the primary coordination spheres of the peripheral Zn sites of Zn_8:$^{C96T}AB5_4$ are identical to those in Zn_8:$AB5_4$, their hydrolytic activities are considerably perturbed, with a 36% increase in k_{cat} at pH 9 despite an elevation of the pK_a value of Zn−OH_2 by >0.7 pH units to 9.1 (Table 2.3 and Figure 2.15). Zn_8:$^{C96T}AB5_4$ also displays a compensating, higher K_M (likely due to the loss of the small substrate interaction pocket), resulting in similar k_{cat}/K_M values for the two species at pH 9.0 (Table 2.3). In light of the large difference between the active site pK_a values of Zn_8:$^{C96T}AB5_4$ and Zn_8:$AB5_4$, I measured their ampicillin hydrolysis activities at pH 9.5, where the Zn−OH_2 moieties should be predominantly deprotonated in both species. Under these conditions, Zn_8:$^{C96T}AB5_4$ displayed increases of ~20% in k_{cat} and ~60% in k_{cat}/K_M (Figure 2.14f and Table 2.5) over those for Zn_8:$AB5_4$. The k_{cat}/K_M value of 210 min^{-1} M^{-1} is second only to the previously evolved, Zn_8:$^{A104/G57}AB3_4$ variant (350 min^{-1} M^{-1}) among all β-lactamase variants I have examined, and, notably, this catalytic activity is achieved without the benefit of a nearby loop to enhance ampicillin binding interactions. Taken together, these observations on Zn_8:$AB5_4$ and Zn_8:$^{A104}AB3_4$ emphasize the

particular importance of point mutations which exert large effects on local and global protein flexibility in terms of enzyme design and evolvability. Although the *in vivo* hydrolytic activities of both Zn_8:$AB5_4$ and Zn_8:$^{C96T}AB5_4$ are sufficiently high to allow for selection (Figure 2.16), they exhibit considerably higher pK_a values than the physiologically relevant pHs (by2−2.5 units), suggesting that the effective *in vivo* concentration of the active, deprotonated Zn−OH species may be quite low. Therefore, perturbations aimed at lowering the active site pK_a's may be particularly critical in increasing the effectiveness of future directed evolution studies that rely on *in vivo* selection.

Variants	k_{cat} **(min^{-1})**	K_M **(mM)**	k_{cat}/K_M **or** k_2 **(min^{-1} M^{-1})**
AB5	1.5(2)	11(3)	130(20)
^{C96T}AB5	1.8(1)	9(1)	210(20)

Table 2.5. Net β-Lactamase activities of *de novo* Metallohydrolases at pH 9.5.

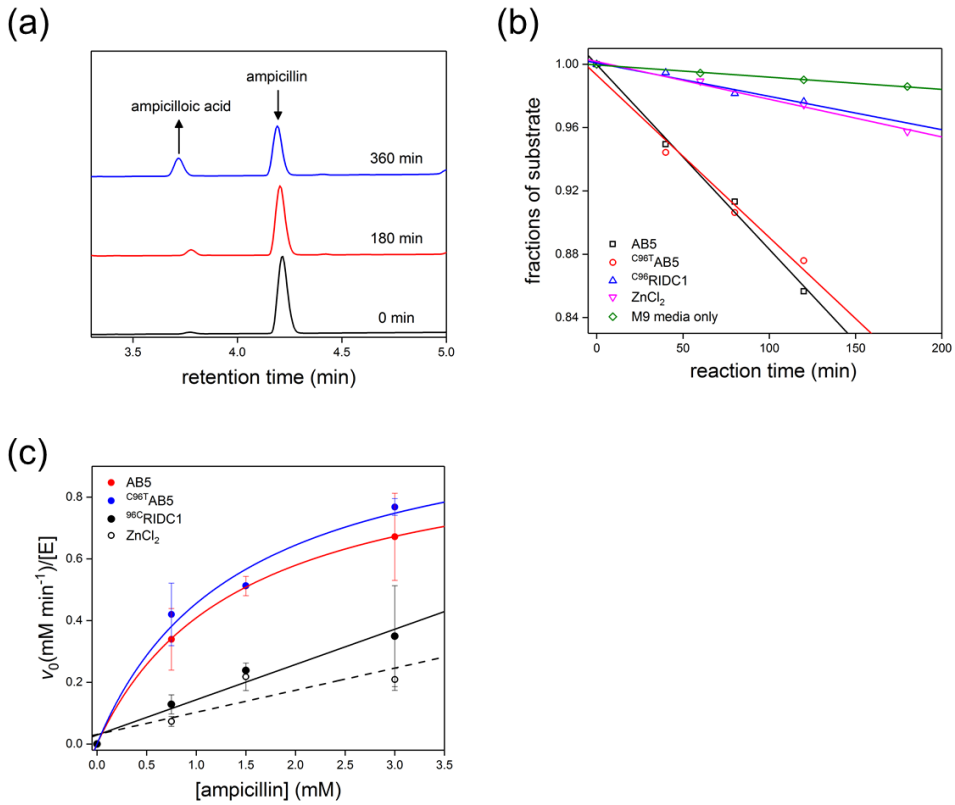


Figure 2.16. Kinetic analyses of *in vivo* β-lactamase activity. (a) Representative time-dependent HPLC traces, displaying the consumption of ampicillin by the hydrolytic activity of *E. coli* cells expressing either AB5. (b) The plot of fraction of substrate versus reaction time for the *E. coli* cells either expressing AB5, ^{C96T}AB5, ^{C96}RIDC1 proteins, or in the presence and absence of $ZnCl_2$ in M9 medium. (c) Ampicillin hydrolysis activity of at various substrate concentrations for the *E. coli* cells either expressing AB5, ^{C96T}AB5, ^{C96}RIDC1 protein, or in the presence of free $ZnCl_2$ in M9 media.

2.3. Conclusion

I have reported here the design, engineering, and evolution of a new protein complex, Zn_8:$AB5_4$, which is the second *de novo* designed protein complex that displays enzymatic activity *in vivo*. Like its predecessor Zn_8:$AB3_4$, the design of Zn_8:$AB5_4$ was based on the metal-templated assembly of a simple four-helix bundle protein cyt cb_{562}, which has no structural or functional homology to any known β-lactamases or metallohydrolases. These findings highlight again the facility with which metal-directed protein assembly can simultaneously generate structural and functional diversity. The exact positioning of the catalytic Zn centers of Zn_8:$AB5_4$ deviates only slightly from those in Zn_8:$^{A104}AB3_4$ in that the two are essentially only half a helix-turn away from one another and share two of the three Zn-coordinating ligands. Yet, this positioning places the catalytic $Zn{-}OH_2$ moieties in entirely disparate microenvironments in the two scaffolds, with different sets of secondary sphere residues and local structure, and different proximal structural elements like surface loops and disulfide bonds. These structural differences are reflected both in the catalytic activities and evolvabilities of Zn_8:$AB5_4$ and Zn_8:$^{A104}AB3_4$. Whereas Zn_8:$^{A104}AB3_4$ required further engineering and directed evolution to attain Michaelis−Menten kinetics for ampicillin hydrolysis, Zn_8:$AB5_4$ displayed such enzyme-like behavior from the start, owing likely to the fortuitous occurrence of a small surface cavity containing the $Zn{-}OH_2$ moiety. Yet, the ampicillin hydrolase activity of Zn_8:$AB5_4$ proved to be considerably less evolvable compared to Zn_8:$^{A104}AB3_4$, which I ascribe to the higher local rigidity and the lack of mutable loops near its active sites. While incorporating new loops near the $Zn{-}OH_2$ centers would be difficult, I have shown that the rational elimination of a proximal disulfide bond can increase local and global flexibility, which resulted in the restructuring of the active site and improved catalytic activity. These findings reaffirm previous observations that the

delicate balance between flexibility[71-74] and stability[75-77] should be taken into account for protein and enzyme design. Finally, the design and evolution studies I carried out here were directed at a specific reaction due primarily to practical reasons (i.e., the ease of screening for β-lactamase activity), which is a general limitation of directed evolution. Excitingly, recent studies of cyt P450 and cyt c mutant libraries have demonstrated that the reactivity scope and evolvability of a protein scaffold can be dramatically widened through the proper choice of screening reactions combined with rational engineering.[29, 31] Along these lines, it will be interesting to explore whether the various metal-templated cyt cb_{562} assemblies I have generated may display high catalytic activities and evolvabilities for other substrates or reactions, particularly when combined with different metal ions.

2.4. Experimental section

Materials and methods for mutagenesis, protein expression and purification. Site-directed mutagenesis of cyt cb_{562} variants was carried out by polymerase chain reaction (PCR) as previously published.[58] The E86D, A89E, Q93H, and A100H mutations were performed using the previously constructed plasmid, kan^{+}/amp^{-}-pET20b(+)/ C96RIDC1.[58] The A89E, Q93H and A100H mutations were made to create a Zn-coordination site. The conservative E86D mutation was additionally made to prevent Zn coordination in an undesired geometry. The sequences of the primers used for PCR are listed in Table 2.3. After mutagenesis and digestion with DpN I for 1 h at 37 ºC, the plasmids were transformed to XL1-Blue or DH5α *E. coli* competent cells. After cell growth overnight at 37 ºC on LB/agar plates containing 50 mg/L kanamycin, single colonies were picked and grown in 5 mL LB media for ~8-10 h at 37 ºC in 200 rpm orbital shaker. The cells were harvested by centrifugation, and the plasmid was extracted using the Miniprep kit (Qiagen) and submitted for sequencing using a T7 promoter as a primer (Retrogen). For protein expression, the plasmid was transformed into competent BL21(DE3) *E. coli* cells, which contained the ccm (cytochrome c maturation) cassette in chloramphenicol resistant vector. The colonies were grown overnight at 37 ºC on LB/agar plates containing 50 mg/L kanamycin and 30 mg/L chloramphenicol. The starter culture was prepared by inoculating a single colony in LB media containing the same amounts of antibiotics. When the optical density of the culture at 600 nm (OD_{600}) reached ~0.7-0.8 (typically after ~5-8 h), the cell culture was inoculated into 12-16 1L flasks containing LB media supplemented with the same amount of antibiotics. Culture were grown for ~16-20 h; proteins were expressed by auto-induction without the need for chemical inducers (e.g., IPTG). Protein purification was performed as described previously.[58] Briefly, the cells were lysed by repeated cycles of sonication (30 s on and 59 s off cycle for 15 min)

after addition of lysozyme in 10 mM sodium phosphate (NaPi) buffer (pH 8.0) in an ice bath. The pH of the lysate was lowered to 6 to precipitate proteins other than the cb_{562} variants. After centrifugation at 8000 rpm for 20 min, the pH was adjusted to 8, and the supernatant was loaded onto a Q sepharose column at 22 ºC, which was pre-equilibrated with the 10 mM NaPi buffer (pH 8). The column was eluted using a step-gradient in the range of 0-1 M NaCl, and the eluted protein was concentrated to less than 50 mL using Amicon with 10 kDa cutoff membranes. Further purification was carried out on a High Q-cartridge column attached to a Biologic DuoFlow workstation (BioRad), using a linear gradient of 0–1 M NaCl in 10 mM NaPi (pH 8.0). After concentrating the protein-containing fractions using Amicon, excess sodium dithionite was added to reduce the proteins. Subsequently, the protein solutions were further purified by size-exclusion chromatography on a Superdex 75 column, using a running buffer solution of 20 mM Tris/HCl (pH 7.0) and 150 mM NaCl. The purity of the protein was monitored in every step of the purification process by SDS-PAGE. The protein was treated with 10-fold molar excess of EDTA overnight at 4 ºC to prepare a metal-free protein stock solution. Excess EDTA was removed by using a 10DG column (BioRad), using a buffer solution containing 20 mM Tris/HCl (pH 7.0) and 150 mM NaCl. The protein samples were kept frozen at -80 ºC after concentrating to a ~1 mM. Extraction of AB5 protein from the periplasm was carried out using a previously described procedure.[58]

Mutations	Primers
Preparation of AB5 variant from ^{C96}RDC1	
E86D/A89E	5'-GTA AAA GAT GCG CAG GAA GCT GCA GAG CAA C-3' 5'-GTT GCT CTG CAG CTT CCT GCG CAT CTT TTA C-3'
Q93H	5'-CAG GCT GCT GCA GAG CAT CTG AAA TGC-3' 5'-GCA TTT CAG ATG CTC TGC AGC AGC CTG-3'
A100H	5'-CCT GCA ACC ATT GCC ACC AGA AGT ATC G-3' 5'-CGA TAC TTC TGG TGG CAA TGG TTG CAG G-3'
Construction of single mutant library by saturated mutagenesis	
K85X	5'-CAA ATG AAG GTA AAG TAN NKG ATG CGC AGG AAG C-3' 5'-GCT TCC TGC GCA TCM NNT ACT TTA CCT TCA TTT G-3'
E92X	5'-GCG CAG GAA GCT GCA NNK CAT CTG AAA TGC-3' 5'-GCA TTT CAG ATG MNN TGC AGC TTC CTG CGC-3'
Q103X	5'-GCA ACC ATT GCC ACN NKA AGT ATC GTT AAT TC-3'

	5'-GAA TTA ACG ATA CTT MNN GTG GCA ATG GTT GC-3'
K104X	5'-CAA CCA TTG CCA CCA GNN KTA TCG TTA ATT CCT C-3' 5'-GAG GAA TTA ACG ATA MNN CTG GTG GCA ATG GTT G-3'
Removal of disulfide bonds	
C96T	5'-GAG CAT CTG AAA ACC ACC TGC AAC CAT TG-3' 5'-CAA TGG TTG CAG GTG GTT TTC AGA TGC TC-3

Table 2.6. DNA primers used for site-directed mutagenesis.

Preparation of Zn-bound AB5 assemblies and determination of Zn-binding properties of AB5 variants by ICP-OES and competitive titrations. Desired volumes of the AB5 protein stock solution were mixed with appropriate amounts of metal stock solutions (100 mM $ZnCl_2$ in double-distilled H_2O). After incubation for ~2-3 h, the protein samples were applied to 10DG column to remove unbound metal ions. The Zn-to-protein stoichiometry was determined by using a strong chelating indicator, PAR (4-(2-pyridylazo) resorcinol), as described previously.[58] Briefly, a metal-free, PAR stock solution (1.44 mM, 200 μL) was prepared in 20 mM MOPS (pH 7.0), 150 mM NaCl and 5 M guadinine hydrochloride (GdnHCl, for protein denaturation). The solution containing the Zn-complexed protein (0.1–1 mM, 1 μL) was added to the PAR solution and its absorption was measured at 500 nm after incubation for ~3 min. 1 μL of a 100 mM EDTA stock solution was then added to the PAR and protein solution. The difference in the absorption before and after the addition of EDTA was then measured and used to determine the overall Zn concentration, using ε_{500nm} for $(PAR)_2Zn$ of 54,800 cm^{-1} M^{-1}.[55] The protein concentration was independently determined at 415 nm (ε = 148,000 cm^{-1} M^{-1}), using a buffer solution not containing GdnHCl. With the concentrations of the protein and Zn ions thus determined independently, the ratio of Zn to the protein monomer was calculated. Alternatively, the metal (Fe and Zn) content was determined by inductively coupled plasma-optical emission spectroscopy (ICP-OES). Standard calibration curves for ICP-OES were prepared using 0–6 ppm Zn and Fe standard stock solutions (Ricca Chemical Company) prior to the analysis of protein solutions. The absorption was monitored at least three different wavelengths and in triplicate. Iron concentration, originating from the covalently bound c-type heme cofactor, was used as an internal standard to determine protein concentration.

To determine the Zn-binding affinities of AB5, a Zn-responsive fluorophore, Newport Green DCF diacetate, was used as reported

previously.[55, 58] Briefly, AB5 protein solutions were pre-treated with 10-fold molar excess of EDTA for ~1 h at 4 ºC, and the resulting apo-AB5 protein was isolated by 10 DG column chromatography using pre-Chelexed 20 mM MOPS (pH 7.0) buffer in 150 mM NaCl. Apo-AB5 protein (~20 μM as a monomer) was then mixed with ~5 μM Newport Green DCF diacetate in a Chelex-treated, 20 mM MOPS (pH 7.0) buffer solution containing 150 mM NaCl at 25 ºC in a cuvette. Before and after the addition of aliquots of AB5-fluorophore mixture, the absorbance was monitored at 504 and 526 nm until no change was observed. The changes in absorbance in response to changes in Zn concentration were fit to 4+4 site model using DynaFit software, yielding two discrete Zn-binding affinities.[78] A representative plot is shown in Figure 2.3.

Oligomerization states of the variants by analytical ultracentrifugation. Analytical ultracentrifugation experiments were carried using ~5-μM protein solutions prepared in 20 mM Tris/HCl (pH 7.0) and 150 mM NaCl. The measurements were made on a Beckman XL-A instrument using an An-60 Ti rotor. For each sample, 250–400 scans were obtained at 25 ºC at 41,000 rpm. All data were processed using Hetero-analysis (University of Connecticut, Analytical Ultracentrifugation Facility) and SEDFIT software[79] with the following fixed parameters: buffer density (r) = 1.0049 g/mL; buffer viscosity = 0.010214 poise; V_{bar} = 0.73084.

X-ray structure determination. Crystals of AB5 assemblies (Zn_4:$AB5_4$, Zn_8:$AB5_4$, and Zn_8:$^{C96T}AB5_4$) were obtained by sitting-drop vapor diffusion at 298 K, using protein stock solutions (2 mM) pre-incubated with 1 or 2 equiv of $ZnCl_2$. The precipitant solutions are listed in Table 2.1. The sitting drops contained 1 μL each of the protein stock solutions and precipitant solutions, and the well contained 300 μL of the precipitant solution. The crystals were frozen in liq. N_2 and diffraction data were collected at SSRL (see Table 2.1 for data collection and refinement statistics). Diffraction data

were processed using CCP4[80] including iMosflm[81] and Scala.[80] Molecular replacement was carried out using either out using either Molrep[82] or Phaser[83] whereby the structure of the ^{C96}RIDC1 monomer (PDB 3IQ6) was used as a search model. Rigid-body and restrained refinement with REFMAC5[84] or Phenix[85] along with manual model rebuilding and metal/water/ligand placement with COOT[86] produced the final models. For the structures containing more than one monomer in asymmetric unit cell, non-crystallographic symmetry (NCS) restraints were applied throughout the refinement process. All figures with crystal structures were produced with PYMOL (www.pymol.org).[87] The angles between the α-helices were determined using PyMol.

In vitro **activity assays for pNPA and ampicillin hydrolysis.** Esterase activity assays using *p*-nitrophenyl acetate (pNPA) and ampicillin as substrates were carried out as reported previously (Figures 2.5-6, and 2.11).[58] Briefly, time-dependent optical spectra were monitored at 405 nm upon the mixing of Zn-complexed proteins (~7 μM) with various concentrations of pNPA (dissolved in CH_3CN) in 0.1 M borate buffer (pH 9.0) at 22 ^{o}C. The background reaction rates were measured with the apo protein in the absence of $ZnCl_2$, and the net activities originating from the catalytic Zn sites in AB5 protein were estimated by calculating the differences between the Zn and apo samples. To monitor pH-dependent esterase activities, the pH of the borate buffer was adjusted accordingly. For the pH 7-7.5 range, MOPS was used as the buffering agent. To calculate the esterase rates at various pHs, I used the pH-dependent extinction coefficients of p-nitrophenolate reported in literature.[88]

To measure β-lactamase activities, AB5 protein (~7 μM) in 0.1 M borate buffer (pH 9) was mixed with various concentrations of ampicillin; the hydrolysis of ampicillin was monitored at 220 nm on HPLC using a linear gradient of H_2O and CH_3CN (9:1 to 1:9 over 20 min) in 0.1% trifluoroacetic

acid. The reaction was monitored for ~4–5 h at 22 ºC and the sample (100 μL) was injected every ~60 min to the C18 column to quantify the ampicillin concentrations. The basal level (i.e., uncatalyzed) decomposition of ampicillin was determined by monitoring the reaction in the buffer solution (pH 9). The apo AB5 protein did not display ampicillin hydrolysis activity above this basal level.

Autodock simulation of Zn_8:$AB5_4$ with ampicillin. The coordinate file for ampicillin was obtained using the PRODRG server.[89] The ampicillin coordinates were set to have the freedom of rotation at every carbon atom except for those comprising the β-lactam and the arene ring. Autodock Vina[90] was utilized to estimate potential binding sites for ampicillin to the AB5 protein surface. The size of grid box was set to 22×24×28 $Å^3$. Among the multiple possible conformations, the one with the shortest distance between the carbonyl group of the β-lactam and the oxygen atom of the Zn-bound water, is shown as a possible docking conformation (Figure 2.2).

Periplasmic extraction and characterization of AB5 protein. AB5 protein was expressed in the periplasm and extracted by osmotic shock as described previously.[1] Oligomeric state and metal occupancy of the protein were characterized by the gel filtration chromatography and ICP-OES, respectively (Figure 2.16), as described previously.[58]

Construction of saturated mutant libraries of AB5 at 85, 92, 103, and 104 residues. Saturated mutagenesis of the selected residues was carried out as reported previously.[58] Briefly, primers are designed such that each amino acid in residue 85, 92, 103, and 104 can be individually randomized using NNK codon (Table 2.6). After 25–30 cycles of PCR runs, the product (5 μL) was transformed into DH5α *E. coli* competent cells (100–200 μL). After overnight growth at 37 ºC on the LB/agar plate containing 50 μg/mL kanamycin, the number of colonies grown on the plates were counted to estimate the size of the mutant library.

After collecting > 100 colonies, the cells were grown in LB media with 50 μg/mL kanamycin at 37 ºC for ~5–6 h, and the mixtures of the plasmids were extracted using MiniPrep or MidiPrep kit (Qiagen). The randomized codons were confirmed by sequencing in Retrogen.

***In vivo* screening of cells containing artificial β-Lactamase.** The mixtures of plasmids were then transformed to BL21 (DE3) with ccm, and grown in LB/agar plate containing 50 μg/mL kanamycin and 30 μg/mL chloramphenicol at 37 ºC overnight. The colonies were harvested and grown in the LB media and 50 μM $ZnCl_2$ under the identical antibiotic conditions for ~10-14 h until the cell pastes become red upon the protein expression. The *E. coli* cells containing the mutants were then plated on the LB/agar plate containing 50 μg/mL kanamycin, 30 μg/mL chloramphenicol, 50 μM $ZnCl_2$, and ~1 μg/mL ampicillin. The plates were incubated at 37 ºC overnight, and each colony was individually grown for Miniprep extraction and sequencing. To obtain more accurate statistics of the screening results, each codon, rather than amino acid, were counted so that there is no bias for the amino acids that are encoded by more than one codon. The dominant and minor outputs for each amino acid were shown with red and white bars, respectively in Figure 2.16. The original sequence of AB5 protein was shown with an asterisk.

Kinetic analysis of *in vivo* β-lactamase activity. Three BL21 (DE3) *E. coli* strains, expressing either ^{C96}RIDC1, AB5, or ^{C96T}AB5 proteins, were prepared in 10 mL LB media in order to produce Zn_4:^{C96}RIDC1$_4$, Zn_8:AB5$_4$, or Zn_8:^{C96T}AB5$_4$ assemblies in the periplasm. After harvesting the cell pellets by centrifugation (5,000 rpm for 10 min), followed by washes with 10 mM Tris/HCl, 30 mM NaCl buffer (pH 7.3) 3 times, the cell pastes were suspended in M9 minimal media (pH 7.0) to monitor the vivo hydrolytic activities. The assays were initiated by the addition of 0-3 mM ampicillin and carried out under constant shaking (200 rpm) at 37 ºC. Aliquots of the cell cultures (1 mL) were centrifuged every 20-40 min and the supernatant was injected to

HPLC to monitor the consumption of ampicillin, as described above for in vitro activity assays. To test the baseline activity of free Zn^{2+} ions under *in vivo* assay condition, I used 1.5 μM $ZnCl_2$, which is the estimated protein concentration used for *in vivo* activity assays. The effective protein concentrations used for *in vivo* activity assays were determined by the lysis of the cell cultures, followed by the measurement of the protein Soret peak or by calculating the average protein yields from the cell growth. Both calculation methods yield a rough estimation of ~1.5 μM protein per culture media.

2.5. References

(1) Jensen, R. A., *Annu. Rev. Microbiol.* **1976,** *30* (1), 409-425.

(2) Wolfenden, R.; Snider, M. J., *Acc. Chem. Res.* **2001,** *34*, 938-945.

(3) Sterner, R.; Höcker, B., *Chem. Rev.* **2005,** *105*, 4038-4055.

(4) Nanda, V.; Koder, R. L., *Nat. Chem.* **2010,** *2*, 15-24.

(5) Lu, Y.; Yeung, N.; Sieracki, N.; Marshall, N. M., *Nature* **2009,** *460*, 855-862.

(6) Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K., *Nature* **2012,** *485* (7397), 185-194.

(7) Heinisch, T.; Ward, T. R., *Acc. Chem. Res.* **2016,** *49* (9), 1711-1721.

(8) Yu, F.; Cangelosi, V. M.; Zastrow, M. L.; Tegoni, M.; Plegaria, J. S.; Tebo, A. G.; Mocny, C. S.; Ruckthong, L.; Qayyum, H.; Pecoraro, V. L., *Chem. Rev.* **2014,** *114* (7), 3495-3578.

(9) Nastri, F.; Chino, M.; Maglio, O.; Bhagi-Damodaran, A.; Lu, Y.; Lombardi, A., *Chem. Soc. Rev.* **2016,** *45* (18), 5020-5054.

(10) Schwizer, F.; Okamoto, Y.; Heinisch, T.; Gu, Y.; Pellizzoni, M. M.; Lebrun, V.; Reuter, R.; Köhler, V.; Lewis, J. C.; Ward, T. R., *Chem. Rev.* **2018,** 118 (1), 142-231.

(11) Cook, S. A.; Hill, E. A.; Borovik, A. S., *Biochemistry* **2015,** *54*, 4167-4180.

(12) Gibney, B. R.; Mulholland, S. E.; Rabanal, F.; Dutton, P. L., *Proc. Natl. Acad. Sci. U. S. A.* **1996,** *93*, 15041-15046.

(13) Ghirlanda, G.; Osyczka, A.; Liu, W.; Antolovich, M.; Smith, K. M.; Dutton, P. L.; Wand, A. J.; DeGrado, W. F., *J. Am. Chem. Soc.* **2004,** *126*, 8141-8147.

(14) Faiella, M.; Andreozzi, C.; Rosales, R. T. M. d.; Pavone, V.; Maglio, O.; Nastri, F.; DeGrado, W. F.; Lombardi, A., *Nat. Chem. Biol.* **2009,** *5*, 882-884.

(15) Zastrow, M. L.; Peacock, A. F.; Stuckey, J. A.; Pecoraro, V. L., *Nat.*

Chem. **2012,** *4*, 118-123.

(16) Rufo, C. M.; Moroz, Y. S.; Moroz, O. V.; Stöhr, J.; Smith, T. A.; Hu, X.; DeGrado, W. F.; Korendovych, I. V., *Nat. Chem.* **2014,** *6* (4), 303-309.

(17) Farid, T. A.; Kodali, G.; Solomon, L. A.; Lichtenstein, B. R.; Sheehan, M. M.; Fry, B. A.; Bialas, C.; Ennist, N. M.; Siedlecki, J. A.; Zhao, Z.; Stetz, M. A.; Valentine, K. G.; Anderson, J. L.; Wand, A. J.; Discher, B. M.; Moser, C. C.; Dutton, P. L., *Nat. Chem. Biol.* **2013,** *9* (12), 826-33.

(18) Park, H.-S.; Nam, S.-H.; Lee, J. K.; Yoon, C. N.; Mannervik, B.; Stephen, J. B.; Kim, H.-S., *Science* **2006,** *311*, 535-538.

(19) Yeung, N.; Lin, Y.-W.; Gao, Y.-G.; Zhao, X.; Russell, B. S.; Lei, L.; Miner, K. D.; Robinson, H.; Lu, Y., *Nature* **2009,** *462*, 1079-1082.

(20) Khare, S. D.; Kipnis, Y.; Greisen, P. J.; Takeuchi, R.; Ashani, Y.; Goldsmith, M.; Song, Y.; Gallaher, J. L.; Silman, I.; Leader, H.; Sussman, J. L.; Stoddard, B. L.; Tawfik, D. S.; Baker, D., *Nat. Chem. Biol.* **2012,** *8*, 294-300.

(21) Hyster, T. K.; Knörr, L.; Ward, T. R.; Rovis, T., *Science* **2012,** *338*, 500-503.

(22) Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H., *Science* **2013,** *339*, 307-310.

(23) Tian, S.; Liu, J.; Cowley, R. E.; Hosseinzadeh, P.; Marshall, N. M.; Yu, Y.; Robinson, H.; Nilges, M. J.; Blackburn, N. J.; Solomon, E. I.; Lu, Y., *Nat. Chem.* **2016,** *8* (7), 670-677.

(24) Key, H. M.; Dydio, P.; Clark, D. S.; Hartwig, J. F., *Nature* **2016,** *534* (7608), 534-537.

(25) Oohora, K.; Kihira, Y.; Mizohata, E.; Inoue, T.; Hayashi, T., *J. Am. Chem. Soc.* **2013,** *135* (46), 17282-17285.

(26) Ohashi, M.; Koshiyama, T.; Ueno, T.; Yanase, M.; Fujii, H.; Watanabe, Y., *Angew. Chem. Int. Ed.* **2003,** *42* (9), 1005-1008.

(27) Tyagi, V.; Fasan, R., *Angew. Chem. Int. Ed.* **2016,** *55* (7), 2512-2516.

(28) Coelho, P. S.; Wang, Z. J.; Ener, M. E.; Baril, S. A.; Kannan, A.; Arnold, F. H.; Brustad, E. M., *Nat. Chem. Biol.* **2013,** *9*, 485-487.

(29) Prier, C. K.; Zhang, R. K.; Buller, A. R.; Brinkmann-Chen, S.; Arnold, F. H., *Nat. Chem.* **2017,** *9* (7), 629-634.

(30) Dydio, P.; Key, H. M.; Nazarenko, A.; Rha, J. Y.-E.; Seyedkazemi, V.; Clark, D. S.; Hartwig, J. F., *Science* **2016,** *354* (6308), 102-106.

(31) Kan, S. B. J.; Lewis, R. D.; Chen, K.; Arnold, F. H., *Science* **2016,** *354* (6315), 1048-1051.

(32) Wierenga, R. K., *FEBS Lett.* **2001,** *492* (3), 193-198.

(33) Sterner, R.; Höcker, B., *Chem. Rev.* **2005,** *105* (11), 4038-4055.

(34) Nanda, V., *Nat. Chem. Biol.* **2016,** *12* (1), 2-3.

(35) Jeschek, M.; Reuter, R.; Heinisch, T.; Trindler, C.; Klehr, J.; Panke, S.; Ward, T. R., *Nature* **2016,** *537* (7622), 661-665.

(36) Köhler, V.; Wilson, Y. M.; Dürrenberger, M.; Ghislieri, D.; Churakova, E.; Quinto, T.; Knörr, L.; Häussinger, D.; Hollmann, F.; Turner, N. J.; Ward, T. R., *Nat. Chem.* **2013,** *5* (2), 93-99.

(37) Letondor, C.; Humbert, N.; Ward, T. R., *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102* (13), 4683-4687.

(38) Srivastava, P.; Yang, H.; Ellis-Guardiola, K.; Lewis, J. C., *Nat. Commun.* **2015,** *6*, 7789.

(39) Heinisch, T.; Pellizzoni, M.; Dürrenberger, M.; Tinberg, C. E.; Köhler, V.; Klehr, J.; Häussinger, D.; Baker, D.; Ward, T. R., *J. Am. Chem. Soc.* **2015,** *137* (32), 10414-10419.

(40) Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; III, C. F. B.; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D., *Science* **2008,** *319*, 1387-1391.

(41) Bergdoll, M.; Eltis, L. D.; Cameron, A. D.; Dumas, P.; Bolin, J. T., *Protein Sci.* **1998,** *7* (8), 1661-1670.

(42) Armstrong, R. N., *Biochemistry* **2000,** *39*, 13625-13632.

(43) Goodsell, D. S.; Olson, A. J., *Annu. Rev.Biophys. Biomol. Struct* **2000,** *29* (1), 105-153.

(44) Brodin, J. D.; Ambroggio, X. I.; Tang, C.; Parent, K. N.; Baker, T. S.; Tezcan, F. A., *Nat. Chem.* **2012,** *4*, 375-382.

(45) Brodin, J. D.; Carr, J. R.; Sontz, P. A.; Tezcan, F. A., *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111* (8), 2897-2902.

(46) Song, W. J.; Sontz, P. A.; Ambroggio, X. I.; Tezcan, F. A., *Annu. Rev. Biophys* **2014,** *43*, 409-431.

(47) Lai, Y.-T.; Reading, E.; Hura, G. L.; Tsai, K.-L.; Laganowsky, A.; Asturias, F. J.; Tainer, J. A.; Robinson, C. V.; Yeates, T. O., *Nat. Chem.* **2014,** *6* (12), 1065-1071.

(48) Sontz, P. A.; Bailey, J. B.; Ahn, S.; Tezcan, F. A., *J. Am. Chem. Soc.* **2015,** *137* (36), 11598-11601.

(49) Burton, A. J.; Thomson, A. R.; Dawson, W. M.; Brady, R. L.; Woolfson, D. N., *Nat. Chem.* **2016,** *8* (9), 837-844.

(50) Smock, R. G.; Yadid, I.; Dym, O.; Clarke, J.; Tawfik, D. S., *Cell* **2016,** *164* (3), 476-486.

(51) King, N. P.; Bale, J. B.; Sheffler, W.; McNamara, D. E.; Gonen, S.; Gonen, T.; Yeates, T. O.; Baker, D., *Nature* **2014,** *510* (7503), 103-108.

(52) Churchfield, L. A.; George, A.; Tezcan, F. A., *Essays In Biochemistry* **2017,** *61* (2), 245-258.

(53) Salgado, E. N.; Faraone-Mennella, J.; Tezcan, F. A., *J. Am. Chem. Soc.* **2007,** *129*, 13374-13375.

(54) Salgado, E. N.; Lewis, R. A.; Faraone-Mennella, J.; Tezcan, F. A., *J. Am. Chem. Soc.* **2008,** *130*, 6082-6084.

(55) Brodin, J. D.; Medina-Morales, A.; Ni, T.; Salgado, E. N.; Ambroggio, X. I.; Tezcan, F. A., *J. Am. Chem. Soc.* **2010,** *132*, 8610-8617.

(56) Salgado, E. N.; Ambroggio, X. I.; Brodin, J. D.; Lewis, R. A.;

Kuhlman, B.; Tezcan, F. A., *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107*, 1827-1832.

(57) Medina-Morales, A.; Perez, A.; Brodin, J. D.; Tezcan, F. A., *J. Am. Chem. Soc.* **2013,** *135*, 12013-12022.

(58) Song, W. J.; Tezcan, F. A., *Science* **2014,** *346*, 1525-1528.

(59) Churchfield, L. A.; Medina-Morales, A.; Brodin, J. D.; Perez, A.; Tezcan, F. A., *J. Am. Chem. Soc.* **2016,** *138* (40), 13163-13166.

(60) Obexer, R.; Godina, A.; Garrabou, X.; Mittl, P. R. E.; Baker, D.; Griffiths, A. D.; Hilvert, D., *Nat. Chem.* **2017,** *9* (1), 50-56.

(61) Li, C.; Hassler, M.; Bugg, T. D. H., *ChemBioChem* **2008,** *9* (1), 71-76.

(62) Jochens, H.; Hesseler, M.; Stiba, K.; Padhi, S. K.; Kazlauskas, R. J.; Bornscheuer, U. T., *ChemBioChem* **2011,** *12* (10), 1508-1517.

(63) Kannan, K. K.; Ramanadham, M.; Jones, T. A., *Annals of the New York Academy of Sciences* **1984,** *429* (1), 49-60.

(64) Christianson, D. W.; Fierke, C. A., *Accounts of Chemical Research* **1996,** *29* (7), 331-339.

(65) Rees, D. C.; Lewis, M.; Lipscomb, W. N., *Journal of Molecular Biology* **1983,** *168* (2), 367-387.

(66) Koike, T.; Takamura, M.; Kimura, E., *J. Am. Chem. Soc.* **1994,** *116*, 8443-8449.

(67) Der, B. S.; Edwards, D. R.; Kuhlman, B., *Biochemistry* **2012,** *51*, 3933-3940.

(68) Zastrow, M. L.; Pecoraro, V. L., *J. Am. Chem. Soc.* **2013,** *135*, 5895-5903.

(69) Reetz, M. T.; Carballeira, J. D., *Nat. Protocols* **2007,** *2* (4), 891-903.

(70) Reetz, M. T., *Angew. Chem. Int. Ed.* **2013,** *52* (10), 2658-2666.

(71) Tomatis, P. E.; Fabiane, S. M.; Simona, F.; Carloni, P.; Sutton, B. J.; Vila, A. J., *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105*, 20605-20610.

(72) Tokuriki, N.; Tawfik, D. S., *Science* **2009,** *324*, 203-207.

(73) Marsh, J. A.; Teichmann, S. A., *PLoS Biol.* **2014,** *12*, e1001870.

(74) Preiswerk, N.; Beck, T.; Schulz, J. D.; Milovník, P.; Mayer, C.; Siegel, J. B.; Baker, D.; Hilvert, D., *Proceedings of the National Academy of Sciences* **2014,** *111* (22), 8013-8018.

(75) Bloom, J. D.; Labthavikul, S. T.; Otey, C. R.; Arnold, F. H., *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103*, 5869-5874.

(76) Khersonsky, O.; Kiss, G.; Röthlisberger, D.; Dym, O.; Albeck, S.; Houk, K. N.; Baker, D.; Tawfik, D. S., *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109*, 10358-10363.

(77) Wang, X.; Minasov, G.; Shoichet, B. K., *J. Mol. Biol.* **2002,** *320*, 85-95.

(78) Kuzmič, P., *Analytical Biochemistry* **1996,** *237* (2), 260-273.

(79) Schuck, P., *Biophysical Journal* **2000,** *78*, 1606-1619.

(80) Collaborative Computational Project, N., *Acta Crystallogr.* **1994,** *D50*, 760-763.

(81) Battye, T. G. G.; Kontogiannis, L.; Johnson, O.; Powell, H. R.; Leslie, A. G. W., *Acta. Cryst.* **2011,** *D67*, 271-281.

(82) A.Vagin; A.Teplyakov, *J. Appl. Cryst.* **1997,** *30*, 1022-1025.

(83) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., *J. Appl. Cryst.* **2007,** *40*, 658-674.

(84) Murshudov, G. N.; A.A.Vagin; E.J.Dodson, *Acta Cryst.* **1997,** *D53*, 240-255.

(85) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., *Acta Cryst.* **2010,** *D66*, 213-221.

(86) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., *Acta Cryst.* **2010,** *D66*, 486-501.

(87) Schrodinger, LLC, 2015.

(88) Nomura, A.; Sugiura, Y., *Inorg. Chem.* **2004,** *43*, 1708-1713.

(89) Schüttelkopf, A. W.; Aalten, D. M. F. v., *Acta. Crystallogr.* **2004,** *D60*, 1355-1363.

(90) Trott, O.; Olson, A. J., *J. Comput. Chem.* **2009,** *31*, 455-461.

Chapter 3.
Symmetry-related Residues as Promising Hotspots for the Evolution of *De Novo* Oligomeric Enzymes

*The part of this chapter was reprinted with permission from **J. Yu**; J. Yang; C. Seok; W. J. Song, *Chem. Sci.* 2021, 12, 5091, copyright 2021, Royal Society of Chemistry.

3.1. Introduction

Natural enzymes have evolved throughout numerous rounds of selection.[1, 2] Accumulating evidence indicates that proteins have adapted to altered chemical environments by sequence modification. For example, it is proposed that the Great Oxygenation Event (GOE) is correlated with the elevated frequency of metalloenzymes with air-stable metal ions.[3, 4] The introduction of natural and synthetic herbicides and antibiotics also led to the appearance of enzymes that react with these substances, resulting in resistance to these substances.[5, 6] Analogously, specific chemical pressures can be applied to randomized mutant libraries in a laboratory. Numerous novel proteins have been created, yielding novel catalytic activity,[7-10] altered substrate selectivity,[11-13] and/or elevated thermal stability.[14, 15]

Despite significant progress in protein design and evolution,[16-18] there is no standard rule prioritizing the residues for sequence optimization. The amino acid residues in the vicinity of active sites are often the first and primary candidates. B-Factor and sequence conservation analyses have also been carried out to generate focused mutant libraries.[19, 20] However, our current state-of-the-art understanding of the protein sequence–structure–function relationship is incomplete in that seemingly trivial or distant mutations are essential to induce the desired properties,[21, 22] while even proximal residues to the active site may cause no detectable changes in structure and function.[23] In addition, multiple mutations often exhibit non-additive effects so that iterative mutations and screenings are necessary,[24] requiring at least mid to high throughput screening.[25-30] Therefore, an alternative guideline to pinpoint key residues for sequence optimization would be advantageous to facilitate enzyme design and to elucidate intertwined protein sequence networks.

In the exploration of this complicated yet, essential question, *de novo* enzymes could be a versatile model system. Their sequence and structure are

not strongly correlated with the nascent function, and thus, enzyme evolution may occur with little underlying interrelation to the intrinsic nature of proteins. Previously, numerous protein scaffolds have been transformed into artificial enzymes, particularly metalloenzymes.[31] A large portion of these examples is found as homomeric proteins, having catalytic metal-sites on the protein–protein interfaces.[32-34] One of these examples is an artificial metallo-β-lactamase, AB5, where a catalytic center, Zn–OH_2/OH unit, was created on the protein–protein interfaces of an α-helical homo-tetramer (Figure 3.1a).[35] Intriguingly, I found that a seemingly trivial mutation (C96T) yielded substantial structural alterations (Figure 3.1b), resulting in the conversion of a tetramer with a large void space into a closely packed one. Besides, hydrolytic activities with ampicillin were enhanced. In contrast, several attempts to optimize K85, E92, Q103, and K104 residues, close to the Zn-active sites (Figure 3.1a), exerted no significant enhancement in the hydrolytic activities. Based on these observations, I surmised that C96 residue is a more effective hotspot than four proximal residues (K85, E92, Q103, and K104) because of its location on the C_2-rotation axis. If the mutations of residues related to symmetry operations such as rotation may produce significant impacts on overall protein structure and function, possibly using the fluxional protein–protein interface, I further inquired whether other residues on rotational axes can be efficiently targeted for directed evolution of *de novo* designed oligomeric enzymes, despite being distantly located from active sites and seemingly unrelated to enzyme catalysis.

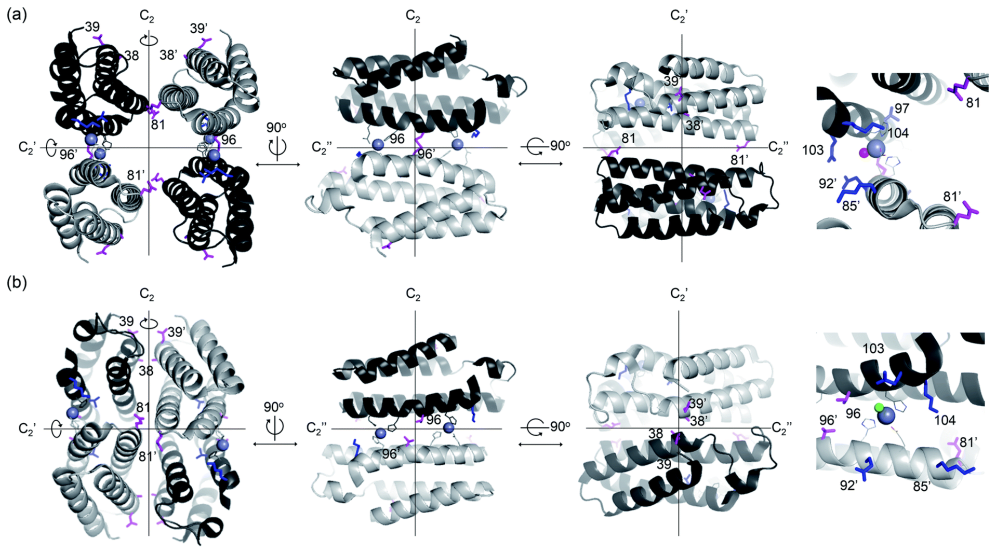


Figure 3.1. X-ray crystal structures of *de novo* Zn-dependent β-lactamases related to this study. (a) AB5 protein with C96 residue (PDB 5XZI). (b) C96T variant (PDB5XZJ). Residues located at the C2 rotational axes, such as 96, 38, 39, and 81, are shown with magenta sticks. Residues located at proximal to the active sites, such as 85, 92, 97, 103, and 104, are shown with blue sticks. Zn ions located at the catalytic sites and metal-ligating residues are shown with grey spheres and lines, respectively. The metal-bound water or chloride ions in catalytic sites are depicted with red and green spheres, respectively. The Zn ions at the structural sites and crystal packing interactions are omitted for clarity.

3.2. Results and discussion

3.2.1. Construction and screening of symmetry versus proximity-related libraries.

To demonstrate whether symmetry-guided residues are potential hotspots for the construction of mutant libraries and screening, I selected three residues located on the C_2-rotation axes, A38, D39, and E81 (Figure 3.1a). They are located 17.2–24.1 Å from the active site (Tables 3.1), and even the closest symmetry-related residue to the active site, C96, is not in direct contact with the catalytic unit, a Zn–OH_2/OH species. As a parallel study, I also selected two non-symmetry-related but proximal residues to the active site, T97 and K104, which lie at 9.6 and 6.4 Å, respectively. I then individually randomized the symmetry-related and proximal residues by saturated mutagenesis using the NNK codon (Table 3.1 and Figure. 3.2).

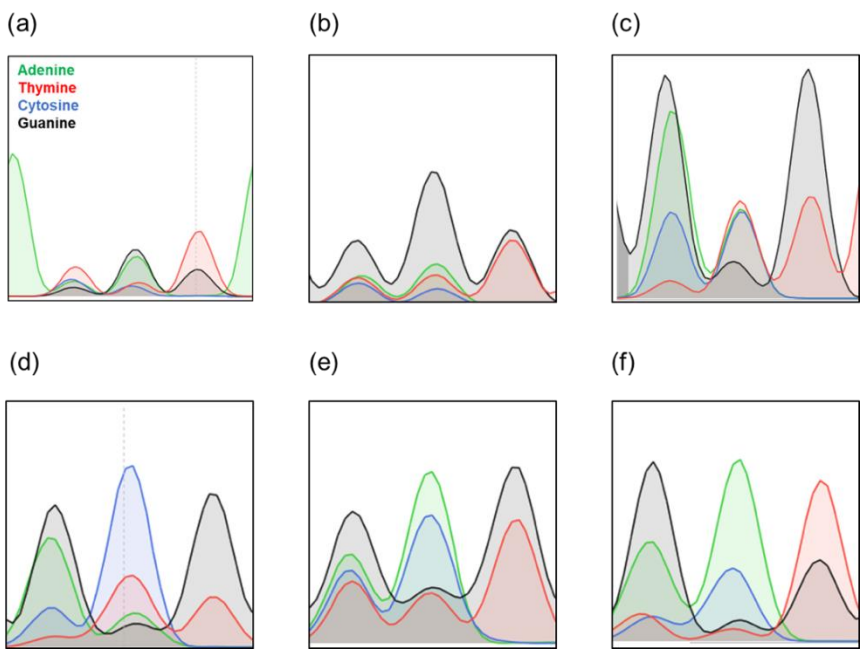


Figure 3.2. Representative sequencing chromatograms of AB5 mutant libraries randomized at (a) C96, (b) T97, (c) K104, (d) A38, (e) D39, and (f) E81 positions.

Protein	Library type	Selected residue	Distance to side chain (Å)
AB5	Proximity-guided	K85	14.0
		E92	5.3
		T97	9.6
		Q103	4.5
		K104	6.4
	Symmetry-guided	A38	22.3
		D39	23.9–24.1
		E81	17.2–18.1
		C96	7.9–9.4
^{C96I}AB5	Proximity-guided	T97	6.8–7.4
		K104	5.2–8.0
	Symmetry-guided	A38	19.0–19.1
		D39	22.8–23.1
		E81	10.2–10.4

Table 3.1. Distances between active sites and selected residues for evolution of AB5 and ^{C96I}AB5. A distance was determined from oxygen atom of zinc coordinated water to the terminal atom, except for hydrogen, of side chain of selected residues.

Both symmetry and proximity-related mutant libraries were screened for the presence of whole-cell hydrolytic activities with an antibiotic β-lactam substrate, ampicillin. For full coverage with a 95% confidence level,[36] greater than 94 colonies for each single-site saturated mutant library and 1,953 colonies in total were screened (Figure 3.3a). For more efficient and quantitative cell-based screening, I slightly modified the method from previous studies[35] so that I measured the cell-growth rates upon the addition of a high concentration (10–15 mg L^{-1}) of ampicillin instead of using LB/agar

plates containing a relatively low concentration of ampicillin. Then, mutations detrimental to β-lactamase activity would lead to zero or negative cell-growth rates upon the addition of antibiotic substrates, if the cells no longer grow or get ruptured by the cell wall-targeting antibiotics, respectively. In contrast, beneficial mutations would lead to the rates faster than that of the parent protein. When the individual cells of each mutant library were sorted in increasing order of cell-growth rate relative to that of the parent protein, the fitness effect, which can be estimated by the magnitude of alterations upon mutations, irrespective of the occurrence of beneficial or detrimental effects upon mutation, followed the order T97X $\approx$ K104X $\ll$ E81X $\leq$ A38X $\approx$ C96X < D39X (Figure 3.3b, c), where X indicates the 20 randomized amino acid. Notably, the four libraries comprising mutants randomized at the symmetry-related residues, 96, 38, 39, and 81 (Figure 3.3b), exhibited considerably greater magnitudes of alterations in cell-growth rate than those of the two libraries saturated at residues 97 and 104 (Figure 3.3c), indicating that symmetry-related residues exert greater degrees of perturbation to the protein than proximal residues. The distribution of fitness effects[37-39] was also depicted by box charts (Figure 3.3c inset), where the mutations at symmetry-related residues exhibit much more outspread cell-growth rates than those of proximal residues. The fastest-growing cells from each library were sequenced, resulting in T97T (parent), K104S, E81G, A38D, C96I, C96K, and D38E mutants.

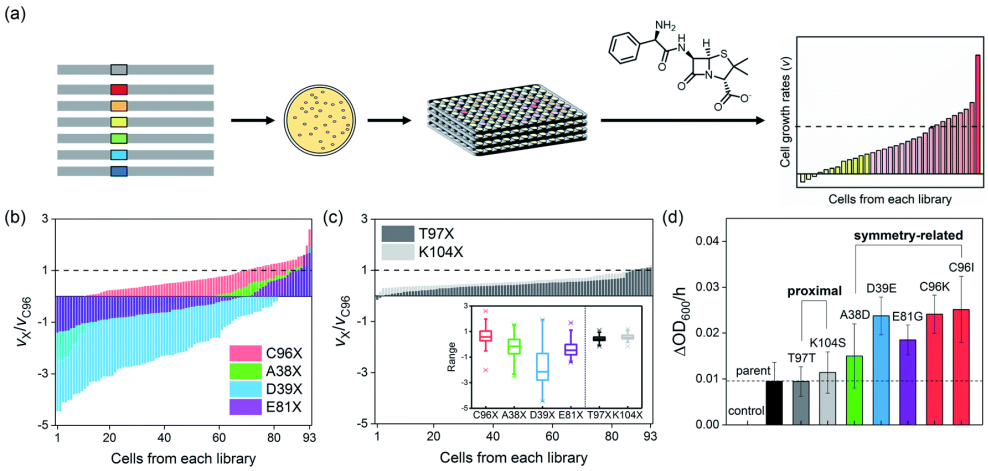


Figure 3.3. The screening of the first-generation libraries. (a) A scheme for β-lactamase activity-based screening of single-site randomized mutant libraries with ampicillin. Representative screening results of (b) symmetry- and (c) proximity-related mutants. The relative cell-growth rates of 93 colonies to the AB5 protein (C96) are plotted in the increasing order. In the (c) inset, a box chart per each mutant library is included, where the boundaries indicate the 25 and the 75 percentiles, and the lines in the box represent the median values. The caps represent the 1 and 99 percentiles, and whiskers indicate the minimum (below) and maximum (above) cell-growth rates. (d) The cell-growth rates of the best one or two hits from each mutant library. Control and parent indicate ^{C96}RIDC1 and AB5 proteins, respectively, where the former exhibits no catalytic Zn-site. The error bars in (d) indicate the standard deviations of three runs of the experiments.

3.2.2. Characterization of the screened single variants

To validate whether the accelerated cell-growth rates were indeed derived from enhanced β-lactamase activities, I measured the cell-growth rates of the sequenced ones, resulting in the order C96 ≈ T97 ≈ K104S ≤ A38D ≤ E81G ≤ D39E ≈ C96K ≈ C96I (Figure 3.3c). The order was similar to that of the degrees of perturbations described above, implying that a mutant library with a higher fitness effect is likely to possess more beneficial mutations. Therefore, these results indicate that residues located at symmetry-related positions are promising hotspots for enzyme evolution, potentially exhibiting higher evolvability. These data are consistent with the recent studies of insertion and deletion mutagenesis, where both desirable and strongly deleterious mutations were co-isolated,[38] suggesting that a well-focused library might exhibit drastic fitness effects in either directions. In contrast, no hit was observed from the proximity-related library with substantially exceeded cell-growth rates than that of the parent protein; one of the best mutants was indeed the parent (T97), and the other was K104S, and the latter was previously characterized to be only marginally better than the parent protein in terms of catalytic activities.[35]

For more accurate kinetic measurements of the single-site mutants, I carried out in vitro steady-state kinetic analysis of the purified Zn-complexed β-lactamase AB5 variants. By applying various substrate concentrations, the Michaelis– Menten kinetic parameters of the C96K and C96I variants, the best hits from the first-round screening, were obtained (Figure 3.4a, b). These variants indeed exhibited up to 2.9- and 2.7-fold increased k_{cat} and k_{cat}/K_M values, respectively relative to those of the parent AB5 protein (Tables 3.2). The results contrasted with the previous attempts to evolve the protein by randomizing the residues at the active sites, 85, 92, 103, and 104.[35]

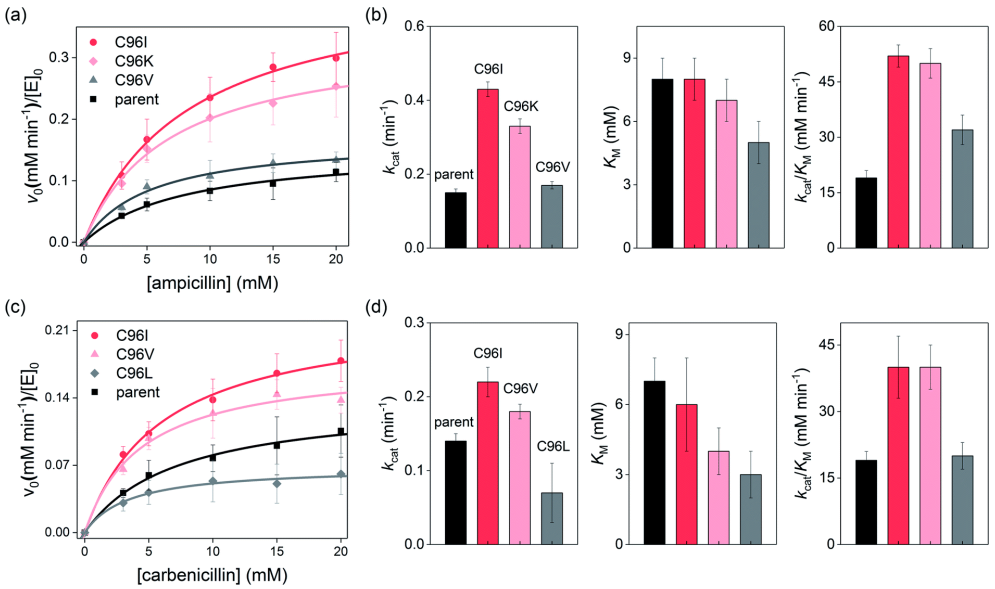


Figure 3.4. Characterization of the positive hits from the first-round screening. (a) Michaelis–Menten plots and (b) kinetic parameters with ampicillin as a substrate. (c) Michaelis–Menten plots and (d) kinetic parameters with carbenicillin as an alternative β-lactam substrate. The error bars indicate the standard deviations of three runs of the experiments.

Substrate	**Protein**	**k_{cat} (min^{-1})**	**K_M (mM)**	**k_{cat}/K_M or k_2 (min^{-1} M^{-1})***
Ampicillin	C96	0.15(1)	8(1)	19(2)
	C96V	0.17(1)	5(1)	32(4)
	C96K	0.33(2)	7(1)	50(4)
	C96I	0.43(2)	8(1)	52(3)
	C96L	nd*	nd*	13(4)*
Carbenicillin	C96	0.14(1)	7(1)	19(2)
	C96V	0.18(1)	4(1)	40(5)
	C96K	0.33(1)	8(1)	42(2)
	C96I	0.22(2)	6(2)	40(7)
	C96L	0.07(4)	3(1)	20(3)

Table 3.2. Kinetic parameters of C96X variants (X= C, V, K, I, and L) with ampicillin and carbenicillin. Hydrolysis activities were measured under pH 7 condition. *The second order rate constant (k_2) instead of k_{cat}/K_M was determined due to the absence of saturation behavior. nd: not detected.

Intriguingly, I noted that the enhanced β-lactamase activities were primarily derived from the elevated k_{cat} rather than lowered K_M, where the latter value is related to the substrate- binding affinity or dissociation constant (K_d). These results implicate that the single mutations selectively accelerated turnover rates rather than increasing substrate affinity. The molecular origin of acceleration contrasted previous work with an analogous protein, AB3, where enhanced catalytic activities were derived from elevated substrate-binding affinity.[40] To identify whether the discrete enhancement mechanism was related to the modified screening conditions using higher concentrations of ampicillin, I reduced the concentration of ampicillin (1.8 mg L^{-1}) for the screening of the C96X mutant library (Figure 3.5b). While C96I is still a positive readout, the parent (C96) and C96V became the dominant hits. When I further modified the screening conditions by replacing ampicillin with carbenicillin, another β-lactam antibiotic, at low concentration (Fig. 3.5d), cells containing C96V and C96L grew relatively fast, implying that screening conditions, such as substrate concentrations and structures, alter the selected readouts.

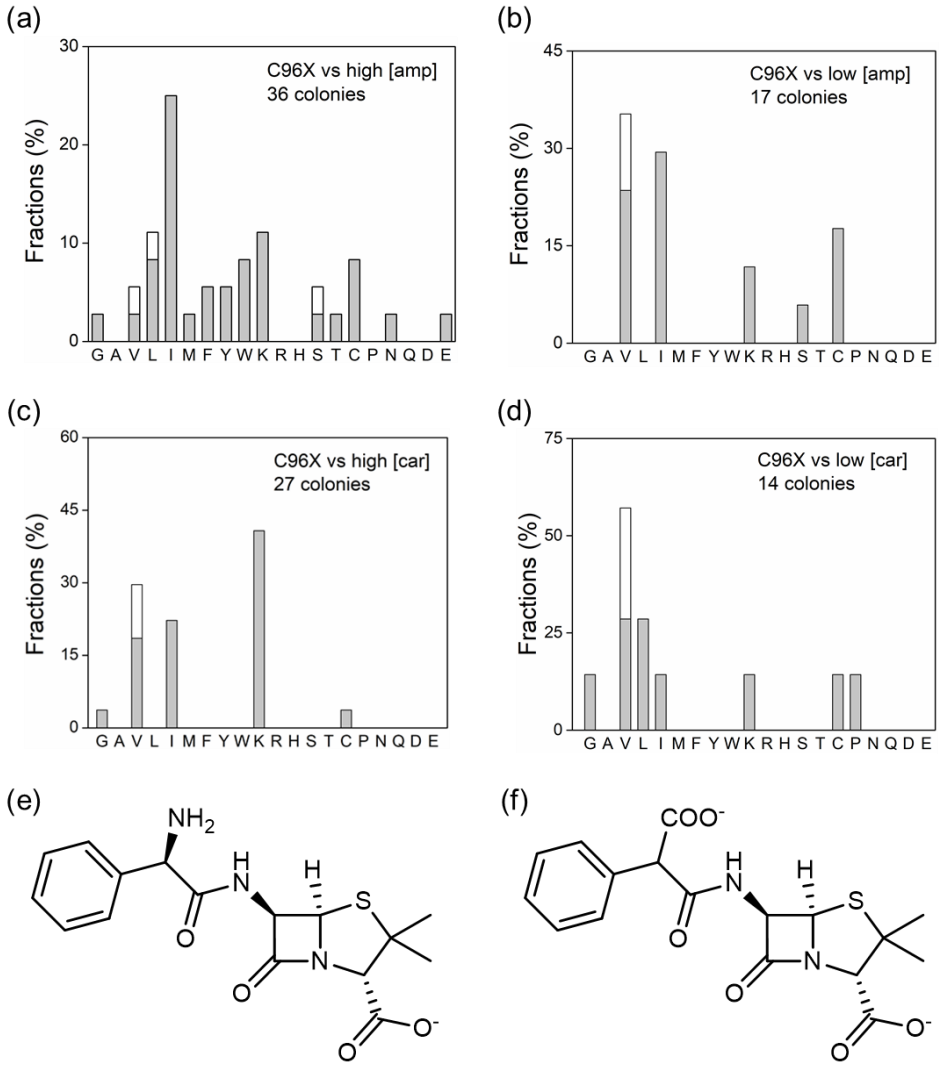
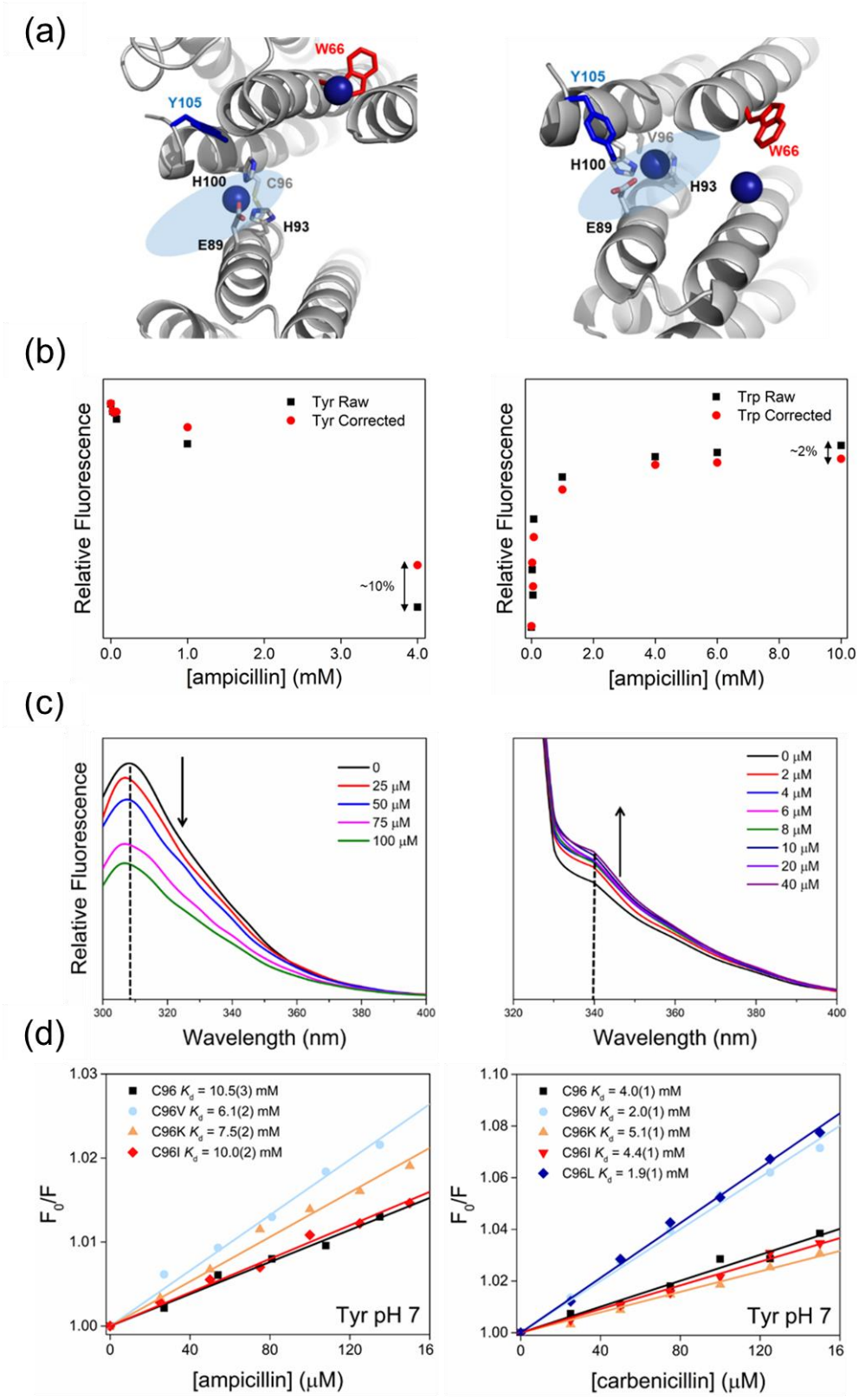


Figure 3.5. The screening readouts of the C96X library with (a) high (10–15 mg/L) and (b) low concentration of ampicillin (1.8 mg/L) and (c) high (20–30 mg/L) and (d) low (1.9 mg/L) concentration of carbenicillin. Two discrete codons that encode the same amino acid in (a)–(b) are shown with different shades. Notably, the fractions of isoleucine and lysine were drastically increased upon the addition of higher concentrations of ampicillin and carbenicillin, respectively. (e) A molecular structure of (left) ampicillin and (right) carbenicillin.

More importantly, the kinetic parameters of the newly detected variants, C96V and C96L, exhibited substantially lower K_M values for ampicillin and/or carbenicillin, respectively, than did the parent protein or previously detected active variants, C96I and C96K (Figure 3.4b, d). These data indicated that the tighter substrate-binding affinity obtained by C96V and C96L mutations is likely to be beneficial when low concentrations of β-lactams are applied for activity-based screening. In contrast, C96I and C96K variants with higher turnover rates and catalytic efficiency are likely to be more advantageous when the substrate concentration and binding affinity are no longer limited.

The correlation between the kinetic parameters and the screening conditions was further supported by measuring the substrate-binding affinity or dissociation constant (K_d) directly from intrinsic fluorescence assays (Figure 3.6and Tables 3.3–4). The Stern–Volmer plots[41, 42] of the C96 variant demonstrate that the thermodynamic parameters were consistent with the K_M values from kinetic analysis, suggesting that the screening conditions and the chemical properties of the selected products might have promoted specific evolutionary trajectories beneficial for the applied chemical pressures.


(a)
W66
Y105
H100
C96
H93
E89
Y105
V96
W66
H100
H93
E89
(b)
Relative Fluorescence
Tyr Raw
Tyr Corrected
~10%
[ampicillin] (mM)
Trp Raw
Trp Corrected
~2%
[ampicillin] (mM)
(c)
Relative Fluorescence
0
25 μM
50 μM
75 μM
100 μM
Wavelength (nm)
0 μM
2 μM
4 μM
6 μM
8 μM
10 μM
20 μM
40 μM
Wavelength (nm)
(d)
F0/F
C96 Kd = 10.5(3) mM
C96V Kd = 6.1(2) mM
C96K Kd = 7.5(2) mM
C96I Kd = 10.0(2) mM
Tyr pH 7
[ampicillin] (μM)
C96 Kd = 4.0(1) mM
C96V Kd = 2.0(1) mM
C96K Kd = 5.1(1) mM
C96I Kd = 4.4(1) mM
C96L Kd = 1.9(1) mM
Tyr pH 7
[carbenicillin] (μM)

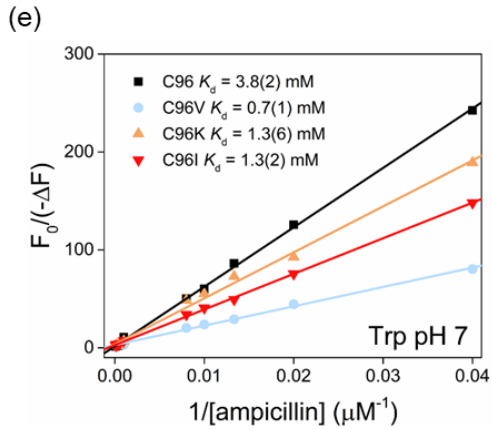


Figure 3.6. Determination of substrate binding affinity by intrinsic fluorescence. (a) Tyrosine and tryptophan residues at the vicinity of the active sites in (left) C96 and (right) C96V protein. (b) Inner filter effect with the intrinsic fluorescence with (left) free tyrosine and (right) tryptophan by addition of ampicillin. (c) The emission profiles of (left) tyrosine and (right) tryptophan upon the addition of ampicillin. (d) The Stern–Volmer linear-fit analysis of tyrosine fluorescence changes with (left) ampicillin and (right) carbenicillin. (e) The modified Stern-Volmer fit analysis of tryptophan fluorescence changes with ampicillin.

Protein	K_M (mM)	$K_{d,\,Tyr}$ (mM)	$K_{d,\,Trp}$ (mM)	$f_{a,\,Trp}$
C96	8(1)	10.5(3)	3.8(2)	6.8–7.4
C96V	5(1)	6.1(2)	0.7(1)	5.2–8.0
C96K	7(1)	7.5(2)	1.3(6)	19.0–19.1
C96I	8(1)	10.0(2)	1.3(2)	22.8–23.1

Table 3.3. Determination of substrate binding affinities of C96X variants for ampicillin (X= C, V, K, and I).

Protein	K_M (mM)	$K_{d,\,Tyr}$ (mM)
C96	8(1)	10.5(3)
C96V	5(1)	6.1(2)
C96K	7(1)	7.5(2)
C96I	8(1)	10.0(2)
C96L	3(1)	1.9(1)

Table 3.4. Determination of substrate binding affinities of C96X variants for carbenicillin (X= C, V, K, I, and L).

To further monitor the effects of single mutations in symmetry-related residues, X-ray crystal structures of three variants from the first-round screening, C96I, C96K, and C96V, were determined (Figure 3.7and Table 3.5). The C96I, C96K, and C96V proteins were isolated as tetramers, consistent with the oligomeric states determined in solution (Figure. 3.8a). They were also similar in that all tetramers possess two sets of Zn-binding sites, structural and catalytic ones (Figures 3.9 and 3.10). Notably, N-terminal residues, surface-exposed acidic residues, and/or H59 and E81 also formed Zn-binding sites, although they are likely to be catalytically irrelevant and generated due to the crystal packing interactions in the presence of excess Zn ions for crystallization (Figure. 3.11).

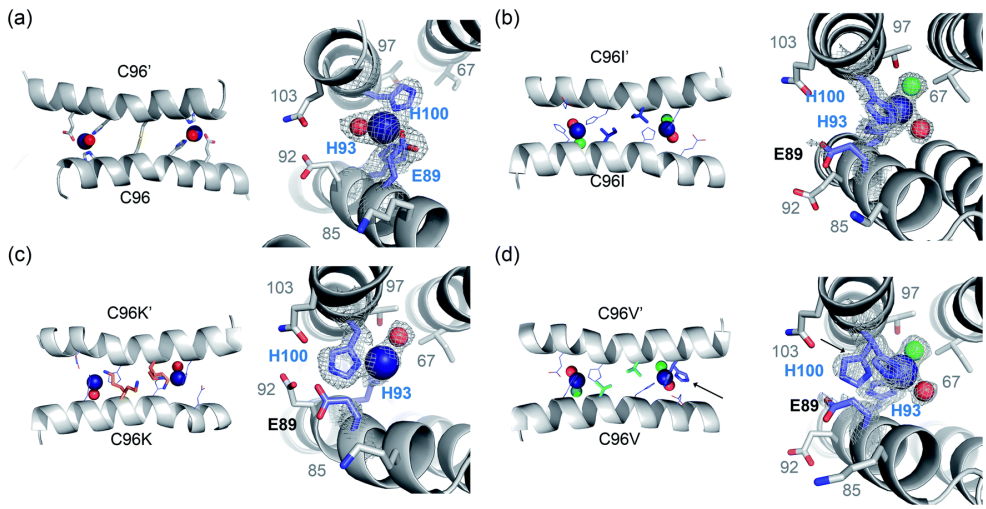


Figure 3.7. X-ray crystal structures of C96 variants from the first-round screening. The catalytic Zn sites of (a) the parent protein, AB5 (PDB 5XZI) C96I (c) C96K and (d) C96V variants. Zn ion and metal-binding residues are shown with navy spheres and navy sticks, respectively. Metal-bound water molecules and chloride ions are shown with red and green spheres, respectively. Two rotamers of H100 residue in C96V protein are highlighted with black arrows. The first metal-coordination sites are shown with $2F_o$–F_c electron density contoured at 1.0 σ overlaid.

Protein	C96I	C96K	C96V	C96I/A38S
crystallization condition	25% Polyethylene glycol 3350 (PEG 3350) in 100 mM HEPES pH 7.5, 200 mM NaCl	45% (+/-)-2-methyl-2,4-pentanediol (MPD) in 100 mM bis-tris pH 6.5, 200 mM NaCl	36% Polypropylene glycol (P400) in 100 mM bis-tris pH 6.5, 200 mM $MgCl_2$	21% Polyethylene glycol 400 (PEG 400) in 100 mM bis-tris pH 6.5, 200 mM ammonium acetate
data collection location	PAL/PLS SB I 7A	PAL/PLS micro-MX 11C	PAL/PLS micro-MX 11C	PAL/PLS micro-MX 11C
unit cell dimensions	53.7×53.7×250.8 $\alpha = \beta = 90^o, \gamma = 120^o$	53.8×53.8×253.8 $\alpha = \beta = 90^o, \gamma = 120^o$	53.8×53.8×250.7 $\alpha = \beta = 90^o, \gamma = 120^o$	53.4×53.4×251.6 $\alpha = \beta = 90^o, \gamma = 120^o$
symmetry group	$P6_122$	$P6_122$	$P6_122$	$P6_122$
resolution (Å)*	28.4–1.98 (2.01–1.98)*	28.62–2.353 (2.39–2.35)*	46.55–2.00 (2.04–2.00)*	46.55–2.45 (2.49–2.45)*
X-ray wavelength (Å)	0.979	0.979	0.979	0.979
number of unique	15931	9893	15537	14826

reflections				
redundancy*	40.0 (40.2)*	39.4 (41.6)*	30.2 (32.5)*	34.2 (37.3)*
Completeness (%)*	99.8 (100)*	99.7 (100)*	100 (100)*	99.9 (100)*
< I /σ>*	105.8 (45.75)*	95.02 (20.88)*	24.22 (15.38)*	88.55 (54.71)*
R_{symm}* (%)	12.9 (32.3)*	10.8 (43.3)*	57.6 (118.6)*	11.8 (17.7)*
R_{work}/R_{free} (%)	18.1/22.9	17.3/24.5	19.3/24.3	20.1/25.8
Number of atoms	1949	1859	2007	1868
Protein Ligands/ions Water	1658 97 194	1672 98 89	1669 94 244	1666 95 107
B-factors (Å^2)	24.0	52.0	21.0	27.7
R.m.s deviations Bond lengths (Å) Bond angles (°)	 0.0146 1.7092	 0.0107 1.6093	 0.0134 1.5992	 0.0114 1.8137
PDB accession code	6LDG	6LDF	6LDE	7DCL

Table 3.5. Crystallographic data collection and refinement statistics of the variants from the 1st and 2nd rounds of screening. *Values for the highest resolution shell are described in parenthesis.

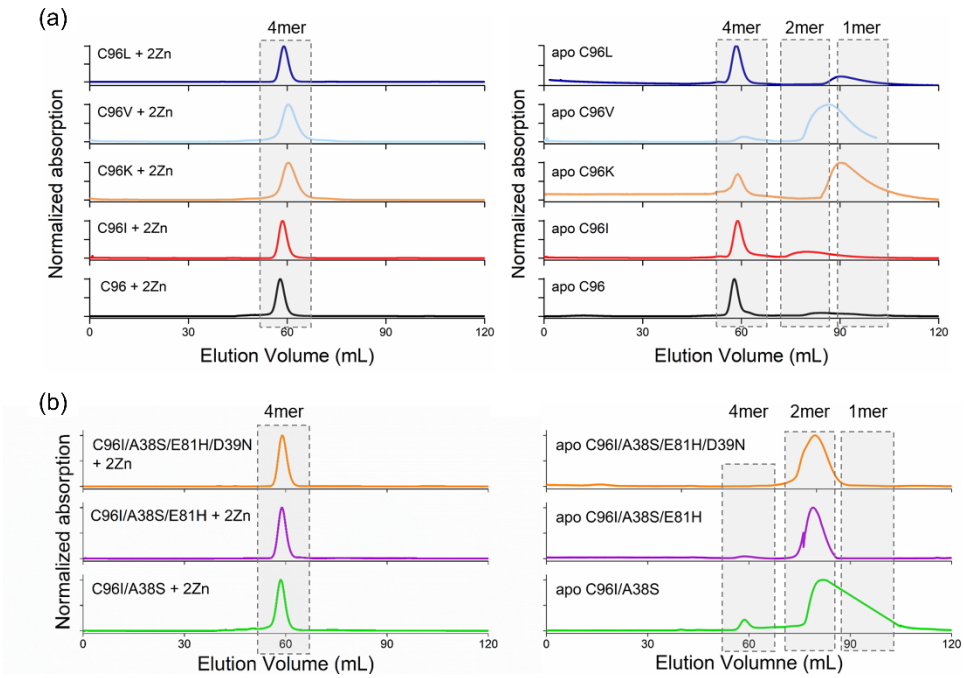


Figure 3.8. Determination of oligomerization states of the C96X variants by size exclusion chromatography. (a) Chromatograms of (left) Zn-complexed and (right) apo C96, C96I, C96K, C96V, and C96L proteins and (b) of (left) Zn-complexed and (right) apo C96I/A38S, C96I/A38S/E81H and C96I/A38S/E81H/D39N.

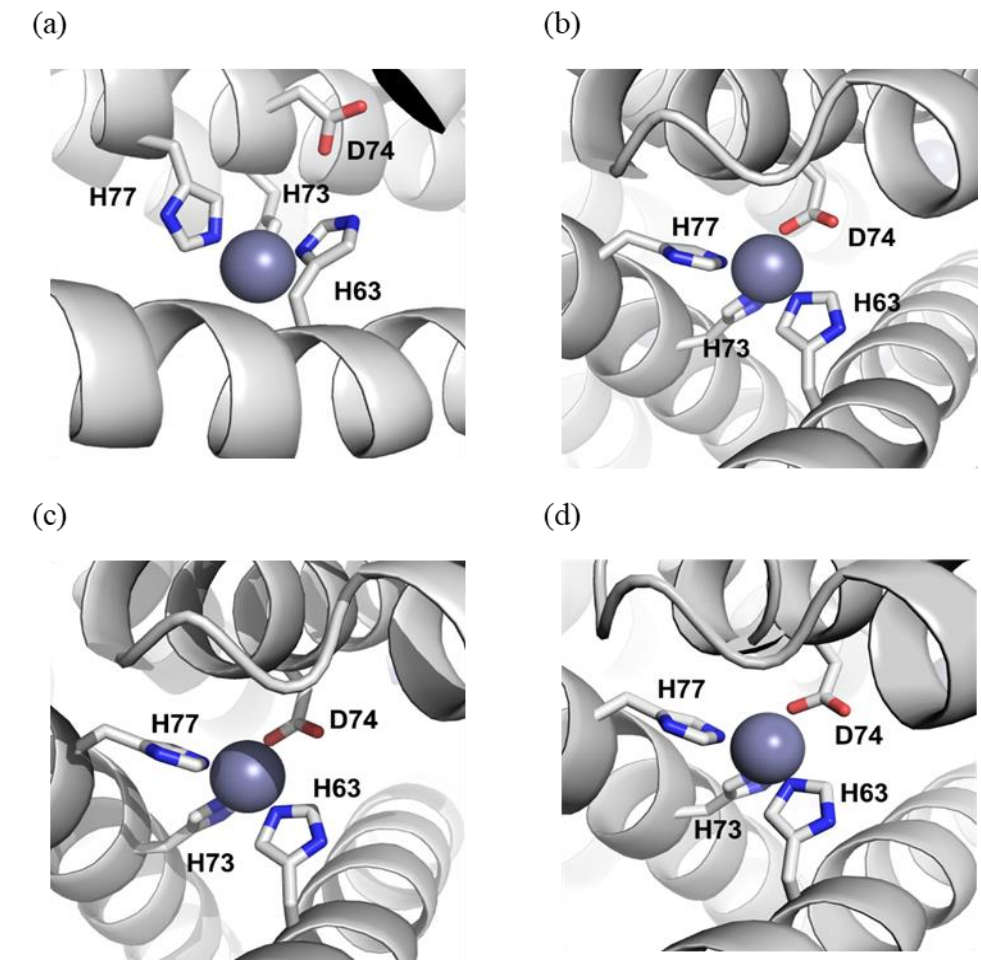


Figure 3.9. Structural Zn-binding sites of C96X variants (X= C, I, K, and V). (a) C96, (b) C96I, (c) C96K, and (d) C96V.

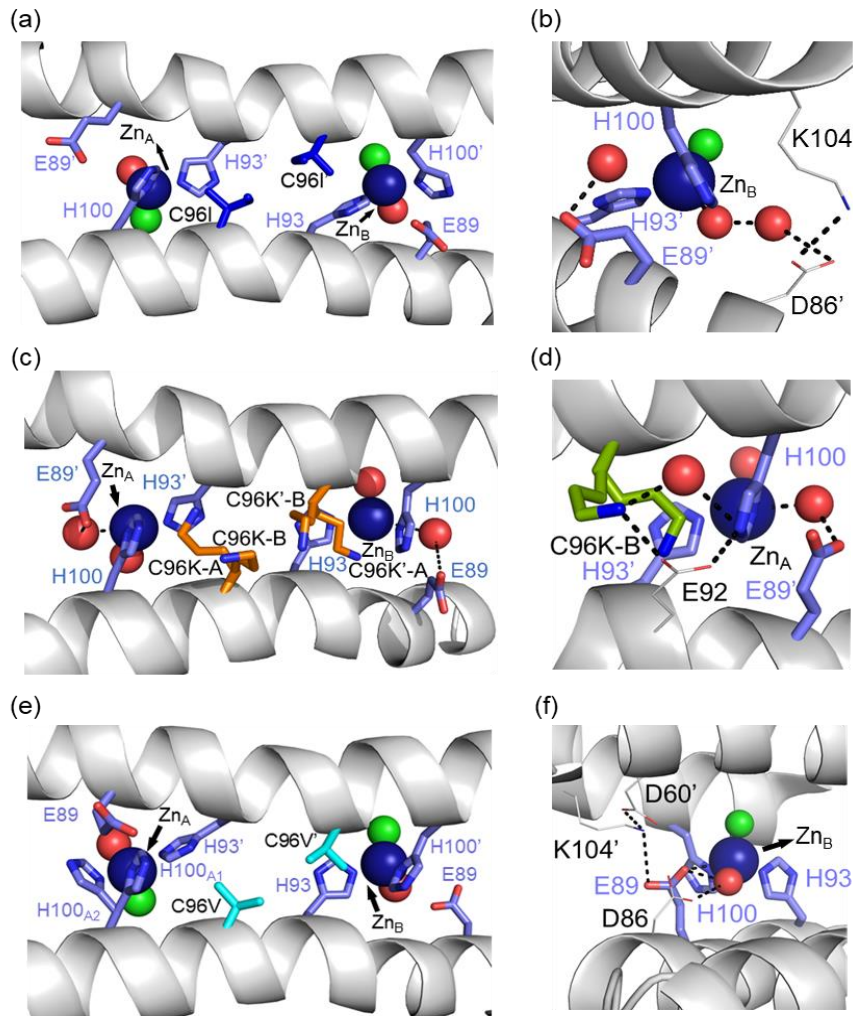


Figure 3.10. Catalytic Zn-binding sites of C96X variants (X= I, K, and V). Depending on the angle of N (H93)-Zn-N (H100), the discrete metal-binding sites are labeled as Zn_A and Zn_B, respectively. (a) The side-view and (b) Hydrogen bonding interaction around the Zn_B-active site in C96I variant. (c) The side-view and (d) Hydrogen bonding interactions of E89 and C96K residues with water molecules and/or interacting adjacent residues in the Zn_A-active site of C96K variant. (e) The side-views and (f) E89 and D86 residues interacting with D60, K104 and D86 residues at the Zn_B-active site in C96V variant. Two discrete conformations of E89 residue were observed in C96V and C96K structures. One of two H100 residues in C96V structure exhibited two rotameric states (H100-A and H100-B). Both C96K residues in C96K protein were also in two alternative conformations.

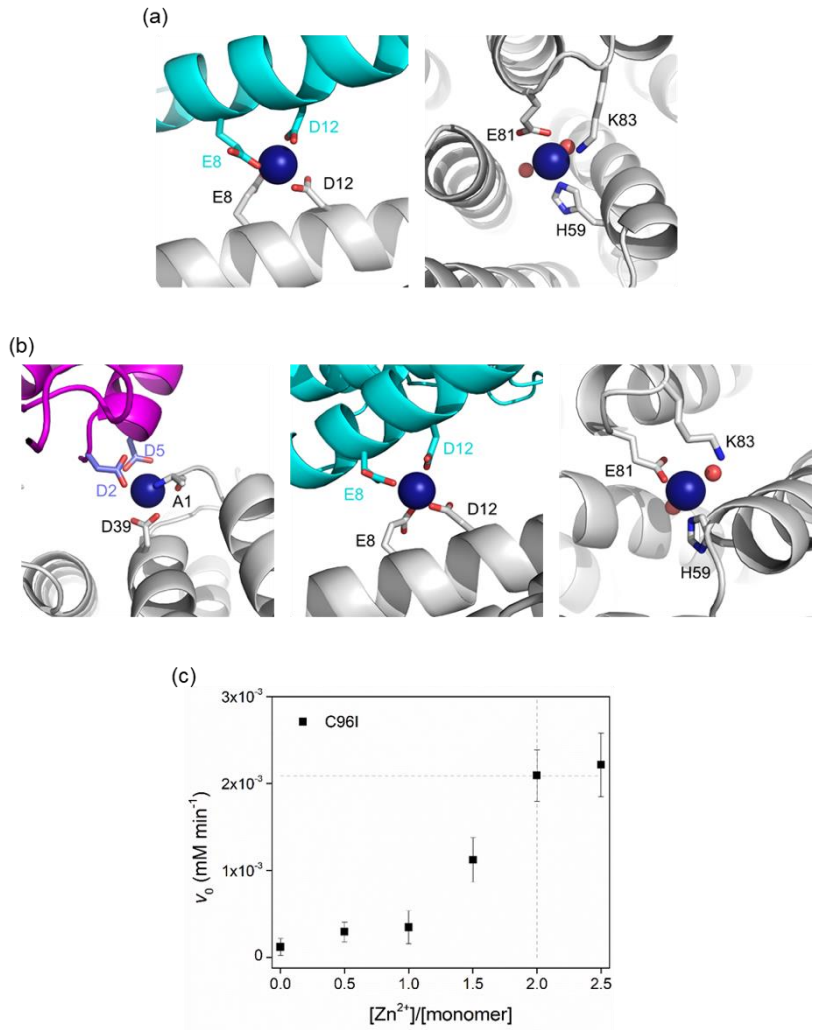


Figure 3.11. Additional Zn-binding sites shown in the crystal structures of (a) C96I and (b) C96K variants. Each tetramer forming crystal packing interactions is shown in different colors. (c) The Zn-dependent hydrolysis activity of C96I with 20 mM ampicillin. The maximum hydrolytic activities were observed when the stoichiometry of the Zn site to protein is 2:1 rather than 2.5:1 or 1.5:1, which would be the case if all (structural, catalytic, and the additional Zn-binding sites) or only structural and the third Zn-binding sites were formed, respectively. Therefore, these additional Zn-binding sites were likely to be observed due to the crystal-packing interactions and nonrelated to the hydrolytic activities of the proteins.

All three variants were closely packed tetramers, similar to the C96T mutant (Figure. 3.1b) and different from the parent protein with the C96 residue (Figure. 3.1a). Consequently, both the structural and catalytic Zn-binding sites were drastically altered by the single mutations. In particular, the directionality of the catalytic Zn–OH_2/OH species in all C96 variants was flipped towards I67 and T97, instead of E92 and Q103 (Figure 3.7). In addition, the first coordination sphere of the Zn center was greatly altered. Whereas the C96 parent protein exhibits a catalytic Zn ion ligated by 2His/1Glu (E89, H93, H100) and a buffer-derived molecule/ion (Figure 3.7a), E89 was no longer coordinated to the catalytic Zn ions in the C96I variant (Figures 3.7b and 3.10a, b). Instead, solvent-derived anions or molecules were bound to Zn, creating the distinct first metal-coordination spheres, the secondary microenvironments, and substrate-binding pockets, which would have enhanced metal-dependent enzyme catalysis. Notably, significantly shorter distances between Zn ions and ligating N atoms of H93 and H100 were observed in C96I protein (1.79–2.03 Å) relative to the parent protein (2.17–2.34 Å) (Tables 3.6). In addition, two discrete angles of N_ε (H93)–Zn–N_δ (H100), 91.3° and 122.7°, were observed, one of which significantly deviated from the parent protein (92.8°), creating two asymmetric bis-His motifs for hydrolytically active Zn-centers.

Proteins (PDB code)	Location of Zn-center	Ligating residues	Interatomic distances between Zn^{2+} and N atom (Å)
C96 (5XZI)	Peripheral/ catalytic	E89, H93, H100	2.17–2.34
C96V (6LDE)		H93, H100	1.93–2.54
C96K (6LDF)		H93, H100	1.94–2.42
C96I (6LDG)		H93, H100	1.79–2.03
Proteins (PDB code)	**Location of Zn-center**	**Ligating residues**	**Interatomic distances between Zn^{2+} and N atom (Å)**
C96 (5XZI)	Core/ structural	H63, H73, H77	1.83–2.04
C96V (6LDE)		H63, H73, D74, H77	1.98–2.04
C96K (6LDF)		H63, H73, D74, H77	1.99–2.02
C96I (6LDG)		H63, H73, D74, H77	1.94–2.03

Table 3.6. Ligating residues and distances between Zn^{2+} and N atoms at core/structural and peripheral/catalytic Zn sites of C96X variants. In C96K and C96I, E89 residue was not coordinated to Zn^{2+} ion and the interatomic distance was measured to be 5.2 Å and 5.1 Å, respectively.

The C96K mutation also altered the first coordination spheres of the catalytic Zn-binding sites (Figure 3.7c). The coordination spheres were composed of H93 and H100 residues with one water molecule, and Zn–OH_2/OH was pointed towards I67 and T97. E89 was also dissociated from the Zn ion and was hydrogen-bonded to E92, Q103, and one ordered water molecule near the catalytic Zn-binding sites (Figure 3.10c, d). Two discrete

angles of two histidine residues and zinc ion were again observed as 103.3° and 125.1°, even further deviated from those in the parent protein. In addition, two conformations of the C96K side chains were observed, forming hydrogen bonds with H93, possibly tuning the chemical properties of the Zn site, such as the pK_a value of catalytically critical residues, the nucleophilicity of Zn–OH species, and Zn-binding affinity.

In the C96V protein, Zn ions were ligated by H93 and H100, and E89 was again no longer directly coordinated to Zn upon C96 mutation (Figure 3.4d). Instead, E89 was hydrogen-bonded to a metal-bound water molecule or was pointed away from the metal-coordination site. Consequently, two solvent-derived molecules, such as Cl and H_2O, were coordinated to the Zn ion in a tetrahedral geometry. In addition, one of the two H100 residues in the asymmetric unit was in two discrete rotameric states, which was not observed in other variants (Figure 3.10e, f). As a result, the angle of N_ε (H93)–Zn–N_δ (H100) became 95.3° on one side and 90.2° and 148.6° on the other side, resulting in the mutant possessing either nearly symmetric or the most asymmetric Zn sites on the α-helical protein–protein interface.

To explore whether these geometric perturbations are related to catalytic activities, I carried out substrate-docking simulations using Galaxydock[43, 44] (Figure 3.12). The data suggest that the discrete Zn-binding sites induced by C96 mutations may favor alternative modes in substrate-binding. Because C96I, C96K, and C96V variants exhibit the unique kinetic parameters or substrate-binding affinity, even the slightly modified Zn-coordination sites might not be an artifact of the crystalline packing interactions but related to the dynamic snapshots of the protein structure in solution. Then, these results indicate that seemingly trivial single mutations on the rotational axis can modify the chemical properties of catalytically active Zn sites.

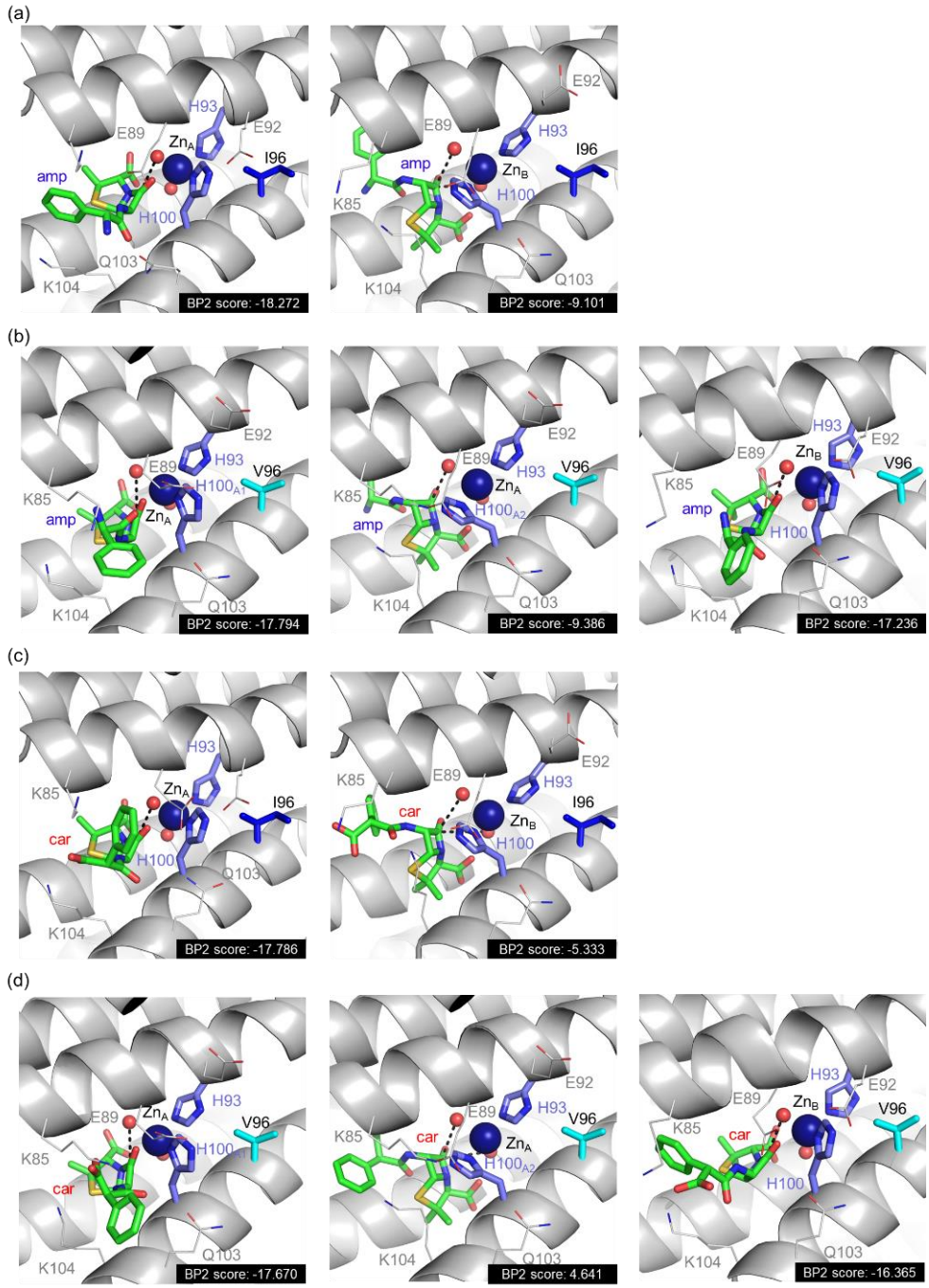


Figure 3.12. The predicted binding poses of β-lactam substrates to C96I or C96V variants. Ampicillin to (a) C96I and (b) C96V mutants. Carbenicillin to (c) C96I and (d) C96V mutants. The BP2 score estimates binding energy in arbitrary energy units, suggesting the most favorable binding mode by respective catalytic Zn sites.

After structural characterization, I further measured the pH-dependent hydrolytic activities to estimate the pK_a value of catalytically critical residues or Zn–OH_2 species (Table 3.7and Figure 3.13). The parent protein exhibited a pK_a value of 8.4(1),[35] whereas those of C96I, C96K, and C96V were estimated to be 8.6(1), 9.3(1), and 9.2(2). These data indicate that a single mutation at a symmetry-related position alters the net concentration of catalytically active species by tuning the chemical environments of catalytic sites.

Round	Mutation	k_{cat} (min^{-1})	K_M (mM)	k_{cat}/K_M or k_2* (min^{-1} M^{-1})	Rate enhancement k_{cat}/k_{uncat}†	Catalytic proficiency ($k_{cat}/K_M)/k_{uncat}$
0	C96	0.15(1)	8(1)	19(2)	$5.0(3)\times10^4$	$6.3(7)\times10^6$
1	C96K	0.33(2)	7(1)	50(4)	$1.1(1)\times10^5$	$1.7(1)\times10^7$
	C96I	0.43(2)	8(1)	52(3)	$1.4(1)\times10^5$	$1.7(1)\times10^7$
2	C96I/A38S	ND§	ND§	23(1)	ND§	$7.7(3)\times10^6$
3	C96I/ A38S/ E81H	0.98(6)	14(2)	71(4)	$3.3(2)\times10^5$	$2.4(1)\times10^7$
4	C96I/ A38S/ E81H/ D39N	1.2(2)	13(3)	100(10)	$4.0(7)\times10^5$	$3.3(3)\times10^7$

Table 3.7. Catalytic activities of the evolved variants at pH 7.0 *For non-saturated behavior, a second-order rate constant (k2) was obtained from a linear fit. †The uncatalyzed rate constant of ampicillin hydrolysis at pH 7 condition was measured to be 3.0(0.1)×10-6 min^{-1}. §ND: not determined due to the lack of saturation behavior.

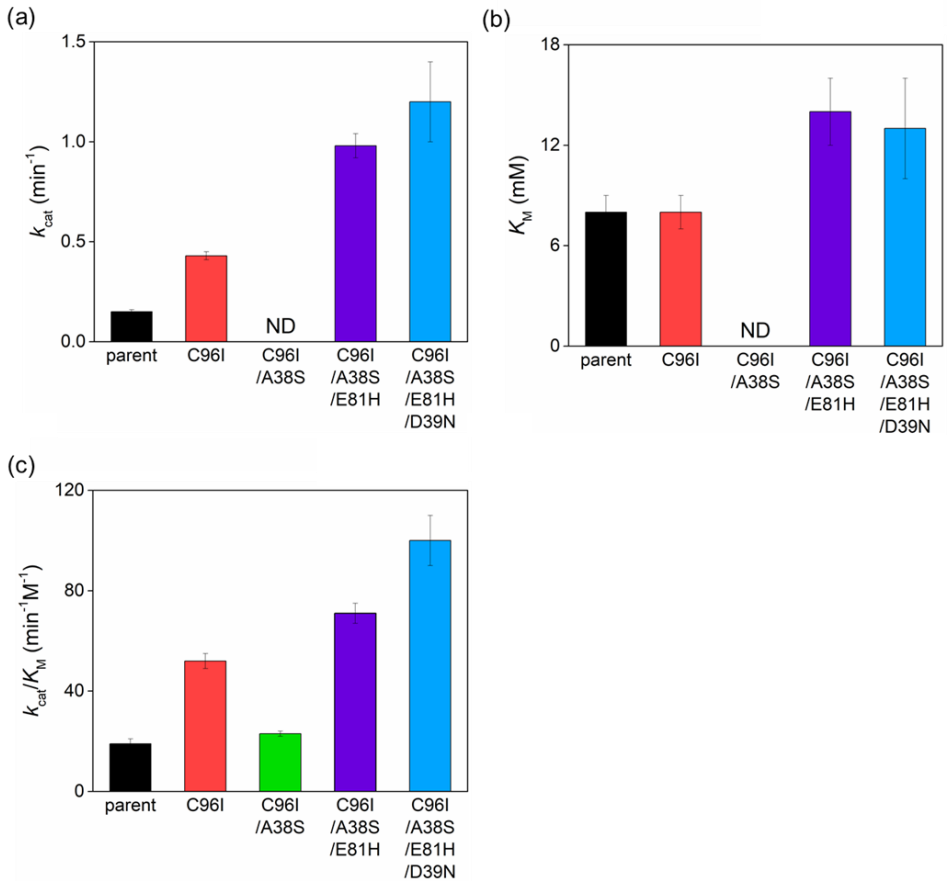


Figure 3.13. Michaelis-Menten kinetic parameters of the variants evolved by symmetry-guided evolution. (a) Catalytic rate constant (k_{cat}) (b) Michaelis constant (K_M) (c) catalytic efficiency (k_{cat}/K_M). In C96I/A38S, a second-order rate constant (k_2) was used to compare to k_{cat}/K_M values of other variants.

3.2.3 Construction and screening of second-, third, and fourth- round libraries

Because even a single mutation at a symmetry-related site gave rise to a significant modification in catalytic activities, protein structures, and biochemical properties, I iteratively constructed second-round libraries with C96I as the template. Again, I constructed two groups of single-site mutant libraries, one by altering symmetry-related residues, such as A38, D39, and E81, and the other by randomizing proximity-related residues, such as T97 and K104. For full coverage with a 95% confidence level,[36] greater than 94 colonies for each single-site saturated mutant library and 1,395 colonies in total were screened. The variation in the relative cell-growth rate upon single mutation, or fitness effect, was observed to follow the order of K104X $\leq$ T97X $\ll$ D39X $\leq$ E81X < A38X (Figure 3.14a), as also represented by box charts (Figure 3.14b inset), again demonstrating that mutations of residues located on the C_2 rotational axes, such as A38, D39, and E81, give rise to more drastic impacts on cell-growth rate than do T97 and K104 mutations. When one or two cells with the highest growth rates from each library were sequenced, A38S, A38I, D39E, E81N, T97M, and T97S mutants were obtained, while in the K104X library, only the template, C96I variant having K104 residue, was the most active cell. The measured specific whole-cell activities of the screened cells were in the following order: T97M $\approx$ T97S $\approx$ C96I $\ll$ E81N $\leq$ D39E $\leq$ A38I $\leq$ A38S (Figure 3.14c and 3.15).

Because C96I/A38S double variants yielded the highest catalytic efficiency, I constructed a third round of the mutant library by randomizing the E81 residue of C96I/A38S as the template (Figure 3.14d). The outputs exhibit variations in cell-growth rates by ca. 3- and 5-fold in positive and negative directions, respectively, and the best-hit was sequenced to be E81H, resulting in a triple variant, C96I/A38S/E81H (Figure. 3.14c). Then, I further

randomized and screened the last symmetry-related residue, D39. The randomized single-mutant library yielded significantly altered cell-growth rates by up to ca. 5-fold relative to that of the C96I/A38S double mutant (Figure 3.14d). The sequencing results indicate that the quadruple variant, C96I/ A38S/E81H/D39N, displayed the fastest cell-growth rate, resulting in 8.1-, 3.2-, and 2.2-fold enhancements relative to that of the best hit from the first, second, and third round of screening, C96I, C96I/A38S, and C96I/A38S/E81H, respectively. These data suggest that the iterative sequence optimizations were carried out efficiently by targeting symmetry-related residues.

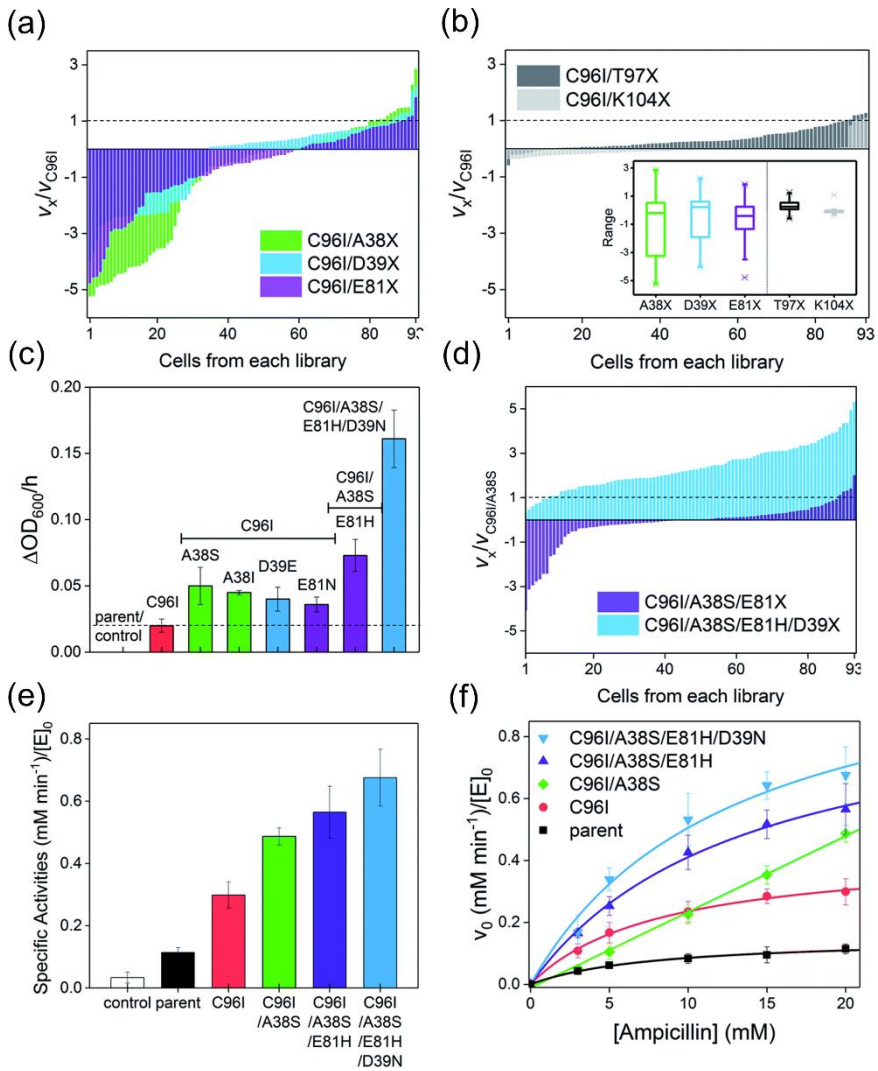


Figure 3.14. The screening and characterization of the hits from the second to fourth rounds of screening. The second round of representative results with (a) symmetry- and (b) proximity-related mutant libraries. The relative cell-growth rates to C96I variant are plotted for each mutant library. In the (b) inset, box chart per each mutant library was included. (c) The cell-growth rates of the best hits from the second-round library in (a). (d) The third and fourth rounds of the screening with symmetry-related mutant libraries. (e) Specific *in vitro* activities of the selected variants. (f) Michaelis–Menten kinetic analysis of the best hits. The error bars in (c), (e), and (f) indicate the standard deviations of three runs of the experiments.

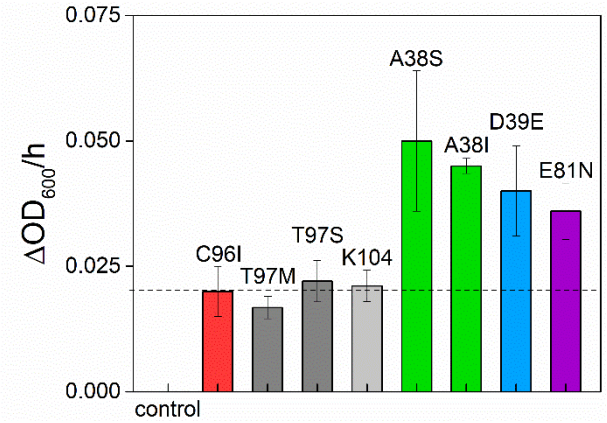


Figure 3.15. Cell growth rates of best hits during the second round of evolution.

3.2.4 Characterization of the second-, third-, and fourth-round mutants.

To demonstrate that the increased cell-growth rates in the presence of antibiotic substrates are derived from enhanced β-lactamase activities, I isolated the best hits from the second, third, and fourth rounds of screening. Then, I measured the steady-state specific activities with 20 mM ampicillin (Figure. 3.14e). The specific β-lactamase activities consecutively increased, yielding a 5.9-fold increase relative to that of the parent (AB5 protein) upon quadruple mutations. In contrast, proximity-related mutations attempted in the first and second rounds of optimizations gave rise to negligible enhancements in hydrolytic activities, consistent with the *in vivo* cell-based activities.

The Michaelis–Menten kinetic analysis of the best hits from each round was also carried out (Figure 3.13, Figure 3.14f and Table 3.7). The overall activities were consecutively elevated throughout every round of evolution, resulting in higher k_{cat} and k_{cat}/K_M values than those of the variants from the previous rounds of screening. Relative to the uncatalyzed rate (k_{uncat}) of $3.0(1)\times10^{6}$ min^{-1}, the activities of the quadruple mutant accounted for

4.0(7)×10^5 and 3.3(3)×10^7 M^{-1} in terms of rate enhancement (k_{cat}/k_{uncat}) and catalytic proficiency ($k_{cat}/K_M/k_{uncat}$), respectively (Table 3.9 and Figure 3.13). Notably, no improvement in K_M was observed throughout the evolution, and C96I/A38S even lost a saturation behavior with increasing substrate concentration, resulting in a second-order rate constant (k_2), instead of k_{cat}/K_M. The weaker substrate-binding affinities of the screened mutants might be attributed to the screening conditions, where substantially high concentrations of antibiotics were applied. Then, the microscopic catalytic properties of the evolved variants might be the result of β-lactamases discretely evolved to elevate turnover rates rather than to lower substrate-binding affinity.

In addition, I measured the pH-dependent hydrolytic activities of the quadruple variant. The pK_a value was 8.4(1) (Table 3.8 and Figure 3.16), which differs by 0.6 from that of C96I/A38S. Notably, I observed an inverse correlation between the pK_a values and kinetic parameters (such as k_{cat} and k_{cat}/K_M) of the mutants having compact tetrameric structures, C96T before the screening, C96I, C96I/A38S, C96I/A38S/E81H, and C96I/A38S/E81H/D39N from evolution, implying that distant mutations increased the net concentration of catalytically essential species, such as Zn–OH species, over that of Zn–OH_2.

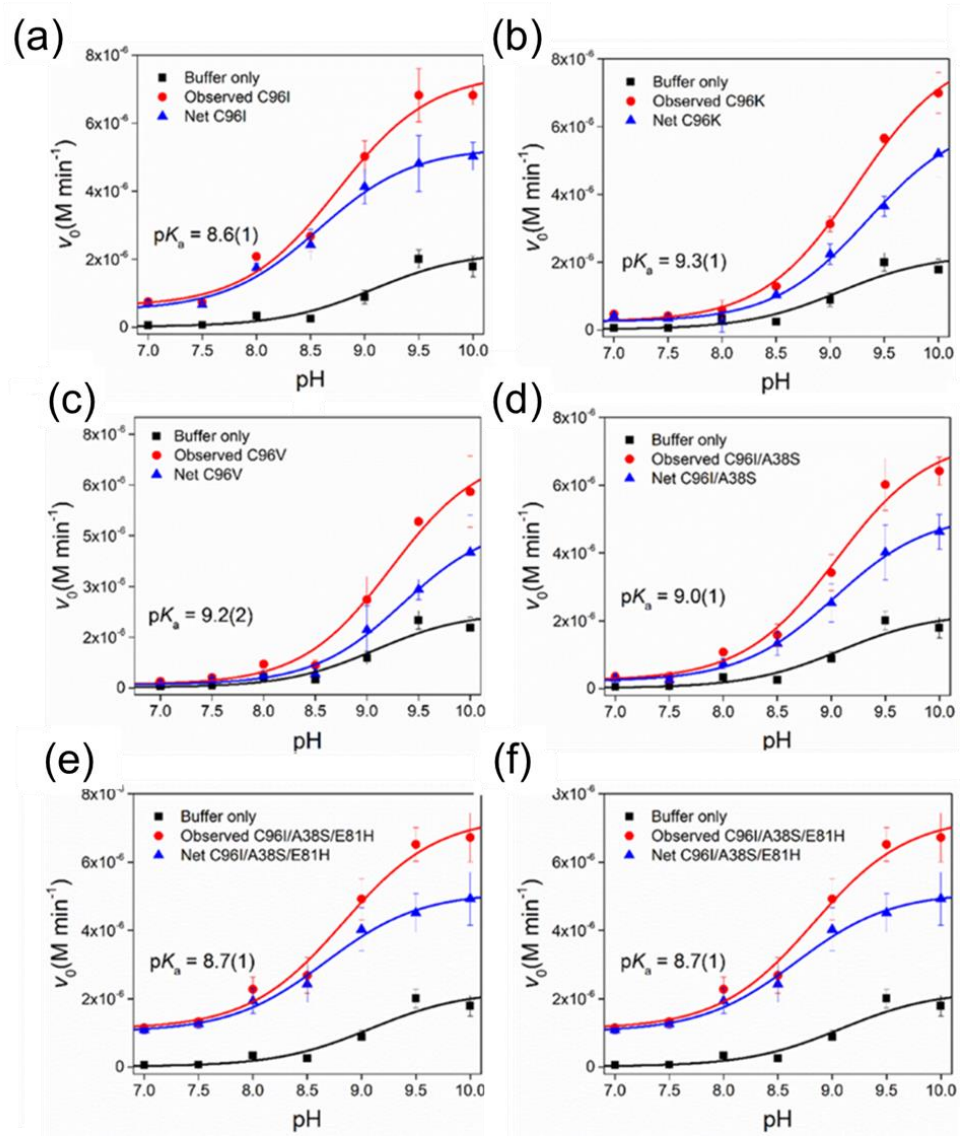


Figure 3.16. The pH-dependent ampicillin hydrolysis activity assays to determine pK_a values of the C96X variants (X = I, K and V) from 1st round of evolution. (a) C96I, (b) C96K, (c) C96V, and further evolved variants (d) C96I/A38S, (e) C96I/A38S/E81H/D39N, and (f) C96I/A38S/E81H/D39N.

Evolution round	Proteins	pK_a
Parent	C96	8.4(1)
1	C96V	9.2(2)
	C96K	9.3(1)
	C96I	8.6(1)
2	C96I/A38S	9.0(1)
3	C96I/A38S/E81H	8.7(1)
4	C96I/A38S/E81H/D39N	8.4(1)

Table 3.8. The pH-dependent hydrolytic activities of the evolved variants. The catalytically critical pK_a values were estimated from non-linear fit analysis. *A second-order rate constant (k_2) was used instead of k_{cat}/K_M. §ND: not determined.

The structural features of the variants were evaluated by size exclusion chromatography. The parent protein exhibits two disulfide bonds formed via four C96 residues in tetramer form,[35] which enables the formation of a tetramer even in the absence of Zn ions at the protein–protein interfaces. Upon the removal of the disulfide bond, Zn-free apo-protein was isolated as a mixture of monomers and dimers, similar to C96T[35] (Figure 3.8b). The isolated mutants from the screening, C96V, C96K, C96I/A38S, C96I/A38S/E81H, and C96I/A38S/E81H/ D39N, exhibit analogous Zn-dependent oligomerization, exclusively forming tetramers even after a series of mutations on the protein–protein interfaces. Intriguingly, exceptions were detected for the C96I and C96L variants from the first-round of screening. They form nearly exclusively tetramers even in the absence of Zn ions. The hydrophobic residues in the two opposing α-helical domains are analogous to those in leucine zippers,[45, 46] where repeated hydrophobic amino acids form noncovalent interactions, inducing oligomerization. Although it is unclear whether the enhanced protein–protein interfaces kinetically promote the *in vivo* activity of C96I and C96L by forming tetramers prior to metal binding and increasing the effective concentrations of catalytically active species inside the cells, the results indicate that the impact of single-site mutations at symmetry-related spots is indeed substantial in both structural and functional aspects, and can therefore give rise to significant impacts on protein evolution.

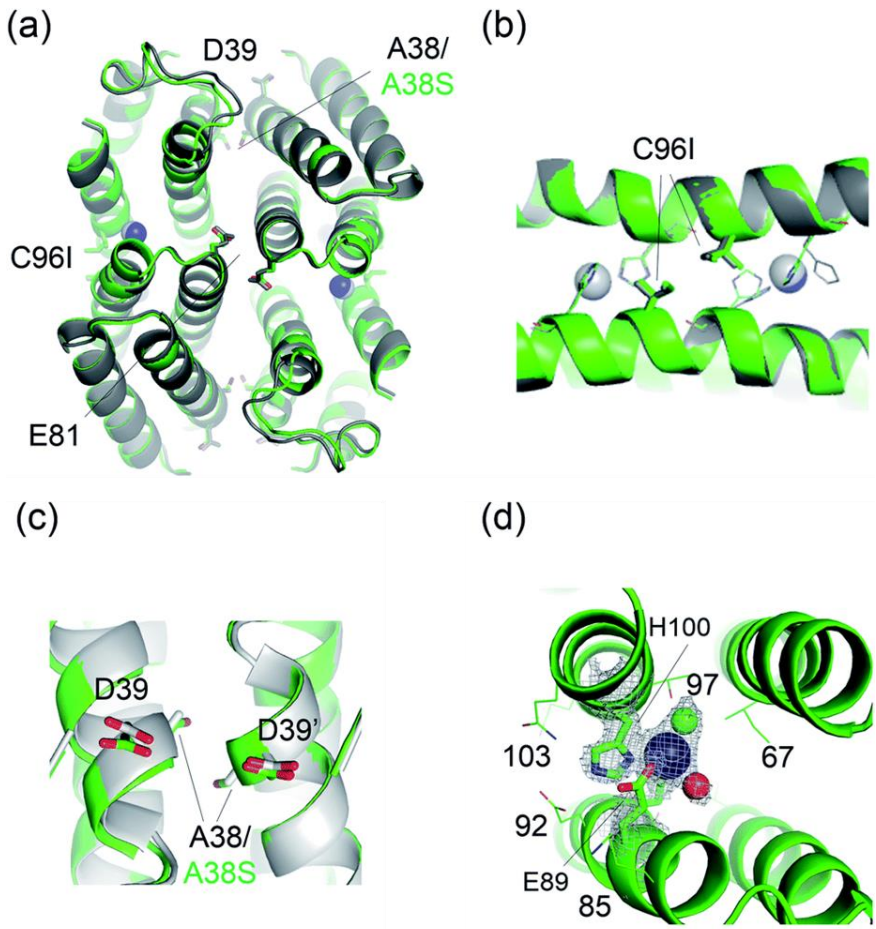


Figure 3.17. X-ray crystal structure of C96I/A38S protein. (a–c) The superimposed structures of C96I/A38S protein with C96I variant, colored in green and grey, respectively. Zn ions in the catalytic sites are shown with navy and grey spheres, respectively. Symmetry-related residues located at the C_2 rotational axes, A38 or A38S, D39, E81, and C96I, are shown with sticks. (d) The enlarged catalytic Zn site in C96I/A38S double variant. A Zn ion, a metal-bound water molecule, and a chloride ion are shown with navy, red, and green spheres, respectively. Metal-binding (H93 and H100) and weakly interacting residue (E89) are shown with sticks. The first metal-coordination site in (d) is shown with $2F_o–F_c$ electron density contoured at 1.0 σ overlaid.

Finally, one of the evolved variants was characterized by X-ray crystallography (Figure 3.17). The C96I/A38S variant from the second-round screening was a compact tetramer, similar to the C96I, C96K, and C96V variants. The catalytic Zn-binding sites were also similar in that they are composed of Zn ions coordinated to H93, H100 (Zn–N = 2.0–2.3 Å), and one or two water molecules. Notably, the angle of 3N (H93)–Zn–δN (H100) was measured to be 92.2° and 94.5°, resulting in more symmetric metal-centers located at the α-helical domains than in the C96I single variant. The subtle yet significant geometric perturbation at the first metal-coordination sphere, induced by A38S mutation, might be associated with the enhanced catalytic activities, possibly by adjusting the nucleophilicity of the Zn–OH species, optimizing noncovalent interactions with the secondary coordination spheres, and/or shaping the internal active site pocket. Notably, A38S is 17 Å distant from the catalytic Zn center. Therefore, the mutation effect exerted at the rotational axis is likely to be transferred through fluxional protein–protein interfaces, suggesting that distant and beneficial mutations can be created efficiently by targeting symmetry-related residues for protein evolution.

3.3. Conclusion

I demonstrated that mutation of residues located on rotational axes can give rise to substantial alterations in the structure and function of a *de novo* metallo-β-lactamase. Because numerous artificial metalloenzymes and even natural proteins are homo-oligomers possessing rotational axes, targeting symmetry-related residues can be a novel strategy in the construction of well-focused mutant libraries. In addition, this approach is orthogonal to the commonly adapted criteria used in the design of targeted libraries, such as distance from active sites, B-factor, and sequence conservation, therefore providing diverse routes to explore protein sequence networks. These results might be related to the evolutionary advantages of protein oligomers that might multiply and propagate mutation effects as well.[47-50] In addition, the *in vitro* characterization of the outputs indicates that discrete evolutionary chemical pressures lead to the emergence and divergence of proteins with discrete catalytic properties. The chemical pressure-dependent evolution suggests that artificial enzymes or whole-cell biocatalysts can be created to aim for specific kinetic and/or thermodynamic properties at a microscopic level, increasing the accuracy and predictability of directed evolution. Therefore, a more efficient design strategy for mutant libraries and screening might expedite enzyme evolution and the exploration of protein sequence–structure–function relations with greater accuracy.

3.4. Experimental section

Construction of mutant libraries. Saturated mutagenesis was carried out using primers containing the NNK degenerate codon as described previously.[36] PCR was performed using custom-designed primers (Table 3.9). After DpN I (Enzynomics) digestion for 1.5 h at 37 ºC, the PCR mixtures were transformed to DH5α *E. coli* competent cells. Greater than 100 colonies were selected for each single-site randomized library, and they were inoculated to LB medium containing 50 mg L^{-1} kanamycin. After the overnight cell-growth at 37 ºC, plasmids were extracted using a mini-prep kit and sent out for sequencing (Macrogen or Bionics). When the selected position was not fully randomized, additional PCRs with the redesigned primers were performed. Representative sequencing chromatograms for each library are shown in Figure. 3.2.

Template	Primer name	Primers	Calculated T_m (°C)	Annealing T (°C) in PCR
C96	C96X	5'-CA GAG CAT CTG AAA NNK ACC TGC AAC CAT TG-3' 5'-CA ATG GTT GCA GGT MNN TTT CAG ATG CTC TG-3'	60.4–66.4	53.4, 55.4, 59.4, 61.4
C96	T97X	5'-G CAT CTG AAA TGC NNK TGC AAC CAT TGC CAC C-3' 5'-G GTG GCA ATG GTT GCA MNN GCA TTT CAG ATG C-3'	63.1–67	55.1, 56.1, 57.1, 58.1, 59.1, 60.1, 61.1, 62.1
C96I	T97X	5'-G CAT CTG AAA ATT NNK TGC AAC CAT TGC CAC C-3' 5'-G GTG GCA ATG GTT GCA MNN AAT TTT CAG ATG C-3'	60.5–64.4	54.5 55.4, 56.4, 57.4, 58.4, 59.4, 60.4, 61.4
C96, C96I	K104X	5'-C AAC CAT TGC CAC CAG NNK TAT CGT TAA TT-3' 5'-AA TTA CG ATA MNN CTG GTG GCA ATG GTT G-3'	57.5–61.6	52.6, 53.6 54.6, 55.6, 56.6, 57.6, 58.6, 59.6
C96, C96I	A38X	5'-CG GCC GCA GCG NNK GAT GCG TGG AGC GCA AC-3' 5'- GT TGC GCT CCA CGC ATC MNN CGC TGC GGC CG-3'	71–75	63, 64, 65, 66, 67, 68, 69, 70
C96, C96I	D39X	5'-G GCC GCA GCG GCG NNK GCG TGG AGC GCA ACG-3' 5'- CGT TGC GCT CCA CGC MNN CGC CGC TGC GGC C-3'	73.4–77.2	65.2, 66.2, 67.2, 68.2, 69.2, 70.2, 71.2, 72.2
C96, C96I, C96I/A38S	E81X	5'-CAC CTG GCA AAT NNK GGT AAA	59.3–63.1	53.3, 54.3, 55.3, 56.3,

		GTA AAA GAT GC-3’ 5’- GC ATC TTT TAC TTT ACC MNN ATT TGC CAG GTG-3’		57.3, 58.3, 59.3, 60.1
C96I/A38S/ E81H	D39X	5’-G GCC GCA GCG TCG NNK GCG TGG AGC GCA ACG-3’ 5’- CGT TGC GCT CCA CGC MNN CGA CGC TGC GGC C-3’	72.3–76.3	64.2, 65.2, 66.2, 67.2, 68.2, 69.2, 70.2, 71.2

Table 3.9. DNA primers used for saturated mutagenesis.

Next-generation-sequencing of C96X libraries. To validate the quality of the randomized mutant libraries, I carried out next-generation-sequencing (NGS) of the C96X library as a representative (Macrogen). Amplicon samples were prepared by attaching sequencing primers and the pre-adaptor sequences (Table 3.10 and Figure 3.9a). The size of the PCR products was validated by 1% agarose gel electrophoresis, prior to the sequencing (Figure. 3.18b). The predicted and NGS results are shown in Figure. 3.18c.

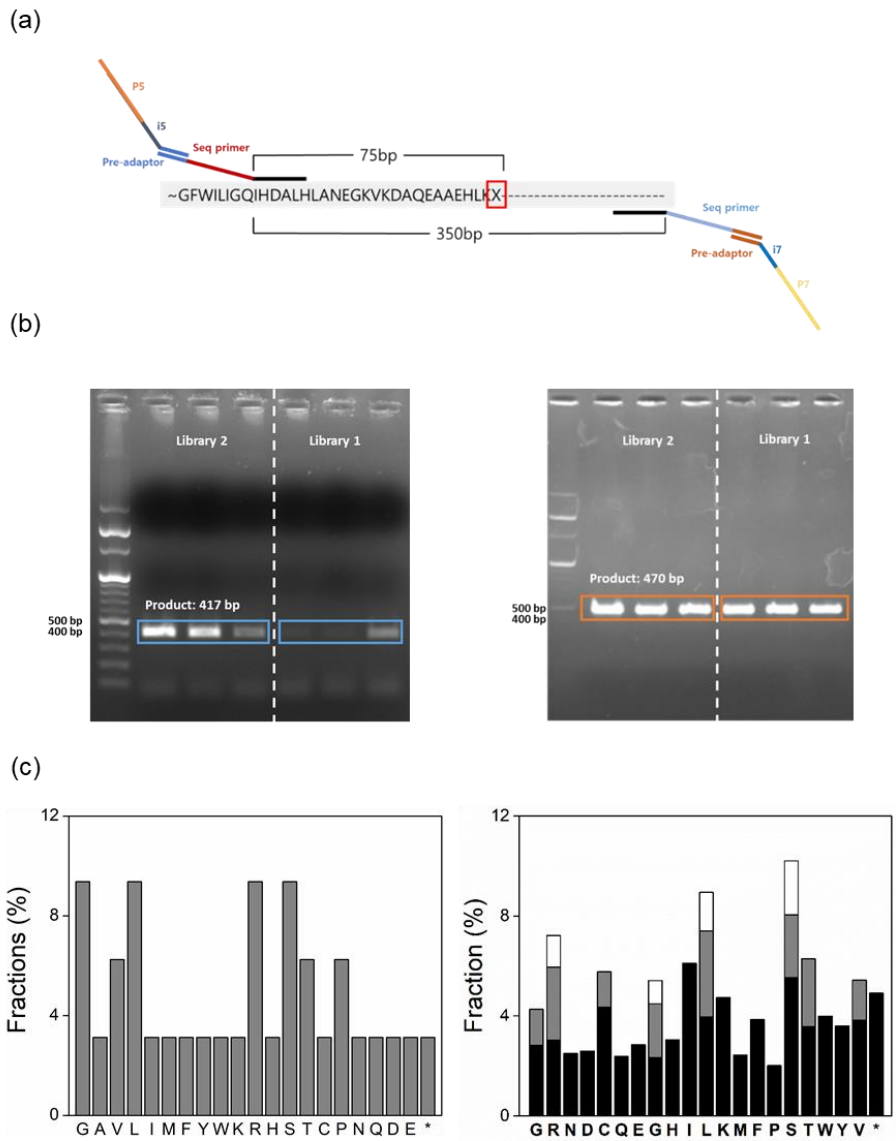


Figure 3.18. The next generation sequencing (NGS) of C96X libraries. (a) Fragment PCR (b) DNA agarose-gel electrophoresis at (left) the 1st and (right) the 2nd step (c) Plots of (left) theoretical fractions of amino acids encoded by NNK degenerated primer and (right) the NGS results. Discrete codons encoding the same amino acids are colored in different grey scales. Asterisks indicate a TAG stop codon.

Primer name	NGS_1st_fwd	NGS_1st_rev	NGS_2nd_fwd (i5+P5)	NGS_2nd_rev (i7+P7)
Pre-adaptor	TCGTCGGCAGCGTC	GTCTCGTGGGCTCGG		
Seq primer	AGATGTGTATAAGAGACAG	AGATGTGTATAAGAGACAG		
Specific locus primer	aTTCACGACGCGCTGCACCTg	ccaatccggatatagttcctc		
Primer sequence	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGaTTCACGACGCGCTGCACCTg-3'	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGccaatccggatatagttcctc-3'	5'-AATGATACGGCGACCACCGAGATCTACAC[i5]TCGTCGGCAGCGTC-3'	5'-CAAGCAGAAGACGGCATACGAGAT[i7]GTCTCGTGGGCTCGG-3'
T_m (°C)	75.2	73.5	73.1	71.8
Annealing T (°C)	68.2, 70.2	66.5, 68.5	66.1, 68.1	64.8, 66.8

Table 3.10. DNA primers for NGS analysis.

Screening of the mutant libraries. The plasmid mixtures were transformed to BL21 (DE3) *E. coli* competent cells containing the cytochrome c maturation cassette (ccm)[51] and were grown overnight on LB/agar containing 50 mg L^{-1} kanamycin and 30 mg L^{-1} chloramphenicol. Over 100 colonies for each single-site randomized library was inoculated in a 96-well microplate containing 200 mL of LB medium with the antibiotics described above. After overnight growth at 37 ºC with constant shaking at 200 rpm, the cultures were diluted 10-fold with 200 mL of LB medium containing 50 mg L^{-1} kanamycin, 35 mg L^{-1} chloramphenicol, and50 mM $ZnCl_2$. Increasing concentrations of ampicillin (10 mg L^{-1} in the first round, 15 mg L^{-1} in the second round, and 17.5 mg L^{-1} in the third and fourth round of the screening) were added throughout the screening when the optical density at 600 nm (OD_{600}) reached approximately 0.4. Notably, the initial cell density was critical in the determination of relative cell-growth rates as illustrated by the Pearson correlation coefficient,[52] and the value of the screening condition (~0.08) was sufficiently lower than the reported threshold, ~0.3 (Figure 3.19). After monitoring the cell-growth rates at 37 ºC for 3 h, the fastest-growing cells were selected for sequencing (Figure 3.5). At least 2–4 sets of a 96-well plate with the colonies of 93 mutants and 3 parent or template proteins were screened for each library and representative results are included in Figure 3.3 and Figure 3.15.

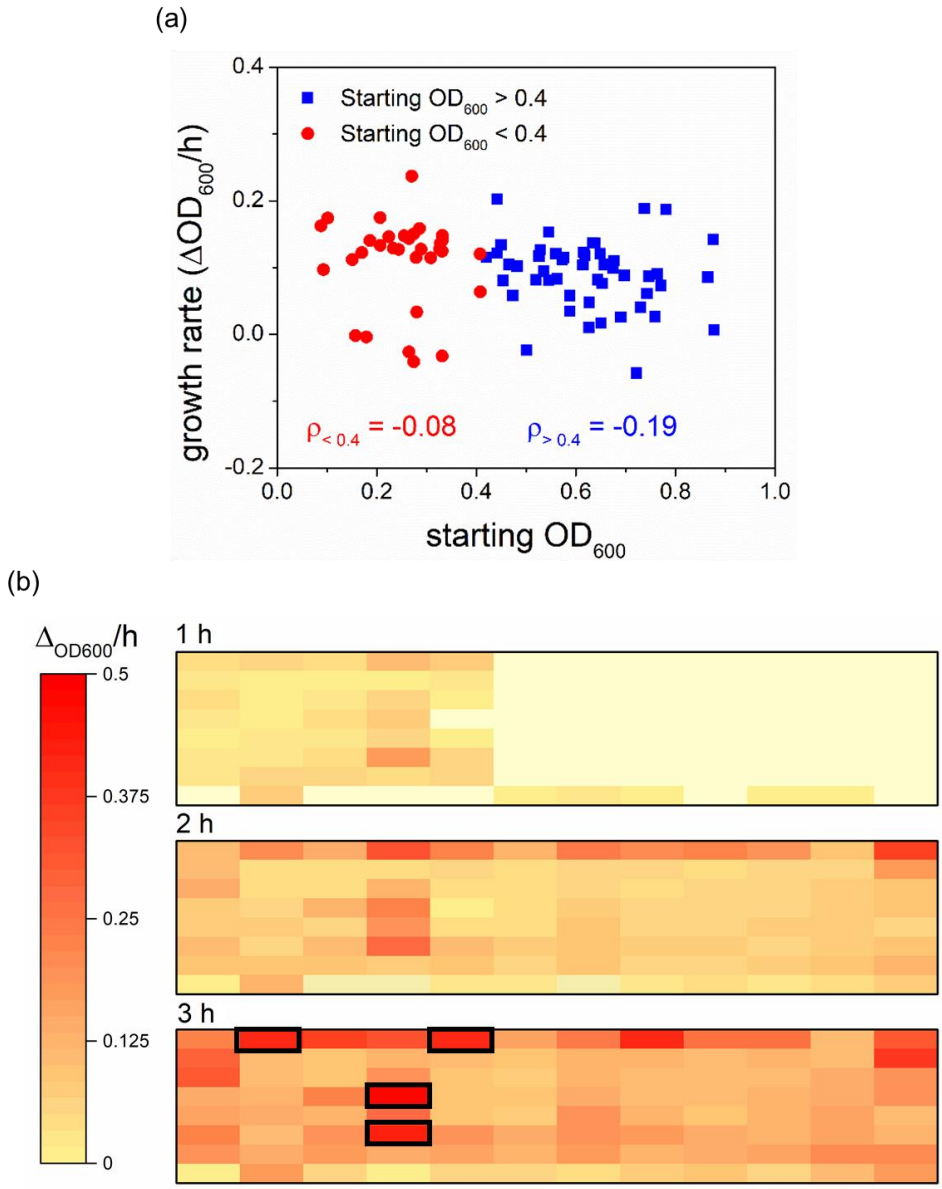


Figure 3.19. The whole-cell screening upon the addition of ampicillin (10–15 mg/L). (a) The correlation of the starting OD_{600} values and cell growth rates. (b) Representative screening result of the mutant library randomized at 96 position. The cell growth rates were obtained by monitoring the optical density at 600 nm for 3 h. The fastest growing cells were highlighted as black boxes.

Protein expression, purification, and biochemical characterization. Protein expression and purification of the positive hits were carried out as described previously.[35] In short, pET20b(+) plasmid for encoding the *ab5* gene was transformed to BL21 (DE3) competent cells. The cells were grown in LB medium with 150 rpm shaking at 37 ºC for 18 h. Cell pellets were harvested by centrifugation at 5000 rpm (4,715 g) at 7 ºC for 10 min. After sonication of the cell pellets in 10 mM sodium phosphate (NaPi) pH 8.0 buffer with pulse on/off = 9/9 s for 40 min in an iced bath, HCl (25%) solution was added to the cell lysates up to pH 5.0. After centrifugation at 13,000 rpm (18,800 g) at 7 ºC for 30 min, the supernatants were adjusted to pH 8.0 by adding 2 M NaOH solution. Then, the solutions were manually loaded onto a Q-sepharose column pre-equilibrated with 10 mM NaPi pH 8.0 buffer. By applying the step gradients of 0–1 M NaCl in NaPi buffer, red fractions due to the presence of heme cofactor were collected and concentrated using stirred cells (Amicon) with 10 kDa cutoff membranes. Then, the samples were loaded onto a HiTrap Q HP anion exchange column (GE Healthcare Life Sciences) pre-equilibrated with 10 mM NaPi pH 8.0 buffer at 4 ºC using FPLC (AKTA pure). A linear gradient of 0–1 M NaCl was applied, and the fractions with $A_{415}/A_{280} > 4$ measured using a UV-Vis spectrophotometer (Agilent Cary 8454) were collected and concentrated.

Then, the protein sample was loaded onto a HiLoad 16/600 Superdex 75 pg column (GE Healthcare Life Sciences) pre- equilibrated with 20 mM Tris/HCl pH 7.0 buffer with 150 mM NaCl. After elution, pure fractions were collected and concentrated. To prepare the metal-free, apo protein, 10-fold EDTA to the protein was treated for 1 h at 4 ºC, and the excess EDTA was removed by 10DG desalting column (Biorad) chromatography. Metal content was measured by a colorimetric assay using 4-(2-pyridylazo) resorcinol (PAR) as described previously.[53, 54] The purified metal-free, apo-protein was concentrated up to 1.0 mM and stored at -80 ºC until further use. The protein concentration was determined by using $\varepsilon_{415\ nm} = 148,000\ cm^{-1}\ M^{-1}$.

***In vitro* hydrolytic activity assay with ampicillin.** *In vitro* hydrolytic activities of Zn-complexed positive hits were measured by time-dependent HPLC analysis as reported previously.[35, 40] The reaction was initiated by the addition of either ampicillin or carbenicillin into Zn^{2+}-bound protein (7 μM in 100 μL of reaction volume) in 100 mM MOPS pH 7.0 at 25 ºC. The reaction mixture (2 μL) was injected onto the C18 column in HPLC (Agilent Infinity 1260) and eluted with a linear gradient, starting from 90% H_2O/10% CH_3CN to 10% H_2O/90% CH_3CN for 20 min. Trifluoroacetic acid (TFA) was added to the elution solvents, H_2O and CH_3CN, at 0.05% and 0.1% (v/v) as final concentrations, respectively. The substrate consumption rates were measured by monitoring the absorbance changes at 220 nm. The initial rates were determined from the reactions with various concentrations of the applied substrate. Michaelis–Menten parameters such as k_{cat} and k_{cat}/K_M or second-order rate constants, k_2, were determined from iterative non-linear or linear plots, respectively, using Origin 2016 software (Figure 3.4, 3.13, 3.14f, and 3.20; Tables 3.2 and 3.7). and the hydrolytic activity of C96I/A38S and C96I/A38S/E81H for ampicillin by equivalent Zn ions were measured and displayed in Figure 3.21.

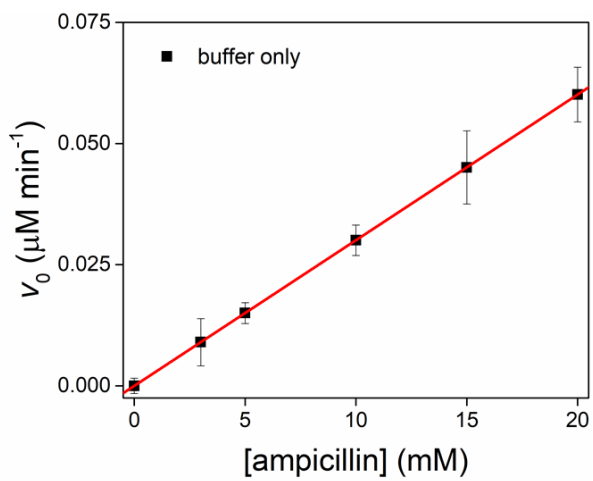


Figure 3.20. The uncatalyzed hydrolysis of ampicillin monitored in 100 mM.

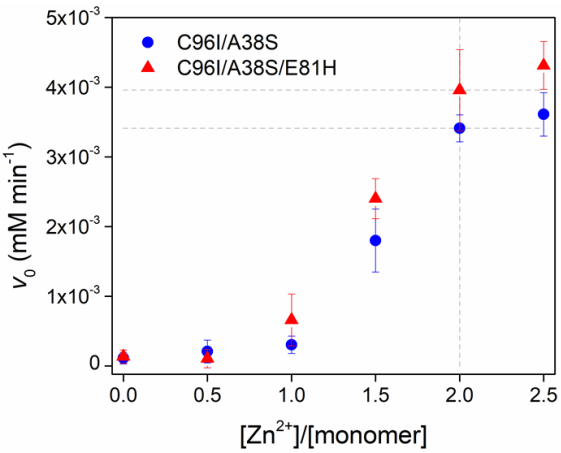


Figure 3.21. The Zn-dependent ampicillin hydrolysis activity assays of C96I/A38S and C96I/A38S/E81H.

The pH-dependent hydrolytic activities of Zn-complexed positive hits were monitored by measuring the specific activity with 3 mM ampicillin under various pH conditions; 100 mM MOPS buffer for pH 7.0–7.5 and 100 mM sodium borate buffer for pH 8.0–10.0 ranges (Table 3.8 and Figure 3.16). Then, the following equation was used for iterative non-linear fitting using Origin 2016 software; $y = (k_{max} * 10^{(pH\text{-}pKa)})/(1 + 10^{(pH\text{-}pKa)})$.

Determination of substrate-binding affinities with intrinsic fluorescence. Intrinsic tryptophan and tyrosine fluorescence changes were monitored using a microplate reader (Biotek Synergy H1) at 25 °C. The parent protein (AB5 with C96 residue) possesses a tryptophan (W66) and a tyrosine (Y105) in the vicinity of the active site (Figure 3.14a). Various concentrations of ampicillin or carbenicillin (0–10 mM) were added to Zn^{2+}-bound protein complexes and were incubated for3 min either in 100 mM MOPS pH 7.0. Notably, ampicillin and carbenicillin have no considerable absorption at 260 and 295 nm, and the inner filter effect can be negligible up to 4 mM for tyrosine and 10 mM for tryptophan residues (Figure. 3.6b). The intrinsic tyrosine and tryptophan fluorescence changes were observed at 315 nm and 350 nm, respectively, upon the addition of ampicillin when 260 nm and 290 nm excitation were

applied, respectively (Figure. 3.6c). The relative fluorescence changes from tyrosine were fit to a linear Stern–Volmer equation, $F_0/F = 1 + (K_a \times [substrate])$. For tryptophan fluorescence changes, a non-linear modified Stern–Volmer equation, $F_0/(-\Delta F) = 1/(K_a \times f_a \times [substrate]) + (1/f_a)$ was applied to account for the presence of multiple tryptophan residues, where K_a and f_a values represent the Stern–Volmer constant and accessible fraction of tryptophan, respectively (Table 3.3 and 3.4; Figure 3.14d, e).

Determination of oligomeric states of the proteins. The oligomerization states of the parent protein and the evolved variants were determined by size exclusion chromatography. The proteins (150 mM) were loaded on a HiLoad 16/600 Superdex 75 pg column (GE Healthcare and Life Sciences). The retention volume for the monomer, dimer, and tetramer was determined by a linear fit analysis. The oligomeric states of the best hits from each round of screening, C96I/A38S, C96I/A38S/E81H, and C96I/A38S/E81H/D39N variants, were also similarly determined by size exclusion chromatography (Figure. 3.8b)

Crystallization and determination of X-ray structures. For crystallization, additional $ZnCl_2$ (0.2 equiv per monomer) was added to the Zn-complexed tetrameric proteins. Single crystals of Zn^{2+}-bound C96X variants (X = I, K, and V) and C96I/A38S double mutants were obtained by sitting-drop vapor diffusion at room temperature using temperature using 1 mL of protein stock (1.0 mM) and 0.5 mL of precipitants listed in Table 3.5. The data were collected in the Pohang Accelerator Laboratory (PAL) using either 7A or 11C beamline. Diffraction data were processed with HKL 2000[55] and CCP4i.[56] Molecular replacements were performed with Molrep[57] by using a structure of ^{C96}RIDC1 monomer (PDB 3IQ6) as a search model. Rigid-body and restrained refinements were carried out using REFMAC5[58] along with manual inspection and refinements with COOT[59] (Table 3.5). The resolutions of C96I, C96K, C96V, and C96I/A38S protein structures were determined as

1.98, 2.35, 2.00, and 2.45 Å, respectively. Structural and catalytic Zn-binding sites in C96X variants (X = I, K, and V) are shown in Figures 3.9–11. The geometric parameters of the Zn-binding sites are summarized in Table 3.5 and 3.6

Docking simulation using GalaxyDock. A flexible protein–ligand docking program GalaxyDock[43] was used to predict the binding poses and energies of ampicillin and carbenicillin to the C96X variants. To simulate catalytic binding modes, a close distance between the catalytic water oxygen and the β-lactam ring carbonyl carbon of the substrate was favored during docking by adding a harmonic penalty for a large distance. Side-chains in the active site were allowed to move except for the Zn-coordinating side chains. After filtering out poses with oxygen-carbon distances longer than 3.5 Å, binding stabilities of the predicted binding poses were estimated using the GalaxyDock BP2 score.[44] Predicted binding poses are displayed in Figure. 3.12.

3.5. References

(1) Davidi, D.; Longo, L. M.; Jablonska, J.; Milo, R.; Tawfik, D. S., *Chem. Rev.* **2018,** *118* (18), 8786-8797.

(2) Tokuriki, N.; Tawfik, D. S., *Science* **2009,** *324* (5924), 203-207.

(3) Anbar, A. D., *Science* **2008,** *322* (5907), 1481.

(4) Dupont, C. L.; Butcher, A.; Valas, R. E.; Bourne, P. E.; Caetano-Anollés, G., *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107* (23), 10567.

(5) Seffernick, J. L.; Wackett, L. P., *Biochemistry* **2001,** *40* (43), 12747-12753.

(6) Orencia, M. C.; Yoon, J. S.; Ness, J. E.; Stemmer, W. P. C.; Stevens, R. C., *Nat. Struct. Biol.* **2001,** *8* (3), 238-242.

(7) Seelig, B.; Szostak, J. W., *Nature* **2007,** *448*, 828.

(8) Rothlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D., *Nature* **2008,** *453* (7192), 190-5.

(9) Studer, S.; Hansen, D. A.; Pianowski, Z. L.; Mittl, P. R. E.; Debon, A.; Guffy, S. L.; Der, B. S.; Kuhlman, B.; Hilvert, D., *Science* **2018,** *362* (6420), 1285-1288.

(10) Jeschek, M.; Reuter, R.; Heinisch, T.; Trindler, C.; Klehr, J.; Panke, S.; Ward, T. R., *Nature* **2016,** *537* (7622), 661-665.

(11) Siddiq, M. A.; Hochberg, G. K. A.; Thornton, J. W., *Curr. Opin. Struct. Biol.* **2017,** *47*, 113-122.

(12) May, O.; Nguyen, P. T.; Arnold, F. H., *Nature Biotechnology* **2000,** *18* (3), 317-320.

(13) Wang, J.-b.; Li, G.; Reetz, M. T., *Chem. Commun.* **2017,** *53* (28), 3916-3928.

(14) Giver, L.; Gershenson, A.; Freskgard, P.-O.; Arnold, F. H., *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95* (22), 12809-12813.

(15) Minagawa, H.; Yoshida, Y.; Kenmochi, N.; Furuichi, M.; Shimada, J.; Kaneko, H., *Cell. Mol. Life Sci.* **2007,** *64* (1), 77-81.

(16) Zeymer, C.; Hilvert, D., *Annu. Rev. Biochem.* **2018,** *87* (1), 131-157.

(17) Arnold, F. H.; Wintrode, P. L.; Miyazaki, K.; Gershenson, A., *Trends Biochem. Sci.* **2001,** *26* (2), 100-106.

(18) Packer, M. S.; Liu, D. R., *Nat. Rev. Genet.* **2015,** *16* (7), 379-94.

(19) Reetz, M. T.; Carballeira, J. D.; Vogel, A., *Angew. Chem. Int. Ed.* **2006,** *45* (46), 7745-7751.

(20) Sumbalova, L.; Stourac, J.; Martinek, T.; Bednar, D.; Damborsky, J., *Nucleic Acids Research* **2018,** *46* (W1), W356-w362.

(21) Wu, S.; Acevedo, J. P.; Reetz, M. T., *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107* (7), 2775-2780.

(22) Campbell, E.; Kaltenbach, M.; Correy, G. J.; Carr, P. D.; Porebski, B. T.; Livingstone, E. K.; Afriat-Jurnou, L.; Buckle, A. M.; Weik, M.; Hollfelder, F.; Tokuriki, N.; Jackson, C. J., *Nat. Chem. Biol.* **2016,** *12* (11), 944-950.

(23) Wagner, A., *Nat. Rev. Genet.* **2008,** *9* (12), 965-974.

(24) Miton, C. M.; Tokuriki, N., *Protein Sci.* **2016,** *25* (7), 1260-1272.

(25) Newton, M. S.; Arcus, V. L.; Gerth, M. L.; Patrick, W. M., *Curr. Opin. Struct. Biol.* **2018,** *48*, 110-116.

(26) Romero, P. A.; Arnold, F. H., *Nature Reviews Molecular Cell Biology* **2009,** *10* (12), 866-876.

(27) Bloom, J. D.; Arnold, F. H., *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106* (Supplement 1), 9995.

(28) Esvelt, K. M.; Carlson, J. C.; Liu, D. R., *Nature* **2011,** *472* (7344), 499-503.

(29) Moore, C. L.; Papa, L. J.; Shoulders, M. D., *J. Am. Chem. Soc.* **2018,** *140* (37), 11560-11564.

(30) Cirino, P. C.; Mayer, K. M.; Umeno, D., Humana Press: 2003; Vol. 1179.

(31) Jeong, W. J.; Yu, J.; Song, W. J., *Chem. Commun.* **2020,** *56* (67), 9586-9599.

(32) Burton, A. J.; Thomson, A. R.; Dawson, W. M.; Brady, R. L.; Woolfson, D. N., *Nat. Chem.* **2016,** *8* (9), 837-844.

(33) Zastrow, M. L.; Peacock, A. F. A.; Stuckey, J. A.; Pecoraro, V. L., *Nat. Chem.* **2012,** *4* (2), 118-123.

(34) Rufo, C. M.; Moroz, Y. S.; Moroz, O. V.; Stöhr, J.; Smith, T. A.; Hu, X.; DeGrado, W. F.; Korendovych, I. V., *Nat. Chem.* **2014,** *6* (4), 303-309.

(35) Song, W. J.; Yu, J.; Tezcan, F. A., *J. Am. Chem. Soc.* **2017,** *139* (46), 16772-16779.

(36) Reetz, M. T.; Carballeira, J. D., *Nat. Protoc.* **2007,** *2* (4), 891-903.

(37) Boucher, J. I.; Bolon, D. N.; Tawfik, D. S., *Protein science : a publication of the Protein Society* **2016,** *25* (7), 1219-26.

(38) Emond, S.; Petek, M.; Kay, E. J.; Heames, B.; Devenish, S. R. A.; Tokuriki, N.; Hollfelder, F., *Nat. Commun.* **2020,** *11* (1), 3469.

(39) Santiago, F.; Doscher, E.; Kim, J.; Camps, M.; Meza, J.; Sindi, S.; Barlow, M., *PLOS ONE* **2020,** *15* (1), e0228240.

(40) Song, W. J.; Tezcan, F. A., *Science* **2014,** *346* (6216), 1525.

(41) J.R., L., 3 ed.; Springer: Boston, MA, 2006.

(42) Poveda, J. A.; Prieto, M.; Encinar, J. A.; González-Ros, J. M.; Mateo, C. R., *Biochemistry* **2003,** *42* (23), 7124-7132.

(43) Shin, W.-H.; Seok, C., *Journal of Chemical Information and Modeling* **2012,** *52* (12), 3225-3232.

(44) Baek, M.; Shin, W.-H.; Chung, H. W.; Seok, C., *Journal of Computer-Aided Molecular Design* **2017,** *31* (7), 653-666.

(45) Landschulz, W. H.; Johnson, P. F.; McKnight, S. L., *Science* **1988,** *240* (4860), 1759.

(46) Alber, T., *Current Opinion in Genetics & Development* **1992,** *2* (2), 205-210.

(47) Marianayagam, N. J.; Sunde, M.; Matthews, J. M., *Trends Biochem. Sci.* **2004,** *29* (11), 618-625.

(48) Ali, M. H.; Imperiali, B., *Bioorg. Med. Chem.* **2005,** *13* (17), 5013-5020.

(49) André, I.; Strauss, C. E. M.; Kaplan, D. B.; Bradley, P.; Baker, D., *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105* (42), 16148-16152.

(50) Marsh, J. A.; Teichmann, S. A., *Annu. Rev. Biochem.* **2015,** *84* (1), 551-575.

(51) Braun, M.; Thöny-Meyer, L., *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101*, 12830-12835.

(52) Wang, J., n *Encyclopedia of Systems Biology*, Dubitzky, W.; Wolkenhauer, O.; Cho, K.-H.; Yokota, H., Eds. Springer New York: New York, NY, 2013; pp 1671-1671.

(53) Brodin, J. D.; Medina-Morales, A.; Ni, T.; Salgado, E. N.; Ambroggio, X. I.; Tezcan, F. A., *J. Am. Chem. Soc.* **2010,** *132*, 8610-8617.

(54) McCall, K. A.; Fierke, C. A., *Anal. Biochem.* **2000,** *284* (2), 307-315.

(55) Otwinowski, Z.; Minor, W., n *Methods in Enzymology*, Academic Press: 1997; Vol. 276, pp 307-326.

(56) Collaborative, *Acta Cryst. D* **1994,** *50* (5), 760-763.

(57) Vagin, A.; Teplyakov, A., *Acta Cryst. D* **2010,** *66* (1), 22-25.

(58) Murshudov, G. N.; A.A.Vagin; E.J.Dodson, *Acta Cryst.* **1997,** *D53*, 240-255.

(59) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., *Acta Cryst.* **2010,** *D66*, 486-501.

Chapter 4. Mechanistic Crossover between Serine- and Metallo-β-lactamases by Rational and Directed Evolution

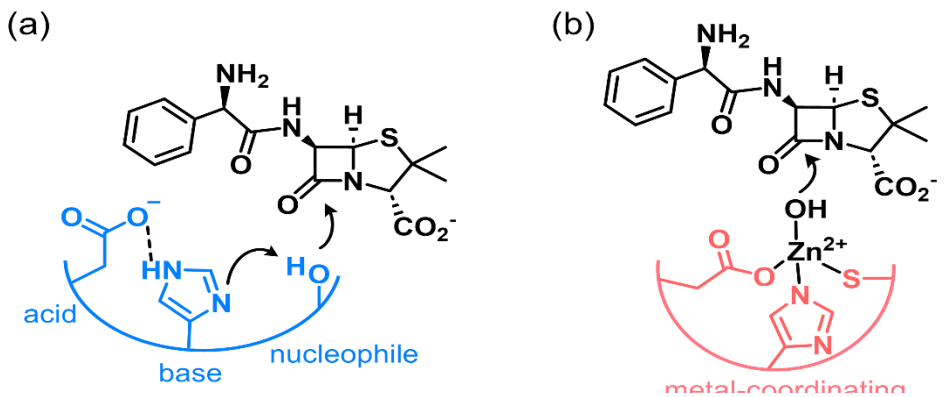
4.1 Introduction

Enzyme sequence space and energy landscape are likely to be determined by the orchestration of chemical pressures exerted into biological systems. Human-made or geographic events, such as the introduction of pesticides[1] and antibiotics[2] or temperature changes resulted in divergent evolution for altered substrate specificity from ancestor genes or the subsequent adaptation of enzyme kinetics, respectively.[3, 4]

Inorganic cofactors are likely to also operate as essential chemical pressures in enzyme adaptation. The concentration of each metal element at specific oxidation varies significantly depending on the local environment and atmospheric compositions. Metal abundance in volcanic or deep oceanic regions differs from the physiological conditions that most of the cultivatable and sequenced organisms have discovered.[5] The relative concentrations of the metal ions before and after the great oxidation event (GOE) also have been drastically fluctuated.[6-9] Consequently, the number of unique metal-binding sequence motifs have also altered accordingly. However, sequence iteration due to the appearance or disappearance of specific metal elements has not been demonstrated directly to date.

One of enzymes related to antibiotic resistance, β-lactamase, has been extensively explored as a model system to monitor the dynamic landscape of enzyme evolution.[10-12] In addition, artificial β-lactamases can be designed by introducing the key components of the active site.[13] It is also intriguing in that β-lactamases can be grouped into two entirely subclasses, metallo-β-lactamases (MβLs) and serine-β-lactamases (SβLs), where the latter can be further divided into three subclasses depending on their sequences and substrate scopes. MβL and SβL show entirely dissimilar overall topology, sequence, active site structure, and chemical mechanism, serving as the representative example for convergent evolution. Most importantly, MβL

requires metal ions for the native activities in the activation of a water molecule and the subsequent formation of a nucleophile to react with β-lactams (Scheme. 4.1b), whereas SβL is independent from the inorganic cofactor. Instead, a catalytic triad, composed of a serine, glutamate, and histidine, forms a nucleophile for the hydrolysis (Scheme. 4.1a). Therefore, their discrete mechanism suggests that they might have emerged and evolved in the discrete chemical environments, particularly in metal concentrations.



Scheme 4.1. The proposed reaction mechanisms of (a) class A type serine-β-lactamase TEM-1 and (b) class B2 type metallo-β-lactamase CphA.

Herein, I have demonstrated that metal elements can be operated as a direct and specific chemical cue for enzyme evolution. I conducted the directed evolution of two types of β-lactamases in the opposite conditions that they might have evolved, MβL in the absence of Zn ions and SβL in the presence of Zn ions. Herein, I employed an artificial MβL (^{I96}AB5) and a natural SβL (TEM-1) and observed whether they can employ the swapped reaction mechanism under the altered cell-growth conditions. The biochemical, kinetic, and structural characterizations of the evolved mutants under discrete conditions explicitly demonstrate that they indeed adapted to the altered chemical conditions, showing crossover between MβL and SβL regardless of their initial sequence and structural dissimilarity.

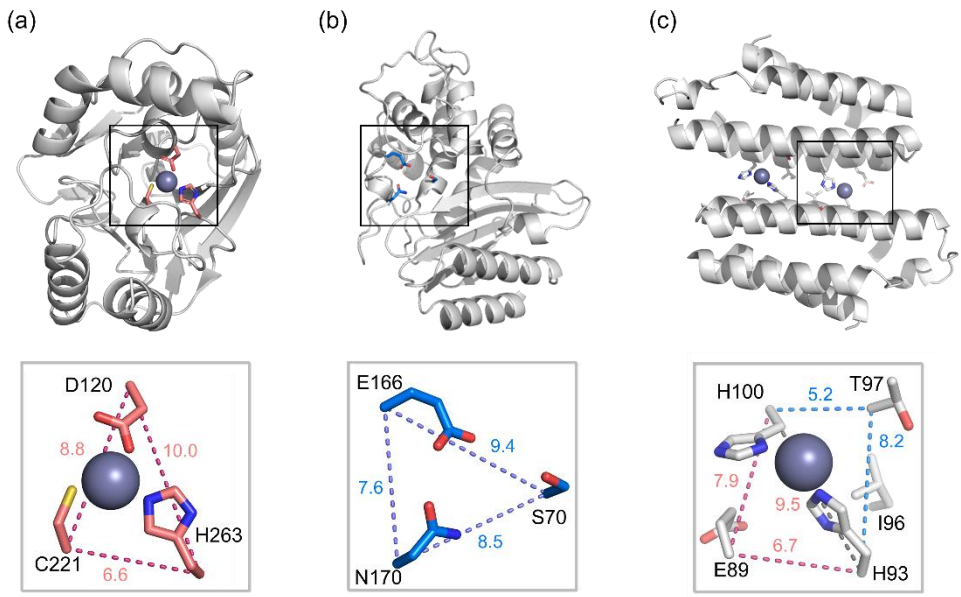


Figure 4.1. X-ray structures of (a) CphA (PDB 1X8G), (b) TEM-1 (PDB 1BTL), and (c) ^{I96}AB5 (PDB 6LDG).

4.2 Results and discussion

4.2.1 Switching from MβL to SβL.

For laboratory evolution under newly introduced chemical environments,

a less specifically evolved primordial enzyme could be advantageous as an initial template because it would exhibit high evolvability.[14] In this regard, I employed an artificial β-lactamase (^{I96}AB5) that I have designed previously (Figure 4.1c); two histidine residues (H93/H100) constitutes a coordinatively unsaturated Zn-binding site within a structurally and functionally innocent α-helical protein scaffold, and an adjacent residue was optimized for β-lactamase (^{I96}AB5). The artificial MβL is catalytically active in the presence of Zn ion, indicating that the Zn-coordinated water molecule operates as a nucleophile for the reaction with β-lactams.

Then, I first constructed a single-site mutant library (Figure 4.1a) by randomizing the amino acid sequence of the T97 position using NNK codon (T97X), as described previously.[15] Because T97 residue is located right next to the C96I position, which is located in the vicinity of two Zn-binding active sites, showing the distance between two histidine residues (H93 and H100) and Zn-bound water molecule of 2.6–4.5 Å. I surmised that T97 is a potential hotspot for sequence optimization. Then, I applied whole-cell activity assays by adding ampicillin as a substrate and measuring the cell-growth rates; only the cells containing the beneficial mutation would survive and grow rapidly. Of note, I conducted the whole-cell activity-based assays in M9 media, where Zn concentration was less than 10 nM, and it is considerably lower than the condition used for the evolution of artificial metallo-β-lactamases, Zn-supplemented LB media with 50 μM Zn ions at the final concentration. Thus, a Zn-deficient condition was introduced for laboratory evolution. The fast-growing cells were sorted and sequenced, yielding T97S mutation. Intriguingly, the simulated structure of ^{I96}AB5+T97S mutant suggests that it may form a putative catalytic triad along with S97, H93 and H100 residues in the absence of Zn ion, implying that the active site may resembles those of SβLs.

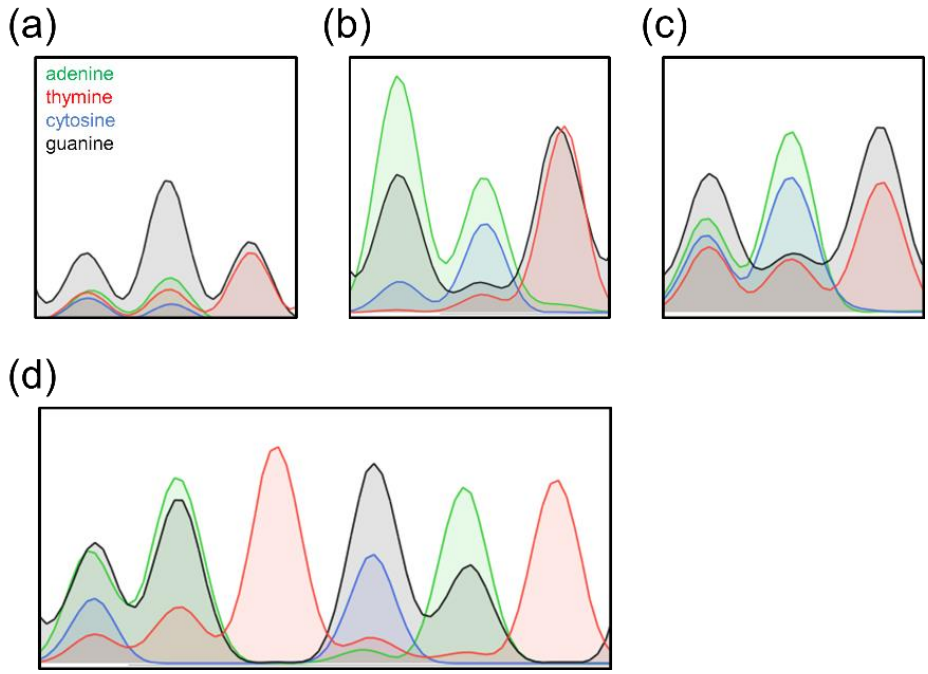


Figure 4.2. Chromatograms of mutant libraries screened for the evolution of SBL variants. (a) T97X, (b) H93X, (c) H100X, and (d) A38X/D39X.

To validate the single mutation effect, I measured the β-lactamase activity of AB5/C96I/T97S mutant with ampicillin under *in vitro* condition, using the purified proteins (Figure 4.3a). The resulting mutant indeed exhibited detectable amounts of an increase in β-lactamase activity in the absence of Zn ion relative to the parent protein (^{I96}AB5) while its catalytic activity is considerably lower in the presence of Zn ions, indicating that the single mutation has switched from MβL into SβL.

To validate that pre-existing H93 and H100 contribute to the catalysis of the readout from the first round, the interacting residues around T97 were analyzed using X-ray structure of SBL-R0 (Figure 4.3a). Interestingly, Q71 residue can be positioned very close to T97 in a distance of 2.5–3.5 Å. I hypothesized that Q71 residue supports the activation of serine on 97th position of SBL-R1 and other readouts from second round to fourth round. I conducted Q71A mutation, and the cell growth rate of SBL-R1+Q71A mutant

was significantly decreased and recovered with Q71H mutation, which possesses the positively charged histidine (Figure 4.3b). This decrease in growth rate by Q71A mutation highly suggests that the SBL variants contain the putative active site that comprises serine and glutamine residues, which probably plays an important role during the hydrolysis reaction, indeed. However, glutamine-containing catalytic moiety of β-lactamases has not been reported yet. I hypothesized that the serine-glutamine dyad can be a *de novo* catalytic motif for SβLs.

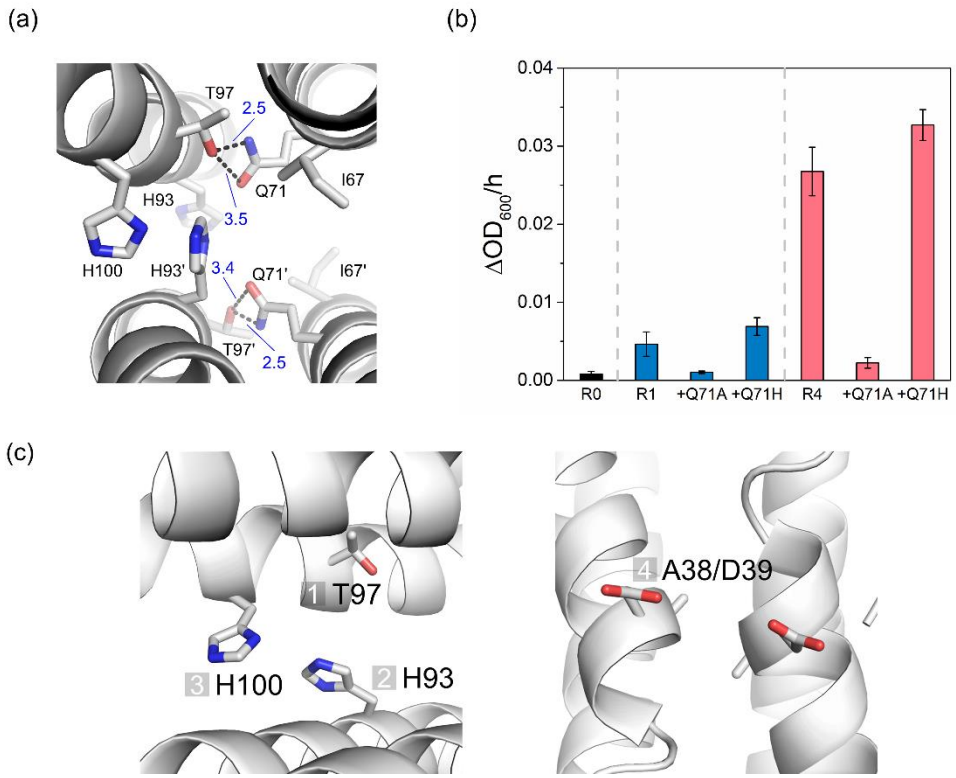


Figure 4.3. Determination of active site conformation via Q71A and Q71 mutation.

To optimize the active site conformation of SBL variants, pre-existing histidine residues (H93 and H100) were iteratively targeted for the evolution (Figure 4.3). I further proceeded into the second round of evolution by randomizing H93 position, which originally constituted the Zn-binding site. The screening of the mutant library acquired H93I mutation. Because

isoleucine is incapable of forming coordinate covalent bonding with Zn ion, the result is consistent with the development of SβL. I also iteratively optimized the H100 position, another Zn-binding residue, resulting in H93I/H100P mutant (Figure 4.4b and Table 4.1)

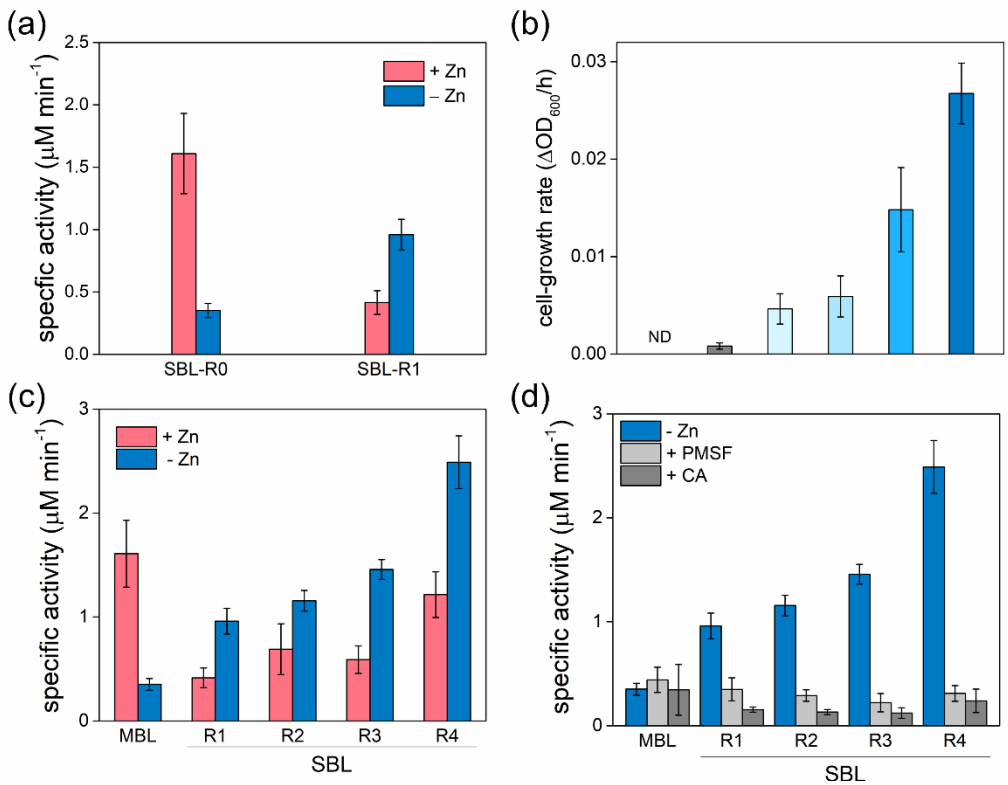


Figure 4.4. Characterization of screened readouts by measuring (a) specific activities of SBL-R0 and -R1 for ampicillin in the presence/absence of Zn^{2+} ions. (b) The cell growth rates under *in vivo* condition. (c) Specific activities for ampicillin hydrolysis in the presence/absence of Zn ions under in vitro condition and (d) specific activities after the addition of SβL inhibitor molecules; PMSF (phenylmethylsulfonyl fluoride) and CA (clavulanic acid).

Variant	Round	Mutation list
SBL-R0	0	-
SBL-R1	1	T97S
SBL-R2	2	T97S/H93I
SBL-R3	3	T97S/H93I/H100P
SBL-R4	4	T97S/H93I.H100P/A38G/D39N

Table 4.1. Mutation list by rounds of evolution for *de novo* serine-β-lactamases.

Interestingly, mutated residues on the 93rd and 100th positions were isoleucine and proline, which cannot contribute to the catalysis of serine-based hydrolases. This result indeed supports the hypothesis that the serine-glutamine dyad can serve as a catalytic motif for β-lactamases. This was unexpected that two histidine residues (H93 and H100) and a glutamate (E89) residue for the pre-existing active site candidate were optimized by non-relevant residues by iterative directed evolution. For SβLs, Serine should be activated by neighboring residues like hydrogen bonding partners and molecular base, respectively. On the other hand, optimized residues like isoleucine and proline cannot contribute to create a catalytic motif of serine-dependent β-lactamases such as S-X-X-K or S-D-N frequently discovered in TEM like SβLs.[16] This result highly suggested that a new potential active site motif with the serine-glutamine dyad can maintain and/or enhance the *de novo* catalytic activity.

For more efficient directed evolution, I randomized two additional residues, A38 and D39, which might be hotspots owing to the proximity to the rotational axis as reported previously[15]; although these residues are located remotely from the active sites, they are located at the C_2 rotational

axis, which possibly governs overall protein symmetry, structure, and possibly function, as reported previously. Indeed, A38G/D39N mutations were obtained as the positive hit. The catalytic activity measured under *in vitro* conditions demonstrated that the mutant from the four rounds of laboratory evolution (AB5/C96I/T97S/H93I/H100P/A38G/D39N) indeed accelerates the β-lactamase activities with ampicillin (Figure 4.4c). These results indicate that MβLs is indeed successfully transformed into SβLs, regardless of no similarity in protein sequence and structure between two classes of β-lactamases in nature.

To further verify the components of catalytic motifs, I measured the activities of mutants after treating with inhibitors, such as phenylmethylsulfonyl fluoride (PMSF) and clavulanic acid (CA), which form a selective covalent bond with a catalytic serine in hydrolases and/or serine-β-lactamases. The pre-incubation with those two inhibitors led to a drastic decrease in the specific activity with ampicillin, indicating that the serine residue might have become inactivated. In contrast, MβL, such as AB5/C96I, shows no alteration in specific activity upon the addition of PMSF and CA. These results indicate that these mutants have evolved to SβLs, although they have no resemblance to the natural SβLs in their overall structure and sequence.

4.2.2 Biochemical characterization of the evolved SβL.

The parent protein ^{I96}AB5 forms a homo-tetramer to serve as an artificial metallo-β-lactamase coordinating Zn^{2+} ions at two distinct Zn-binding sites; one is the structural Zn-binding site at the core of the tetramer, and the other is the catalytic Zn-binding site located at the periphery region of the tetramer. In particular, catalytic Zn-binding sites help to stabilize the tetramer as well as carry out the active sites. The pre-existing catalytic Zn-binding motif comprising H93 and H100 was lost with the screening for the evolution of SBL variants. Thus, I firstly determined the oligomeric state of the SBL-R4

variant, the readout from 4th round of iterative directed evolution, by size exclusion chromatography to determine the catalytically active oligomerization state. Whereas the parent protein (^{I96}AB5) was isolated as a tetramer,[15] the quintuple mutant was primarily isolated as a dimer, indicating that iterative mutations, possibly including the loss of Zn-binding site, greatly altered the stability of tetramer. Although some aggregate species was also detected, the fraction showed much lower β-lactamase activity than that of the dimers, indicating that the catalytically relevant state of the quintuple mutant is a dimer (Figure 4.5).

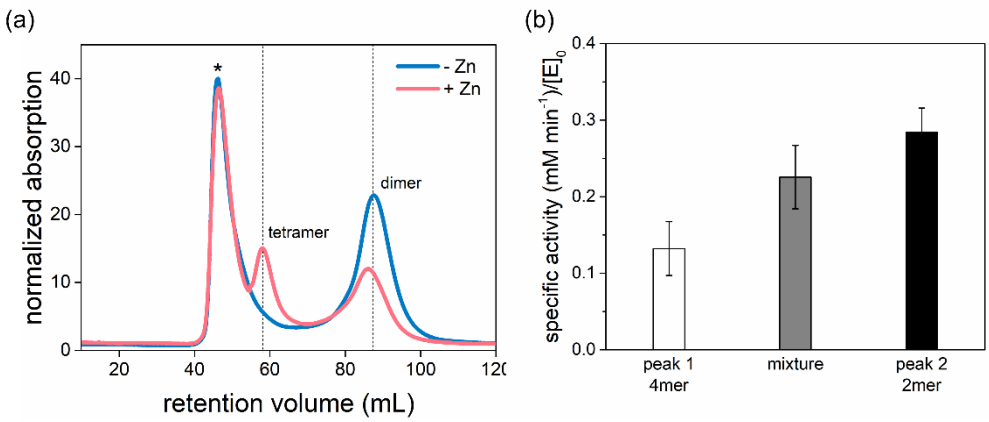


Figure 4.5. Determination of oligomerization state and catalytically active species. (a) Size exclusion chromatograms of apo-SBL-R4 (blue) and Zn added SBL-R4 (red) and (b) specific activity of peak 1 (eluted at 58 mL, tetramer) and peak 2 (eluted at 86-87 mL, dimer) for ampicillin hydrolysis. *higher oligomeric species and/or aggregates were shown with an asterisk.

To demonstrate whether the acquired mutations indeed facilitated the reactions via the altered chemical mechanisms, I measured specific activities of the screened mutants (Figure 4.4c) under the *in vitro* condition. The mutants from the screening with Zn-deficient M9 growth media show elevated Zn-independent β-lactamase activities, indicating that they have evolved to SβLs by switching the components of catalytic moieties.

Additionally, to validate the activated serine at the active site of evolved proteins plays a catalytic role, I measured the activities of variants with inhibiting agents including zinc ion, general serine hydrolase inhibitor phenylmethylsulfonyl fluoride (PMSF), and clavulanic acid (CA), which is known as the inhibitor of the serine-β-lactamases. The variants exhibited 63.6-87.5% and 83.9-90.4% decrease with PMSF and CA, respectively. (Figure 4.4d) This result highly suggests that the catalysis toward ampicillin occurred at the expected active site comprising T97 and Q71. Although the inhibitions by PMSF and CA were measured, the inhibitor-bound CA-SBL (or PMSF-SBL) complex was not detected yet. Because PMSF and CA covalently inhibit the activated serine residue, the enzyme that reacted with the inhibitors gains molecular mass permanently. To examine the increased mass of inhibitor-treated variants, mass analysis using high-resolution tandem mass-spectrometry will be carried out.

For more accurate characterizations, I measured the Michaelis-Menten kinetic parameters of the evolved mutants. The final mutant SBL-R4 exhibited the comparable kinetic parameters to those of Zn complexed ^{I96}AB5 which I reported previously.[15] This variant displayed up to 24.5-fold increased catalytic efficiency (k_{cat}/K_M) to k_2 constant of parent protein apo-^{I96}AB5 which did not exhibit non-saturated behavior. Interestingly, H93I mutation from 2nd round caused 1.7-fold increase of catalytic rate (k_{cat}) and this was almost conserved to 4th round, albeit it was slightly perturbed at 3rd round exhibiting 30% decreased k_{cat} value. Despite the increased catalytic rate, SBL-

R2 variant displayed merely decreased catalytic efficiency (k_{cat}/K_M) due to an increased K_M value. However, iterative screening for the 3rd and 4th rounds yielded variants which possess highly decreased K_M values. This led SBL-R3 and -R4 to display enhanced catalytic efficiencies in comparison to SBL-R1 and SBL-R2 (Figure 4.6 and Table 4.2).

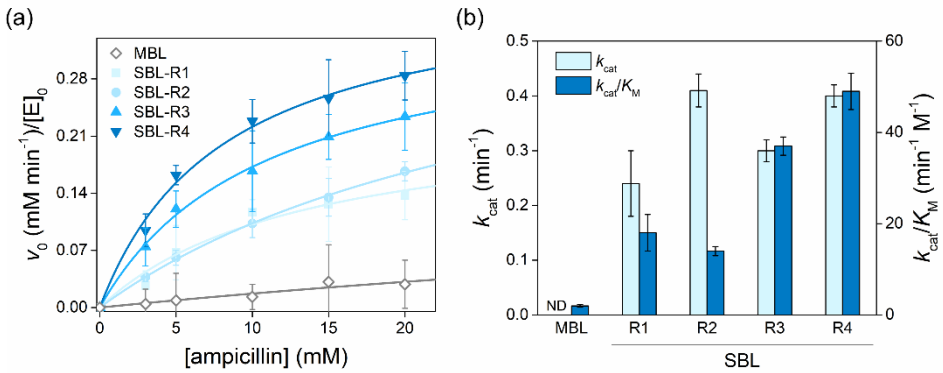


Figure 4.6. Steady-state kinetic analysis of SBL variants. (a) The Michaelis-Menten plots and (b) kinetic parameters for the SBL variants.

Variant	Zn content (equiv)	k_{cat} (min^{-1})	K_M (mM)	k_{cat}/K_M (min^{-1} M^{-1})
SBL-R0	2	0.43(2)	8(1)	52(3)
	0	nd*	nd*	2.0(3)†
SBL-R1	0	0.24(6)	14(6)	18(4)
SBL-R2	0	0.41(3)	30(3)	14(1)
SBL-R3	0	0.30(2)	8(1)	37(2)
SBL-R4	0	0.40(2)	8(1)	49(4)

Table 4.2. Steady-state kinetic parameters of SBL variants by screening rounds. *nd: not determined, †second order rate constant (k_2) was determined instead of catalytic efficiency because of not saturation.

4.2.3 Switching from SβL to MβL.

I adopted TEM1 as the representative SβL to demonstrate that metal-dependent evolution is operative in the opposite direction and even with natural enzymes. TEM1 β-lactamase possesses a serine (S70), a glutamate (E166), and an asparagine (N170) as a catalytic triad (Figure 4.1b), and the wild-type TEM1 shows exceptionally high catalytic activities with ampicillin regardless of the presence and absence of Zn ions (Figure 4.6b). When the pre-exiting catalytic serine was deleted by S70G mutation, TEM-1 S70G mutant exhibited 8,700-fold decreased catalytic efficiency.[17] I designed the metallo-TEM-1 mutant, MT0 by incorporating Zn-ligating histidine residues at the pre-existing active site comprised of S70 to carry out a catalytic role, E166 to serve as a base, and N170 to engage in substrate anchoring via S70H, E166H, and N170H mutations. Design for 3-His catalytic triad was based on the atomic distance between C_α atoms of native Zn-dependent hydrolases.[18] This result implies that the incorporated Zn-binding site was proper to coordinate Zn ion. The active site of TEM1 and other SβLs, such as KPC and CTX-M, can be overlaid with those of MβL, such as CphA and Sfh-I, although they show no similarity in sequence and overall structure (Figure 4.7a).

I measured the hydrolytic activity of the MT0 mutant using ampicillin as a substrate under the *in vivo* condition by measuring the cell growth rate for 5 hr since the addition of ampicillin MT0 exhibited 25% of the cell-growth rate relative to that of the wild-type TEM-1 only in the presence of Zn^{2+} ion, but not with Zn^{2+}-free media. These results indicate that even such minimal mutations converted the natural SβL to show MβL-like reactivity. (Figure 4.7c) Under the *in vitro* condition, moreover, MT0 mutant did not exhibit the hydrolytic activity toward ampicillin in the absence of Zn^{2+} but displayed a basal level of the activity in the presence of Zn ions (Figure 4.7d).

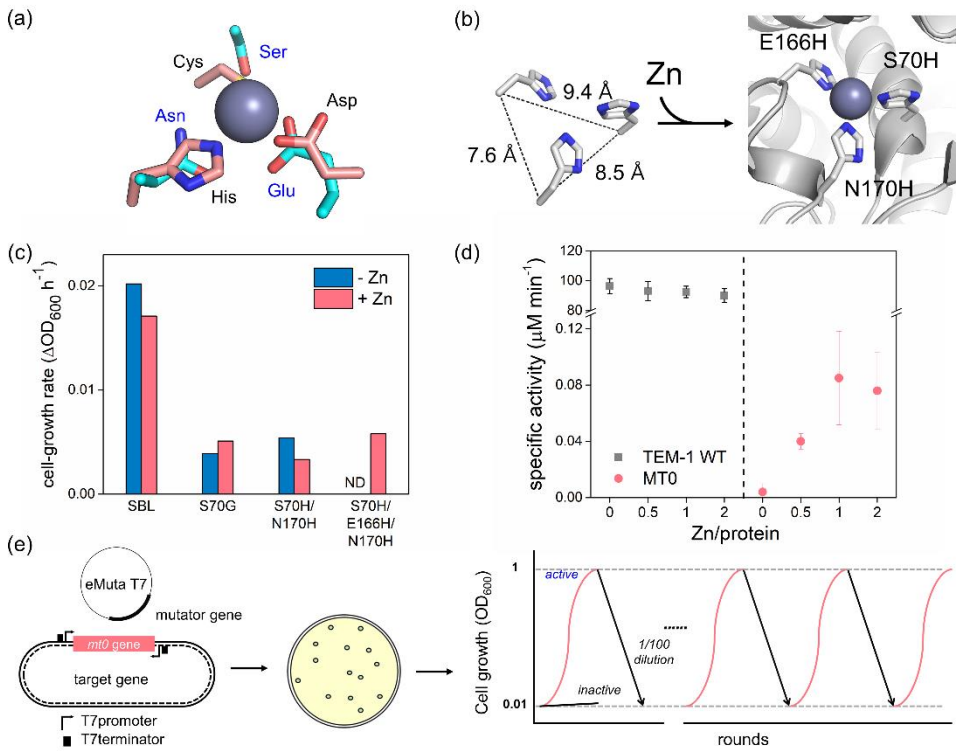


Figure 4.7. The crossover of native SβL to *de novo* MβL. (a) overlaid structure of catalytic triad of SβL and MβL (cyan sticks for TME-1 and pink for CphA) (b) Design for metallo-TEM-1 variant 0 (MT0). (c) Growth rates of TEM-1 and TEM-1 S70G, S70H/N170H and S70H/E166H/N170H mutants by presence/absence of Zn^{2+}. (d) Specific activity of WT TEM-1 and MT0 for ampicillin hydrolysis by Zn equivalence. (e) The scheme for the evolution of MT0 via continuous evolution strategy using eMuta T7 system.

Then, I started the directed evolution of MT0 using Zn-supplemented cell-growth media. Herein, I apply a recently developed continuous evolution method, using eMuta T7[19] for more extensive and efficient sequence optimizations. The MT0 encoded gene and mutator gene were co-transformed to BL21 (DE3) competent cells and grown in LB media, containing isopropyl β-D-1-thiogalactopyranoside (IPTG) and arabinose for the orthogonal inductions of each promotor, respectively. The concentration of ampicillin was increased every three cycles, starting from 0 mg/L ampicillin to 70 mg/L

in 37 cycles (Figure 4.7a).

4.2.4 Characterization of continuously evolved MT variants

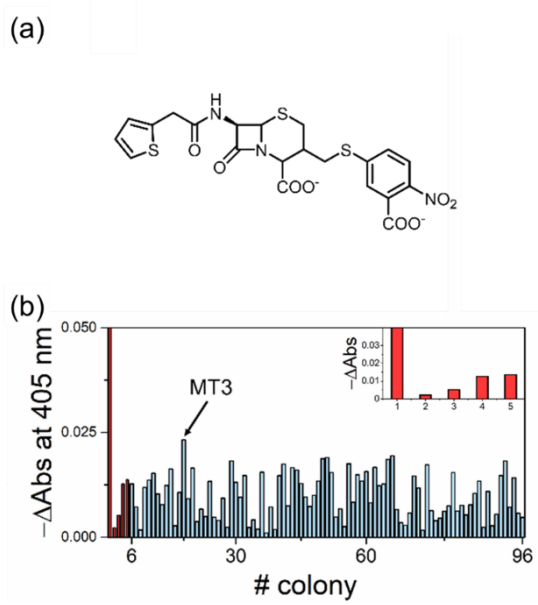


Figure 4.8. Lysate assay for hydrolytic activity toward (a) CENTA, a chromogenic surrogate of β-lactam substrate and (b) assay result readouts from 60 mg L^{-1} ampicillin condition. In the inset, –ΔAbs values for control colonies were displayed (1. WT TEM-1; 2. MT0; 3. MT1; 4. MT2.1; 5. MT2.2).

To investigate the activities of readouts from continuous evolution, lysate assay was carried out using *E. coli* cells harboring these proteins from screening cultures at the ampicillin concentrations of 25 mg/L, 45 mg/L, 60 mg/L, and 70 mg/L. To validate the readouts obtained from lysate assays, top ranked variants were isolated for *in vitro* characterization (Figure 4.8). Indeed, they show elevated whole-cell activities in the presence of Zn ions in growth media, but not in the absence of Zn ions, indicating that they indeed evolved to MβLs.

By applying various and gradually increasing concentrations of ampicillin, I randomized and selected the protein sequences that are beneficial

for cell growth in the presence of ampicillin. By 37 rounds of dilutions, I identified the following mutants, MT1, MT2.1, MT2.2, MT3, MT4.1, and MT4.2 (Figure 4.9a and Table 4.3). I measured specific activity of the isolated ones for ampicillin hydrolysis, and every variant excepting MT0 was remarkably enhanced. However, the increase between variants selected from 25, 45, and 60 mg/L ampicillin condition were trivial, which remains still elusive. Additionally, hydrolysis in the presence of clavulanic acid or EDTA was also measured to validate that the hydrolytic activity of MT variants resulted from the Zn center incorporated. Loss-of-function was observed when EDTA chelated Zn ions at the metal binding center while the activity was maintained with or without excess CA, which is a representative inhibitor for SβLs. This result strongly implies that the hydrolytic activity of MT variants originated from the incorporated Zn-center.

For accurate analysis of the kinetic parameter, I additionally measured steady-state activities of each variant. MT2.1 and MT2.2 variants exhibited similar specific activities, but the saturation behavior was not observed with MT2.1 variant, while other variants were almost saturated at 20 mM ampicillin condition. There were no significant differences in catalytic rates (k_{cat}) and Michaelis constant (K_M) between MT1 and MT2.2 variants, although their chemical pressures were quite different. MT3, the best readout, exhibited up to 20% increased k_{cat} and 11% increased K_M values at the same time, leading to a trivial increase in catalytic efficiency (k_{cat}/K_M) (Figure. 4.9c, d and Table4.4).

Table 4.3. Mutation list from continuous evolution.

Variants	mutations	mut. #	G>A %	Retain %	net mut. #
MT0	-	-	-	-	-
MT1	C>T (6) I56I/R83C/V108V/T133I/G196G/A248V G>A (4) D50N/V159I/A232T/R275Q	10	40	30	7
MT2.1	C>T (9) R61C/H96Y/T140I/T160I/A202V/ P219S/S256F/I261I/A282V G>A (5) E58E/D131N/G196S/G236R/G250G	14	35.7	14.2	12
MT2.2	C>T (11) Y46Y/L51L/S106L/C123C/T128I/A172V/ G196C/A213V/P251L/A268I/A282A G>A (4) V84I/M182I/G238D/A278T	15	26.7	20	12
MT3	C>T (10) G78G/T118I/A125V/P145S/A172V/ T188I/S203F/P219L/T263M/A268V G>A (5) V84I/Q99Q/D176N/G244S/G281D	15	33.3	13.3	13
Mt4.1	C>T (11) P27L/R65C/S82F/T140I/K146K/A172T/ H153H/A224V/I258I/A278V/S283L G>A (6) S53N/R94H/T189T/T195I/G241S/K286K	17	35.3	23.5	13
MT4.2	C>T (12) A36V/P62P/A86V/L91L/H112Y/T140I/D176D/ A202V/P219S/I245I/S256S/A282A G>A (5) V84I/M182I/A237T/T263T/A278T	17	29.4	41.2	10

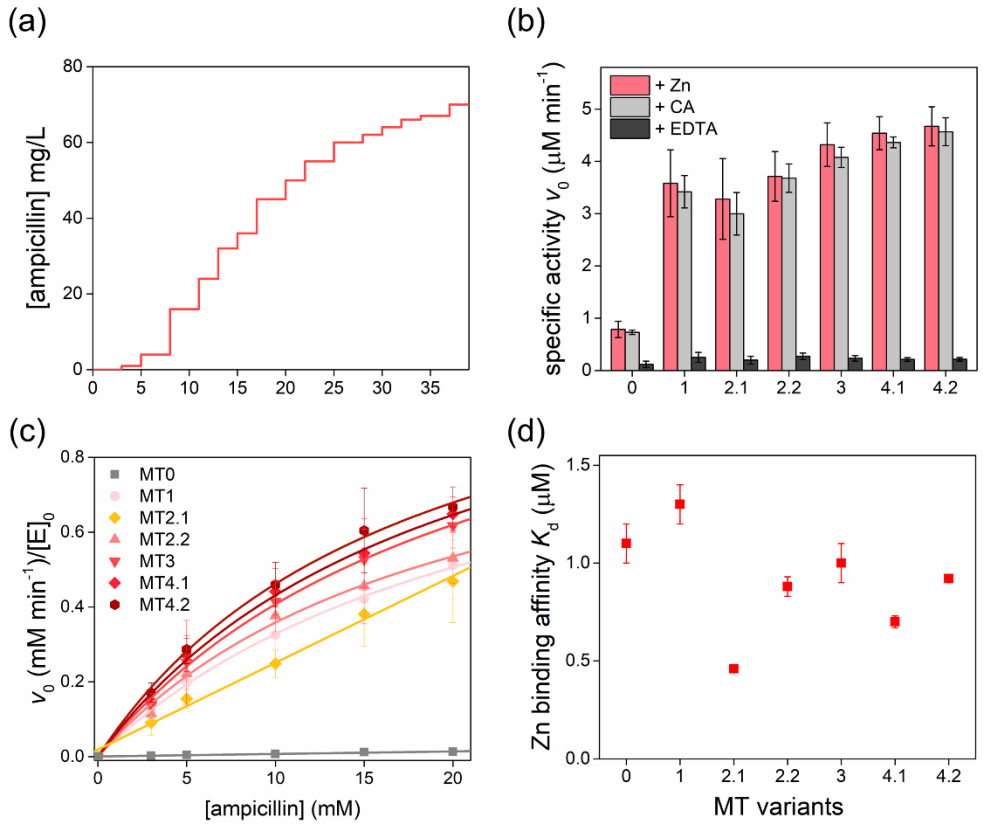


Figure 4.9. Kinetic analysis of MT variants. (a) ampicillin concentrations by cycles during continuous evolution. (b) specific activities of MT variants under *in vitro* condition in the presence of 1 equiv Zn^{2+} ion, clavulanic acid (CA) and ethylenediaminetetraacetic acid (EDTA). (c) Plots for the steady-state kinetics of MT variants. (d) Dissociation constants of MT variants with Zn^{2+} ion.

Variants	k_{cat} (min^{-1})	K_M (mM)	k_{cat}/K_M (min^{-1} M^{-1})
MT0	nd*	nd*	0.67(6)§
MT1	1.1(1)	24(2)	47(2)
MT2.1	nd*	nd*	23(1)§
MT2.2	1.1(1)	19(4)	54(5)
MT3	1.3(1)	21(3)	59(3)
MT4.1	1.3(1)	19(3)	65(4)
MT4.2	1.3(1)	17(2)	72(4)

Table 4.4. Steady-state kinetic parameters of MT variants. *nd: not determined. §instead of k_{cat}/K_M, second order rate constant (k_2) was determined due to the non-saturation behavior.

Then, I performed the competitive titration assay to determine Zn-binding affinities of MT variants (Table 4.5 and Figure 4.10). K_d values of MT variants were highly consistent with those of catalytic Zn site reported previously.[15] K_d values were relatively not perturbed by variants. Although the X-ray structure was not revealed yet, the locations of the mutation by variant can be tracked using the structure of the wildtype TEM-1. Interestingly, most mutations were located apart from the active site, and this might be the reason why K_d values were similar due to this.

Variants	K_d **(μM)**
MT0	1.1(1)
MT1	1.3(1)
MT2.1	0.46(2)
MT2.2	0.88(5)
MT3	1.0(1)
MT4.1	0.70(3)
MT4.2	0.92(6)

Table 4.5. Zn binding affinities of MT variants.

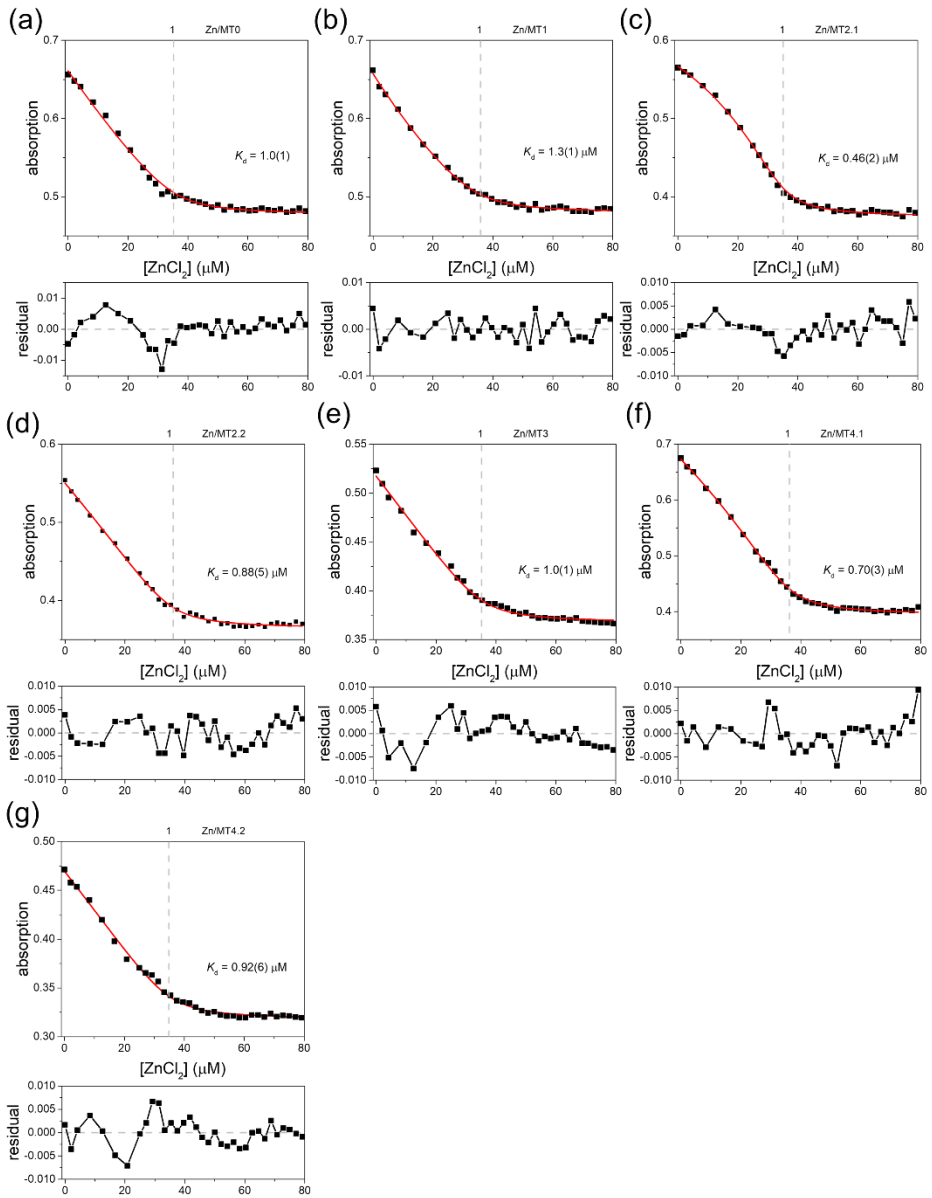


Figure 4.10. Plots for competitive titration assay to determine Zn-binding affinities of MT variants (a) MT0, (b) MT1, (c) MT2.1, (d) MT2.2, (e) MT3, (f) MT4.1, and (g) MT4.2.

4.3 Conclusion

I demonstrated that the interconversions between SβL to MβL were enabled by tuning the concentration of Zn^{2+} ion, which works as a cofactor for the hydrolysis of β-lactam antibiotics. Classification of β-lactamases is mainly dependent upon the origin of catalytic power, activated serine, and the Lewis acid Zn–OH, respectively. Interestingly, those two subclasses were non-relevant in the structures and the reaction mechanisms although SβLs and MβLs were engaged in similar chemical transformations, which means that they have been evolved convergently. Nevertheless, I hypothesized that the crossover between SβLs and MβLs was still possible with consideration that respective catalytic units were exchangeable with each other. Both activated serine motifs and Zn–OH motifs were known to carry out similar roles as nucleophiles to attack carbonyl carbon at the β-lactam ring. In class B2 β-lactamases like CphA, the proposed mechanism depicts that the Zn center anchors substrate only, and active catalysis occurs by the water molecule activated by an adjacent base such as aspartate residue. However, mechanisms that Zn–OH directly conduct nucleophilic attack to carbonyl center of β-lactam was also proposed. Albeit this controversy on the catalytic basis of Zn-mediated β-lactamases, the formation of hydroxide nucleophiles and their roles in the catalysis resembled obviously.

Directed evolution by altering the concentration of metal ion cofactor successfully converted ^{C96I}AB5 protein, Zn-mediated metallo-β-lactamase, into SBL-R4, *de novo* serine-β-lactamase. This result suggests that evolution in nature might also be highly dependent upon environmental conditions like concentrations of metal ions. The *in vitro* characterizations of the screening outputs demonstrate that these approaches have yielded a unique example of conversion of MβLs to SβLs and *vice versa.*

Additionally, I also targeted native serine- β-lactamases to investigate

whether reverse conversion of SβLs to MβLs is also possible. Continuous evolution method led to the evolution of metallo-TEM (MT) variants. This allowed MT0 to be evolved into MT4.1 and MT4.2 variants only after 456 hr of continuous screening. Although the enhanced activity of MT4.2 was trivial in comparison to native TEM-1, MT4.2 displayed ca. 6-fold increased hydrolytic activity toward ampicillin compared to MT0. This is still understandable because of the exchanged reaction mechanism with Zn–OH. *In vitro* characterizations, including analyses for determination of the Zn binding affinity and hydrolytic activity toward ampicillin, strongly suggest that the hydrolysis reaction occurred at designed Zn-coordinated active sites.

Finally, I successfully conducted the mechanistic crossover of β-lactamases, which have convergently evolved from the distinct origins, implies that structurally and mechanistically less-relevant enzymes in nature can also be interconverted along the environmental factors. Moreover, I also suggest that tuning cofactor concentration could be an alternative route for conventional directed evolution approaches adapting substrate concentration only as a chemical pressure. However, additional evidences for serine activation are required to validate the origin of catalytic power. To supplement the elucidation of the activities from SBL variant thoroughly, additional experiments like tandem mass spectrometry, X-ray crystallization, and the rational sequence alterations are ongoing.

4.4 Experimental section

template	Primer	Sequences	Calculated T_m (°C)	Annealing Temp (°C)
C96I	T97X	5'-G CAT CTG AAA ATT **NNK** TGC AAC CAT TGC CAC C-3' 5'-G GTG GCA ATG GTT GCA **MNN** AAT TTT CAG ATG C-3'	60.5–64.4	59.4- (0.2/cycle) 25 cycle
C96I/ T97S	H93X	5'-CAG GAA GCT GCA GAG **NNK** CTG AAA ATT AGT TGC-3' 5'-GCA ACT AAT TTT CAG **MNN** CTC TGC AGC TTC CTG-3'	61.9–65.6	60.6- (0.2/cycle) 25 cycle
C96I/ T97S/ H93I	H100X	5'-A ATT AGT TGC AAC **NNK** TGC CAC CAG AAG TAT C-3' 5'-G ATA CTT CTG GTG GCA MNN GTT GCA ACT AAT T-3'	59.3–63.1	58.1– (0.2/cycle) 25 cycle
C96I/ T97S/ H93I/ H100P	A38X D39X	5'- GCC GCA GCG **NDT NDT** GCG TGG AGC-3' 5'- GCT CCA CGC **AHN AHN** CGC TGC GGC-3'	62.5–69.3	64.3 – (0.3/cycle) 25 cycle
C96I/ T97S C96I/ T97S/ H93I/ H100P	Q71A	5'- TGG ATT CTG ATT GGT **GCG** ATT CAC GAC GCG-3' 5'- CGC GTC GTG AAT **CGC** ACC AAT CAG AAT CCA-3'	64.4	59.4, 57.4
	Q71H	5'- TGG ATT CTG ATT GGT **CAC** ATT CAC GAC GCG-3' 5'- CGC GTC GTG AAT **GTG** ACC AAT CAG AAT CCA-3'	63	58, 56

Table 4.6. Primer sequences for the evolution of SBL variants.

Construction of metal dependent evolution for *de novo* cyt cb_{562} derived SβLs. To prepare single mutant library for metal dependent evolution, site-saturation metagenesis was carried out using degenerated primers such as NNK and NDT codon (Table 4.4). After DpN I (Enzynomics) digestion for 1 hr PCR product was transformed to *E. coli* competent cells and grown on LB/agar plate containing 50 μg/mL kanamycin. Greater 100–430 colonies were collected and inoculated to 5 mL of LB media (Kan^+). After 8–10 hr incubation, Genes were extracted by miniprep kit (Cosmo Genetech) from grown cells and asked for sequencing using T7 promoter as a primer (Macrogen) and chromatograms were displayed in Figure 4.2.

Growth rate-based screening of SBL-R1 derived libraries. The plasmid mixtures were transformed to BL21 (DE3) *E. coli* competent cells containing the cytochrome c maturation cassette (ccm) and were grown overnight on LB/agar containing 50 μg/mL kanamycin and 30 μg/mL chloramphenicol. Over 100 colonies for each single- or double-site randomized library were inoculated in a 96-deep well microplate containing 800 μL of M9 minimal medium with the antibiotics described above. After overnight growth at 37 ℃ with constant shaking at 200 rpm, when the optical density at 600 nm (OD_{600}) of each culture reached ~1, the cultures were diluted 10-fold with 200 mL of M9 minimal medium containing 50 μg/mL kanamycin and 35 μg/L chloramphenicol. Increasing concentrations of ampicillin were added throughout the screening when OD_{600} reached approximately 0.1. Notably, after monitoring the cell-growth rates at 37 ºC for 6 h, the fastest-growing cells were selected for sequencing. At least 2–4 sets of a 96-well plate with the colonies of 93 mutants and 3 parent or template proteins were screened for each library.

Expression and purification for SBL variants. Protein expression and purification of the positive hits were carried out as described previously. Shortly, pET20b(+) plasmid for encoding the ab5 gene was transformed to

BL21 (DE3) competent cells. The cells were grown in LB medium with 150 rpm shaking at 37 ºC for 18 h. Cell pellets were harvested by centrifugation at 5000 rpm (4,715 g) at 7 ℃ for 10 min. After sonication of the cell pellets in 10 mM sodium phosphate (NaPi) pH 8.0 buffer with pulse on/off = 9/9 s for 40 min in an iced bath, HCl (25%) solution was added to the cell lysates up to pH 5.0. After centrifugation at 13,000 rpm (18,800g) at 7 ºC for 30 min, the supernatants were adjusted to pH 8 by adding 2 M NaOH solution. Then, the solutions were manually loaded onto a Q-sepharose column pre-equilibrated with 10 mM NaPi pH 8.0 buffer. By applying the step gradients of 0–1 M NaCl in NaPi buffer, red fractions due to the presence of heme cofactor were collected and concentrated using stirred cells (Amicon) with 10 kDa cutoff membranes. Then, the samples were loaded onto a HiTrap Q HP anion exchange column (Cytiva) pre-equilibrated with 10 mM NaPi pH 8.0 buffer at 4 ℃ using FPLC (ÄKTA pure). A linear gradient of 0–1 M NaCl was applied, and the fractions with $A_{415}/A_{280} \geq 4$ measured using a UV-Vis spectrophotometer (Agilent Cary 8454) were collected using and concentrated.

Then, the protein sample was loaded onto a HiLoad 16/600 Superdex 75 pg column (Cytiva) pre-equilibrated with 20 mM Tris/HCl pH 7.0 buffer with 150 mM NaCl. After elution, pure fractions were collected and concentrated. To prepare the metal-free, apo protein, 10-fold EDTA to the protein was treated for 1 h at 4 ºC, and the excess ethylenediaminetetraacetic acid (EDTA) was removed by 10DG desalting column (Biorad) chromatography. Metal content was measured by a colorimetric assay using 4-(2-pyridylazo) resorcinol (PAR) as described previously. The purified metal-free, apo-protein was concentrated up to 1.0 mM and stored at 80 ºC until further use. The protein concentration was determined by using $\varepsilon_{415\ nm} = 148{,}000\ cm^{-1}\ M^{-1}$.

Determination of oligomerization states of SBL-R4. oligomerization states of SBL-R4 variant was determined by size exclusion chromatography using FPLC (Äkta pure). 500 μL of 100 μM SBL-R4 with or without 2 equivalent $ZnCl_2$ was loaded to HiLoad 16/600 Superdex 75 pg column (Cytiva) pre-equilibrated with 20 mM tris/HCl pH 7.0 buffer containing 150 mM NaCl. SBL variants display distinct retention volume ranges; tetramer at 58–61 mL, dimer 76–82 mL and monomer at > 90 mL. Absorption at 415 nm and 280 nm was simultaneously monitored to identify the oligomerization state of eluted protein.

***In vitro* hydrolytic activity assay with ampicillin.** *In vitro* hydrolytic activities of Zn-free positive hits were measured by time-dependent HPLC analysis as reported previously.[13] The reaction was initiated by the addition of either ampicillin to apo protein (7 μM in 100 μL of reaction volume) in 100 mM MOPS pH 7.0 at 25 °C. The reaction mixture (2 μL) was injected onto the C18 column in HPLC (Agilent Infinity 1260) and eluted with a linear gradient, starting from 90% H_2O/10% CH_3CN to 10% H_2O/90% CH_3CN for 20 min. Trifluoroacetic acid (TFA) was added to the elution solvents, H_2O and CH_3CN, at 0.05% and 0.1% (v/v) as final concentrations, respectively. The substrate consumption rates were measured by monitoring the absorbance changes at 220 nm. The initial rates were determined from the reactions with various concentrations of the applied substrate. Michaelis–Menten parameters such as k_{cat} and k_{cat}/K_M were determined from iterative non-linear respectively, using Origin 2016 software (Figure 4.9c and Table 4.4).

To validate the hydrolytic activities of the readouts by rounds are originated from the activated serine residue like natural serine- β -lactamases, specific activity of each MT variant was determined after the addition of phenylmethylsulfonyl fluoride (PMSF) and clavulanic acid (CA) which are the inhibitors for general serine hydrolases and native serine-β-lactamases,

respectively. 10-fold PMSF and CA were respectively added to MT variants followed by overnight incubation at 4 ºC. Then, residual inhibitors were removed with 10DG desalting column. (Biorad) Using HPLC (Agilent Infinity 1260), 7 µM of inhibitor bound MT variants mixed with 20 mM ampicillin substrate in 100 mM MOPS pH 7.0 buffer were analyzed as described above (Figure 4.2). The inhibition mechanisms with PMSF and CA was displayed in Figure 4.11.

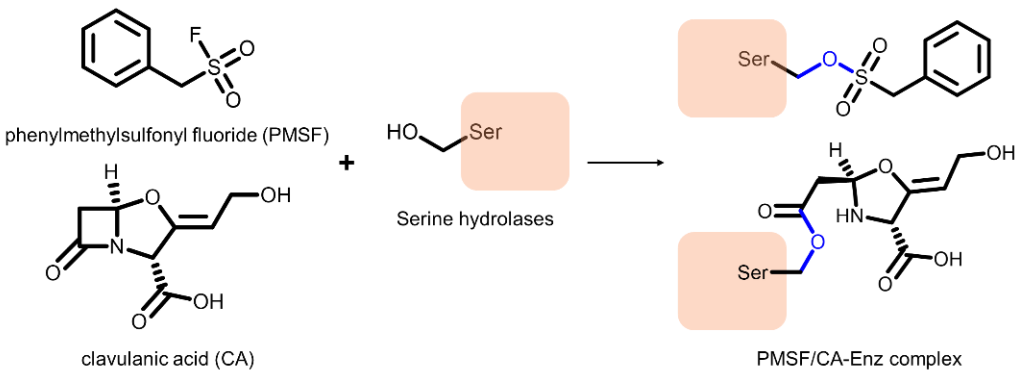


Figure 4.11. Covalent inhibition of serine-dependent hydrolase by PMSF and CA.

Primer	Sequences	Calculated T_m (ºC)	Annealing T (ºC)
IVA_priA	5'-CAT GGT ATA TCT CCT TCT TAA AGT TAA ACA AAA TTA TTT C-3'	58.3	55
IVA_priB	5'-TGA GAT CCG GCT GCT AAC AAA GCC-3'	59.1	55
IVA_priC	5'-G TTT AAC TTT AAG AAG GAG ATA TAC CAT G(vector)//AG TAT TCA GCA CTT CCG TGT CGC CC(TEM-1)-3	69.9	55 (for overlap)
IVA_priD	5'-CTT TGT TAG CAG CCG GAT CTC A(Vector)//CCA ATG CTT AAT CAG TGA GGC ACC TAT CTC AG(TEM-1)-3	72.2	55 (for overlap)
S70G	5'-GAA CCA ATG ATG **GGT** ACT TTT AAA GTT CTG C-3' 5'- G CAG AAC TTT AAA AGT ACC CA T CAT TGG TTC-3'	59.1	55.1, 54.1
S70H	5'-GAA CCA ATG ATG **CAC** ACT TTT AAA GTT CTG C-3' 5'- G CAG AAC TTT AAA AGT **GTG** CA T CAT TGG TTC -3'	59.1	55.1, 54.1
N170H	5'- GAA CCG GAG CTG **CAT** GAA GCC ATA CCA AAC-3' 5'- CGC CTT GAT CGT TGG **CAT** CCG GAG CTG CAT-3'	64.4	59.4, 57.4

E166H	5’- CGC CTT GAT CGT TGG **CAT** CCG GAG CTG CAT-3’ 5’- ATG CAG CTC CGG **ATG** CCA ACG ATC AAG GCG-3’	67.1	62.1, 60.1
Histag insert	5’-CAC CAC CAC CAT CAT CAT CTG GTT CCA CGT GGT AGC cac cca gaa acg ctg gtg aaa g-3’ 5’-CT ACC ACG TGG AAC CAG ATG ATG ATG GTG GTG GTG agc aaa aac agg aag gca aaa tgc c-3	56.7, 56.0 for overlap	55

Table 4.7. Primer sequences used for PCR to prepare TEM-1 derived mutants.

Construction of MT0 gene for the biochemical characterization and continuous evolution. pET28b(+) vector encoding wildtype TEM1 sequence was constructed via *in vivo* assembly[20] with empty pET28b(+) vector and amp^R selectable marker (wildtype TEM-1) of pET20b(+) or site-directed mutagenesis with designed primers and pET28b(+) vector harboring wildtype TEM-1. Using primers listed in table 4.5, PCR was performed to prepare MT0 which contains 3-His Zn-binding motif at original active site of TEM-1. After DpN I (Enzynomics) digestion at 37 ºC for 2 h, PCR product was transformed to DH5α *E. coli* competent cells and grown on LB/agar plate containing 50 mg L^{-1} kanamycin for overnight. Single colony was inoculated to 5 mL of LB media (kan^+) and grown for 8 h. Then, plasmid was extracted using miniprep kit (COSMO Genetech). Extracted plasmid was asked for sequencing. (Macrogen) For continuous evolution, wildtype TEM-1 gene in pGE158 vector was mutated to MT0 with site-directed mutagenesis as described above. Then, 6xHistag and thrombin cleavage sequence was additionally inserted for efficient purification. Sequences used for PCR was listed in Table 4.7

Calculation of mutation rates using eMuta T7. Before screening for continuous evolution of MT0, Mutation rate of eMuta T7 RNAP mutator under the system used in this study. pGE158 vector plasmid anchoring wildtype TEM-1 gene and mutator gene was co-transformed to BL21 (DE3) competent cells and grown for overnight on LB/agar plate containing 6 mg/L tetracycline (tetC) and 35 mg/L chloramphenicol (Cm). Colony was picked and inoculated to LB media containing appropriate amount of tetracycline and chloramphenicol. After growth till OD_{600} reached 1.0, growth culture was 100-fold diluted to OD_{600} was 0.01 followed by addition of arabinose inducer with final concentration of 0.2% to activate mutator gene. Culture was grown by OD_{600} was 1.0 and diluted to fresh LB media containing tetracycline, chloramphenicol and arabinose every 12 hr. by cycle, 5 μL of culture was inoculated to 500 μL of fresh LB media and grown for overnight on LB/agar plate ($tetC^+/Cm^+$). 5 colony each for 7 cycle totals was respectively inoculated

to 20 mL of LB media containing appropriate amount of tetracycline and chloramphenicol, and grown for 8 hr. Then, genes were respectively extracted by miniprep kit and asked for sequencing. (Macrogen) nucleotide and amino acid mutation per cycle was listed in table 4.7.

Mutation rate was calculated as following; mutations per day per 1 kb. After 7 cycles, mutations were counted within 789-bp TEM-1 encoding region. The average number of mutations (5.2 mutations) was divided by the length of TEM-1 gene (0.789 kb) and the screening days (2.33 days). Final mutation rate was 2.829 mutations day^{-1} kb^{-1}.

Cycle #	Ave. # of nt mutated	# of AA mutated
1	0.8	0.2
2	1.4	0.8
3	2.6	1.2
4	3.2	1.8
5	3.6	2
6	4.4	2.6
7	5.2	2.8

Table 4.8. Mutations per cycles.

Screening for continuous evolution using eMuta T7 system. For continuous evolution of MT variants, constructed vector plasmid encoding mt0 gene was co-transformed to BL21 (DE) e coli. competent cells with the plasmid for the eMuta T7 mutator gene and then grown for overnight on LB/agar plate containing 6 mg/L tetracycline (tetC), 30 mg/L chloramphenicol (Cm) and 25 μM $ZnCl_2$. single colony harboring both target and mutator genes were inoculate to 5 mL of LB media ($tetC^+$/Cm^+) and grown for 12 h. When OD_{600} reached ~1.0, the culture was 100-fold diluted

to 500 μL of LB media containing 6 mg/L tetracycline, 30 mg/L chloramphenicol, 0.2 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG), 0.2% arabinose, and 25 μM $ZnCl_2$.

Then, diluted culture was grown at 37 ºC (200 rpm) till OD_{600} reached 1.0 (12 hr). after 2 cycles, 1 mg/L ampicillin was added to the screening culture. Then, higher concentration of ampicillin was added every 2–3 cycle. Screening was repeated to 37th cycle till the concentration of ampicillin reached 70 mg/L (Figure 4.9a). Cells from screened cultures for each cycle was centrifuged at 13,000 rpm and resuspended with LB media containing 20% glycerol then stored at -80 ºC.

Hydrolytic activity assay with lysates of screened culture. Each culture at 25, 45, 60, and 70 mg/L ampicillin condition was respectively inoculated to 5 mL of fresh LB media ($tetC^+/Cm^+$) and grown for 3 hr at 200 rpm / 37 ºC condition followed by incubation on the LB/agar plate ($tetC^+/Cm^+$) containing 0.1 mM IPTG only to avoid unexpected mutation by mutator gene which is activated with arabinose.

For each culture, 90 colonies were picked and inoculated to 800 μL of LB media containing 6 mg/L tetC, 30 mg/L Cm, 0.2 mM IPTG, and 25 μM of $ZnCl_2$ in 96-well deep-well microplate. After the growth at 37 ºC and 200 rpm, 200 μL of each culture was centrifuged with 5,000 rpm. cell pellets were lysed by incubating cells in lysozyme-EDTA cocktail solution (1 mg/ml lysozyme and 100 μM EDTA in 50 mM tris/HCl pH 7.4 buffer) for 30 min at 37 ºC followed by supplementation of 125 mM $ZnCl_2$ to overcome the removal of Zn^{2+} ions from the active sites of MT variants by EDTA. Representative example for lysate assay was expressed in Figure 4.8

For monitoring the hydrolytic activities of lysates, CENTA, which is a chromogenic surrogate of β-lactam antibiotics, was used. (Figure 4.8a) Using microplate reader, absorption losses at 405 nm ($\varepsilon_{CENTA,\ 405\ nm}$ = -6,400 $M^{-1}cm^{-}$

[1]) of 90 wells were measured for 1 hr followed by selection of Top 1–2 fastest wells. (Figure 4.8b) 5 μL of cultures corresponding wells from deep-well plate were inoculated to 20 mL of LB media (tetC+/Cm+) and grown for overnight. Genes were extracted with miniprep kit (COSMO Genetech) and sent for sequencing. (Macrogen) Mutations, which occurred during screening, were listed in Table 4.3.

Expression and purification for MT variants. Sequence-identified mt gene was transformed to BL21 (DE3) *E. coli* competent cells and grown on LB/agar plate. ($tetC^+$) After overnight growth, the subculture was inoculated to 8 L of LB media containing 6 mg/L tetC then grown at 37 ºC and 200 rpm. when OD_{600} reached 0.8, 1 mM IPTG (final conc.) was added followed by 18 hr growth at 25 ºC. Grown culture was centrifuged at 5,000 rpm and then cell pellet was resuspended with 20 mM tris/HCl pH 7.0 buffer and lysed by cell homogenization. (Young Jin Corporation LM20-30 Microfluidizer) cell lysate was centrifuged at 13,000 rpm and supernatant was collected and syringe filtered. Supernatant was loaded to Ni-NTA column (Cytiva HiTrap FF 5 ml) which was pre-equilibrated with lysis buffer containing 10 mM imidazole. MT variants were eluted at ~25% of B buffer. (20 mM tris/HCl pH 7.0 buffer containing 500 mM imidazole) After removal of remaining imidazole from protein sample, proteins were concentrated and stored at -80 ºC.

***In vitro* characterization of MT variants.** The hydrolytic activities of MT variants under *in vitro* condition, specific activities of MT variants toward ampicillin substrate were analyzed in presence/absence of $ZnCl_2$ using HPLC (Agilent infinity 1260) as described above. To validate accurate kinetic characterization, steady-state kinetics were also analyzed (Figure 4.9c and Table 4.4).

Determination of Zn binding affinity. To validate the binding affinities of MT variants toward Zn^{2+} ions, competitive titration assay was conducted to determine dissociated constant toward Zn^{2+} ion using Newport Green DCF diacetate (Thermo Fisher Scientific) which is a Zn-responsive fluorophore. The assay revealed one Zn binding equilibria with a single site binding model using DinaFit4 software. This result suggests that the Zn mediated MT variants possess one Zn-binding sites as incorporated (Figure 4.10 and Table 4.8).

4.5 References

(1) Turusov, V.; Rakitsky, V.; Tomatis, L., *Environmental Health Perspectives* **2002,** *110* (2), 125-128.

(2) Fleming, A., *Br J Exp Pathol* **1929,** *10* (3), 226-236.

(3) Gerlt, J. A.; Babbitt, P. C., *Annual Review of Biochemistry* **2001,** *70* (1), 209-246.

(4) Tawfik, O. K.; Dan, S., *Annual Review of Biochemistry* **2010,** *79* (1), 471-505.

(5) Eom, H.; Song, W. J., *JBIC Journal of Biological Inorganic Chemistry* **2019,** *24* (4), 517-531.

(6) Anbar, A. D., *Science* **2008,** *322* (5907), 1481-1483.

(7) Dupont, C. L.; Butcher, A.; Valas, R. E.; Bourne, P. E.; Caetano-Anollés, G., *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107* (23), 10567.

(8) Lyons, T. W.; Reinhard, C. T.; Planavsky, N. J., *Nature* **2014,** *506*, 307.

(9) Dupont, C. L.; Yang, S.; Palenik, B.; Bourne, P. E., *Proceedings of the National Academy of Sciences* **2006,** *103* (47), 17822-17827.

(10) Tomatis, P. E.; Rasia, R. M.; Segovia, L.; Vila, A. J., *Proceedings of the National Academy of Sciences* **2005,** *102* (39), 13761-13766.

(11) Stiffler, Michael A.; Hekstra, Doeke R.; Ranganathan, R., *Cell* **2015,** *160* (5), 882-892.

(12) Steinberg, B.; Ostermeier, M., *Science Advances 2* (1), e1500921.

(13) Song, W. J.; Tezcan, F. A., *Science* **2014,** *346* (6216), 1525-1528.

(14) van Loo, B.; Bayer, C. D.; Fischer, G.; Jonas, S.; Valkov, E.; Mohamed, M. F.; Vorobieva, A.; Dutruel, C.; Hyvönen, M.; Hollfelder, F., *Journal of the American Chemical Society* **2019,** *141* (1), 370-387.

(15) Yu, J.; Yang, J.; Seok, C.; Song, W. J., *Chemical Science* **2021,** *12* (14), 5091-5101.

(16) Singh, R.; Saxena, A.; Singh, H., *Journal of Biomedical Science* **2009,** *16* (1), 109.

(17) Stojanoski, V.; Adamski, C. J.; Hu, L.; Mehta, S. C.; Sankaran, B.; Zwart, P.; Prasad, B. V. V.; Palzkill, T., *Biochemistry* **2016,** *55* (17), 2479-2490.

(18) Jeong, W. J.; Song, W. J., *Nature Communications* **2022,** *13* (1), 6844.

(19) Park, H.; Kim, S., *Nucleic Acids Research* **2021,** *49* (6), e32-e32.

(20) García-Nafría, J.; Watson, J. F.; Greger, I. H., *Scientific Reports* **2016,** *6* (1), 27459.

국문초록

Exploring the Evolutionary Landscape of Artificial β-lactamases

인공 베타-락탐 가수분해효소의 진화적 다양성 연구

수많은 효소들은 수십 억년의 시간동안 아주 무분별한 반응성을 가지는 조상들로부터 특정 반응들에 전문적으로 반응하는 현대의 효소의 형태로 진화를 거쳐왔다. 이러한 생촉매는 고유의 촉매활성과 기질 특이성을 바탕으로 다양한 화학반응들을 매개한다. 이런 효소의 강력한 활성은 새로운 촉매를 만들고자 했던 많은 연구자들을 매료시키고 영감을 주었다. 특히, 유도진화는 자연적인 진화 과정을 모방하는 방식으로 기존의 반응성을 향상시키거나, 완전히 새로우면서도 자연에서 일어나지 않는 반응을 촉진하는 새로운 효소 단백질들을 만드는데 기여해왔다. 그럼에도 불구하고, 목표로 하는 단백질이 어떻게 진화하는지 혹은, 무엇이 유도진화 과정 중에서 특정 효소를 새로운 효소로 진화하게 하는지 등은 여전히 설명이 필요한 어려운 문제로 남아있다. 효소의 진화 양상과 더 효율적인 진화 연구를 위한 이론적 해석의 이해를 확장하기 위해, 이 논문은 새롭게 설계된 인공 금속-베타-락탐 가수분해효소와 천연 세린-베타-락탐 가수분해효소를 모델로, 그들의 진화적인 풍경을 탐구했다.

1장에서는 유도진화 연구 분야에서 효과적인 변형체 라이브러리를 만들기 위한 관습적인 접근방법들을 간략하게 소개한다. 유도진화의 도입은 우리로 하여금 자연에서 일어나는 진화과정을 모방해 실험실에서도 어떤 반응성도 가지지 않거나, 혹은 다른 반응성을 가지던 특정 단백질을 새로운 효소로 전환할 수 있게 해주었다. 초기에는, 밝혀진 구조를 바탕으로 활성 자리 주변의 '잘 연구된' 자리들에 대해 반복적인 자리 포화 서열 변형법이 주로 이용되었다. 하지만 최근에는 표적 단백질의 잠재적인 진화가능성을 극대화하기 위해 무작위 서열변형법을 더 많이 사용하고 있다. 이 두가지 대표적인 라이브러리 구축 방법의 장단점 등을 해당 장에서 기술하고자 한다.

추가적으로, 베타-락탐 가수분해효소를 모델 시스템으로 이용하는 것에 대한 장점들 또한 이야기한다. 이 효소는 베타-락탐 항생제를 가수분해하는 효소로 촉매 역할을 하는 활성화된 세린 또는, 아연금속 결합자리의 유무를 통해 크게 세린-베타-락탐 가수분해효소와 금속-베타-락탐 가수분해효소, 두가지로 분류할 수 있다. 흥미롭게도, 첫 발견 이후로 더 많은 베타-락탐 항생제들이 발견되거나 새롭게 개발될수록, 베타-락탐 가수분해효소의 의해 더 빠르게 그에 대한 내성이 나타나게 되는데 이러한 특성을 어떻게 진화 연구에 유용하게 이용할 수 있는지에 대해 이야기하고자 한다.

2장은 효소의 진화과정에서 단백질 구조의 유연성과 강직성의 균형을 잘 맞추는 것이 얼마나 중요한가에 대해 설명하고자 한다. 본 연구는 2당량의 2가 아연 이온을 통해 사량체 구조를 형성하는 새로운 다량체 금속-베타-락탐 가수분해효소인 AB5를 모델로 채택해 실험을 진행했다. 하지만 의도와는 다르게, AB5의 활성화

자리 주변에 대한 유도 진화 실험 결과는 본 단백질이 매우 낮은 진화가능성을 가지고 있음을 의미했다. 하지만, 96번 자리에 진행한 하나의 서열변형을 통해 AB5가 매우 놀라운 정도의 구조적, 기능적 변화를 겪게 되었다. AB5는 96번 자리의 시스테인을 트레오닌으로 바꿈으로써 구조적 유연성을 획득하게 된다. 기존의 시스테인 잔기들이 분자간 이황화 결합을 형성하면서 AB5 사량체를 더 견고하게 만들어주었으나, 활성화 자리 부근에 위치하면서 활성자리의 유연성을 낮추고 기질의 접근을 막고 있었다. AB5와 ^{C96T}AB5 단백질들에 대한 생체 내 외 조건의 분석 결과는 새로운 인공 효소를 설계하고 유도 진화 과정에서 단백질의 유연성과 강직성의 균형을 맞추는 것의 중요성을 이야기한다.

3장은 효소의 진화를 위해 서열 변형을 진행할 핫스팟을 찾는 새로운 접근 방식을 제안하다. 유도진화 연구에서 본격적인 무작위 서열변형법의 이용은 초기에 활성자리 주변에 국한되었던 연구들을 넘어 활성화 자리로부터 매우 멀리 떨어진 영역에서 일어나는 서열변형의 중요성을 인식하게 되었다. 그럼에도 불구하고, 변형체 라이브러리 구축 과정에서의 임의성과, 알로스테릭 효과로도 알려진 원거리에서 일어난 서열변형들의 불명확한 영향 등으로 인해, 무엇을 표적으로 하고, 또 어떻게 결과를 분석할지는 여전히 설명이 필요한 어려운 문제들로 남아 있다. 본 연구진은 다량체를 형성하는 단백질의 대칭성과 관련된 잔기들을 서열변형의 핫스팟으로 가정하고 유도 진화를 진행했다. 이 가설은 2장에서 구조적, 기능적 변화를 동시에 야기한 C96T 서열변형 자리가 AB5 단백질의 회전축 상에 존재함을 확인한 것으로부터 출발했다. AB5 단백질은 D_2 대칭성을 갖는 단백질로,

세개의 서로 직교하는 회전축을 가지고 있다. 본 장에서, AB5 단백질의 다른 회전축들의 주변에 존재하는 잔기들을 대상으로 유도진화를 진행하고, 그 결과 얻은 진화된 단백질들의 생화학적, 구조적 분석을 보고할 것이다.

4장에서는 크게 세린-베타-락탐 가수분해효소와 금속-베타-락탐 가수분해효소로 분류되는 두 베타-락탐 가수분해효소들의 나뉘어진 진화적 풍경을 탐험한다. 두가지 타입의 베타-락탐 가수분해효소들은 매우 유사한 기능을 수행하지만 전혀 다른 진화적 기원을 가지는 것으로 알려진, 이른바 수렴 진화의 대표적인 예시이다. 본 장에서는, 두가지 형태의 서로 다른 락탐 가수분해효소들의 금속 이온의 농도와 같은 화학적 환경을 변화시킴으로써, 화학적 반응기전과 서로 다른 활성화자리 구성이 서로 상호 전환이 가능함을 설명한다.

본 연구는 반복적인 자리 포화 서열변형법 또는 연속진화법을 이용해 주어진 화학적 조건에서 근원적으로 다른 베타-락탐 가수분해효소의 반응성이 발생하고 이것이 추가적인 진화를 통해 향상될 수 있음을 설명한다. 이러한 결과로, 본 연구는 금속 이온이 효소의 진화적 풍경을 조절할 수 있는 매우 중요한 열쇠가 될 수 있음을 이야기한다.

주요어: 효소 진화, 베타-락탐 가수분해효소, 단백질 대칭성, 생촉매, 인공 금속 효소

학 번: 2016-27405

Acknowledgement

6년. 지난 2016년 가을부터 정말 6년이란 시간을 꼬박 흘려 보내고 나서야 비로소 이 학위 논문을 제출하기에 이르렀습니다. 돌이켜보면 그 동안의 날들을 들여다보면 마치 영겁의 시간처럼 느껴지다가도 한발짝 떨어져 조망하면 다시 그 긴 시간이 찰나처럼 느껴지는 모순의 감정이 듭니다. 학위를 시작하면서 가졌던 수많은 불확실한 미래에 대한 의문들이 학위를 마치는 지금 모두 다 해소되었다고는 말할 수 없지만, 고독하다고 생각했던 학위기간을 곰곰이 돌이켜보면 수많은 분들이 제 주변을 지켜주셨습니다. 이 자리를 빌려 혼자서는 결코 해낼 수 없었던 제 청춘의 긴 시간을 바친 학위기간 동안 큰 도움을 주신 수많은 분들께 감사의 말씀을 전하고자 합니다.

가장 먼저, 항상 의욕적이지 못하고 많이 부족한 제가 이 연구를 그리고 학위 과정을 무사히 마칠 수 있게끔 지도와 편달을 아끼지 않으셨던 저의 지도교수님, 송윤주 교수님께 진심으로 감사의 말씀을 전합니다. 교수님께서 이끌어 주신 덕분에 여기까지 올 수 있었습니다. 긴 학위기간 동안 교수님께 배운 연구자로서의 마음가짐을 항상 간직하겠습니다. 또, 본 연구에 큰 도움을 주신 석차옥 교수님과 김석희 교수님께도 감사의 말씀을 드립니다. 그리고 저의 학위논문 심사를 맡아주시고, 저의 연구내용에 아낌없이 조언해주신 이동환 교수님, 김석희 교수님, 이형호 교수님, 그리고 홍승우 교수님, 정말 감사드립니다.

실험실 인턴 경험 한 번 없이 무턱대고 대학원 생활을 시작한 저에게 실험실 생활의 전반에 대해 많은 도움을 주었을 뿐만 아니라, 제 연구 내용에 대해서도 함께 고민해주고 조언해준 제 동기

정우재에게 큰 감사의 말을 전합니다. 연구실내에서 같은 분야를 연구하는 동료로서 항상 좋은 영감을 주는 조언을 해주었던 엄현욱, 공동연구를 통해 익숙하지 않았던 분야에 대해 많은 도움을 주신 양진솔 박사님과 새로운 실험 설계에 큰 조언을 준 서대제 학생에게 감사의 말을 전합니다. 또한, 긴 실험실 생활동안 저의 연구내용에 대해 관심을 가져주고 같이 고민해하고 문제를 헤쳐나가준 우리 실험실의 양민우 박사님, 그리고 이재원, 이재희, 박민찬, 손소연, 양정민, 정세민, 주환진, 한민정, 문재원에게도 감사의 말 전합니다. 부사수로서 함께 실험해준 인턴 정승환, 원다예, 박정명, 이동헌 그리고 정웅 학생에게도 감사의 말을 전합니다. 그리고 학위 과정동안 저의 사사로운 일부터 연구내용에 이르기까지 격려해주고 조언해주었으며, 함께 연구자의 꿈을 이야기하고, 저보다 먼저 그 출발을 이루었던 저의 오랜 벗이며 선배 연구자인 손현철 박사님께 큰 감사의 말씀을 전합니다.

끝으로, 많은 동료들이 도움을 준 만큼이나 가족들의 믿음과 응원 덕분에 끝까지 포기하지 않을 수 있었습니다. 긴 시간 동안 저를 믿고 묵묵히 기다려주고 아낌없이 지원해준 우리 가족들, 어머니, 아버지, 그리고 동생에게 진심으로 감사의 말을 전합니다.

긴 시간 동안 정말 많은 것을 배웠고, 정말 많은 좋은 동료들을 만났습니다. 그 덕분에 많이 성장할 수 있었습니다. 모든 날들이 즐거웠다고는 말하지 못하겠지만, 돌이켜본 그 전체의 시간이 저에게 상이었습니다. 학위 과정은 끝이 나지만 이는 곧 또 다른 시작이라는 것을 잘 알고 있습니다. 이 모든 시간 동안의 배움을 자양분으로 앞으로도 더 치열하게, 더 끈기있게 학문에 임하는 좋은 연구자이자 동료, 친구 그리고 가족으로 함께하는 사람이 되도록 하겠습니다. 감사합니다.