Electronic Supplementary Material (ESI) for Chemical Science.
This journal is © The Royal Society of Chemistry 2021

Electronic Supplementary Material (ESI) for

Symmetry-related residues as promising hotspots for the evolution of *de novo* oligomeric enzymes

Jaeseung Yu,[1] Jinsol Yang,[1] Chaok Seok,[1] and Woon Ju Song[1,*]

[1]Department of Chemistry, College of Natural Sciences, Seoul National University, Seoul 08826, Republic of Korea

E-mail: woonjusong@snu.ac.kr

Table of contents

Supplementary Figures

Supplementary Tables

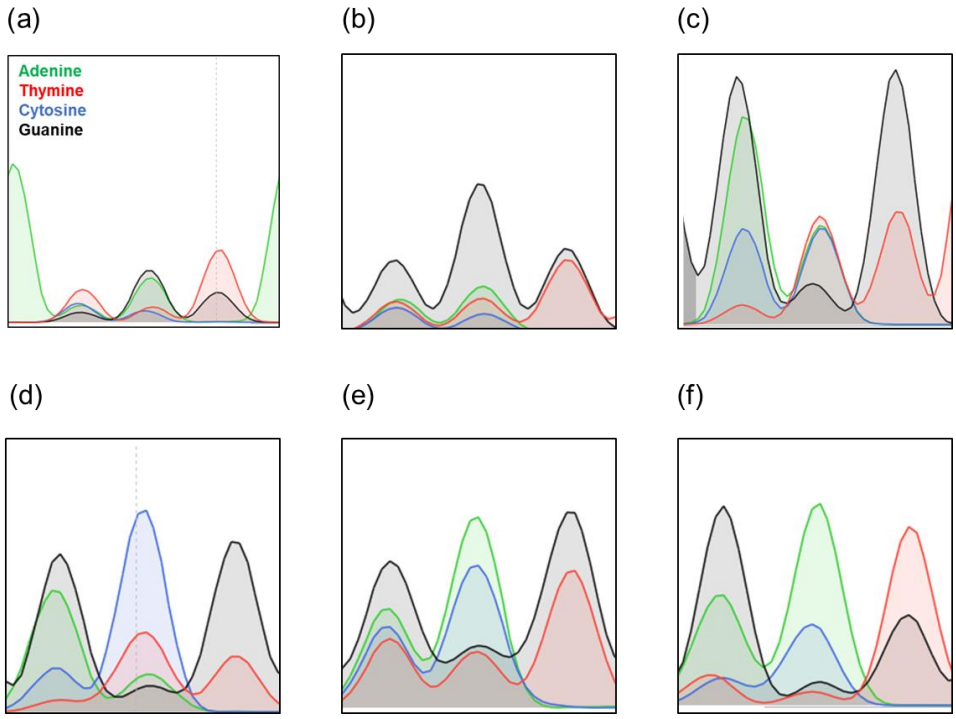


Fig. S1. Representative sequencing chromatograms of AB5 mutant libraries randomized at (a) C96 (b) T97 (c) K104 (d) A38 (e) D39 and (f) E81 positions.

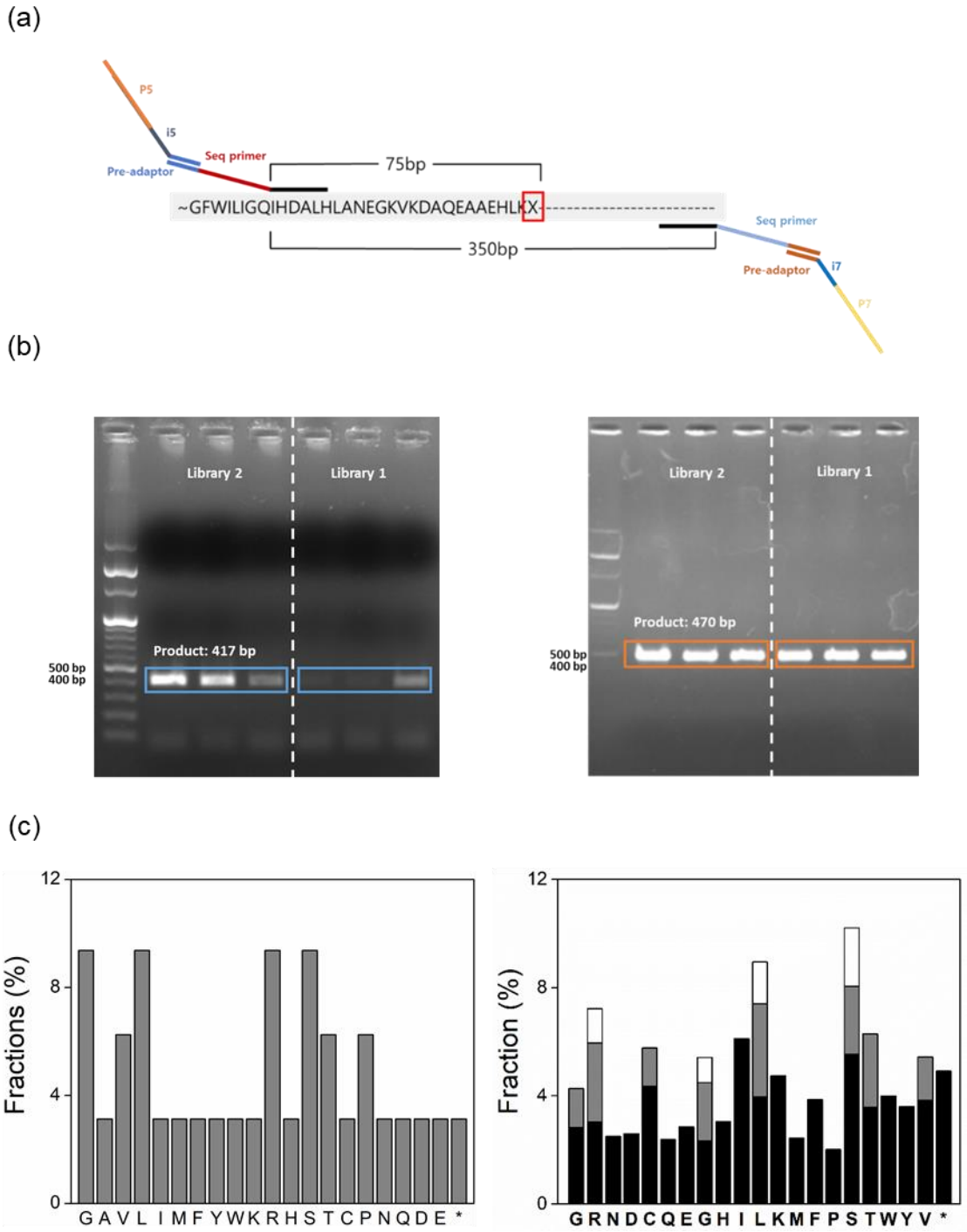


Fig. S2. The next-generation-sequencing (NGS) of C96X libraries. (a) Fragment PCR (b) DNA agarose-gel electrophoresis at (left) the 1st and (right) the 2nd step (c) Plots of (left) theoretical fractions of amino acids encoded by NNK degenerated primer and (right) the NGS results. Discrete codons encoding the same amino acids are colored in different grey scales. Asterisks indicate a TAG stop codon.

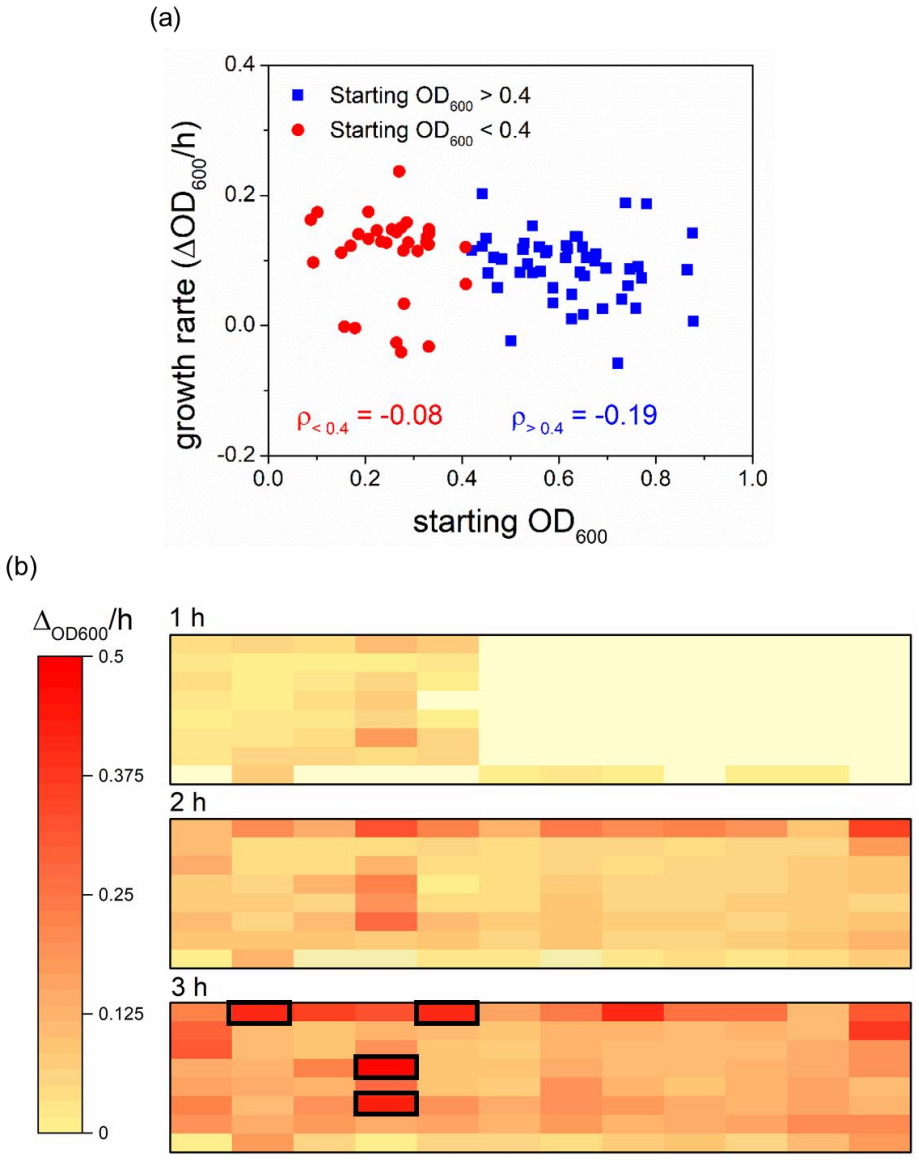


Fig. S3. The whole-cell screening upon the addition of ampicillin (10–15 mg/L). (a) The correlation of the starting OD_{600} values and cell growth rates. (b) Representative screening result of the mutant library randomized at 96 position. The cell growth rates were obtained by monitoring the optical density at 600 nm for 3 h. The fastest-growing cells were highlighted as black boxes.

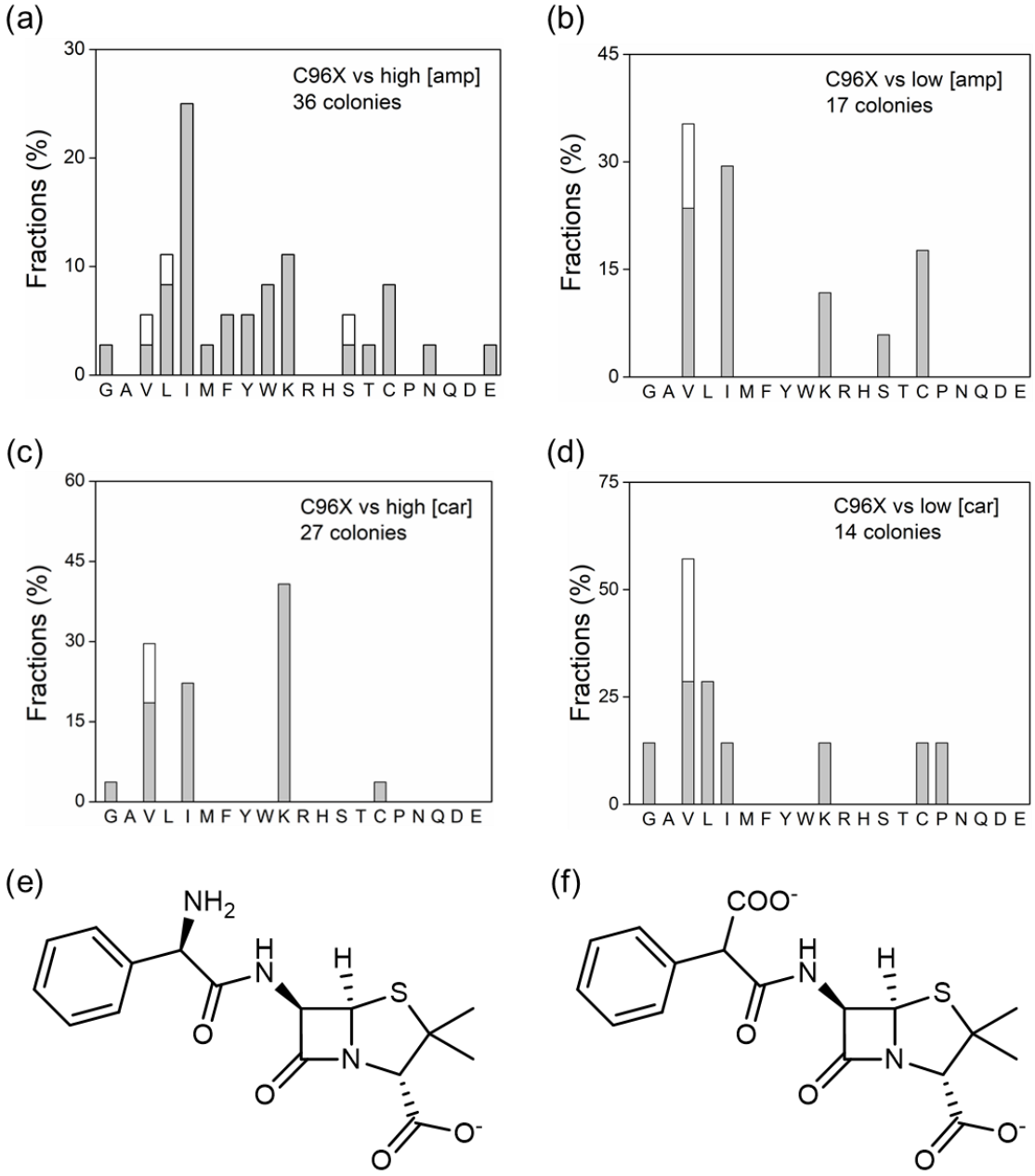
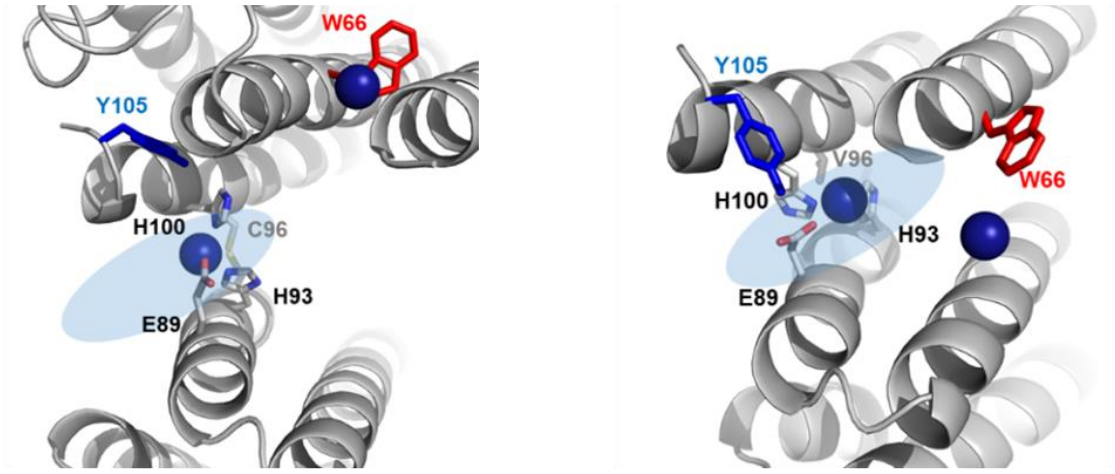
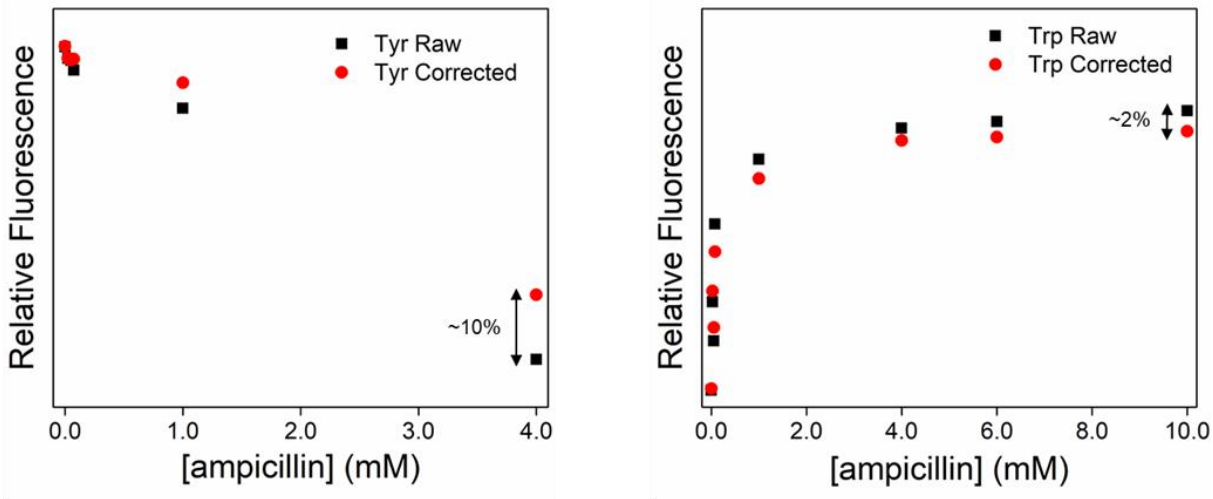


Fig. S4. The screening readouts of the C96X library with (a) high (10–15 mg/L) and (b) low concentration of ampicillin (1.8 mg/L) and (c) high (20–30 mg/L) and (d) low (1.9 mg/L) concentration of carbenicillin. Two discrete codons that encode the same amino acid in (a)–(b) are shown with different shades. 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.

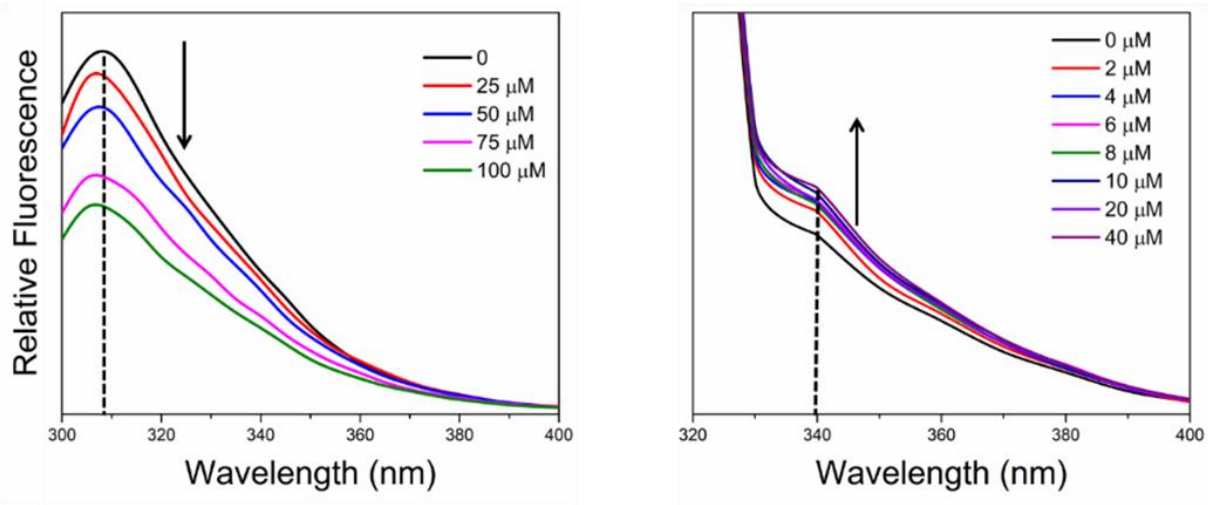
(a)


W66
Y105
H100
C96
H93
E89
Y105
V96
W66
H100
H93
E89

(b)


Tyr Raw
Tyr Corrected
~10%
Relative Fluorescence
[ampicillin] (mM)
0.0
1.0
2.0
3.0
4.0
Trp Raw
Trp Corrected
~2%
Relative Fluorescence
[ampicillin] (mM)
0.0
2.0
4.0
6.0
8.0
10.0

(c)


0
25 μM
50 μM
75 μM
100 μM
Relative Fluorescence
Wavelength (nm)
300
320
340
360
380
400
0 μM
2 μM
4 μM
6 μM
8 μM
10 μM
20 μM
40 μM
Wavelength (nm)
320
340
360
380
400

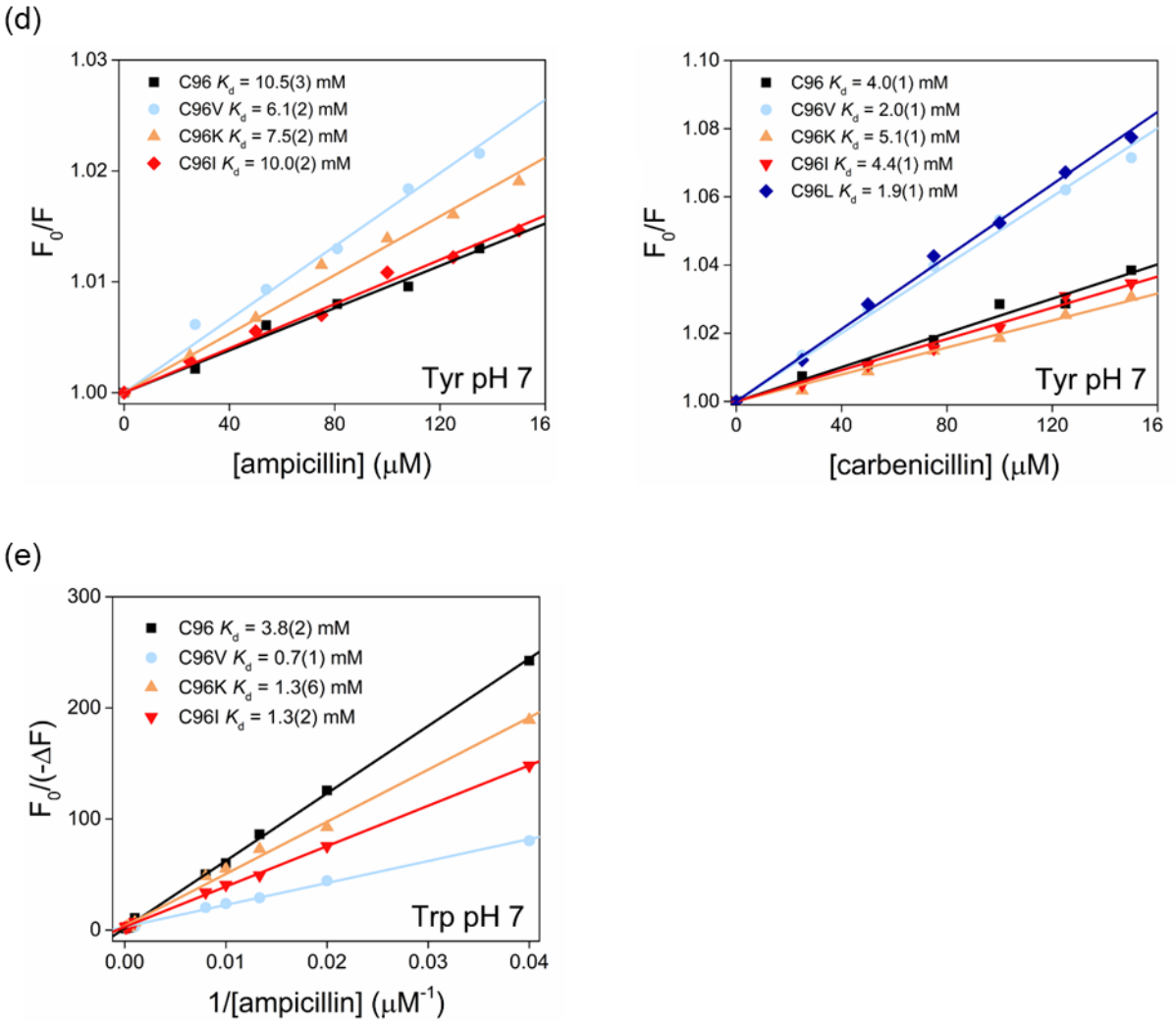


Fig. S5. Determination of substrate binding affinity by intrinsic fluorescence. (a) Tyrosine and tryptophan residues at the vicinity of the active sites in (left) C96 and (right) C96V protein. (b) Inner filter effect with the intrinsic fluorescence with (left) free tyrosine and (right) tryptophan by 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.

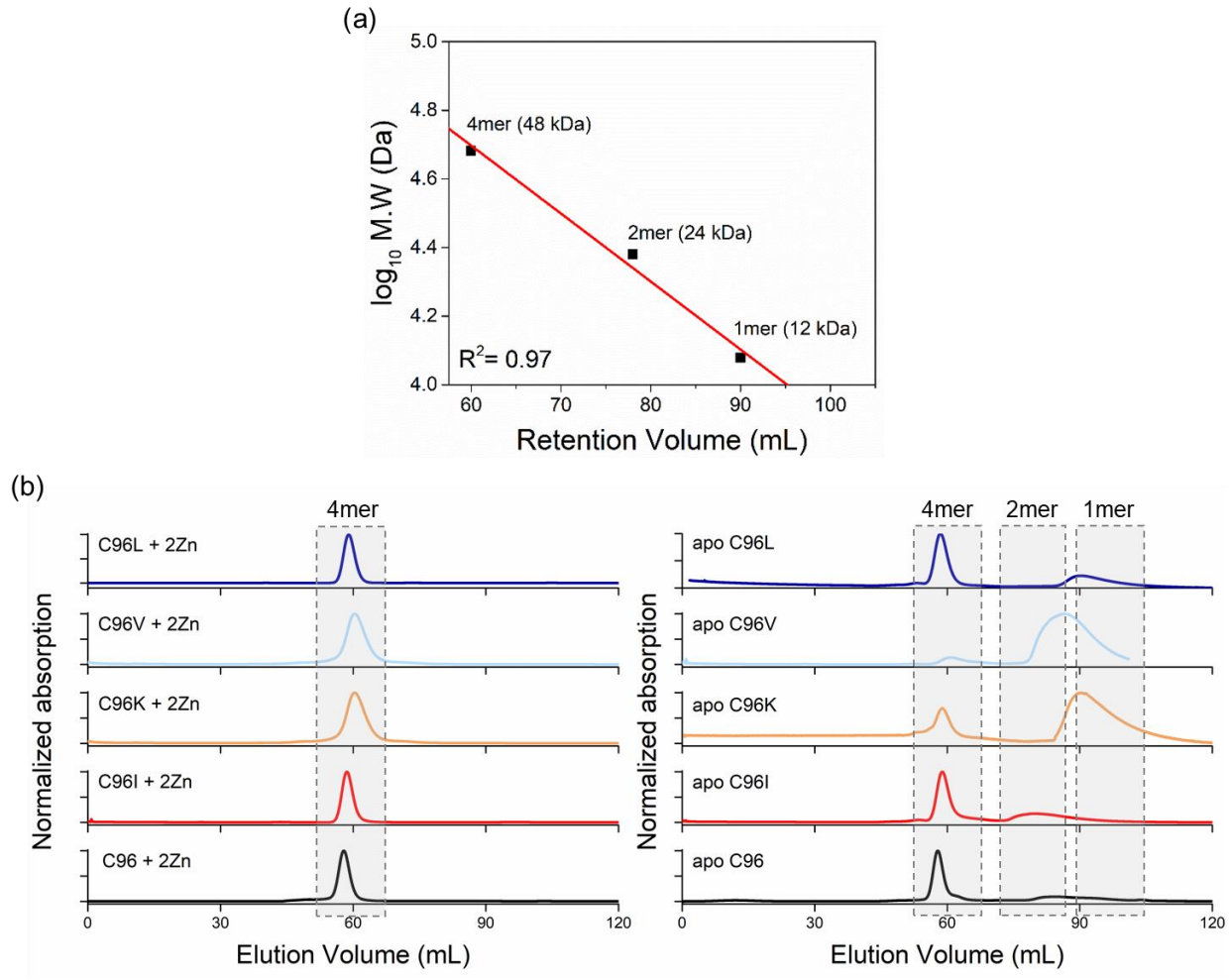


Fig. S6. Determination of oligomerization states of the C96X variants by size exclusion chromatography. (a) A standard curve (b) Chromatograms of (left) Zn-complexed and (right) apo C96, C96I, C96K, C96V, and C96L proteins.

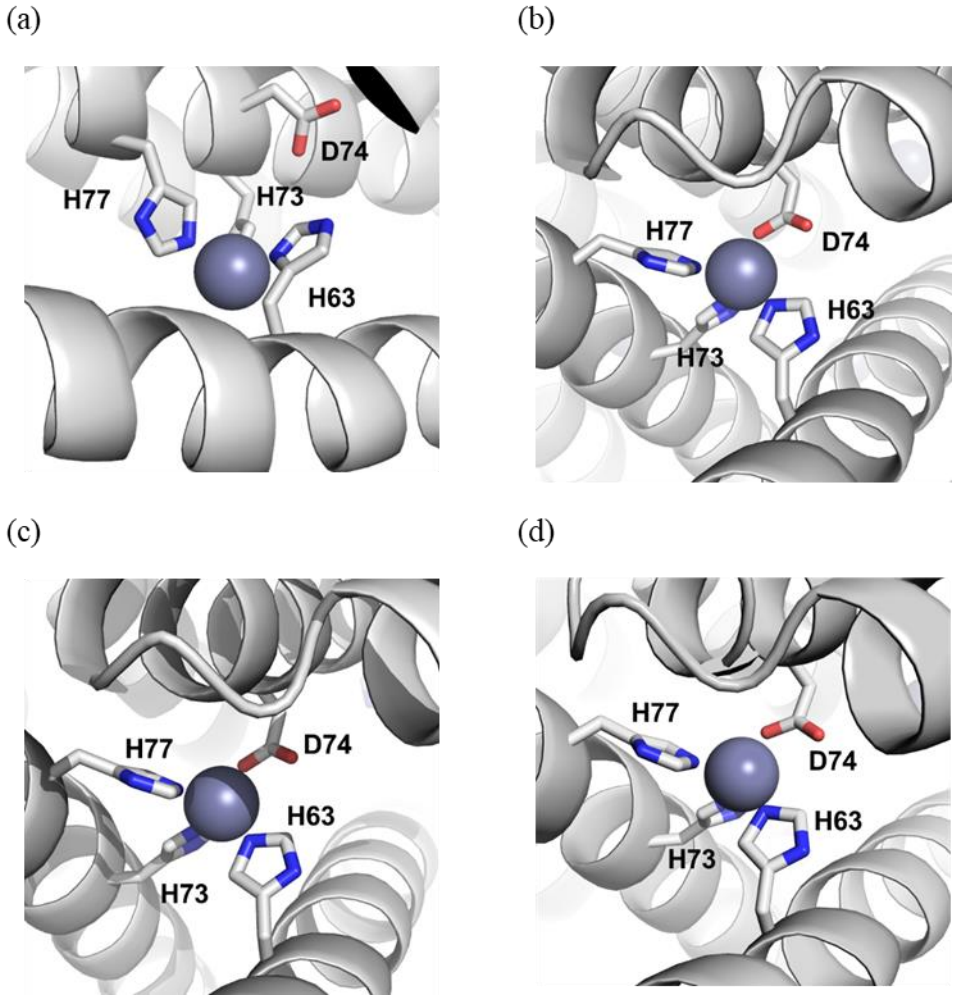


Fig. S7. Structural Zn-binding sites of C96X variants (X= C, I, K, and V). (a) C96 (b) C96I (c) C96K (d) C96V

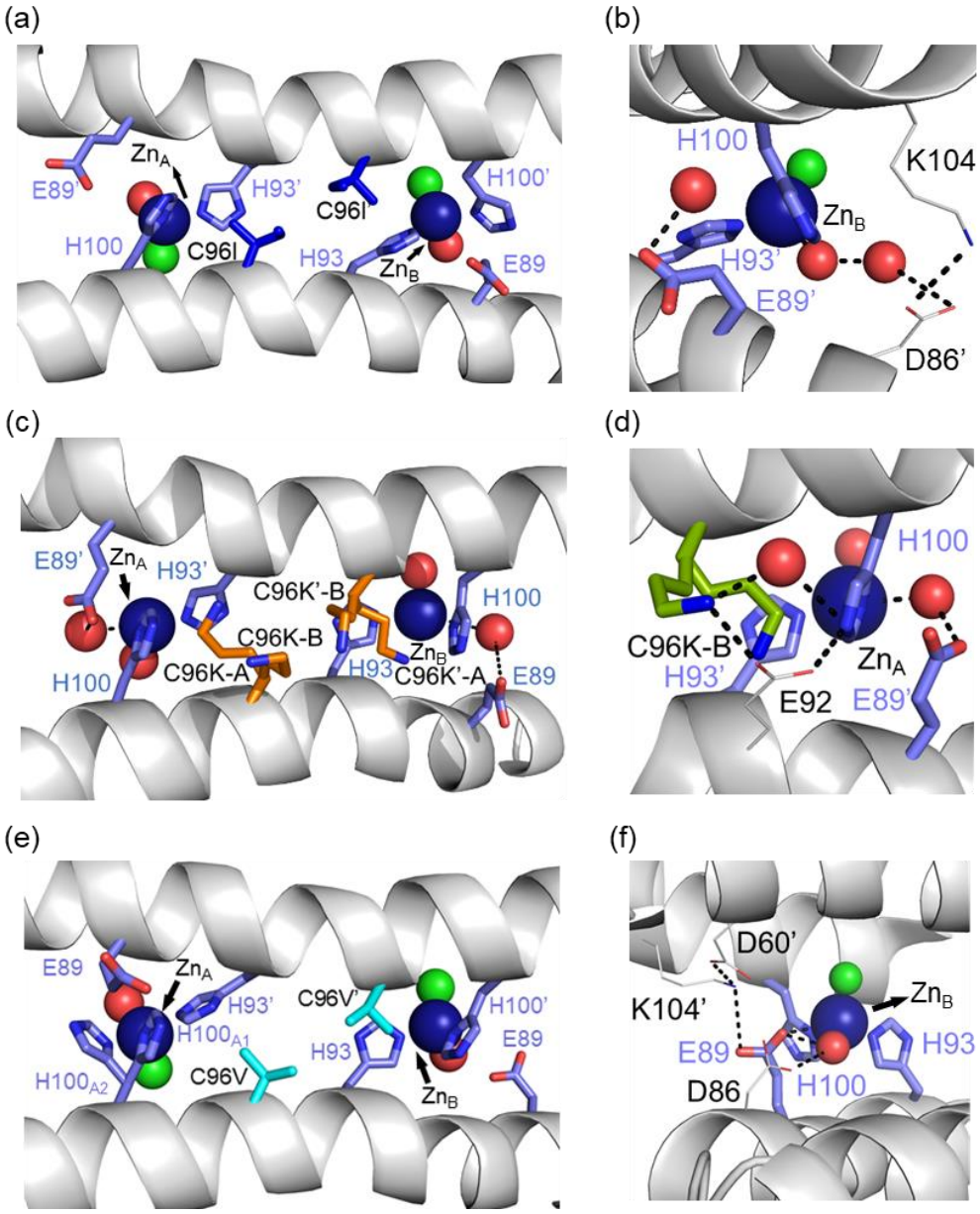


Fig. S8. Catalytic Zn-binding sites of C96X variants (X= I, K, and V). Depending on the angle of N (H93)-Zn-N (H100), the discrete metal-binding sites are labeled as Zn_A and Zn_B, respectively. (a) The side-view and (b) Hydrogen bonding interaction around the Zn_B-active site in the C96I variant. (c) 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 the 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 (C96K-A, C96K-B, C96K'-A, and C96K'-B).

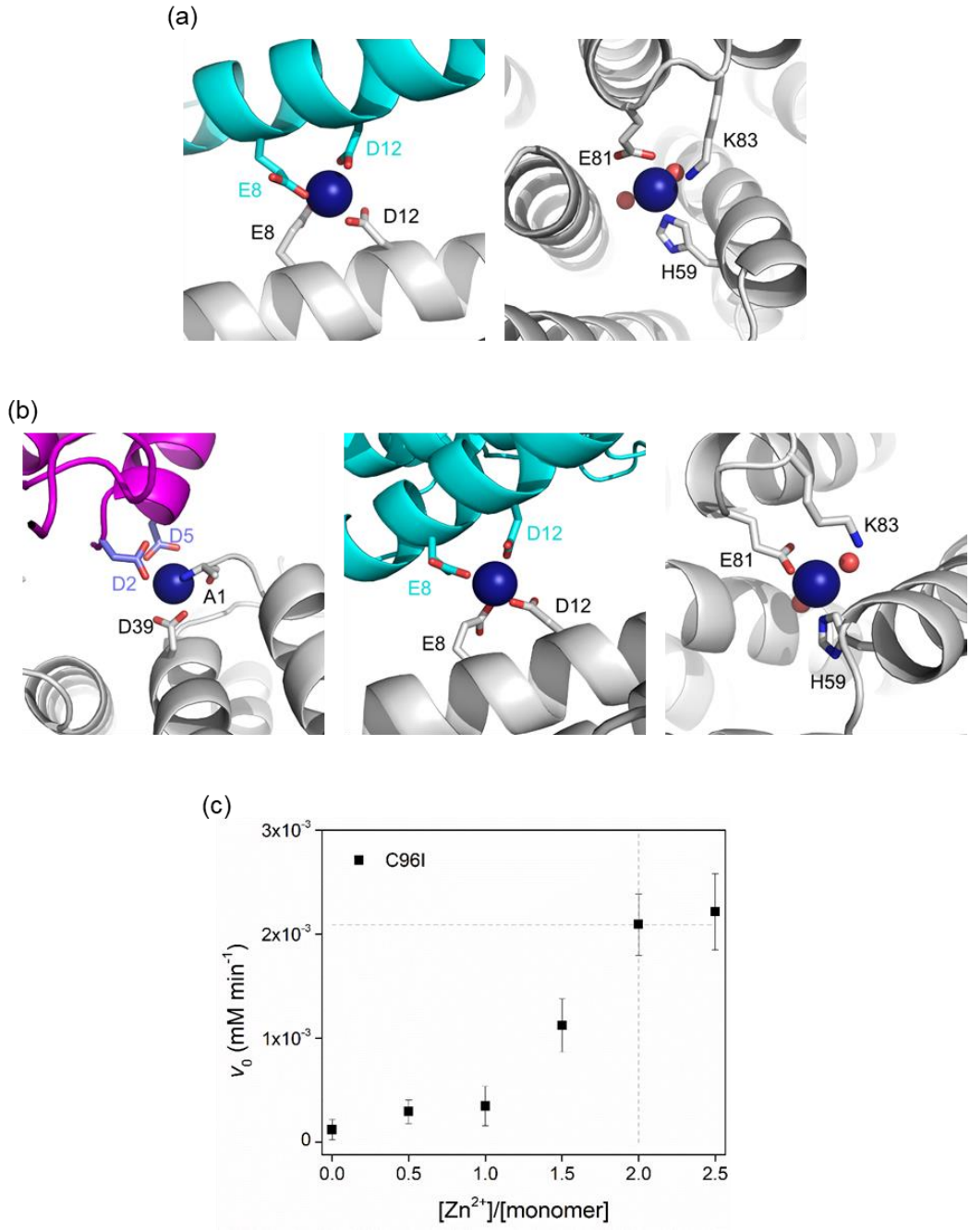


Fig. S9. Additional Zn-binding sites shown in the crystal structures of (a) C96I and (b) C96K variants. Each tetramer forming crystal packing interactions is shown in different colors. (c) The Zn-dependent hydrolysis activity of C96I with 20 mM ampicillin. The maximum hydrolytic activities were observed 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.

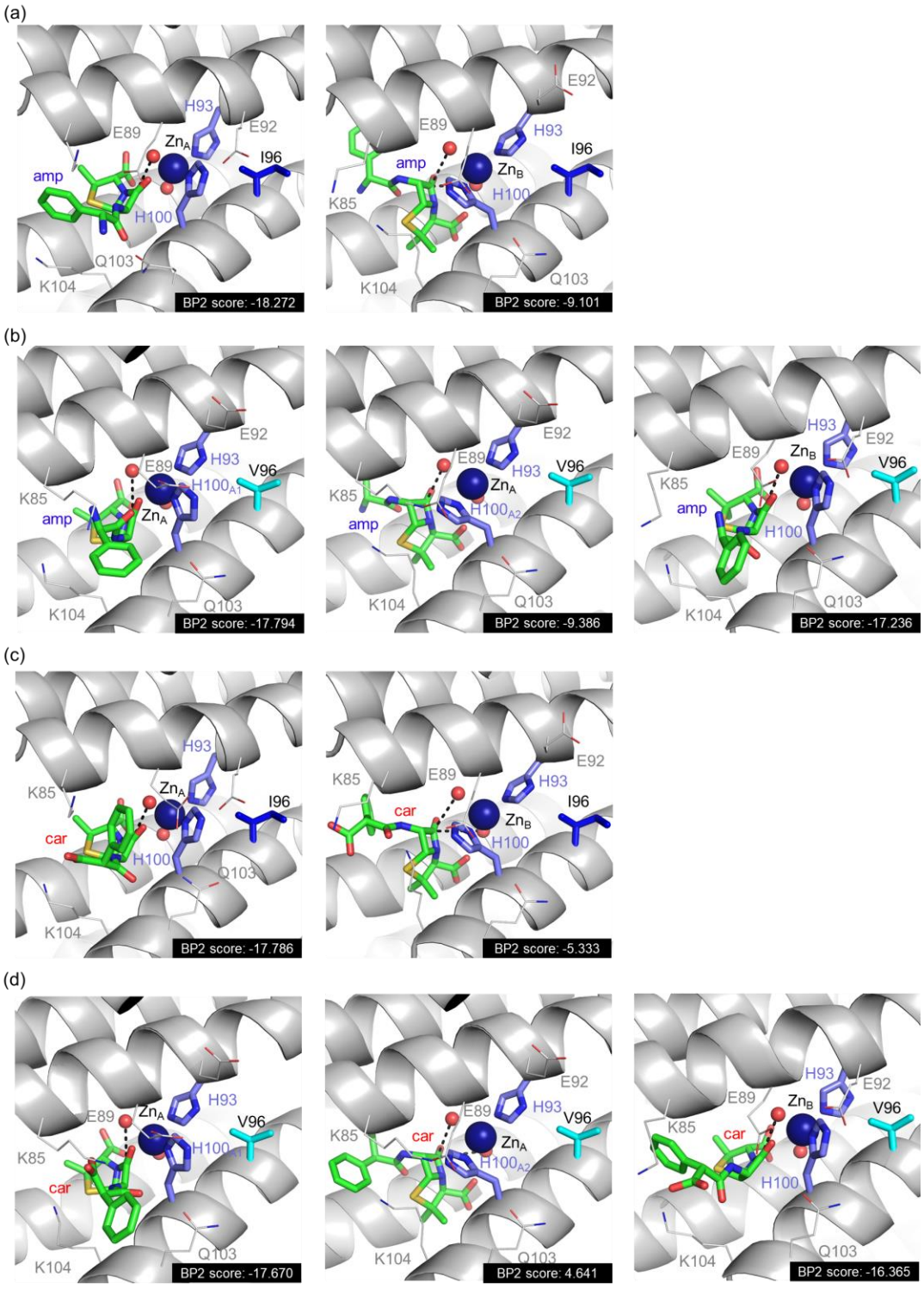


Fig. S10. The predicted binding poses of β-lactam substrates to C96I or C96V variants. Ampicillin to (a) C96I and (b) C96V mutants. Carbenicillin to (c) C96I and (d) C96V mutants. The BP2 score estimates binding energy in arbitrary energy units, suggesting the most favorable binding mode by respective catalytic Zn sites.

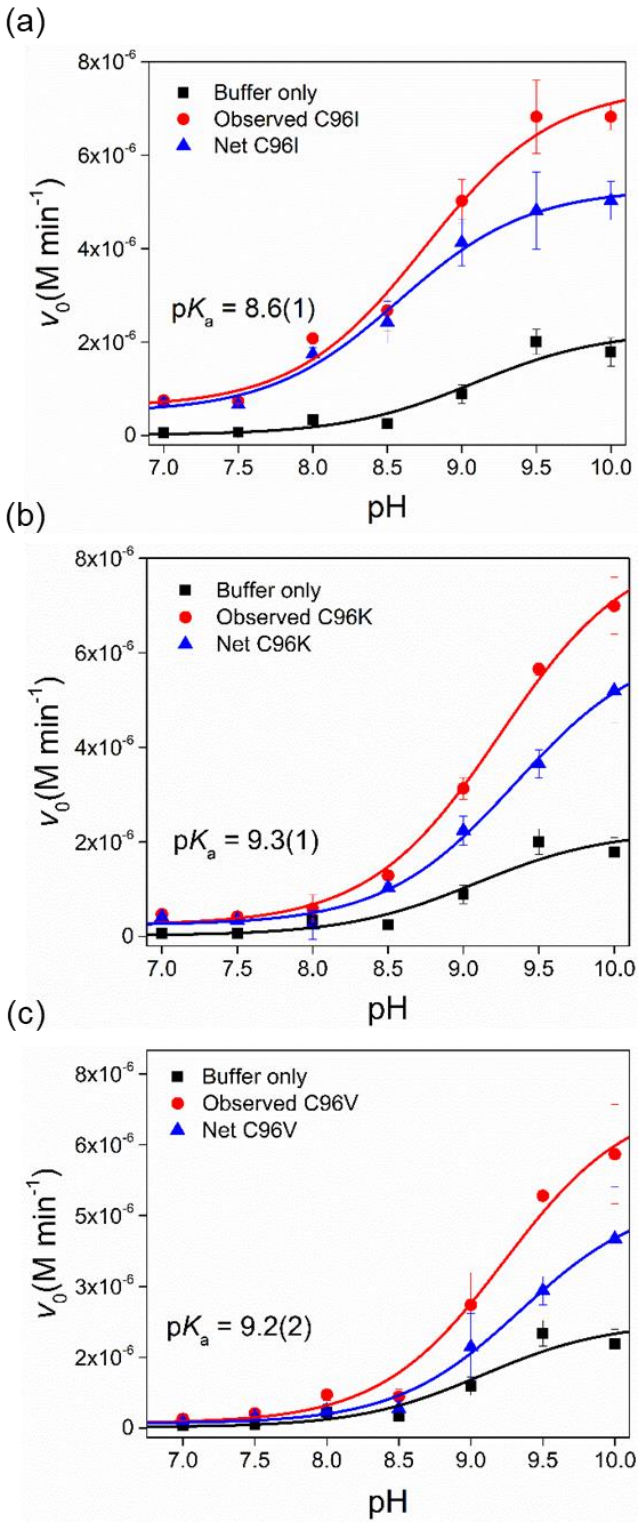


Fig. S11. The pH-dependent ampicillin hydrolysis activity assays to determine pK_a values of the C96X variants (X = I, K, and V) from 1st round of evolution. (a) C96I (b) C96K (c) C96V.

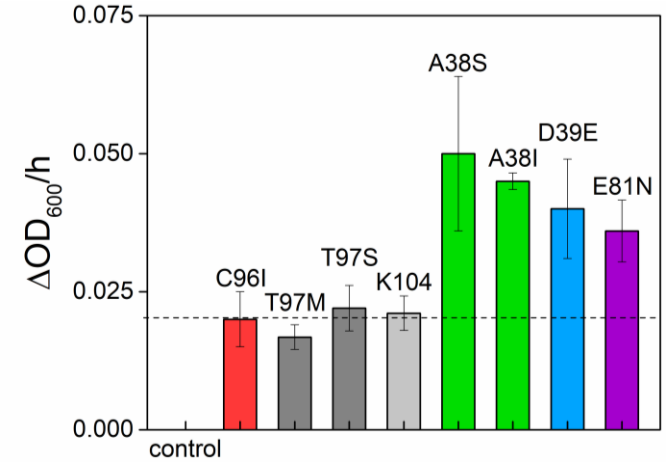


Fig. S12. Cell growth rates of best hits during the second round of evolution.

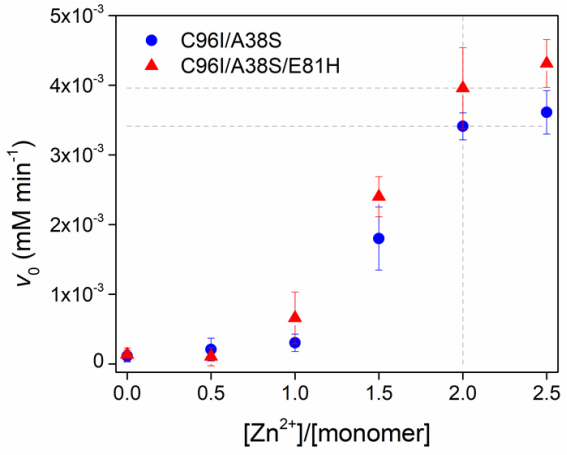


Fig. S13. The Zn-dependent ampicillin hydrolysis activity assays of C96I/A38S and C96I/A38S/E81H.

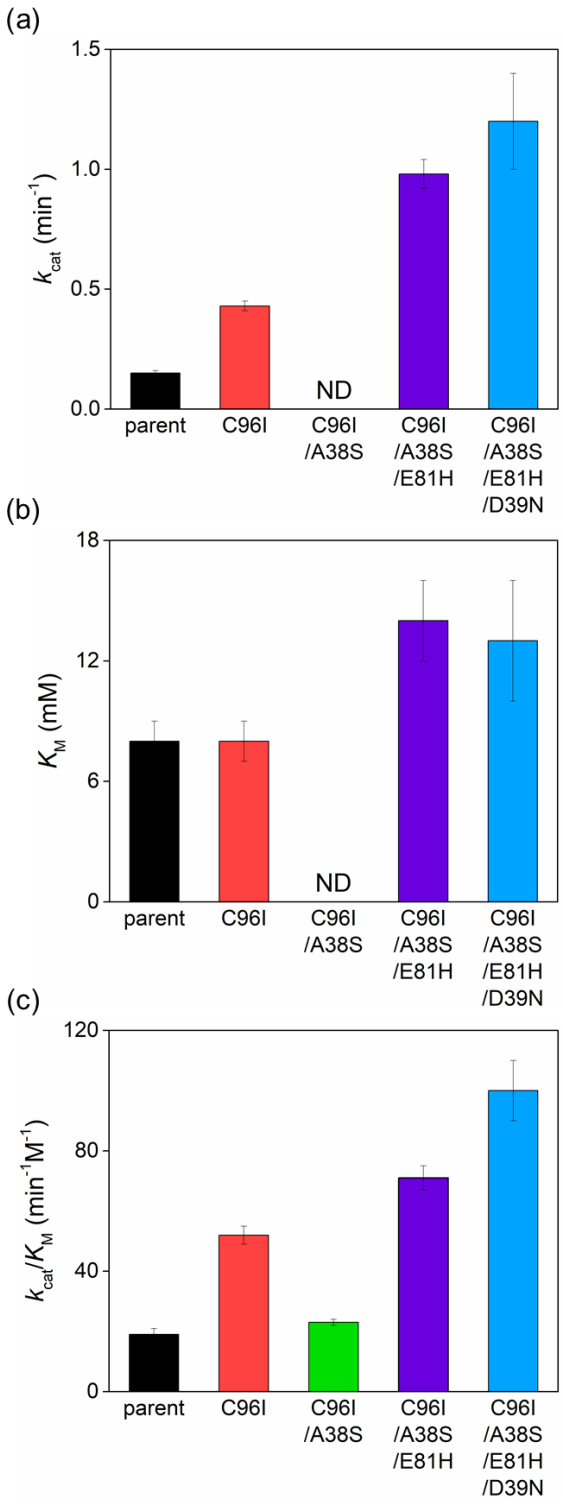


Fig. S14. Michaelis-Menten kinetic parameters of the variants evolved by symmetry-guided evolution. (a) Catalytic rate constant (k_{cat}) (b) Michaelis constant (K_M) (c) catalytic efficiency (k_{cat}/K_M). In C96I/A38S, a second-order rate constant (k_2) was used to compare to k_{cat}/K_M values of other variants.

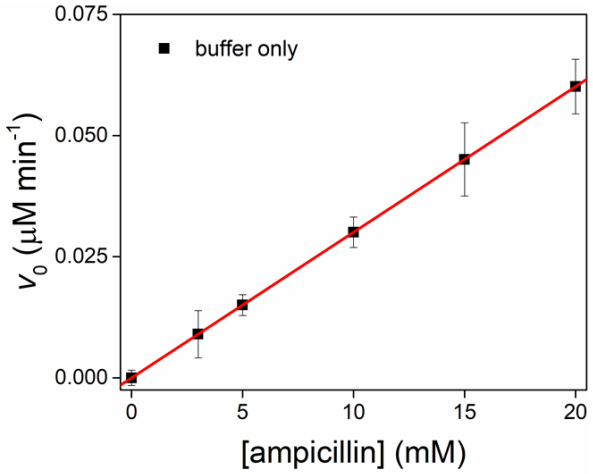


Fig. S15. The uncatalyzed hydrolysis of ampicillin monitored in 100 mM MOPS pH 7.0 buffer at room temperature.

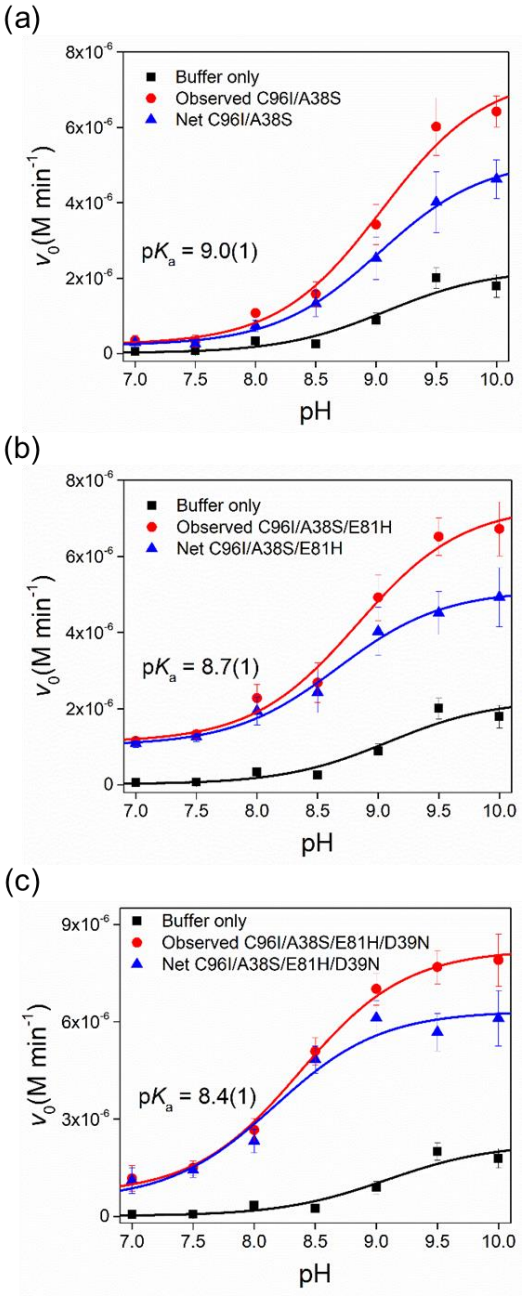


Fig. S16. The pH-dependent ampicillin hydrolytic activity to determine the pK_a values of evolved variants. (a) C96I/A38S (b) C96I/A38S/E81H (c) C96I/A38S/E81H/D39N.

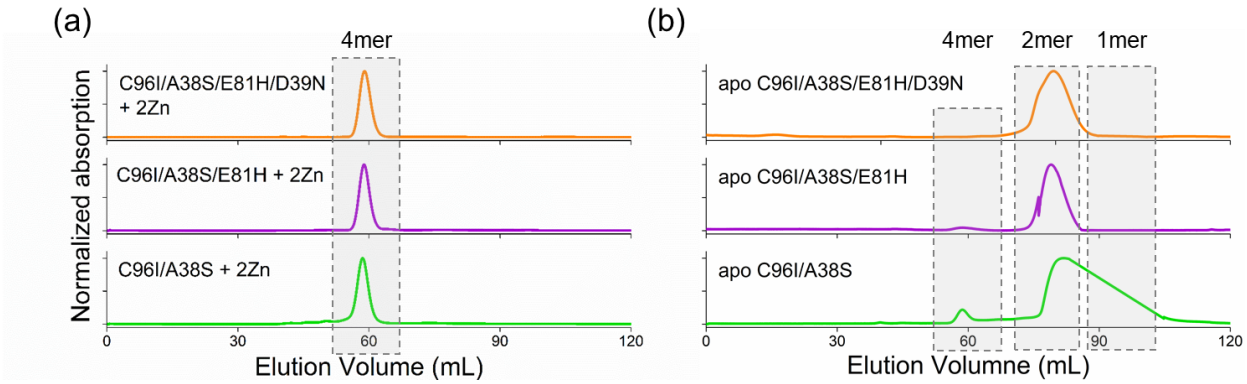


Fig. S17. Determination of oligomerization states of the evolved variants by size exclusion chromatography. Chromatograms of (a) Zn-complexed and (b) Zn-free C96I/A38S, C96I/A38S/E81H, and C96I/A38S/E81H/D39N.

(a)

	Selected residue	Distance to side chain (Å)
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

(b)

	Selected residue	Distance to side chain (Å)
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 S1. Distances between active sites and selected residues for the evolution of AB5 and C96I. A distance was determined from the oxygen atom of zinc-coordinated water molecule to the terminal atom of the selected residues.

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	C97X	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	C97X	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 GTA AAA GAT GC-3' 5'- GC ATC TTT TAC TTT ACC MNN ATT TGC CAG GTG-3'	59.3–63.1	53.3, 54.3, 55.3, 56.3, 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 S2. DNA primers used for saturated mutagenesis.

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 S3. DNA primers for NGS analysis.

(a)

Protein	k_{cat} (min^{-1})	K_M (mM)	k_{cat}/K_M (min^{-1} M^{-1})
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)*

(b)

Protein	k_{cat} (min^{-1})	K_M (mM)	k_{cat}/K_M (min^{-1} M^{-1})
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 S4. Kinetic parameters of C96X variants (X= C, V, K, I, and L) with (a) ampicillin (b) carbenicillin. ND: not detected. Hydrolysis activities were measured under pH 7 conditions. *The second-order rate constant (k_2) instead of k_{cat}/K_M was determined due to the absence of saturation behavior.

(a)

Proteins	K_M (mM)	$K_{d,\ Tyr}$ (mM)	$K_{d,\ Trp}$ (mM)	$f_{a,\ Trp}$
C96	8(1)	10.5(3)	3.8(2)	0.62
C96V	5(1)	6.1(2)	0.7(1)	0.35
C96K	7(1)	7.5(2)	1.3(6)	0.27
C96I	8(1)	10.0(2)	1.3(2)	0.36

(b)

Proteins	K_M (mM)	$K_{d,\ Tyr}$ (mM)
C96	7(1)	4.0(1)
C96V	4(1)	2.0(1)
C96K	8(1)	5.1(1)
C96I	6(2)	4.4(1)
C96L	3(1)	1.9(1)

Table S5. Determination of substrate binding affinities of (a) ampicillin (b) carbenicillin.

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 (A)*	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 reflections	15931	9893	15537	14826
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)*
$< I / \sigma >$*	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	1658	1672	1669	1666
Ligands/ions	97	98	94	95
Water	194	89	244	107
B-factors (Å^2)	24.0	52.0	21.0	27.7
R.m.s deviations				
Bond lengths (Å)	0.0146	0.0107	0.0134	0.0114
Bond angles (^{o})	1.7092	1.6093	1.5992	1.8137
PDB accession code	6LDG	6LDF	6LDE	7DCL

Table S6. 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.

(a)

Proteins (PDB code)	Ligating residues	Interatomic distances between Zn^{2+} and N atom (Å)
C96 (5XZI)	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

(b)

Proteins (PDB code)	Ligating residues	Interatomic distances between Zn^{2+} and ligating (N) atom (Å)
C96 (5XZI)	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

Table S7. Structural properties of C96X variants. (a) Core/structural Zn sites (b) Periphery/catalytic Zn sites. 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.

Round	Protein	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 $(k_{cat}/K_M)/k_{uncat}$ (M^{-1})
0	C96	0.15(1)	8(1)	19(2)	5.0(3) x 10^4	6.3(7) x 10^6
1	C96K	0.33(2)	7(1)	50(4)	1.1(1) x 10^5	1.7(1) x 10^7
	C96I	0.43(2)	8(1)	52(3)	1.4(1) x 10^5	1.7(1) x 10^7
2	C96I /A38S	ND§	ND§	23(1)*	ND§	7.7(3) x 10^6
3	C96I /A38S /E81H	0.98(6)	14(2)	71(4)	3.3(2) x 10^5	2.4(1) x 10^7
4	C96I /A38S /E81H /D39N	1.2(2)	13(3)	100(10)	4.0(7) x 10^5	3.3(3) x 10^7

*For non-saturated behavior, a second-order rate constant (k_2) 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) x 10^{-6} min^{-1}. §ND: not determined due to the lack of saturation behavior.

Table S8. Catalytic activities of the evolved variants at pH 7.0

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)

*A second-order rate constant (k_2) was used instead of k_{cat}/K_M. [§]ND: not determined

Table S9. The pH-dependent hydrolytic activities of the evolved variants. The catalytically critical pK_a values were estimated from the non-linear fit analysis.