이학박사 학위논문

Development of a Gene-specific Mutator for Rapid Continuous Evolution of Proteins *in vivo*

세포 내 단백질의 빠른 진화를 위한 유전자 특이 돌연변이체 개발

2021 년 2 월

서울대학교 대학원

화학부

박 효 진

Development of a Gene-specific Mutator Enzyme for Rapid Continuous Evolution of Proteins *in vivo*

지도교수 김 석 희

이 논문을 이학박사 학위논문으로 제출함

2021 년 2 월

서울대학교 대학원

화학부 생화학 전공

박 효 진

박효진의 박사 학위논문을 인준함

2020 년 12 월

위 원 장 이형하 (인)

부 위 원 장 [illegible] (인)

위 원 [illegible] (인)

위 원 [illegible] (인)

위 원 이현수 (인)

Part I

Abstract

Development of a Gene-specific Mutator for Rapid Continuous Evolution of Proteins *in vivo*

Hyojin Park

Departments of Chemistry

The Graduate School

Seoul National University

Continuous directed evolution has emerged enabling the fast and efficient evolution of proteins in cells. Additionally, diverse *in vivo* mutagenesis methods have been developed. However, they still have limitations such as a low mutation rate, no targeting ability, or narrow editing window. Here, we report a mutator, eMutaT7, with a high mutation rate and and high gene-specificity in *Escherichia coli*. eMutaT7, a cytidine deaminase fused to T7 RNA polymerase, can introduce up to ~4 mutations per 1 kb per day, which is comparable to the rate in traditional *in vitro* mutagenesis.

eMutaT7 enables the rapid continuous evolution of proteins for antibiotic resistance and allosteric activation.

Keywords: Directed evolution, Gene-specific mutagenesis, Cytidine deaminase, T7 RNA polymerase, Continuous directed evolution, *in vivo* mutagenesis

Student Number: 2014-30078

Table of Contents

Lists of Figures

Introduction

Evolution has brought a tremendous diversity to life. Frances Arnold came up with the idea that the same principles could be applied to develop proteins that relieve the biological problems that the problems of humans. In 1993, Arnold performed the first directed evolution of enzymes able to catalyze chemical reactions (1). This finding has fueled great progress in protein chemistry. To date, directed evolution has paved the way to achieving biomolecules with the desired properties in industrial, research, and therapeutic applications (2).

Typical directed evolution requires *in vitro* diversification of the target gene, followed by ligation, transformation, and selection or screening of variants with the desired function (2). Evolutionary success relies largely on the generation of libraries covering all possible candidates, and lots of methods have been developed for gene diversification such as chemical mutagens, error-prone polymerases, mutator strains and several recombination methods (2). But they still suffer from a labor-intensive process, making directed evolution difficult.

Recently, continuous directed evolution (CDE) has emerged, transforming the labor-intensive phase into a single phase *in vivo* (2). In 2011, David Liu's group reported that phage-assisted continuous evolution (PACE), a pioneering example of CDE, overcomes the disadvantages of typical *in vitro* mutagenesis methods by using a mutator plasmid (MP6) and phage, but the system cannot avoid off-target mutations and needs a specialized device (3). In addition to PACE, multiple methods claiming targeted mutagenesis have appeared (e.g. base editor (4), CRISPR-X (5),

EvolvR (6), OrthoRep (7)). However, there was no method that truly targeted a specific gene. In 2018, John Dueber's group developed EvolvR by fusing an error-prone DNA polymerase and nCas9. Though they tried to target specific DNA with the guide RNA, at most, it can edit only 350 nucleotides from the gRNA binding site and even the distribution is uneven (6). Subsequently, Chang Liu's group reported OrthoRep, bypassing the off-target issue by using orthogonal plasmid-polymerase pairs in yeast (7). But in reality, it mutates the entire orthogonal plasmid, and still cannot target specific genes. In the same year, Shoulders' group reported MutaT7, a mutator strain expressing a *rat* cytidine deaminase apolipoprotein B editing complex 1 (rAPOBEC1) fused with a T7 RNA polymerase (T7RNAP) that targets a gene downstream of the T7 promoter. MutaT7 led us to a genuinely targeted continuous evolution era, but it showed a very low mutation rate, introducing 0.34 mutations per day in a 1-kb gene (8).

The ideal *in vivo* mutagenesis methods should cause mutations rapidly and evenly through the target gene. We reasoned that using orthogonal polymerase could be a solution for targeted mutagenesis. We also thought the mutation rate could be accelerated by using more efficient mutating enzyme and keeping the bacteria in a fast-growing state. Here, we report a simple gene-specific *in vivo* mutagenesis system that represents a dramatically high mutation rate in *E. coli*.

Results

eMutaT7 is a mutator plasmid expressing *Petromyzon marinus* cytidine deaminase (PmCDA1) fused to T7 RNA polymerase (T7RNAP). T7RNAP directs the mutating region between the T7 promoter and the T7 terminator and exposes ssDNA, creating an opportunity to be mutated by deaminases. For deaminase, we thought that using PmCDA1 could increase the mutation efficiency since PmCDA1 was previously reported to be more effective deamination rate than rApobec1 in E. coli. eMutaT7 is a fusion of PmCDA1 and T7RNAP (n-PmCDA1-T7RNAP-c) (Figure 1). The mutant gene is under the control of an arabinose inducible promoter. UGI, an inhibitor of uracil- DNA glycosylase (UNG), was inserted downstream of the mutant gene to inhibit the repair of uracil (15), increasing the efficiency of mutation.

We chose *pheS*_A294G gene, a mutant of *pheS* allele, as a model target. *pheS*_A294G-expressing cells induce cytotoxicity by incorporating *p*-chloro-phenylalanine (*p*-Cl-Phe) into its cognate $tRNA^{Phe}$ (16). Since the growth of pheS_A294G-expressing cells is inhibited by *p*-Cl-Phe, mutation caused on *pheS*_A294G can make cells viable (e.g. CAG → TAG stop codon) and the mutation efficiency can be measured by counting viable cells in the presence of *p*-Cl-Phe. To minimize pheS heterogeneity in a single cell, low copy plasmids were used to clone *pheS*_A294G. To be processed by T7 RNA polymerase, *pheS*_A294G gene was inserted between the T7 promoter and the T7 terminator and induced expression with Isopropyl *β*-D-1-thiogalactopyranoside (IPTG).

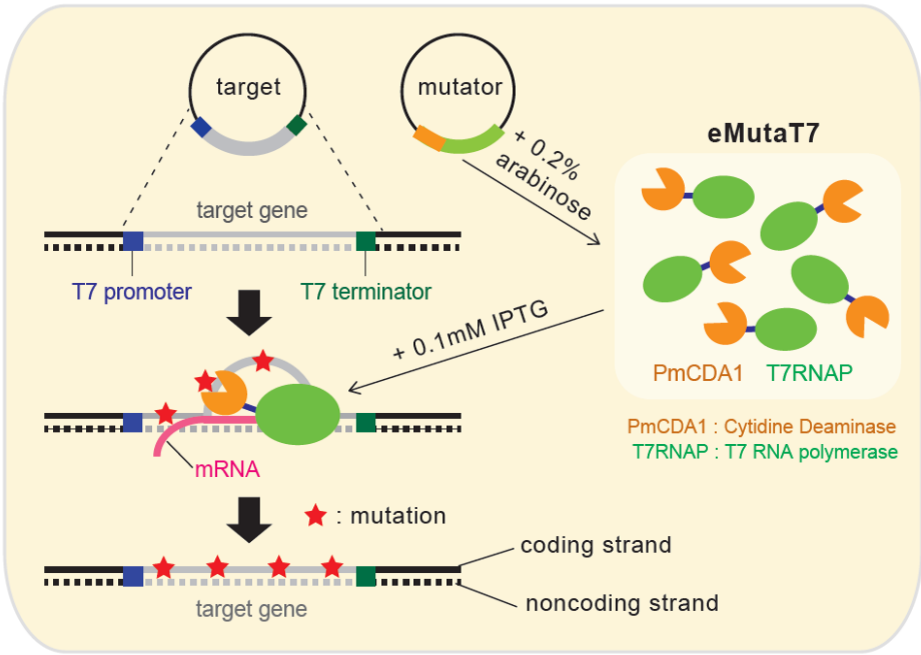


Figure 1. eMutaT7 induces rapid in vivo mutagenesis in the target gene.

Schematic diagram of the eMutaT7 system. The mutator and the target genes are induced with 0.2% arabinose and 0.1mM IPTG, respectively. T7 RNA polymerase of the mutator transcribes the target gene between the T7 promoter and the T7 terminator, exposing a short region of the coding strand to single-stranded DNA during transcription. The cytidine deaminase in the mutator can convert cytosine to uracil, primarily on the coding strand.

To validate *in vivo* mutagenesis, we constructed Δ*ung* strain transformed with the target plasmid and the mutator plasmid (eMutaT7) and performed mutagenesis by adding IPTG and arabinose (Figure 1). We also analyzed the expression of no mutator (empty plasmid), unfused PmCDA1 and T7RNAP and T7RNAP only as negative controls. Cells were diluted 100-fold with fresh medium every 4 hours to avoid reaching the stationary phase (Figure 2A). Cells obtained from each cycle were grown on agar plates with or without *p*-Cl-Phe to determine the level of suppression of toxicity of PheS_A294G (Figure 3). Cells with empty vectors did not express T7RNAP and were excluded from inhibition experiments as they were unable to express PheS_A294G under the control of the T7 promoter. As we expected eMutaT7-expressing cells showed a marked increase in the frequency suppressor over three cycles (12 hours) compared to cells-expressing unfused proteins or T7RNAP only (Figure 2B). The suppressor frequency is nearly saturated after three cycles, indicating that most of the cells either inactivate the pheS_A294G gene or have off-target mutations that suppress the PheS_A294G toxicity. Additional experiments showed that the wild-type strain showed slower increase in suppressor frequency than Δ*ung* strain, and our experimental set-up, growing cells in 96-well plate, did not affect the growing speed compared to growing in Erlenmyer flask and test tube (Figure 4).

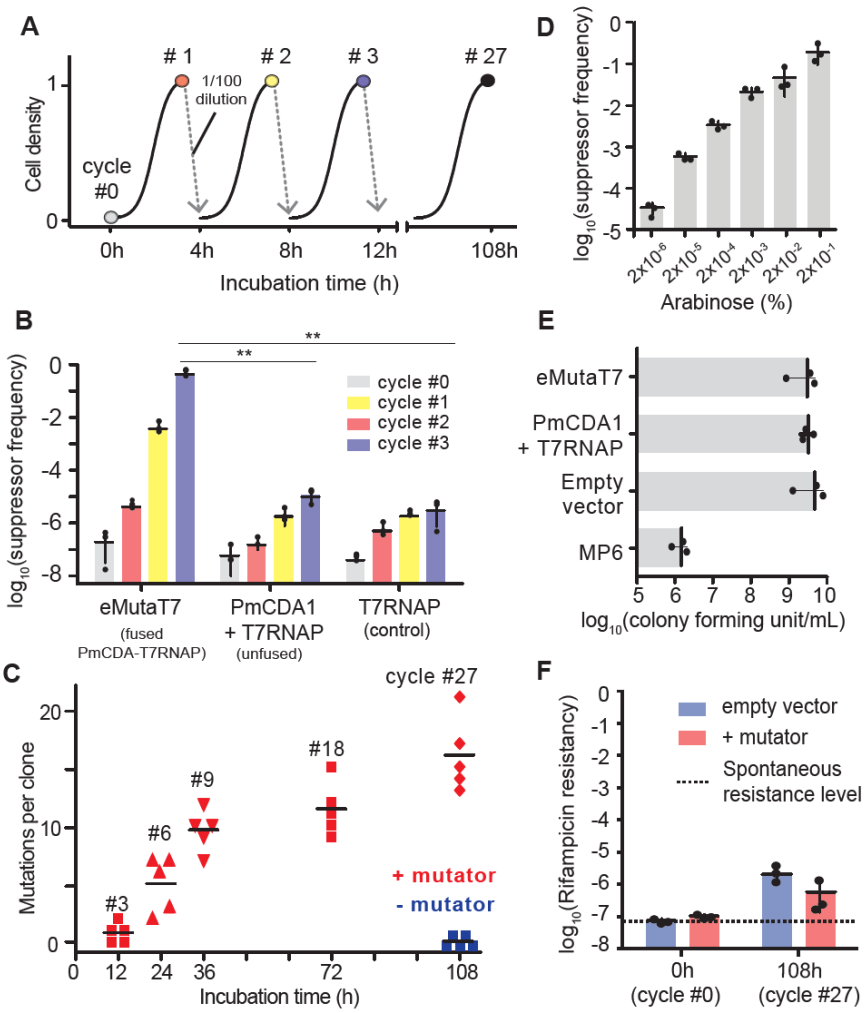


Figure 2. eMutaT7 induces rapid *in vivo* mutagenesis in the target gene.

(A) Schematic illustration of the mutagenesis cycle. (B) PheS_A294G suppressor frequency at each mutagenesis cycle for cells expressing PmCDA1-T7RNAP (eMutaT7), unfused proteins or T7RNAP only. (C) The number of mutations found in five clones in each mutagenesis cycle. (D) PheS_A294G suppressor frequencies with different arabinose concentrations. (E) Viability of cells expressing eMutaT7,

unfused protein, no protein, or MP6. (F) The level of off-target mutations estimated by rifampicin resistance frequency. The dotted line represents the level of spontaneous rifampicin resistance. Data are presented as dot plots of mean (short black line) ± 1 SD. n = 3 (B, D-F) or 5 (C); ** $p < 0.01$ by Student's *t*-test.

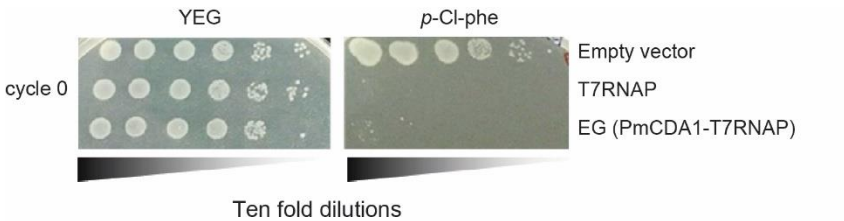


Figure 3. *pheS*_A294G/*p*-Cl-Phe selection system

The Δ*ung* cells carrying a vector expressing no protein, T7RNAP, or eMutaT7 were grown in LB media (with neither arabinose nor IPTG) until OD_{600} ~0.2. Serial 10-fold dilutions of cells were spotted and grown on YEG-agar plates supplemented with or without 1.6 mM *p*-Cl-Phe, 0.2% arabinose, and 0.1mM IPTG. Growth of cells expressing PheS_A294G is inhibited in the presence of *p*-Cl-Phe.

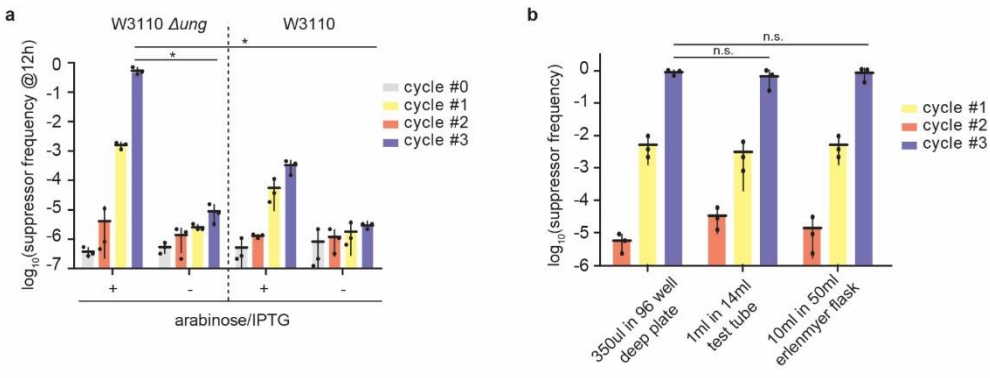


Figure 4. Culture conditions for eMutaT7

(a) eMutaT7 in the Δ*ung* strain makes more suppressors than in the wild-type strain (W3110). Wild-type cells (W3110) or the Δ*ung* cells carrying the mutator plasmid and a target (*pheS*_A924G) plasmid were grown in LB supplemented with (+) or without (-) 0.2% arabinose and 0.1mM IPTG. Suppressor frequency was calculated at the end of each cycle. (b) Cultures were grown in three different containers—350 µl in 96-well deep plate, 1 mL in 14-mL tube, and 10 mL in 50-mL Erlenmeyer flask—at 37°C with shaking at 150 rpm. *pheS*_A924G suppressor frequency reached 0.1-1 in 12 hours under all three conditions. Data are presented as dot plots with averages (short black lines) ± 1 SD (n = 3). **p* < 0.05 by Student's *t*-test.

Next, we checked the targeting ability of eMutaT7. We continued mutagenesis for 27 cycles (108 hours or 4.5 days) without *p*-Cl-Phe to avoid exerting selection pressure. Then, five clones were randomly selected at various cycles to analyze mutations in the target gene via Sanger sequencing. Surprisingly, the mutations accumulated very quickly. Average mutation rate was ~3.7 mutations•day^{-1}•kb^{-1} or ~9.4 x 10^{-5} mutations•base^{-1}•generation^{-1} at ~1.1 kb of the target gene region (Figure 2C and 5A). This mutation rate is 7 to 11 times higher than that of MutaT7 (11), and is comparable to typical *in vitro* mutagenesis methods for directed evolution experiments (typically 2-5 mutations•gene^{-1}) (30). In contrast, no mutation was found in cells grown without eMutaT7 it (Figure 2C). The number of mutations on target gene increased with the number of cycles, and it means eMutaT7 was not affected by itself during the mutagenesis cycles. As anticipated in Figure 1, the C→T mutations (208 mutations, 96%) dominated G→A (8 mutations, 4%). Mutations are quite evenly distributed throughout the target gene region (Figure 5B).

The high mutation frequency in the genome can cause cytotoxicity or surpass the essential mutations (31). However, in eMutaT7 system, since mutator expression is induced by arabinose, the mutation rate can be controlled by altering the arabinose amount. Indeed, the smaller the amount of arabinose, the lower the suppressor frequency, indicating that *in vivo* mutagenesis by eMutaT7 is tunable (Figure 2D). MP6-using mutagenesis generates random mutations in genomes, but reduced cell viability 1000-fold (Figure 2E and 6). However, eMutaT7 represented no significant decrease in cell viability, meaning that eMutaT7 had no special toxicity (Figure 2E). To assess the off-target mutagenesis in genomes, we determined the frequency of rifampicin resistancy in cycle 27 in the presence or absence of

mutator. We revealed that the frequency of rifampicin resistance was similar between cells with and without eMutaT7, suggesting that genome-wide off-target mutagenesis is not frequent in eMutaT7 (Figure 2F). We also tested the fusion of PmCDA1 to the C-terminus of T7RNAP, various linker lengths between PmCDA1 and T7RNAP, T7RNAP variants that could lower transcriptional elongation rate to increase the time taken for deamination (32,33), but the original form was the best among all the variants we have tested (Figure 7).

a

Time(h)-clone #	C-to-T mutations	mut #	G-to-A mutations	mut #	total #	average	mutation rate
12-1		0					
12-2	T126T, T251T	2					1.6 mut•day^{-1}
12-3	-80*	1			4	0.8	
12-4	T221T	1					3.7 x 10^{-5} mut•bp^{-1}•gen^{-1}
12-5		0					
24-1	S2L, L4L, Q41X, T177T, Q200X, P204S	6	M178I	1			
24-2	Q41X, A76V, I133I, Q169X, T221T, T251T, F316F	7					5 mut•day^{-1}
24-3	Q69X, T129T, I188I, R195C, T251T, +32*	6			25	5	
24-4	A56V, T129T, F234F	3					1.1 x 10^{-4} mut•bp^{-1}•gen^{-1}
24-5	L71L, Q200X	2					
36-1	Q41X, N61N, A73V, P131S, Q169X, R195C, I218I	7					
36-2	L4L, A9A, Q69X, L106L, T126I x2, T162I, S219S, T221T	10					
36-3	A13V, A56V, Q66X, H166Y, I175I, R176C, R195C, P249L	8	P249L	1	48	9.6	6.5 mut•day^{-1}
36-4	-42*, A22A, R28C, R53C, A76V, F158F, G172G, R176C, T221T, F295F	10					1.5 x 10^{-4} mut•bp^{-1} gen^{-1}
36-5	-39*, -36*, Q41X, I133I, D135D, F140F, D154D, Q169X, R176C, R195C, N196N, T251T	13					
72-1	-36*, L4L, Q41X, T129T, A151V, Q169X, R176C, Q208X, T221T, P277L	10	M/I	1			
72-2	-39*, S16S, Q41X, R111C, T126T, T129T, F140F, P146L, I218I, P277L, L293L, +35*	12					3.8 mut•day^{-1}
72-3	-39*, -36*, S/S*, A9A, Q41X, A73V, T110I, R115C, T129T, I175I, R192C, N217N, T251T, S314L, L323L	15			57	11.4	8.7 x 10^{-5} mut•bp^{-1}•gen^{-1}
72-4	Q17X, P108L, T110I, P131L, H138Y, P150L, H155Y, S246L, P287L	9					
72-5	A9A, A76V, D114D, Q169X, Q182X, I218I A255V, R313C, F316F	9	E30K	1			
108-1	A22A, T39T, I60I, Q69X, T126T, T129T, I218I, P249S, T251T, N278N, S314L, +3*, +31*	13	A88S	1			
108-2	L4L, Q17X, Q41X, A85V, R111C, P146S, P131L, R176C, Q208X, T216T, N217N, T251T, +24*, +35*	14	L270L	1			
108-3	-39*, A14V, L96L, P108S, L143L, H155Y, R186C, T216T, T221T, L227L, T251T, A255V, L267L, P277L, R313C, S314L	16	-57*	1	81	16.2	3.6 mut•day^{-1}
108-4	-36*, A9A, A22A, Q69X, R99C, P108S, R111C, I113I, R115C, T129T, R164C, R195C, H207Y, T221T R244C, T251T, N278N, P287L, L293L, F316F, +31*	21	A255T	1			8.2 x 10^{-5}mut•bp^{-1}•gen^{-1}
108-5	L4L, A9A, A56V, G104G, P131L, Q169X, R192C, N217N, T221T, H228Y, T251T, I285I, L323L	13					
total (average)		208		8	216		(4.1 mut•day^{-1}) (9.4 x 10^{-6} mut•bp^{-1}•gen^{-1})

Amino acids were numbered based on pheS gene sequence and after stop codon of pheS gene
+/- means nucleotide number before start codon (-) or after stop codon (+) of pheS gene.
Silent mutations are presented in grey and stop codon is abbreviated as X.
* mutations not made on pheS gene orf but made between a T7 promoter and a T7 terminator.

b

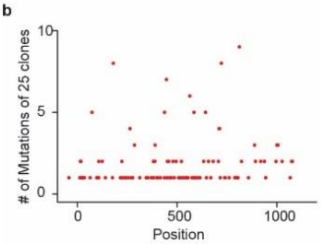


Figure 5. Related to Figure 2C. Mutation analysis by Sanger sequencing

(a) A list of mutations found in samples shown in Figure 1D. (b) Distribution of

mutations from 25 clones listed in Figure 5A. X-axis presents the reading frame of *pheS*_A924G.

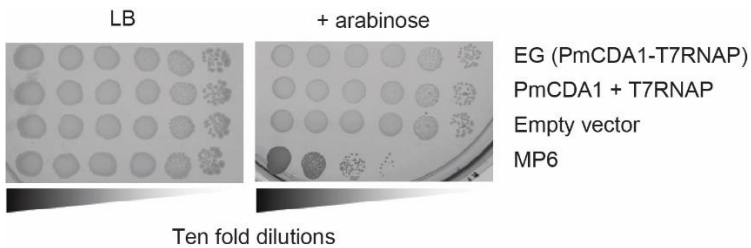


Figure 6. Related to Figure 2E. Toxicity test

The Δ*ung* cells (cHYO057) carrying a vector expressing eMutaT7, unfused PmCDA1 and T7RNAP, no mutator (negative control), or MP6 were grown on LB-agar plate (+35 μg/mL chloramphenicol) supplemented with or without 0.2% arabinose.

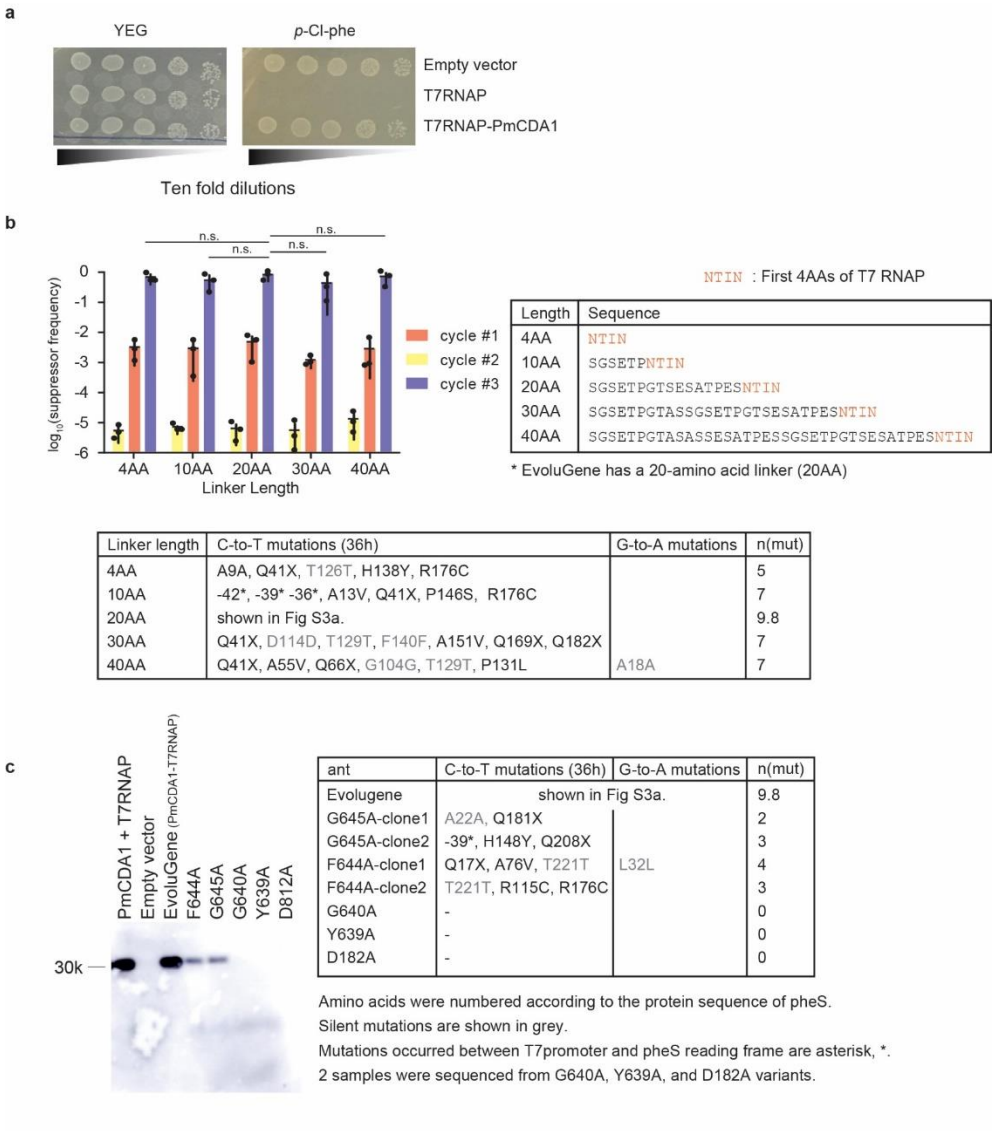


Figure 7. eMutaT7 optimization

(a) PmCDA1 fused to the C-terminus of T7RNAP does not suppress the *pheS*_A924G toxicity (1.6 mM *p*-Cl-Phe), probably because the inward facing of the T7RNAP C-terminus might disrupt proper folding or function of the fusion protein.[40] (b) Variations of linker length between PmCDA1 and T7RNAP (top right) do not

alter the *pheS*_A294G suppressor frequency (top left). One clone from each sample (36 hr) was sequenced and the detected mutations are listed (bottom) except eMutaT7. (c) Transcription efficiency of mutators with less processive T7RNAP variants was assayed indirectly by measuring the amount of expressed FLAG-tagged *PheS*_A294G using western blot with a FLAG-tag antibody (left). Of five less processive variants, two and three variants appear to be less efficient and nonfunctional, respectively. Two clones from each less-efficient variant (36 hr) were sequenced and the detected mutations are listed (right).

To thoroughly analyze gene specificity, we sequenced ~3.3kb DNA around the target gene from cells harvested at cycle 27 by Illumina sequencing. We revealed that of all substitution types, only the C→T and G→A mutations accumulated significantly in the presence of eMutaT7 (1.72% and 0.22%, respectively; Figure 8A). Of the three regions: upstream, target gene and downstream, the target gene region had the highest average frequency (3.6%) of C→T mutations compared to the upstream (0.27%) and downstream (0.96%) regions (Figure 8B). These results indicate that eMutaT7 preferentially modifies the target gene. Though eMutaT7 generally increased the frequency of mutations in all three regions, in particular, the downstream region showed a higher mutation frequency than the upstream region, suggesting there was a leakage of termination. It does not hampered the evolution of target gene and reduced by incorporation of multiple terminators. However, we believe that off-target mutations were not introduced randomly into the genome, as shown by the low rifampicin resistance frequency (Figure 2F). Interestingly, we found that the G→A mutation was mainly abundant in the ~280 bp region at the transcription start site (Figure 8C). One possibility is that the longer the newly synthesized RNA, the better it protects the template DNA strand (non-coding strand) from PmCDA1-mediated deamination. We also found that pyrimidine-C-purine (Y<u>C</u>R) is a preferred but not exclusive sequence for eMutaT7 (Figure 9).

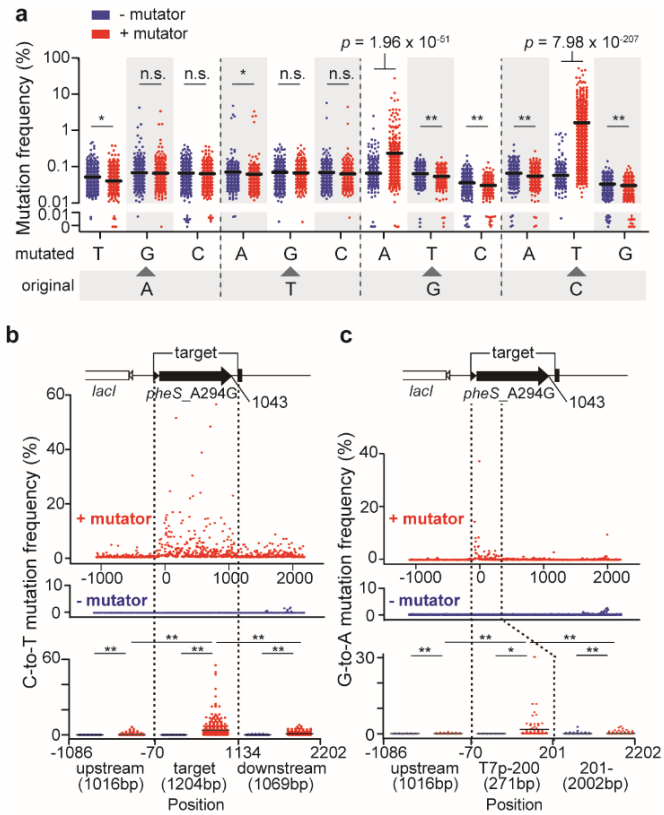


Figure 8. Illumina sequencing demonstrates gene specific mutagenesis of eMutaT7.

(**A**) Mutation frequency of all possible substitution types with or without eMutaT7. Y-axes above and below 0.01% are in log- and linear-scale, respectively. (**B**, **C**) Frequency of all C→T (**B**) and G→A (**C**) mutations in ~3.3 kb DNA around the target gene with or without eMutaT7. Collective mutation frequencies in upstream, target (C→T) or -70–200 (G→A), and downstream regions are shown as dot plots with averages (short black lines). *$p < 0.05$, **$p < 0.0001$ by Mann-Whitney's *t*-test.

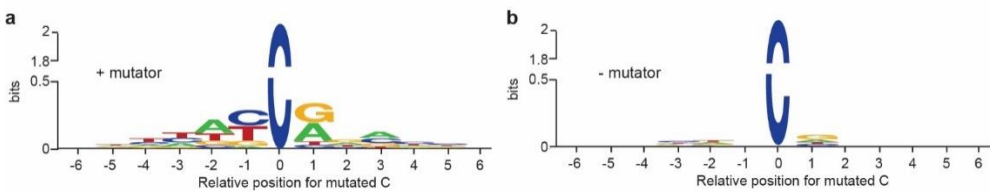


Figure 9. Mutation hotspot analysis

Analysis of the Illumina sequencing data with samples grown with (a) or without (b) the mutator for 27 cycles reveals that eMutaT7 prefers pyrimidine-C-purine (YCR) for mutagenesis. Ratios of the C→T mutation at specific C positions were used to generate Sequence logos (WebLogo, http://weblogo.berkeley.edu).

The above results suggest that the mutation rate of eMutaT7 is faster than that of MutaT7, but the apparent discrepancy may occur depending on the growth conditions, strains and target genes. To directly compare eMutaT7 and MutaT7, we made two plasmids expressing MutaT7 under the control of the araBAD promoter which is used for eMutaT7 or A1lacO promoter which is originally used for MutaT7 system. To efficiently generate the G→A mutations as well as C→T mutations, we used dual promoter/terminator system adopted by Shoulders group, where the target gene is sandwiched by two pair of T7 promoter/terminator in the reverse direction (Figur 10A) (11). This approach suppressed pheS_A294G toxicity rivalling the previous approach, but introduced more G→A mutations (24%; Figure 11). We express the eMutaT7 or MutaT7 from any promoter which has applied to three independent cultures of the three strains prepared, and each of the nine strains period of 4 hours and the growth was diluted 100 times (36 hours). We randomly selected two colonies from each batch, and sequenced the target *pheS*_A294G gene (total six sequences in each strain). P_{BAD}_eMutaT7, P_{BAD} _MutaT7 P_{A1lacO}_MutaT7 produced 6.8, 1.0, and 0.33 mutations, respectively on average. (Figure 10B and 12). It demonstrates that eMutaT7 actually has a higher mutation rate than MutaT7.

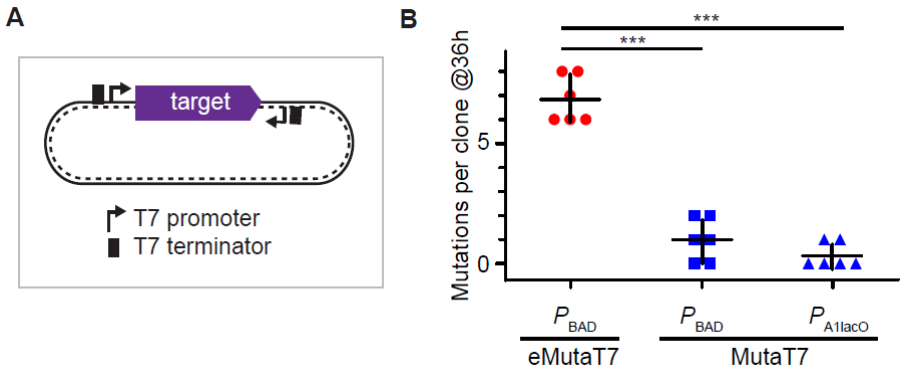
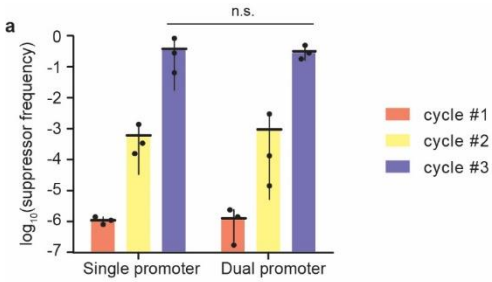


Figure 10. Comparison of eMutaT7 and MutaT7.

(A) Dual promoter/terminator system. (B) eMutaT7 produces more mutations compared to MutaT7. Cells carrying P_{BAD}_eMutaT7, P_{BAD}_MutaT7 and P_{A1lacO}_MutaT7 were subjected to evolution experiments for 36 hours. The number of mutations found in the 6 clones is shown as a dot plot of the mean (short black line) ± 1 SD. ***P <0.001 by Student's *t*-test.

a



b

Clone #	C-to-T mutations (36h)	n(C/T)	G-to-A mutations	n(G/A)	total
1	-39*, L4L, L106L, I133I, T170T, R195C, T216T, T303I	9		0	9
2	A5V, Q208X	3	A63A, L71L, W159X	3	6
3	L4L, A5V, A88V, +32*	4	L71L, D135N, W159X	3	7
4	A22A, R176C, Q41X, T163T, H207Y, I218I, T251T, I285I	8	V259I	1	9
5	T/S*, A22A, Q69X, A180A, F295F	5	A11A, E37K	2	7
total		29		9	38
average		5.8		1.8	7.6

Amino acids were numbered based on pheS gene sequence and after stop codon of pheS gene
+/- means nucleotide number before start codon (-) or after stop codon (+) of pheS gene.
Silent mutations are presented in grey and stop codon is abbreviated as X.
* mutations not made on pheS gene orf but made between a T7 promoter and a T7 terminator.

Figure 11. Dual promoter system generates more G→A mutations without compromising mutation rate

(a) The *pheS*_A924G suppressor frequency of the dual promoter system is comparable to that of the single promoter system. Statistical analysis was performed by Student's *t*-test. (b) A list of identified mutations in five clones after 36-hour *in vivo* mutagenesis with dual promoter system.

Mutator	clone	C-to-T mutations	G-to-A mutations	total #	average	mutation rate
eMutaT7	1	Q182X, T221T, T251T	L32L, R46H, E77K A255A, E300A	8		
(P_{BAD}-PmCDA1-T7RNAP)	2	R195C, T216T, T303I	L71L, D135N, E300A	6		
	3	T44I, F119F, T170T, F234F	E30K, A63A, G98S	7	6.8	4.5 mut•day^{-1}
	4	-61*, T44I, T221T	E30K, R46H, W159X	6		
	5	S2L, L4L, A73V	A23T, G298G, E300A	6		
	6	-12, Q69X, T163T, H207Y, I218I	V259I, F295F, E300A	8		
P_{BAD}-MutaT7	1	L4L, T44I	-	2		
(P_{BAD}-rApobec1-T7RNAP)	2	F119F	-	1		
	3	-	-	0	1.0	0.67 mut•day^{-1}
	4	-	D93N	1		
	5	T44I	R53H	2		
	6	-	-	0		
MutaT7	1	-	-	0		
(P_{A1lacO}-rApobec1-T7RNAP)	2	-	R46H	1		
	3	-	-	0	0.33	0.22 mut•day^{-1}
	4	I133I	-	1		
	5	-	-	0		
	6	-	-	0		

Amino acids were numbered based on pheS gene sequence and after stop codon of pheS gene
+/- means nucleotide number before start codon (-) or after stop codon (+) of pheS gene.
Silent mutations are presented in grey and stop codon is abbreviated as X.
* mutations not made on pheS gene orf but made between a T7 promoter and a T7 terminator.

Figure 12. Related to Figure 10. Mutation list.

A list of mutations found in samples shown in Figure 4.

Next, we tested applicability of eMutaT7 for the evolution of proteins using a dual promoter/terminator system. We evolved a class A β-lactamase, TEM-1, for resistance to the third generation cephalosporin antibiotic, cefotaxim (CTX) and ceftazidim (CAZ) (Figure 13A). We tested multiple antibiotic concentrations at the same time and the cells grown in the highest concentration were subjected to the next round of evolution in fresh media with higher amounts of antibiotics (Figure 13A). The minimum inhibitory concentration (MIC) increased from 0.06 μg/ml to 400–800 μg/ml in 32 h *in vivo* CDE for CTX and from 0.2 μg/ml to 2000–4000 μg/ml in 24 h for CAZ (Figures 13B, 13C and 14). These results indicate a much faster evolution of antibiotic resistance than previous reports using conventional directed evolution methods. 3-5 rounds of *in vitro* mutagenesis, transformation, and selection for CTX (34, 35) and 3-4 rounds for CAZ (36, 37). Sanger sequencing of five colonies shows that all samples share two previously reported mutations (E102K and G236S for CTX, E102K and R162H for CAZ) located near the substrate binding site, which can alter the substrate specificity of the enzyme. Turned out (Figures 13D and 13E). We also tried to evolve AmpC, a class C β-lactamase, for resistance against carbenicillin (CB), and the MIC increased from 16 μg/ml to 2000–4000 μg/ml in 28 hours (Figure 15). Although most clinical isolates resistant to ampicillin carry promoter-region mutation (38), some mutations (e.g. G214R/S, D217N, and E196K) were found near the substrate binding site in our experiments (Figure 15).

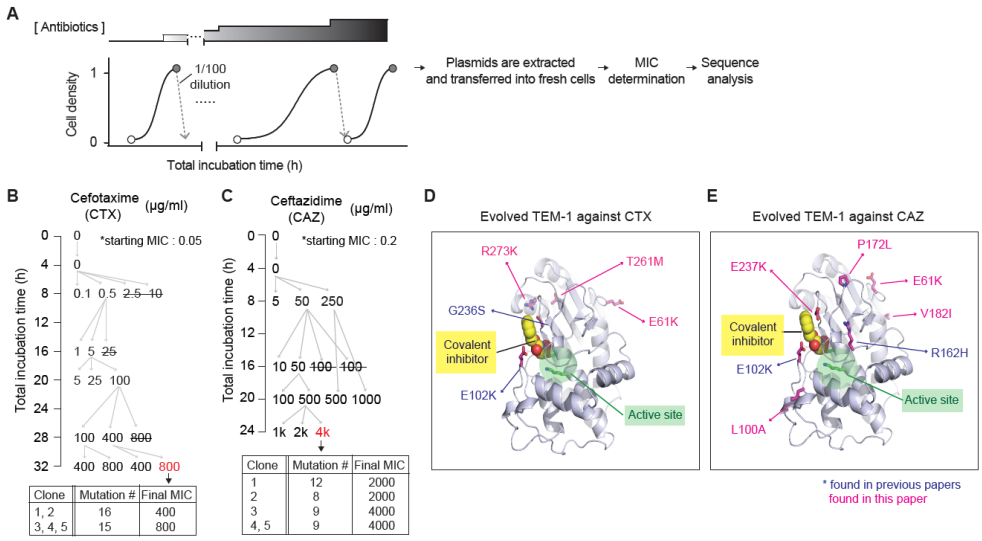


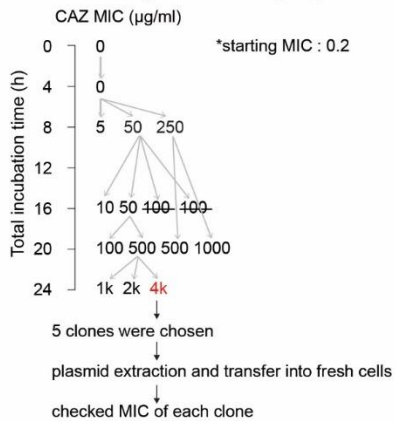
Figure 13. eMutaT7 promotes rapid continuous directed evolution of model proteins.

(A) TEM-1 evolution workflow. **(B, C)** Evolutionary pathway of TEM-1 for antibiotic resistance. Each number indicates a CTX **(B)** or CAZ **(C)** concentration in a culture. Strikethrough indicates no growth. Grey indicates that two cultures out of three did not grow. **(D, E)** Structure of TEM-1 (PDB, 1axb) with a covalent inhibitor (yellow), the active site (green), and mutations found in the evolved TEM-1 (blue and pink).

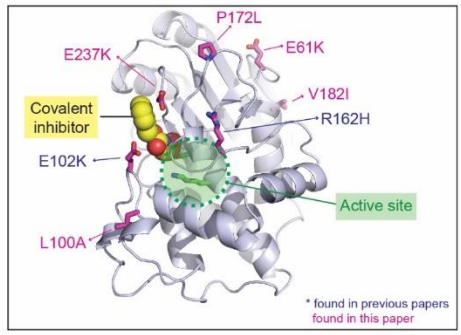
a Mutation list of evolved TEM-1 against cefotaxime (CTX)

Clone (#)	C-to-T mutations	n(C/T)	G-to-A mutations	n(G/A)	n(total)	MIC (μg/ml)
1, 2	V8V*, P12L*, L49L, S80S, A84A, L89L, L137L, T261M	8	L47L, E61K, L100L, E102K, S221S, G236S, R273K, E277E	8	16	400
3, 4, 5	V8V*, P12L*, L49L, S80S, A84A, L89L, L137L, T261M	8	L47L, L100L, E102K, S221S, G236S, R273K, E277E	7	15	800

b TEM-1 evolution against ceftazidime (CAZ)



d



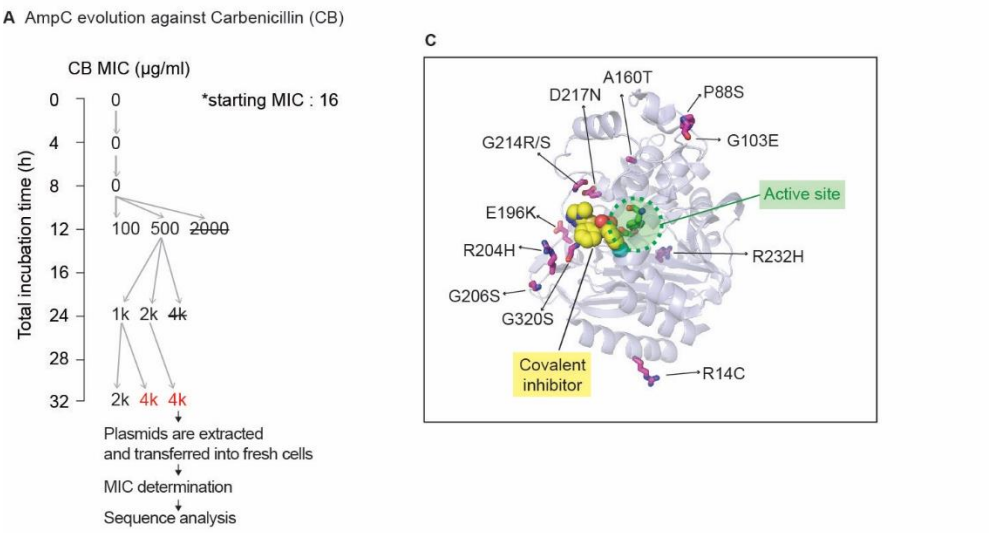
c

Clone (#)	C-to-T mutations	n(C/T)	G-to-A mutations	n(G/A)	n(total)	MIC (μg/ml)
1	I45I, L49L, L89L	3	L38L, L47L, A77A, E102K, R162H, K109K, V182I, L282L, E237K	9	12	2000
2	I45I, L49L, L137L, L219L	4	L100A, E102A, R162H, E237K	4	8	2000
3	T147T, A235A	2	L47L, E61K, L100A, E102K, K109K, L137L, R162H	7	9	4000
4, 5	I45I, L49L, L89L, P172L	4	L38L, E102K, R162H, V182I, E237K	5	9	4000

Amino acids were numbered according to the protein sequence of TEM-1 including signal sequence.
Silent mutations, grey ; previously known mutations, blue ; newly found in this paper, pink.
* mutations made on signal sequence (24AA).

Figure 14. TEM-1 evolution by eMutaT7

(**A**) Related to Figure 13B. A list of mutations found in five clones of TEM-1 variants evolved against cefotaxime (CTX). (**B**) Related to Figure 13C. A list of mutations found in five clones of TEM-1 variants evolved against ceftazidime (CAZ).



B Mutation list of evolved AmpC against carbenicillin (CB)

Clone(#)	C-to-T mutations	n(C/T)	G-to-A mutations	n(G/A)	n(total)	MIC(ug/ml)
1	T81T, I83I, S153S	3	A160T, G214R, T278T, G320S	4	7	4000
2	A49A, T97T, I155I, A231A	4	A160T, G214R, D217N, I243I	4	8	4000
3	S66S, I155I, A231A	3	G214R, D217N, V234V, I243I	4	7	4000
4	T90I, A135A	2	G103E, L168L, G214R, D217N	4	6	4000
5	R14C, P88S	2	L168L, E196K, R204H, G206S, G214S, D217N, R232H	7	9	2000

Amino acids were numbered according to AmpC sequence.
Silent mutations are depicted as grey.

Figure 15. AmpC evolution by eMutaT7

(**A**) AmpC evolution workflow. (**B**) A list of mutations found in five clones of AmpC variants evolved against carbenicillin (CB). (**C**) Structure of AmpC (PDB, 6dpt) with a covalent inhibitor (yellow), the active site (green), and mutations found in the evolved TEM-1 (black).

Finally, we tried to find an allosteric mutation on DegP, a major heat shock protease in bacterial periplasm. The proteolytic activity of DegP should be carefully regulated to maintain cellular fitness under misfolded protein stress. Lower and higher activity of DegP reduce bacterial fitness by misfolded protein stress and hyperactive proteolysis, respectively (16). Many allosteric mutations that lower activity have been found, but only one activating mutation has been reported (16). We begins with a less active variant, DegP_A184S, which reduced cell viability at high temperature and evolved new variants that restore the cellular fitness even at high temperatures by gradually increasing the temperature during the growth cycle (Figure 16A). We assumed that the new variants should have an activating mutation, increasing lowered proteolytic activity of DegP_A184S. DegP genes from five clones were sequenced and a common mutation P231L was found (Figure 16A and 17). Indeed, adding just P231L to DegP_A184S restored cell viability at high temperatures (43 ℃), and P231L alone reduced cell viability (Figure 16B). In addition, *in vitro* enzymatic assay showed that the addition of P231L increased the basal activity of DegP_A184S to wild-type DegP levels, indicating a rebalance of proteolytic activity (Figure 16C). P231 residue is not directly in contact with the substrate binding site near the center of the trimer DegP to suggest that the P231L positioned to increase the allosteric activation (Figure 16D). Collectively, we have demonstrated that mutations allow rapid in vivo directed evolution of target proteins given the conditions of continuous selection.

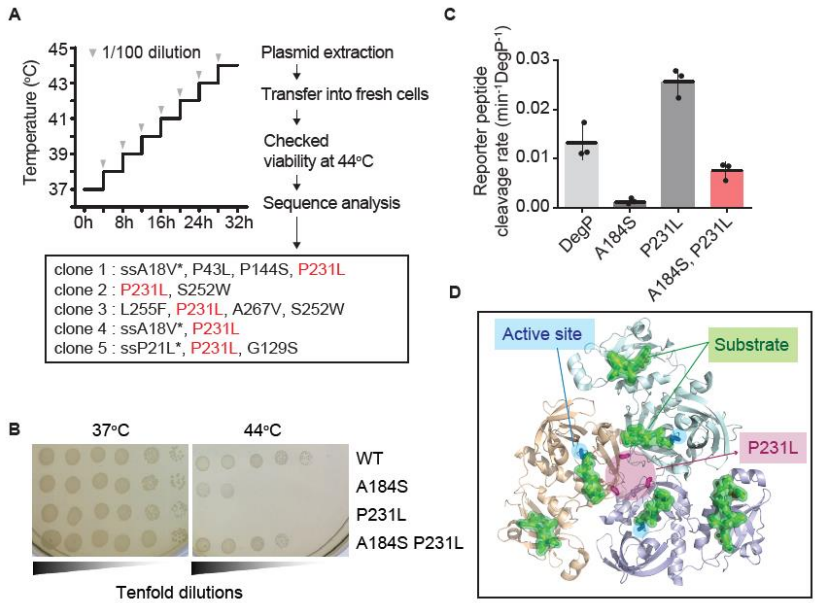


Figure 16. Identification of an activating mutation in the DegP protease by eMutaT7.

(**A**) Workflow of the DegP_A184S evolution. Mutations found in five clones are listed and a common mutation, P231L, is represented as red. ssA18V* and ssP21L* are mutations in the signal sequence, which is required for protein translocation to the periplasm and cleaved after translocation. (**B**) Addition of P231L in DegP_A184S restores the cell viability at high temperature (44ºC), while P231L alone reduces cell viability. (**C**) Basal activities of various DegP mutants (10 μM) were determined by cleavage of the reporter peptide (100 μM). Data are presented as dot plots with averages (short black lines) ± 1 SD (n = 3). (**D**) Structure of the DegP trimer (PDB, 3otp) with active sites (blue), substrates (green), and the P231 residue (magenta).

e DegP Mutation list from dual promoter system

Clone(#)	C-to-T mutations	n(C/T)	G-to-A mutations	n(G/A)	n(total)
1	Signal sequence, P43L, F46F, P144S, I228I, P231L	6		0	6
2	I228I, P231L, I242I, S252W	4		0	4
3	L247L, L255F, P231L(2), A267V, T251T, S252W	7		0	7
4	Signal sequence, P231L, L294L	3		0	3
5	Signal sequence, P231L	2	G129S	1	3

Figure 17. A mutation list of DegP_A184S evolution by eMutaT7

Related to Figure 16A. A list of mutations found in five clones of DegP_A184S variants evolved against heat stress.

Finally, we attempted to find an allosteric mutation of DegP, the major heat-shock protease in bacterial periplasm. The proteolytic activity of DegP is carefully controlled to maintain cellular fitness under heat stress: the lower and higher activity reduce bacterial fitness by misfolded protein stress and hyper-proteolysis, respectively (16). Although many activity-lowering allosteric mutations have been found, only one activating mutation was reported (16). We started with a less active variant, DegP_A184S, which reduced the cell viability at high temperature, and evolved new variants that support better cellular growth at high temperature, by increasing temperature stepwise during growth cycles (Figure 6A). We expected that the new variants should have a compensating mutation that increases the proteolytic activity of DegP_A184S. Sanger sequencing of the *degP* gene from five colonies revealed P231L as the only common mutation (Figure 6A and Supplementary Figure 11). Indeed, the addition of P231L in DegP_A184S restored the cell viability at high temperature (43ºC), and P231L alone reduced cell viability (Figure 6B). Also the *in vitro* enzymatic assay showed that P231L raises the basal activity, and the addition of P231L increased the basal activity of DegP_A184S to the level of wild-type DegP, indicating the rebalancing of proteolytic activity (Figure 6C). The P231 residue is located near the center of the trimeric DegP without direct contact with the substrate, suggesting that P231L allosterically increases activity (Figure 6D). Collectively, we demonstrated that our mutator allows rapid *in vivo* directed evolution of a target protein, given the continuous selection condition.

Discussion

We developed eMutaT7 that allows rapid continuous directed evolution of proteins through gene-specific *in vivo* mutagenesis by more efficient cytidine deaminase fused to an orthogonal RNA polymerase. To our knowledge, eMutaT7 has the highest mutation rate among the targeted *in vivo* mutagenesis methods for CDE of the protein. Also, our method is (1) simple: Two plasmids expressing the mutator and the target are sufficient for gene-specific CDE, and this method can be used in any laboratory with a basic molecular biology setup. These plasmids can be transformed to any strain which needs for the specific condition. (2) Expandible: Multiple genes under the control of the T7 promoter can be targeted. It can be applicable to metabolic engineering. (3) Tunable: The mutation rate can be controlled by the arabinose concentration. The main limitation of our method is the narrow mutation spectrum, predominated by C→T and G→A mutations. However, 15 of 20 amino acid can be replaced by C→T and G→A mutations and we believe that combination with other DNA-modifying enzymes would help to expand the substitution types. Alternatively, gene libraries built by *in vitro* mutagenesis could be a starting point for CDE, where our *in vivo* mutagenesis method explores an additional layer of sequence space.

In addition, the selection of target proteins is limited because it is difficult to find a suitable method of selection. eMutaT7 has already reached near maximum mutation rates (2-5 mutations per cycle) during evolution , other mutations do not neutralize essential mutations. Thus, eMutaT7 can replace the usual labor-intensive

in vitro diversification method, and the selection step can also be replaced by a screening method that can be applied to more proteins.

Materials and Methods

Plasmid cloning and materials

E.coli strains, plasmids, and primers used in this study are listed in Supplementary Table 1, 2, and 3, respectively. The mutator plasmids expressing eMutaT7 (pHyo094) and MutaT7 (pHyo099) were obtained as follows: The deaminase gene that expresses *Petromyzon marinus* cytidine deaminase (PmCDA1) and *rat* apolipoprotein B mRNA editing catalytic polypeptide-like (rApobec1) were amplified from AIDv2 (Addgene #79620) and pCMV-BE3 (Addgene #73021), respectively (3,13). T7 RNA polymerase gene, in which NdeI site was deleted using inverse PCR (14), was amplified from pNL001 (15). Two DNA fragments were linked by overlapping PCR with a linker gene for the 16-amino acid XTEN linker (3) between two genes. The fused DNA expressing PmCDA1-XTEN-T7RNAP was inserted into pSK633 (16) between the NdeI and PstI sites. The XTEN linker of pHyo099 was exchanged into the linker of MutaT7 (pHyo287) and the araBAD promoter of pHyo287 was exchanged into the A1lacO promoter (pHyo288) by inverse PCR.

All target plasmids (pHyo182, *pheS*_A294G; pHyo245, *pheS*_A294G with the dual promoter system; pHyo249, *degP*_A184S with the dual promoter system; pHyo253, *TEM-1* with the dual promoter system) were constructed in a low copy-number plasmid (one or two copy per a single cell; the parent plasmid is pNL001 (15)). To make *pheS*_A294G gene transcribed with T7 RNA polymerase (pHyo175, pET28b-*pheS*_A294G), *pheS*_A294G gene was amplified from the SK324 strain (17) and inserted into pET28b vector using NdeI and XhoI sites. DNA fragment from *lacI*

to *pheS*_A294G gene containing T7-lac promoter and T7 terminator (lacI-T7-lac promoter-*pheS*_A294G-T7 terminator) was amplified from pHyo175, and inserted in the pNL001 vector between BamHI and EcoRI sites to make pHyo182 (pVS133-*pheS*_A294G). To make dual promoter system (pHyo245), a new T7-lac promoter and T7 terminator sequences were inserted in front of T7 terminator and between T7 promoter and *lacI* gene region, respectively, in the reverse direction.

For *degP* evolution experiments, *pheS*_A294G was replaced with *degP*_A184S amplified from pSK735 (16) and wild-type *degP* amplified from p7 (18) to make pHyo249 (*degP*_A294G) and pHyo246 (*degP*), respectively, using NEBuilder Hifi DNA assembly master Mix (New England Biolabs, USA).

For *TEM-1* evolution experiments, ampicillin resistance gene in the target plasmid was replaced with tetracycline resistance gene amplified from pREMCM3 (19) and *TEM-1* gene was replaced with *pheS*_A294G gene to make pHyo253, using NEBuilder Hifi DNA assembly master Mix.

All plasmids harboring variants of the mutator or targets (mutation, deletion, and insertion) were constructed using inverse PCR. eMutaT7 variants (low processive variants or linker length variants) were made on pHyo094. The *degP* variants were made on pHyo246 and pHyo249 to make pHyo254 (*degP*_P231L) and pHyo255 (*degP*_A184S, P231L), respectively, for *in vivo* viability assay, and on p7 and pSK735 to make pHyo256 (*degP*_P231L) and pHyo257 (*degP*_A184S_P231L), respectively, for *in vitro* enzymatic assay.

All PCR experiments were conducted with KOD Plus neo DNA polymerase (Toyobo, Japan). Restriction enzymes, T4 polynucleotide kinase and T4 DNA ligases were purchased from Enzynomics (South Korea). Plasmids and DNA fragments were prepped with LaboPass™ plasmid DNA purification kit mini, LaboPass™ PCR

purification kit, and LaboPass™ Gel extraction kit (Cosmogenetech, South Korea). Sequences of genes in plasmids constructed in this study were confirmed by Sanger sequencing (Macrogen, South Korea). Antibiotics (ampicillin, chloramphenicol, kanamycin, and tetracycline), arabinose, and IPTG were purchased from LPS solution (South Korea). Cefotaxime and ceftazidime were purchased from Tokyo chemical industry (Japan). H-*p*-Chloro-DL-Phe-OH (*p*-Cl-Phe) was purchased from Bachem (Switzerland).

E.coli strain construction

All the strains in this study were constructed by λ-Red-mediated recombineering with the pSIM6 plasmid.[34] To construct the Δ*ung* strain (cHYO057), kanamycin resistance gene region including 100-bp upstream and 100-bp downstream was amplified with primers 180 and 181, and the open reading frame (ORF) of wild-type *ung* in W3110 was replaced by the kanamycin resistance gene. To construct the Δ*degP* Δ*ung* strain (cHYO059), SK345 (Δ*degP*) (31) was used as a recipient cell. Proper gene deletion or insertion were confirmed by colony PCR using 2X TOP Simple™ DyeMIX-Tenuto (Enzynomics), and 30 μg/mL of kanamycin was used for selection.

Western blot

Cells expressing N-terminal FLAG-tagged *pheS*_A924G were grown to OD_{600} ~ 0.6 in LB at 37°C. 30-μL samples of the cultures were mixed with 4X SDS sample buffer, heated at 95 °C for 15 min, and centrifuged at 13,000 × g for 10 min. Samples (7 μL) were separated by SDS/PAGE and transferred to PVDF membrane (ATTO, Japan).

The membrane was sequentially incubated in Ezblock protein-free Blocking buffer (ATTO), anti-OctA (FLAG octapeptide) antibody (2,000:1 dilution; SantaCruz, USA), and anti-rabbit IgG-HRP (2,000:1 dilution; SantaCruz), and then blotted with the WestSaveUp ECL solution (AbFrontier, South Korea). Image was obtained using Amersham imager 600 (GE healthcare, USA).

Mutation cycle

The Δ*ung* cells (cHYO057) harboring a mutator plasmid (pHyo094) and a target plasmid (pHyo182 for single promoter and pHyo245 for dual promoter) were grown overnight in LB broth supplemented with 100 μg/mL ampicillin and 35 μg/mL chloramphenicol (cycle #0). The overnight cultures (3.5 μl) were mixed with 350 μl of LB broth supplemented with 100 μg/mL ampicillin, 35 μg/mL chloramphenicol, 0.2% arabinose, and 0.1mM IPTG in a 96-deep well plate (Bioneer, South Korea) and incubated at 37ºC for 4 hours (cycle #1). The dilution-incubation cycle was repeated up to 27 times. At the end of each cycle, a fraction of cells were stored at -80ºC as 15% glycerol stocks.

To determine the number of accumulated mutations on the *pheS*_A294G gene, cells at the selected mutagenesis cycles (cycle #3, #6, #9, #18, and #27) were streaked on LB-Agar plate containing 100 μg/mL ampicillin. Five colonies were randomly chosen for plasmid isolation and the sequence determination. Mutations were counted in the region from 80-bp upstream to 35-bp downstream of the *pheS* gene (total 1099 bp). Primers 507 and 529 was used for sequencing of target plasmids that have dual promoter system, and universal sequencing primers were used for sequencing of plasmids that have single promoter system (primers T7 promoter and

T7 terminator in Table S3).

PheS_A294G suppression assay

Samples obtained at the endpoint of each cycle (overnight culture for cycle #0) were diluted to OD_{600} ~ 0.2. Serial 10-fold dilutions of cells (5μl) were spotted on YEG-agar plates and grown on either selective (1.6 mM *p*-Cl-Phe, 0.2% arabinose, and 0.1 mM IPTG) or nonselective (no additives) YEG-agar plates at 37ºC overnight. Viable cells on each condition were counted and the suppressor frequency was calculated as N_1/N_0 (N_1 : colony forming unit (CFU) in selective plates and N_0 : CFU in nonselective plates).

Viability assay

Overnight cultures expressing eMutaT7, unfused PmCDA1 and T7RNAP, no mutator, or MP6, were diluted 100-fold in LB broth supplemented with 35 μg/mL chloramphenicol, grown to a log phase (OD_{600} = 0.2-0.5) at 37°C, and diluted to OD_{600} ~ 0.2. Serial 10-fold dilutions of cells (5μl) were spotted on LB-agar plates supplemented with 0.2% arabinose and 35 μg/mL chloramphenicol and grown at 37°C for 16 hours. Viable cells were counted to calculate CFU/mL (Figure 1f and S4).

To determine the rifampicin resistance, samples after cycle #0 and #27 were grown to log phase in LB broth supplemented with 100 μg/mL ampicillin and 35 μg/mL chloramphenicol, and subjected to viability assay on selective (80 μg/mL rifampicin) or nonselective (no rifampicin) plates.

High-throughput sequencing and data analysis

Cells taken at cycle 27 were grown in 15 mL of LB broth without arabinose and IPTG, and the target plasmids were extracted with Plasmid Miniprep Kit. The 3289-bp DNA fragments containing the *pheS*_A924G gene were amplified using primers 512 and 513 covering a region ranging from 1016 bp upstream from T7 promoter to 1069 bp downstream including T7 terminator. Samples were then prepared with TruSeq Nano DNA Kit and sequenced on Hi-seq sequencer (Illumina, USA; operated by Macrogen) in 2 x 101 paired-end runs using the manufacturer's regents following the manufacturer's protocol to determine the mutation pattern occurred on target plasmids.

Raw reads were trimmed to remove adapter and low quality end sequences using Trimmomatic v.0.36 palindrome mode with followed option; ILLUMINACLIP:Adapter.fa:2:30:10:8:true LEADING:15 TRAILING:15 SLIDINGWINDOW:4:15 MINLEN:36.[10] Trimmed reads were aligned against the target sequence using Burrows-Wheeler Alignment tool (BWA) v0.7.17 with mem mode and BAM files generated by mapping were sorted using SAMtools v.1.6.[11-12] Sorted bam files were subject to perform mpileup using SAMtools mpileup with maximum depth option, which was set as total number of trimmed reads, and output tag list option consisting of DP, DP4 and AD. BCFtools v1.6, which was a set of utilities of SAMtools package, was used to call allele for each locus with multiallelic-caller option. Allele count for each allele and ratio (each allele count / Total allele count) were calculated based on AD information of VCF files.

Mutation rate calculation

Mutation rate caused by eMutaT7 was calculated and presented as two styles: (i) mutations per base per generation or (ii) mutations per day per 1 kb. Mutations were counted in the 1099-bp region of *pheS* for Sanger Sequencing. The number of generations per one cycle was calculated as 6.6 ($\log_2 100$) because cell density increases 100-fold during one cycle (4 hours). To present in the first style, the number of mutations was averaged for chosen clones and divided by the number of nucleotides (1099) and generations (6.6 x cycle number). To present in the second style, the number of mutations was averaged and divided by incubation days (6 cycles·day^{-1}) and the length of the gene region (1.099 kb). To compare the values with those of MutaT7, the number of mutations were counted from Figure S10 of the Shoulders group's report (11). Mutation rates of TRACE were estimated from Figure 2a and Supplementary Figure 3b of the Chen group's report (12).

Statistical analysis of high-throughput sequencing data

For high-throughput sequencing data (Figure 2), Mann-Whitney test (unpaired Wilcoxon test) was used to assess the significance of the mutation frequency caused by eMutaT7 system. Calculation was conducted using Stata (USA). Statistically significance were determined with p values defined as $*p < 0.05$, $**p < 0.0001$ for this experiment. For other data, we assumed the data will follow normal distribution and performed Student's t-test.

TEM-1 evolution

For TEM-1 and AmpC evolution, all the experiments were performed with LB supplemented with 6 μg/mL tetracycline, 35 μg/mL chloramphenicol, 0.2% arabinose, and 0.1mM IPTG. We tested multiple antibiotic concentrations at the same time, of which the cells grown at the highest concentration were subjected to the next round of growth in fresh media with equal or higher amounts of antibiotics. The concentration of each antibiotic (cefotaxime, ceftazidime, or carbenicillin) was gradually increased as indicated in Figure 5A, Supplementary Figure 10. All the experiments were conducted in triplicate and if two out of three samples did not grow, one sample that was grown was inoculated into three samples for the next cycle. After the evolution, plasmids were extracted in bulk from evolution mixture grown in LB media supplemented only with 6 μg/mL tetracycline and inserted into fresh Δ*ung* cells (cHYO057) harboring the T7 RNA polymerase-expressing plasmid (pHyo183). Ten colonies were chosen to determine MICs, and plasmids from the 5 colonies with high MIC values (400-800 μg/mL for CTX, 2000-4000 μg/mL for CAZ and CB) were analyzed by Sanger sequencing. At the end of each cycle, cells were stored at -80°C as 15% glycerol stocks.

MIC determination

Each strain was grown overnight without selection pressure, inoculated at 1/10,000 dilution into fresh LB broth with increasing concentrations of antibiotics in 96-well culture plates (SPL, South Korea), and grown at 37°C with shaking (150 rpm) overnight with aluminum foil cover to prevent too much evaporation. Cell density (OD_{600}) was measured with M200 microplate reader (TECAN, Switzerland).

DegP_A184S evolution

For *degP* evolution, all the experiments were performed with LB supplemented with 100 μg/mL ampicillin, 35 μg/mL chloramphenicol, 0.2% arabinose, and 0.1mM IPTG. At the end of each cycle, cells were stored at -80ºC as 15% glycerol stock. We grew Δ*degP* cells carrying pHyo249 (DegP_A184S-expressing plasmid) and pHyo094 (eMutaT7-expressing plasmid) at 37ºC. The growth temperature increased by 1ºC every 4-hr cycle and no cell death was detected while increasing to 44ºC (Figure 6A). After the evolution was over, the plasmids were extracted and retransformed into a fresh cells (cHYO059). Ten colonies were chosen and the cell viability was tested on LB-Agar plates at 44ºC, where originally *degP*_A184S could not grow. Plasmids from five colonies grown at 44ºC were analyzed by Sanger sequencing (Supplementary Figure 11).

The *degP*_P231L and *degP*_A184S P231L variants were freshly constructed (pHyo254 and 255 for cell viability test and pHyo256 and 257 for *in vitro* enzymatic assay) to remove possible mutations that could have occurred in other regions of the evolved plasmid.

To test viability of the Δ*degP* cells (cHYO059) expressing *degP*_A184S variants, cells were grown on selective (44ºC) and nonselective (37ºC) LB-agar plates for 16 hours.

DegP protein preparation and *in vitro* enzyme activity assay

DegP and its variants (P231L, A184S, P231L/A184S) were expressed and purified as previously described[32]. *In vitro* enzymatic assays were performed as previously described[32, 39]. *In vitro* enzymatic assays were performed at 37°C in 50 mM sodium

phosphate (pH 8) and 100 mM NaCl. Basal activities of DegP variants (10 µM) were monitored by increase of fluorescence after cleavage of the reporter peptide (100 µM; excitation, 320 nm; emission, 430 nm) using Infinite F200Pro microplate reader (TECAN).

Strain	Description	Reference
W3110		
SK324	W3110 *degP::PheS_A294G-kan^R*	(2)
SK345	W3110 *ΔdegP*	(2)
cHYO057	W3110 *Δung::kan^R*	This study
cHYO059	W3110 *ΔdegP, Δung::kan^R*	This study

Table 1. *E.coli* strains used in this study

Plasmid	Construct	Description	Reference
pNL001		T7 RNA polymerase gene/cloning vector	(3)
AIDv2		Source of PmCDA1 gene	(4)
pCMV-BE3		Source of rApobec1 gene	(5)
pREMCM3		Tetracycline resistance gene	(6)
pBAD33		Experimental control/cloning vector	
pET28b		Cloning vector	
pSK735	pET15b-DegP_A184S	DegP_A184S overexpression	(7)
p7	pET15b-DegP	DegP overexpression	(8)
pHyo094	pBAD33-PmCDA1-T7RNAP, ugi	eMutaT7	This study
pHyo099	pBAD33-rApobec1-T7RNAP, ugi	rApobec1 version of eMutaT7	This study
pHyo175	pET28b-pheS_A294G	Cloning intermediate	This study
pHyo182	pVS133-lacI, pheS_A924G	Target plasmid (pheS_A924G)	This study
pHyo183	pBAD33-T7RNAP, ugi	Control	This study
pHyo188	pBAD33-PmCDA-GX-T7RNAP, ugi	Linker length 30AA	This study
pHyo189	pBAD33-PmCDA-GGX-T7RNAP, ugi	Linker length 40AA	This study
pHyo191	pBAD33-PmCDA, T7RNAP, ugi	Control	This study
pHyo203	pBAD33-PmCDA-X-	Low processive mutatnt	This study

	T7RNAP (I810S), ugi		
pHyo223	pBAD33-PmCDA-X-T7RNAP (F644A), ugi	Low processive mutatnt	This study
pHyo227	pBAD33-PmCDA-X-T7RNAP (G645A), ugi	Low processive mutatnt	This study
pHyo237	pBAD33-PmCDA-X-T7RNAP (Y639A), ugi	Low processive mutatnt	This study
pHyo238	pBAD33-PmCDA-X-T7RNAP (G640A), ugi	Low processive mutatnt	This study
pHyo239	pBAD33-PmCDA-X-T7RNAP (D812A), ugi	Low processive mutatnt	This study
pHyo240	pBAD33-PmCDA-6aa-T7RNAP, ugi	Linker length 10AA	This study
pHyo241	pBAD33-PmCDA-0aa-T7RNAP, ugi	Linker length 4AA	This study
pHyo243	pVS133-PT7_FLAG_ pheS_A924G	For western blot	This study
pHyo245	pVS133-dualT7_ pheS_A924G	Dual promoter system	This study
pHyo246	pVS133-ss-DegP		This study
pHyo249	pVS133-ss-DegP_A184S	Evolution target (degP_A184S)	This study
pHyo253	pVS133-dualT7_ss-TEM-1, tetR	Evolution target (TEM-1)	This study
pHyo254	pVS133-ss-DegP_P231L		This study
pHyo255	pVS133-ss-DegP_A184S, P231L		This study
pHyo256	pET15b-DegP_P231L	Overexpression of DegP_P231L	This study
pHyo257	pET15b-DegP_A184S,	Overexpression of	This study

	P231L	DegP_A184S P231L	
pHyo287	pBAD33-rApobec1-T7RNAP	With the linker used in MutaT7	This study
pHyo288	pBAD33-PA1lacO-rApobec1-T7RNAP (araBAD promoter-deleted)	araBAD promoterr was exchanged with PA1lacO	This study

* ss : signal sequence

Table 2. Plasmids used in this study

Oligonucleotides	Sequence (5'→3')	Description
T7promoter	TAATACGACTCACTATAGGG	Universal sequencing primer
T7terminator	GCTAGTTATTGCTCAGCGG	Universal sequencing primer
pBAD-F	ATGCCATAGCATTTTTATCCA	Universal sequencing primer
pBAD-R	GATTTAATCTGTATCAGG	Universal sequencing primer
157_pSK633_NdeI_fw	CATATGggcaaaaaaaccacattagc	Deletion of NdeI site on vector for cloning
158_NdeI_pSK633_rv	CCATGGtgaattcctcctGag	Deletion of NdeI site on vector for cloning
159_PmCDA1_Pst1_rv	atactgcagttaaacggctggagacttagtg	Cloning arabinose-inducible PmCDA1 (pHyo101)
165_PmCDA1_XTEN_rv	ggtgtggcggactctgaggtcccgggagtctcgctAacggctggagacttagtgg	Amplification of PmCDA from AIDv2 gene to make T7RNAP fusion linked with XTEN linker (pHyo094)
167_rApobec1_XTEN_rv	ggtgtggcggactctgaggtcccgggagtctcgcTttcaacccggtggccc	Amplification of PmCDA from AIDv2 gene to make T7RNAP fusion linked with XTEN linker (pHyo099)
170_XTEN_T7RNAP_fw	gggacctcagagtccgccacaccgaaagt**a**Aacacgattaacatcgctaagaacg	Amplification of T7RNAP from pNL001 to make PmCDA1 fusion linked with XTEN linker (pHyo094)
175_T7RNAP_D812_fw	GAC TCC TTC GGT ACC ATT CCG	Cloning low-processive T7RNAP mutants (pHyo203, 209)

176_T7RNAP_I810S_rv	GTG AGC CAG TGC AAA AGA TTC GAT TCC GTA C	Cloning low-processive T7RNAP mutants (pHyo203)
177_T7RNAP_H811A	ATC AAT CAG TGC AAA AGA TTC GAT TCC G	Cloning low-processive T7RNAP mutants (pHyo209)
180_ung-up44_kanR-up100_fw	GAAGCAGTTAAGCTAG GCGGATTGAAGATTCG CAGGAGAGCGAGCGA AAACTCACGTTAAGGG ATTTTG	Amplification of kanamycin resistance gene from pET28b for construction of ung k/o strain
181_kanR-down100_ung-up44_rv	ATCAGCCGGGTGGCAA CTCTGCCATCCGGCAT TTCCCCGCAAATGGCA CTTTTCGGGGAAATGT G	Amplification of kanamycin resistance gene from pET28b for construction of ung k/o strain
184_PstI_ugi_fw	atactgcagaccaacctttccgac atcatag	Amplification of ugi gene from AIDv2 for insertion of UGI into plasmids (pHyo094, pHyo191)
185_ugi_HindIII_rv	ataaagcttttacagcatcttgatct tgttctctc	Insertion of UGI into plasmids (pHyo094, pHyo191)
187_T7RNAP_PstI_rv	atactgcagttacgcgaacgcga agtcc	Cloning of mutator plasmid
188_PstI_10bp_ugi_fw	atactgcagcccggggatcctcta gagtcgacctgcgatgaccaacc tttccgacatcatag	Amplification of ugi gene from AIDv2(4), for insertion of UGI into plasmids (pHyo094, pHyo191)
200_T7RNAP_dNdeI_fw	Ctatgagtcttgtgatgtactggct g	Deletion of NdeI site on T7RNAP gene for cloning
201_T7RNAP_d	gtgtcaaccatagtttcgcg	Deletion of NdeI site on

NdeI_rv		T7RNAP gene for cloning
219_F_ov-lacI_dw_fw	gtgaagttaccatcacggaaaaaggttatgctgcttCCACCGGAAGGAGCTG	Amplification of F plasmid pNL001 (3) Insertion of pheS_A924G into F plasmid (pHyo182)
220_pET28b-RBS_rv	GAAGGAGATATACCATGGGCAGC	Amplification of F plasmid pNL001 (3)
221_T7t_up_fw	TAAGATCCGGCTGCTAACAAAGC	Insertion of T7terminator in upstream of T7 promoter
222_T7t_down_ov_F_rv	gtagcgaccggcgctcagttggaattcATCCGGATATAGTTCCTCCTTTCAG	
223_F_ov_rv	aagcagcataacctttttccgt	Amplification of F plasmid
224_F_ov_fw	gaattccaactgagcgcc	Amplification of F plasmid
227_F_AmpR_down_fw	cttggttaccgtgaagttaccatc	
229_Ptrc_lac_dw_atg_rv	CATggtctgtttcctgtgtg	
230_XTEN_rv	actttcgggtgtagcggactc	
281_FLAG_pmcda_fw	GAT TAC AAG GAT GAC GAC GAT AAGacagacgccgagtac	
283_RBS_down_rv	ggTGCTGCTGCCCATG	
284_PmCDA-fw	Acagacgccgagtacg	
285_pBAD_start_rv	ATGggtgaattcctcctGag	
286_T7_ov_pBAD_rv	cctgcaggtcgactctagaggatcccgggttacgcgaacgcgaagtc	

287_Pm_ov_pBAD_rv	cctgcaggtcgactctagaggatcccgggttaaacggctggagactagtgg	
289_pBAD_end_fw	cccggggatcctctagag	
374_T7RNAP_I810S_rv	GTGAGACAGTGCAAAAGATTCGATTCCGTAC	Cloning low-processive T7RNAP mutants (pHyo203)
375_T7RNAP_H811A_rv	CGCAATCAGTGCAAAAGATTCGATTCCG	Cloning low-processive T7RNAP mutants (pHyo209)
377_FLAG_taa_pBAD_fw	GATTACAAGGATGACGACGATAAGtaaCTGCAGgcatgcaagc	
378_FLAG_taa_ugi_fw	GATTACAAGGATGACGACGATAAGtaaCtgcagccgcgggga	
442_T7RNAP_F644A_fw	GCTggcttccgtcaacaagtg	Cloning low-processive T7RNAP mutants (pHyo223)
443_T7RNAP_G645A_fw	ttcGCTttccgtcaacaagtgctgg	Cloning low-processive T7RNAP mutants (pHyo227)
444_T7RNAP_Q643_rv	ctctttggacccgtaagcc	Cloning low-processive T7RNAP mutants (pHyo223, 227)
466_T7RNAP_S813_fw	tccttcggtaccattccggc	Cloning low-processive T7RNAP mutants (pHyo239)
467_T7RNAP_D812A_rv	CGCgtgaatcagtgcaaaagatcgat	Cloning low-processive T7RNAP mutants (pHyo239)
468_T7RNAP_Y639A_fw	GCTgggtccaaagagttcggc	Cloning low-processive T7RNAP mutants (pHyo237)
469_T7RNAP_G640A_fw	tacGCGtccaaagagttcggcttcc	Cloning low-processive T7RNAP mutants (pHyo238)

470_T7RNAP_A638_rv	agccagcgtcatgactga	Cloning low-processive T7RNAP mutants (pHyo237, 238)
493_F_ov_Ptet_fw	cttatgtcgtaattctcatgtttgacagcttatca	
494_Ptet_ov_F_rv	agctgtcaaacatgagaattacgacataagtccatcagttcaacg	
495_tet_ov_F_rv	acttggtctgacagtggagtggtgaatccgttagc	
496_F_ov_tet_fw	ttcaccactccactgtcagaccaagtttactcaac	
497_T7t_up_rv	cttgaggggttttttgctgaagaattcTTATcaactgagcgccgg	
502_FLAG_pheS_fw	GGTGATTACAAGGATGACGACGATAAGCATATGTCACATCTCGCAGAAC	Cloning of FLAG-tagged pheS_A924G plasmid from pHyo182 (pHyo243)
505_T7p_back_rv	CACTATAGGGGAATTGTGGCTTTGTTAGCAGCCG	Insertion of another T7promoter before T7terminator into pHyo182 (pHyo244)
506_T7p_back_fw	AGTCGTATTAATTTCGGAAAGGAAGCTGAGTTGG	Insertion of another T7promoter before T7terminator into pHyo182 (pHyo244)
507_T7p_lacO_fw	CACTATAGGGGAATTGTGAGC	Sequencing primer for dual promoter plasmids
512_lacI_mid_forNGS_fw	GATATTTATGCCAGCCAGCC	Amplification of DNA containing pheS_A924G gene for NGS
513_sopB_end_forNGS_rv	gacaggtctcgtgttccaac	Amplification of DNA containing pheS_A924G gene for NGS
514_T7t_50up_	GCCCCAAGGGGTTATG	Insertion of another

T7p_fw	CTAGCAACCGCACCTG	T7terminator before T7promoter into pHyo244 (pHyo245)
515_T7t_50up_T7p_rv	CTCTAAACGGGTCTTGAGGGGTTTTTTGCTGGCGCCTATATCG	Insertion of another T7terminator before T7promoter into pHyo244 (pHyo245)
516_RBS_ov_DegP_fw	TTTTGTTTAACTTTAAGAAGGAGATATACCATGAAAAAAACCACATTAGCACTG	Cloning of DegP gene into F plasmid by Hifi-assembly (pHyo246)
517_RBS_rv	GGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTC	Cloning of DegP gene into F plasmid by Hifi-assembly (pHyo246)
518_DegP_ov_pET28b_rv	TCAGTGGTGGTGGTGGTGGTGCTctcgaGgTTACTGCATTAACAGGTAGATGGTG	Cloning of DegP gene into F plasmid by Hifi-assembly (pHyo246)
519_pET28b_end_fw	cCtcgagAGCACCACC	Cloning of DegP gene into F plasmid by Hifi-assembly (pHyo246)
529_dual_T7terminator_rv	GGTTATGCTAGTTATTGCTCAGC	Sequencing primer for dual promoter plasmids
530_RBS_ov_AmpR_fw	TTTTGTTTAACTTTAAGAAGGAGATATACCAtgagtattcaacatttccgtgtc	Cloning of TEM-1 gene into F plasmid by Hifi-assembly (pHyo250)
531_pET28b_ov_AmpR_rv	CAGTGGTGGTGGTGGTGGTGCTctcgaGttaccaatgcttaatcagtgagg	Cloning of TEM-1 gene into F plasmid by Hifi-assembly (pHyo250)
534_DegP_P231L_fw	CTGGACGGCGGCAACATC	Cloning of DegP mutants (pHyo254, 255, 256, 257)
535_DegP_A23	TGCGAGGATCGCGGTG	Cloning of DegP mutants

0_rv		(pHyo254, 255, 256, 257)
525_DegP_A184S_fw	GGGATTGTCTCTTCTCTGGGGCGTAGC	Cloning of DegP mutants (pHyo259)
526_DegP_A184S_rv	GGAAGTTACCGTCTCGCC	Cloning of DegP mutants (pHyo259)
643_Apobec_MutaT7linker_rv	CACTCCCTCCGCTACCGCCACTCCCTCCGCTACCGCCTTTCAACCCGGTTGCCC	Exchanging XTEN linker with the linker used in MutaT7 (pHyo287)
635_T7RNAP_MutaT7linker_fw	GGAGCTCAAGAGGATACCATATGAACACGATTAACATCGCTAAGAAC	Exchanging XTEN linker with the linker used in MutaT7
636_PA1lacO_fw	GGATAACAATTTCACACAttttatcgcaactctctactgtttc	Exchanging araBAD promoter with the pA1lacO promoter used in MutaT7 (pHyo287)
637_PA1lacO-1_rv	GTAATTGTTATCCGCTCACAAATAAACACTCTttggtaacgaatcagacaattgac	Exchanging araBAD promoter with the pA1lacO promoter used in MutaT7 (pHyo288)
638_PA1lacO-2_rv	GCTCACAATTGAATCTAATTGTAATTGTTATCCGCTCACAAATAAAC	Exchanging araBAD promoter with the pA1lacO promoter used in MutaT7 (pHyo288)

Table 3. Primers used in this study

Part II.

Tripodal Lipoprotein Variants with C-Terminal Hydrophobic Residues Allosterically Modulate Activity of the DegP Protease

Abstract

DegP, a member of the highly conserved HtrA protease family, performs elaborately regulated degradation of misfolded proteins in the periplasm of Gram-negative bacteria. The allosteric conversion between the inactive and active conformations is a principal mechanism for careful controlling the proteolytic activity of DegP and maintaining optimal cellular fitness, but few molecules except the substrates are known as the allosteric modulators of DegP. Here, we show that a mutant variant of the outer membrane lipoprotein, Lpp^{+Leu}, can act as a novel allosteric effector that alters the dynamic range of DegP activity. The three leucines at the C-termini of the Lpp^{+Leu} trimer are key components for activity regulation. Selection experiments using the Lpp variant library show that Lpp variants with varying sequences at or near the C-terminus, especially Lpp variants with hydrophobic residues at the C-terminus, function similarly to Lpp^{+Leu}. Interestingly, Lpp variants with different residues at the C-terminal, penultimate, or antepenultimate position exhibit dramatically different patterns of activation and inhibition effects, indicating that their interactions with DegP differentially stabilize distinct active or inactive forms. We suggest that the tripodal structure with three hydrophobic ends that mimics the Lpp^{+Leu} can be a new platform for allosteric effectors, which can be applicable to develop new antibiotics against DegP or homologous HtrA protease.

Keywords: Allostery, Protease, Protein quality control, heat shock protein, HtrA

Student Number: 2014-30078

Introduction

Intracellular proteolysis should be carefully regulated though the accumulation of misfolded proteins and unnecessary degradation of functional proteins can be toxic to cells (41). Allosteric regulation of proteases, where effector interactions can alter their activity at a distance, often plays an important role in regulating proteolysis. For example, the degradation of misfolded proteins in the cytoplasm around *E. coli* is primarily mediated by the DegP protease, a member of the highly conserved HtrA protease family, where the trimers function as basic allosteric units (42-47). Proteases homologous to DegP in bacteria, plants and humans are mostly associated with bacterial pathogenesis and various human diseases, and share a trimeric structure, indicating that the 'trimer' is a defining feature of HtrA protease (42, 48, 49).

DegP proteolysis shows several distinct features in regulating allosteric activity. First, DegP preferably interacts with a substrate that contains two distinct motifs, a hydrophobic C-terminal degron that binds to the PDZ1 domain and a cleavage site degron (50, 51) that binds to the active site of DegP. This bipartite-substrate interaction not only helps to differentiate suitable substrates for degradation, but also provides a positive cooperative mode of substrate binding and cleavage that can react immediately to slight increases in misfolded proteins.

Second, the binding of substrates activates DegP by structural changes, where the equilibrium population of inactive and active species determines the total proteolytic activity (46, 50-52). At the molecular level, substrate binding leads to the formation of standard Ser-His-Asp catalytic triad and oxyanion holes and structural

remodeling of several loops near the active site (53-57). Highly conserved residues clustered at the interface of two adjacent subunits in a trimer are important for allosteric activation by linking the active sites and substrate binding sites (58, 59). These dynamic morphological changes can allow DegP in its active state only in the presence of an appropriate substrate and prevent unnecessary proteolysis. Third, substrate binding also leads to the assembly of cage-like polyhedrons containing four, six or eight trimers. DegP cage can function as a proteolysis chamber, where the sequestered substrates in the chamber are efficiently degraded by active sites located on the inner surface, while the folded proteins cannot easily access the proteolysis chamber. Thus, cage assembly may protect cells from excessive proteolytic activity (47, 50, 51, 60–62).

The allosteric control of DegP is important for balancing proteolytic activity and maximizing bacterial fitness under the misfolded protein stress (22). Mutations that allosterically increase or decrease the population of active DegP species reduce their ability to tolerate misfolded protein stress, and especially the combination of the two mutations stabilizing active DegP and disabling the cage formation produces a hyperactive DegP variant, $DegP^{R207P/Y444A}$, killing bacteria by excessive protein degradation.

This lethality could be suppressed by the intragenic mutations that stabilize inactive form of DegP species, or by a mutant variant of the outer-membrane lipoprotein that carries additional leucine to the C-terminus, Lpp^{+Leu}, which has an inhibitory effect *in vitro* (62, 63). Thus, the intrinsic activity level of DegP appears to be tuned evolutionarily by controlling equilibrium populations of inactive and active species.

As mentioned above, the allosteric properties of DegP have relatively been studied well, but little is known what molecules other than substrates can modulate the activity of DegP. Here, we show that Lpp^{+Leu} is not a simple active site inhibitor, but functions as a novel allosteric activity modulator. Selection experiments using the Lpp variant libraries show that $DegP^{R207P/Y444A}$ toxicity could be suppressed by Lpp variants with a much wider variety of sequences, especially with hydrophobic C-terminal residues.

Interestingly, Lpp^{+Leu} and selected Lpp variants exhibit dramatically different activation and inhibition patterns, indicating that different C-terminal sequences of Lpp can stabilize a distinct set of conformational populations with different levels of activity. We suggest that a tripodal moluecule with three hydrophobic ends that mimics the Lpp^{+Leu} can be a common platform for new allosteric effectors, which might become useful in developing new antibiotics for DegP or homologous HtrA proteins (48).

Results

Lpp^{+Leu} contains additional leucines at the C-termini of trimeric coiled-coil of Lpp^{WT} (64, 65). Thus, we deduced that the additional leucines of Lpp^{+Leu} play a key role in inhibition. To verify this suggestion, we purified several types of Lpp variants and analyzed their ability to inhibit $DegP^{R207P/Y444A}$ cleavage of the p23 substrate, a model substrate derived from lysozyme that tightly interacts with and activates DegP (46). The activity of DegP can be quantitatively monitored by fluorescence increase, as the p23 cleavage separates the fluorophore and a quencher in the p23 peptide.

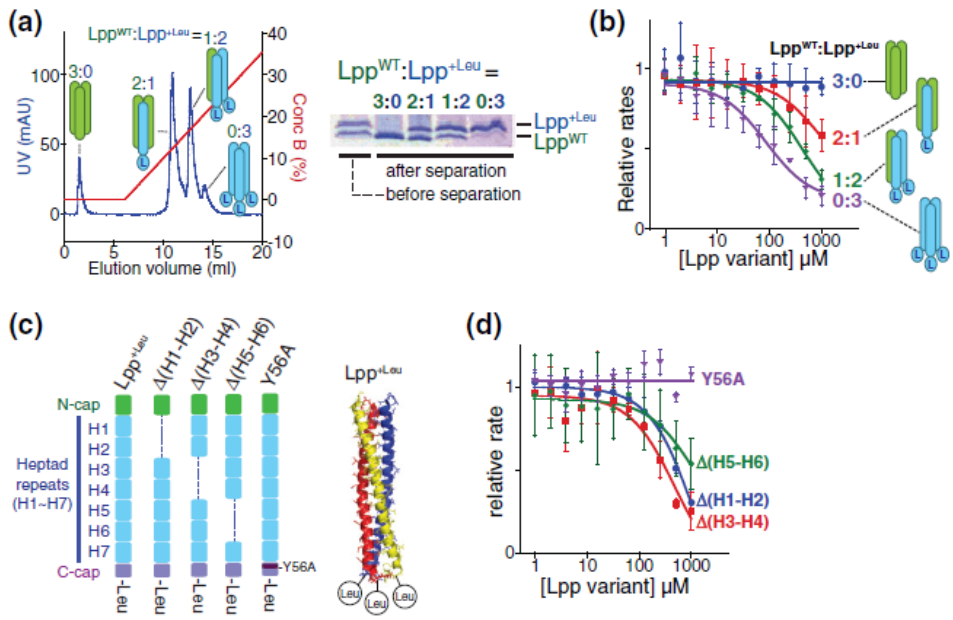


Figure 18. Three leucines at the C-termini of the trimeric Lpp+Leu are critical for inhibition of the DegPR207P/Y444A activity.

(a) Hybrid trimers of non-Flag-tagged Lpp^{WT} and Flag-tagged Lpp^{+Leu}, in which Lpp^{WT}:Lpp^{+Leu} stoichiometry is 3:0, 2:1, 1:2, or 0:3, were separated by anion-exchange chromatography (left). Separation of different species was confirmed by Coomassie-stained SDS-PAGE gel (right). (b and d) Initial rates of $DegP^{R207P/Y444A}$ (0.5 μM) cleavage of the p23 substrate (10 μM) were measured in the presence of different concentrations (1 μM–1 mM) of hybrid Lpp trimers (b) or other Lpp variants with truncated coiled-coils or a mutation at C-cap (d). Error bars are averages ± 1 SD (n = 3). (c) Schematic view of Lpp^{+Leu} or Lpp^{+Leu} variants with truncated coiled-coils or a mutation at C-cap.

First, we tested the inhibitory effects of the hybrid trimers of Lpp^{WT} and Lpp^{+Leu} to determine how the number of C-terminal leucines in the Lpp^{+Leu} trimer affects function (Lpp^{WT}:Lpp^{+Leu} stoichiometry is 3:0, 2:1, 1:2 or 0:3). Individual hybrid trimers can be obtained by mixing Flag-tagged Lpp^{WT} with Flag-tagged Lpp^{+Leu} under denaturing conditions, folding again, and separating the mixture by anion exchange chromatography (Figure 18A) (66, 67). We found that the inhibitory effect was maximal in three C-terminal leucines and decreased in less than three leucines (Figure 18B). This result suggests that all three leucines at the C-termini participate in DegP inhibition.

Second, we separated three different variants to determine whether the coiled coil area of the Lpp trimer is also important for suppression. Two adjacent heptad repeats were truncated (Figure 18C) (65). Inhibition assay indicated that all three truncated variants retained the inhibition effect, suggesting that most of the coiled-coil area was not required for inhibition (Figure 18D). The C-terminus of Lpp^{+Leu} is ~Y^{56}RKL and Tyr56 functions as a C-terminal capping motif and reorients the next three residues, ~RKL (65). Lpp^{+Leu}/Y56A, Tyr56 mutated to alanine, lost its inhibiting effect, indicating that the C-terminal residues should be in proper location to inhibit DegP (Figure 18D). Overall, this suggest that the C-terminal leucines of the Lpp^{+Leu} trimer are key components for inhibition of the promiscuous protease, $DegP^{R207P/Y444A}$.

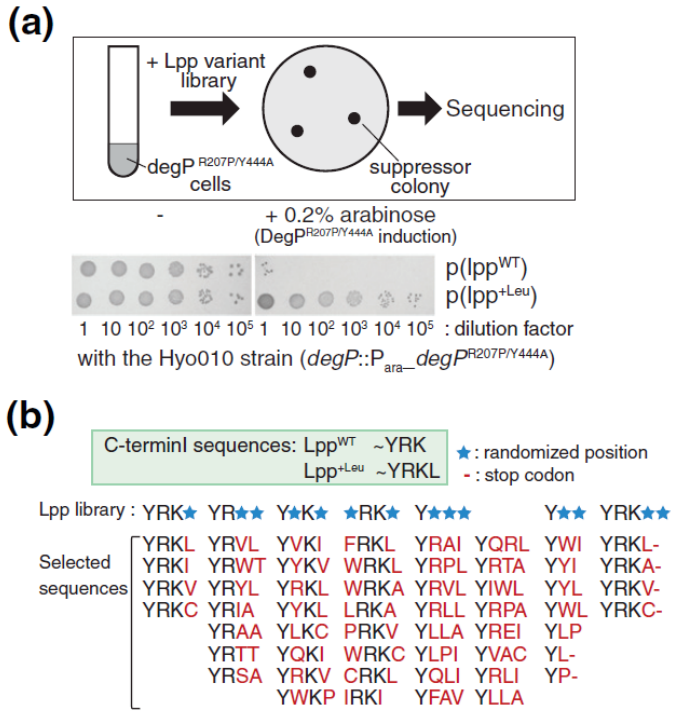


Figure 19. Lpp variants with diverse C-terminal sequences could suppress the lethality of DegP$^{R207P/Y444A}$.

(a) Selection system was constructed with the HYO010 strain, in which cells carrying the *lpp*$^{+Leu}$ allele or others with a similar suppressing effect could survive when DegP$^{R207P/Y444A}$ was induced with 0.2% arabinose. (b) Various Lpp libraries, in which specific positions near the C-terminus were randomized, were subjected to selection for survival in the presence of DegP$^{R207P/Y444A}$ on LB agar plates. C-terminal sequences of Lpp variants from 10 to 20 selected colonies are listed. Hydrophobic residues are highly enriched at the C-terminal position. Selection of the library with longer C-termini (YRK**; * is a randomized position) resulted in those carrying a stop codon at the C-termini.

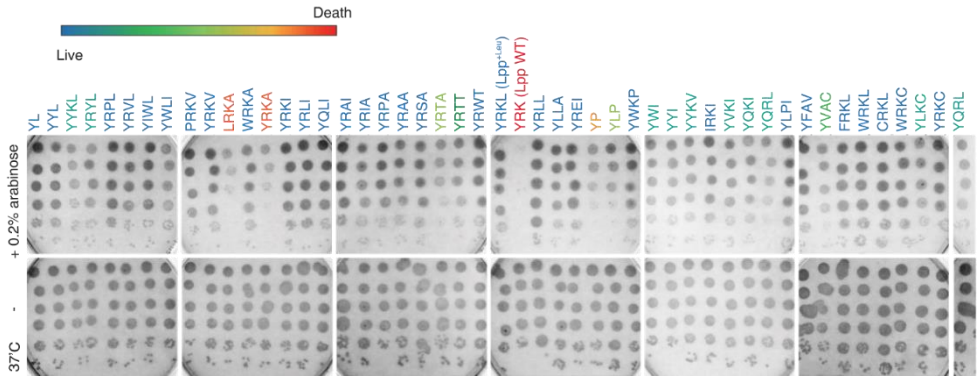


Figure 20. The selected Lpp variants were confirmed to suppress the DegP$^{R207P/Y444A}$ toxicity.

Plasmids were isolated from 10-20 colonies from selection experiments with each Lpp library (YRK∗, YR∗∗, Y∗K∗, ∗RK∗, Y∗∗∗, Y∗∗, YRK∗∗; ∗ is a randomized position near C-terminus), and freshly reintroduced into HYO010. Ten-fold serial dilutions of cells were spotted on LB-agar plates in the absence of presence of 0.2% arabinose, and grown at 37^{o}C. The level of cell viability was presented with differential colors. (Ranging from red, which indicates dead cells, to blue, which indicates viable cells).

The *lpp*$^{+Leu}$ allele containing a single nucleotide substitution in the stop codon of the lpp gene was previously selected through experiments based on spontaneous mutations (62). Analyzing the Lpp libraries with diverse sequences in the C-terminal region may allow identification of more Lpp variants with similar functions. To check new Lpp variants that supress DegP$^{R207P/Y444A}$ toxicity, we constructed an *E. coli* strain (HYO010) in which DegP$^{R207P/Y444A}$ is expressed from bacterial chromosome DNA under control of the arabinose inducible promotor and transformed it with Lpp variants-expressing low copy plasmids under the native promoter. As the arabinose induces toxic DegP$^{R207P/Y444A}$ protease, strains co-expressing Lpp^{+Leu}, but not Lpp^{WT}, can survive (Figure 19A). Various Lpp libraries were constructed in plasmids in which specific positions near the C-terminus (positions C1~C4 from the C-terminus) are randomized, and those suppressed the toxicity of DegP$^{R207P/Y444A}$ on LB agar plates were selected (Figure 19). We isolated plasmids from 10-20 selected colonies in each library, sequenced and confirmed the inhibition of DegP$^{R207P/Y444A}$ toxicity (Figures 19B and 20). We found that Lpp variants with a much wider variety of C-terminal sequences can suppress DegP$^{R207P/Y444A}$ toxicity, albeit with different suppression levels.

While the small number of sequenced Lpp variants limits generalization, one striking pattern is that hydrophobic residues (e.g., Leu, Ile, Val, and Ala) are highly abundant at the position C1 (C-terminal residues).

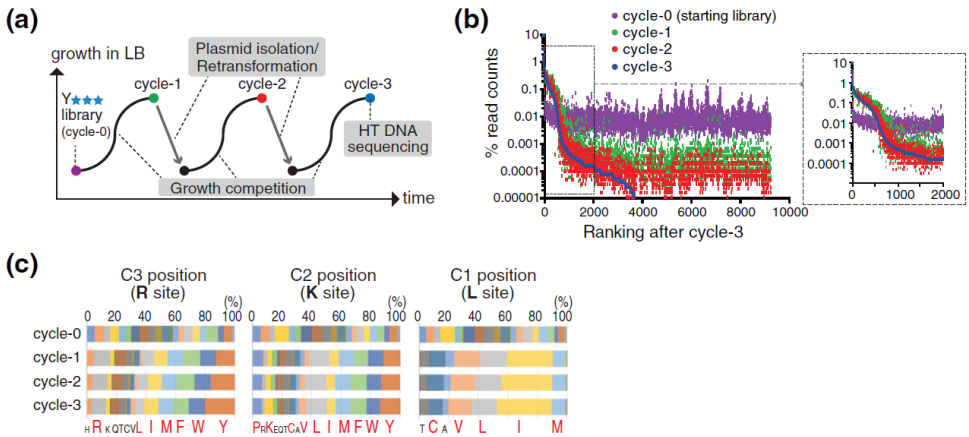


Figure 21. Growth competition experiments show much more Lpp strains that inhibit the mortality of DegP$^{R207P/Y444A}$.

(a) Cells carrying the Lpp library, in which the C-terminal three residues are randomized, were subjected to three growth competition cycles in LB broth at 37 °C in the presence of 0.2% arabinose. To prevent the outgrowth of undesirable suppressor strains, plasmids were isolated after each competition cycle and reintroduced into fresh HYO010 cells. C-terminal sequences were obtained by high-throughput amplicon sequencing. (b) The 9261 unique sequences were sorted by the % read counts after the third growth competition cycle and plotted to show abundance after each cycle. A single growth competition was sufficient for the major selection event. (c) The distribution of amino acids at each position near C-terminus before the growth competition or after each competition cycle. Residues colored in red were found above the original occurrence. The hydrophobic residues, such as Val, Leu, Ile, and Met, were highly enriched at C1 position compared to the less-restrictive C2 and C3 positions.

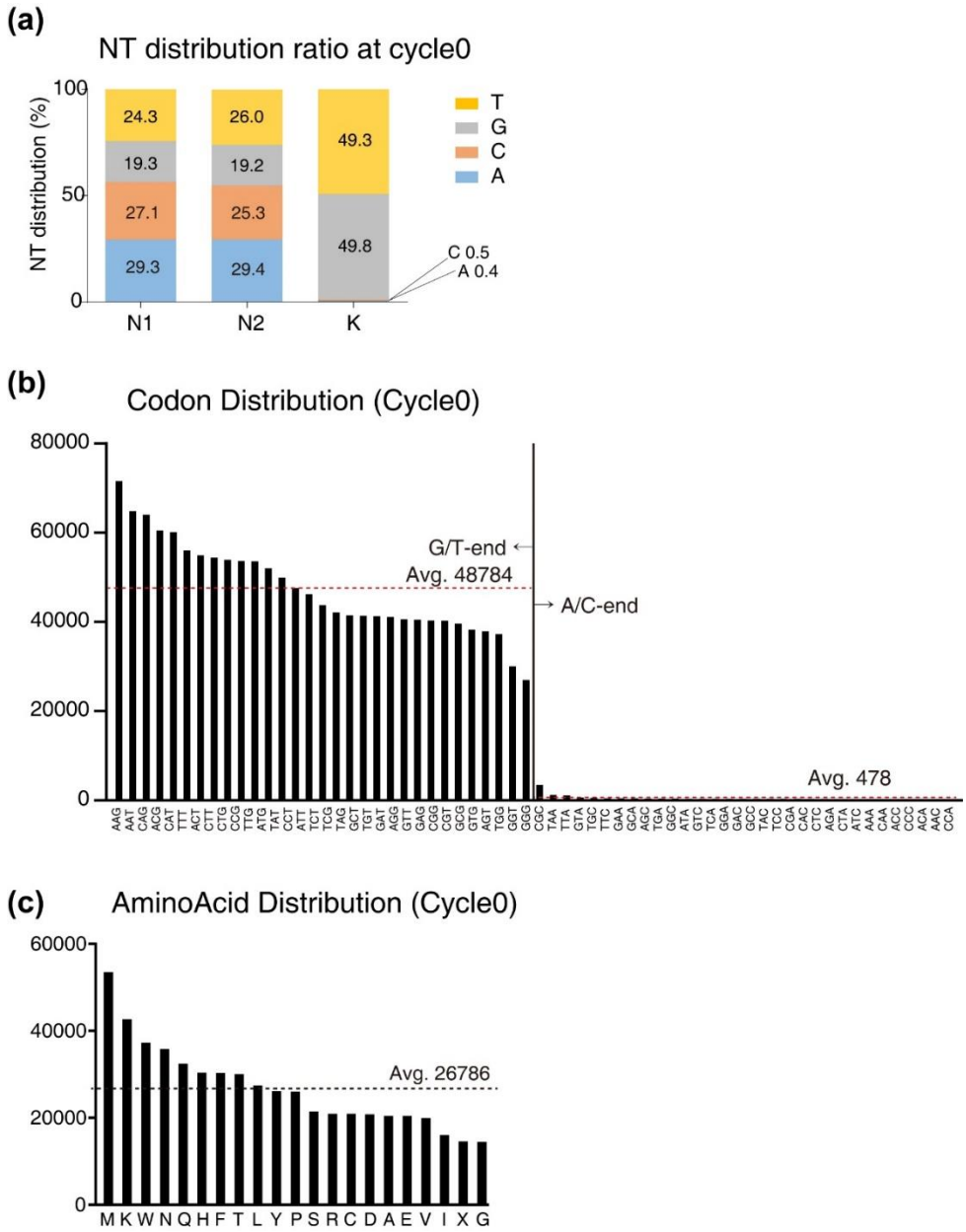


Figure 22. Analysis of sequences from cycle 0.

To verify that the Lpp library prior to the growth competition experiment is uniformly randomized with NNK, we analyzed the sequences in the following ways. (A) The distribution of nucleotides at each position in a codon, NNK, is well dispersed. The last position randomized with K (G and T) has two major nucleotides, G and T (>99%), with very low levels of C and A nucleotides (0.5% for C and 0.4% for A). (B) Codons end with G/T are found ~100 times more than those that end with A/C, and distributed well. (C) The numbers of 20 amino acids and a stop codon (X) normalized with codon counts were distributed well.

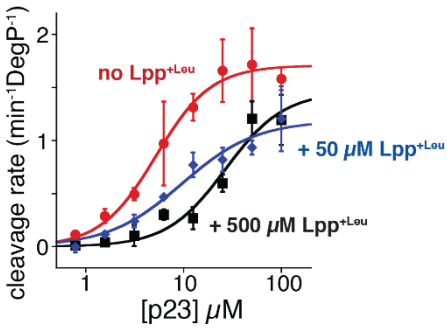


Figure 23. Lpp^{+Leu} inhibits the p23 cleavage by DegP$^{R207P/Y444A}$.

Cleavage rates of the p23 substrate (0.8 μM ~ 100 μM) by DegP$^{R207P/Y444A}$ (0.5 μM) were measured in the absence or presence of Lpp^{+Leu} (50 or 500 μM). Fitting to the Hill form of the Michaelis-Menten equation (rate = k_{cat}•(S)^n/(K_M^n + (S)^n)) gave k_{cat} values of 1.7±0.064, 1.2±0.13, and 1.5±0.32, K_M values of 5.2±0.54, 9.6±2.9, and 27±11, and Hill coefficients (n) of 1.5±0.20, 1.1±0.27, and 1.5±0.51, for data with 0, 50, and 500 μM Lpp^{+Leu}, respectively. Error bars are averages ±1 SD (n = 3).

To obtain a more thorough analysis on efficient Lpp variants with inhibitory effect, we carried out a growth competition test, where the starting Lpp library was randomized at positions C1-C3 and underwent three growth competition cycle in the presence of $DegP^{R207P/Y444A}$ (Figures 21 and 22). Plasmids were isolated after each competition cycle and retransformed into fresh bacterial cells for the next competition round in order to avoid enrichment of unwanted suppressor strains.

High-throughput DNA sequencing allows 100-200 reads for the 9261 (=213; 20 amino acids and stop codons) distinct sequences. A single growth competition test resulted in ~600 sequences (~6%) higher than the average number of reads, and was found to be high numbers after additional two cycles, showing that a single growth competition was sufficient for major selection event (Figure 21B). Among them, 185 sequences were enriched (~2%) more than 10-fold after the third cycle, supporting the idea that more various sequences than Lpp^{+Leu} could suppress $DegP^{R207P/Y444A}$ toxicity.

To determine the sequence features for suppressing toxicity, we analyzed the amino acid composition at the C1-C3 position of the Lpp variants obtained after the third cycle (Figure 21C). In particular, only five amino acids at C1 position (C, I, L, M, and V) were found to be above the original occurrence. Hydrophobic residues (I, L, M, and V) were highly abundant. Many different amino acids were found at C2 and C3 positions, where hydrophobic residues (V, L, I, and M; 1.8- and 1.6- fold) moderately enriched and aromatic residues were highly abundant (F, W, and Y; 2.7- and 3.8- fold at C2 and C3 positions, respectively). This also support the suggestion that one of the key factors generating inhibitory effects is the hydrophobic feature at the C-terminus of Lpp.

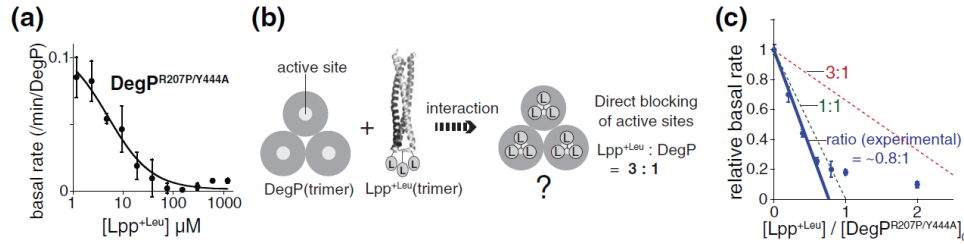


Figure 24. Lpp^{+Leu} inhibits the basal activity of DegP$^{R207P/Y444A}$.

(a) The basal proteolytic activity of DegP$^{R207P/Y444A}$ (10 μM) monitored by the cleavage of the reporter peptide (100 μM) was inhibited with increasing concentrations of Lpp^{+Leu}. Error bars are averages ± 1 SD (n = 3). (b) One model for the interaction between Lpp^{+Leu} and DegP, in which three trimeric Lpp^{+Leu} molecules sterically block three active sites of a trimeric DegP. (c) The basal activity of DegP$^{R207P/Y444A}$ (100 μM) was measured by the cleavage of reporter (100 μM) in the presence of different concentrations of Lpp^{+Leu} (0–200 μM). The stoichiometry of interaction between Lpp^{+Leu} and DegP$^{R207P/Y444A}$ is closer to 1:1 than to 3:1. Error bars are averages ± 1 SD (n = 3).

Lpp+Leu inhibits the underlying activity of the hyperactive DegP variant DegP, which is converted between the active and inactive conformations by interacting with the substrate. Since the p23 peptide is a tightly binding substrate, inhibition of DegPR207P/Y444A by Lpp+Leu (Figure 18B and 23) reveals that Lpp+Leu can inhibit DegPR207P/Y444A active state, but Lpp+Leu also suppresses the inactive state. Previously, short peptides (reporters) were used as model substrates to monitor the underlying activity of DegP, because it does not bind tightly to or activate DegP (50). To determine how Lpp+Leu affects DegPR207P/Y444A the inactive state, we monitored reporter cleavage in plus Lpp+Leu condition and found that Lpp+Leu also suppresses the basal activity of DegPR207P/Y444A with half maximal inhibition at about 5 μM (Figure 24A).

One simple inhibition model is that Lpp+Leu binds to the active site of DegP and directly blocks the access of the substrate. Since the three active sites of the DegP trimer are located distantly, the stoichiometry of the interaction between Lpp^{+Leu} and DegP is 3:1 in this case (Figures 24B). To determine the stoichiometry, we measured the basal activity of $DegP^{R207P/Y444A}$ at high concentration (100 μM; about 20-fold higher than the one at half of the maximal inhibition) was monitored in the presence of diverse concentrations of Lpp^{+Leu} (Figure 24C). Intriguingly, the inhibition pattern at the substoichiometric concentration of Lpp^{+Leu} showed that the inhibitory stoichiometry between Lpp^{+Leu} and $DegP^{R207P/Y444A}$ was closer to 1:1 than to 3:1. These results suggest that Lpp^{+Leu} does not simply block the active site of DegP.

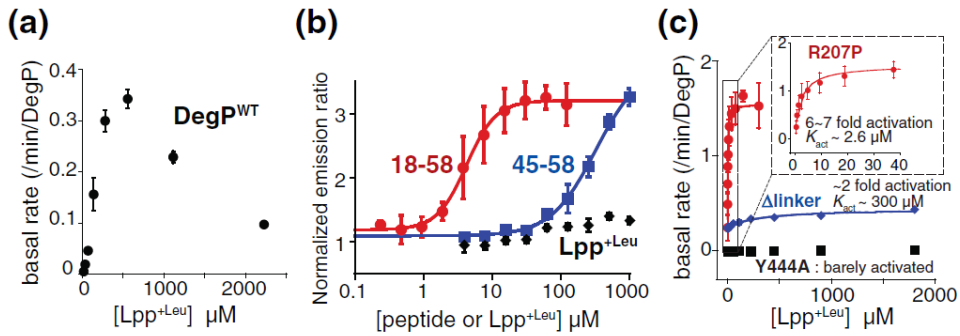


Figure 25. Lpp^{+Leu} activates the basal activity of DegP^{WT} and its variants.

(a) The basal rate of DegP^{WT} (1 μM) monitored by the cleavage of reporter (100 μM) was increased with up to ~500 μM Lpp^{+Leu} and decreased with higher concentrations of Lpp^{+Leu}. Error bars are averages ± 1 SD (n = 3). (b) DegP assembly was monitored with dye-labeled DegP variants (0.2 μM) and different concentrations of Lpp^{+Leu} or one of two model substrate peptides, 18–58 or 45–58. Lpp^{+Leu} barely triggered DegP assembly. Ratios of two emissions were normalized to those without Lpp^{+Leu} or a peptide. Error bars are averages ± 1 SD (n = 3). (c) The basal rates of various DegP variants (DegP^{R207P}, DegP$^{\Delta linker}$, and DegP^{Y444A}) were monitored by the reporter cleavage (100 μM) with different concentrations of Lpp^{+Leu}. Lpp^{+Leu} differentially affects their basal activities. Error bars are averages ± 1 SD (n = 3).

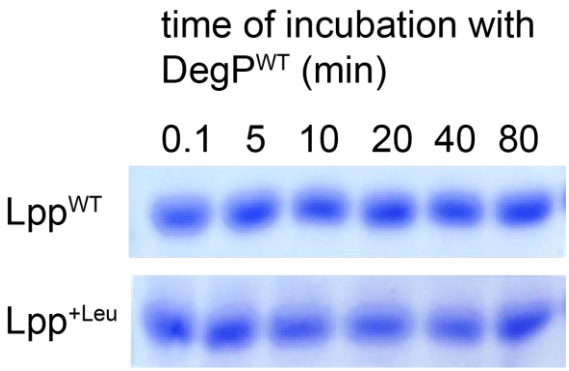


Figure 26. Lpp^{WT} and Lpp^{+Leu} are not substrates of the DegP protease.

Lpp^{WT} and Lpp^{+Leu} (10 μM) were incubated with DegP^{WT} (1 μM) at 37^{o}C. Aliquots of the reaction solutions at different time points were taken for SDS-PAGE.

Lpp^{+Leu} has an allosteric effect on DegP. The vast majority of protease inhibitors kill the protease activity by simply blocking the substrate access, but several experimental data suggest that Lpp^{+Leu} is an allosteric modulator of the DegP protease.

First, Lpp^{+Leu} can activate the basal activity of wild-type DegP and some variants. When testing how Lpp^{+Leu} affects $DegP^{WT}$-mediated reporter cleavage, the $DegP^{WT}$ basal cleavage rate increased significantly, up to ~500 μM Lpp^{+Leu}, and decreased with higher Lpp^{+Leu} concentrations (Figure 25A). Thus, Lpp^{+Leu} differentially regulates the basal activity of the DegP variants: activation on $DegP^{WT}$ and inhibition on $DegP^{R207P/Y444A}$. The activation effect is only possible by some type of allosteric effect and again suggests that Lpp^{+Leu} does not simply block the active site of DegP. The activation effect of Lpp^{+Leu} is not the result of increased cage assembly as Lpp^{+Leu} did not trigger assembly significantly, whereas the postsynaptic density protein (PDZ)-binding motif (45-58) and a good model substrate peptide (18-58) could assemble higher oligomers, as analyzed by fluorescence resonance energy transfer (FRET) of dye-labeled DegP variants (Figure 25B) (50). We also confirmed that Lpp^{+Leu} is not a substrate for DegP (Figure 26). We further tested the effect of Lpp^{+Leu} on other DegP variants, $DegP^{R207P}$, $DegP^{\Delta\ linker}$, and $DegP^{Y444A}$, which showed different basal activities and assembly behaviors (Figure 25C) (46). Lpp^{+Leu} also activated the basal cleavage rates of $DegP^{R207P}$ and $DegP^{\Delta linker}$ in different concentration ranges, but we confirmed that the allosteric effect of Lpp^{+Leu}, not one of $DegP^{Y444A}$, could be very different on the DegP variants.

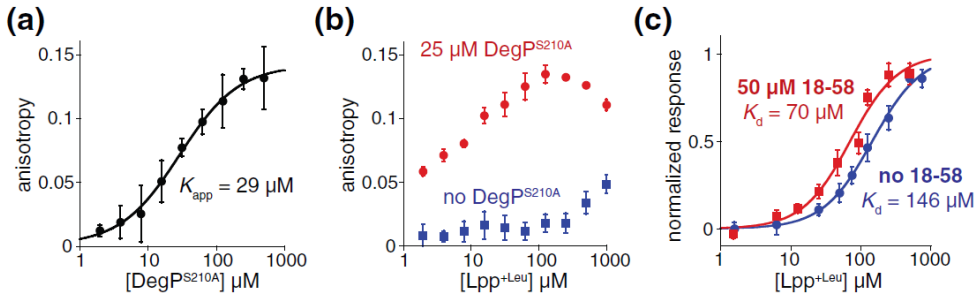


Figure 27. Lpp^{+Leu} modulates substrate affinity to DegP.

(a) Fluorescence anisotropy of the fluorescently labeled model substrate (flC18–58; 0.05 μM) was monitored in the presence of increasing amounts of a catalytically inactive DegP variant (DegP^{S210A}). Data were fitted to a hyperbolic equation with the resulting apparent dissociation constant (K_{app}) of 29 μM. Error bars are averages ± 1 SD (n = 3). (b) Fluorescence anisotropy of flC18–58 (0.05 μM) was monitored with DegP^{S210A} (0 or 25 μM) in the presence of increasing amounts of Lpp^{+Leu}. Error bars are averages ± 1 SD (n = 3). (c) Binding of Lpp^{+Leu} to DegP^{S210A} was monitored by the BLI experiments in two conditions, without 18–58 or with a saturating amount of 18–58 (50 μM). Fitting to a hyperbolic equation gave the dissociation constants (K_d) of 146 and 70 μM, respectively. Error bars are averages ± 1 SD (n = 3).

Second, Lpp^{+Leu} modulates substrate affinity for DegP. As previously reported (50), binding of a fluorescently labeled model substrate (flC18–58) can be monitored by measuring the FL-anisotropy in the presence of diverse amounts of the catalytically inactive DegP variant ($DegP^{S210A}$) (Figure 27A). To test the effect of Lpp^{+Leu} on substrate binding, we monitored the FL-anisotropy of flC18-58 with an increase in the amount of Lpp^{+Leu} in two conditions, one condition without $DegP^{S210A}$ (65) and the other condition with 25 μM $DegP^{S210A}$ that the half of model peptides are in a bound form (Figure 27B).

Interestingly, substrate affinity increased up to ~ 100 μM Lpp^{+Leu}, evidence of positive cooperativity in the allosteric system, and decreased slightly with higher Lpp^{+Leu} concentrations, whereas Lpp^{+Leu} alone increased only slightly above 100 μM Lpp^{+Leu} (Figure 27B). This pattern of concentration-dependent affinity changes, in parallel with changes in basal activity (Figure 25A), suggests that Lpp^{+Leu} stabilizes the active form with higher substrate affinity.

We also directly measured Lpp^{+Leu} affinity for $DegP^{S210A}$ using bio-layer interferometry (BLI) experiments (Figure 27C). Here, $DegP^{S210A}$ was covalently attached to the end surface of the AR2G biosensor chip and incubated with different concentrations of Lpp^{+Leu}. The relative amount of bound Lpp^{+Leu} was determined by measuring wavelength shift of the interference pattern of reflected light (68). The Kd values from Lpp^{+Leu} to $DegP^{S210A}$ were determined under two conditions (Figure 27C). One condition is the absence of activators (18–58), where $DegP^{S210A}$ mostly exists in an inactive state, resulting in Kd = 146 μM. The other condition is with saturation amount 18-58 (50 μM), in which $DegP^{S210A}$ is mainly in the active state. If Lpp^{+Leu} blocks the active site of DegP without allosteric effect, 18–58 in the latter

condition competes directly with Lpp^{+Leu} to bind the active site and reduces the Lpp^{+Leu} affinity. However, we found that the Lpp^{+Leu} affinity doubled in the presence of 18–58 (Kd = 70 μM). This also suggests that the interactions between the three components (DegP, substrate (18-58), and Lpp^{+Leu}) are correlated with positive cooperativity, which is a clear signature of allostery.

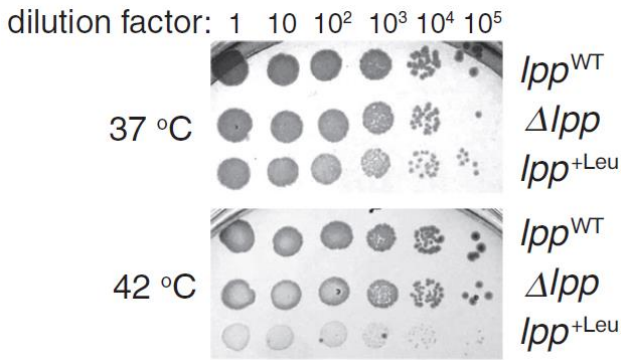


Figure 28. Lpp^{+Leu} changes the growth fitness of the *degP*^{WT} cells in misfolded protein stress.

The *degP*^{WT} cells carrying the *lpp*^{WT}, *Δlpp*, or *lpp*$^{+Leu}$ allele were serially diluted, spotted on LB agar plates, and grown at 37 °C or 42 °C. The *degP*^{WT} *lpp*$^{+Leu}$ strain showed slow growth.

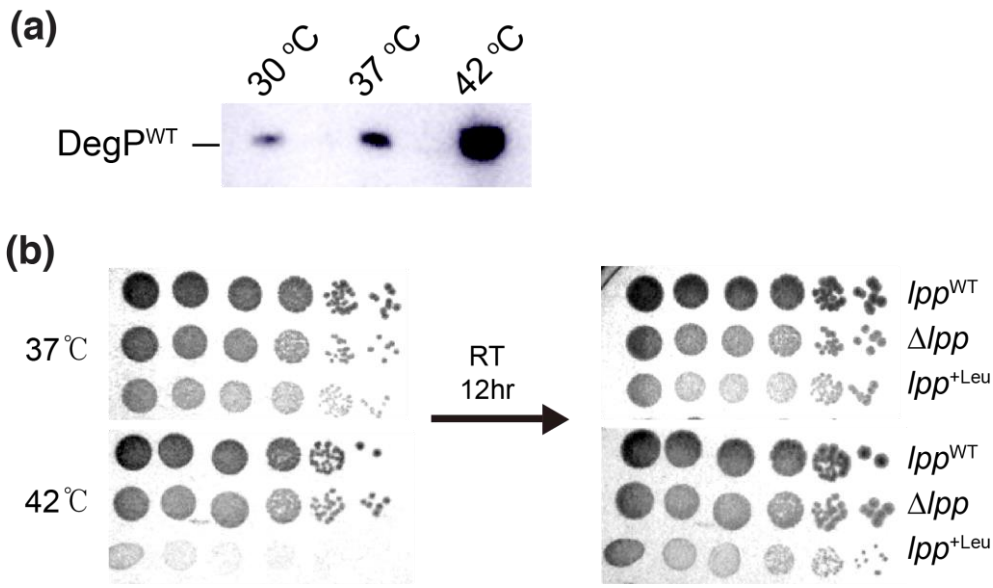


Figure 29.

(a) Wild-type cells (W3110) were grown in LB broth at 30°C, 37°C, or 42°C, and same numbers of cells were subjected to the Western blotting using an anti-DegP antibody. Expression level of DegP is highest at 42°C and lowest at 30°C. (b) Growth patterns of serially diluted *degP*^{WT} strains carrying the *lpp*^{WT}, Δ*lpp*, or *lpp*$^{+Leu}$ allele were monitored at 37°C or 42°C on LB agar plates. The *degP*^{WT} *lpp*$^{+Leu}$ strain showed a seriously slowed colony formation at high temperature (42°C), but eventually grew to normal colonies when they were incubated for additional 12-hr at lower temperature (room temperature).

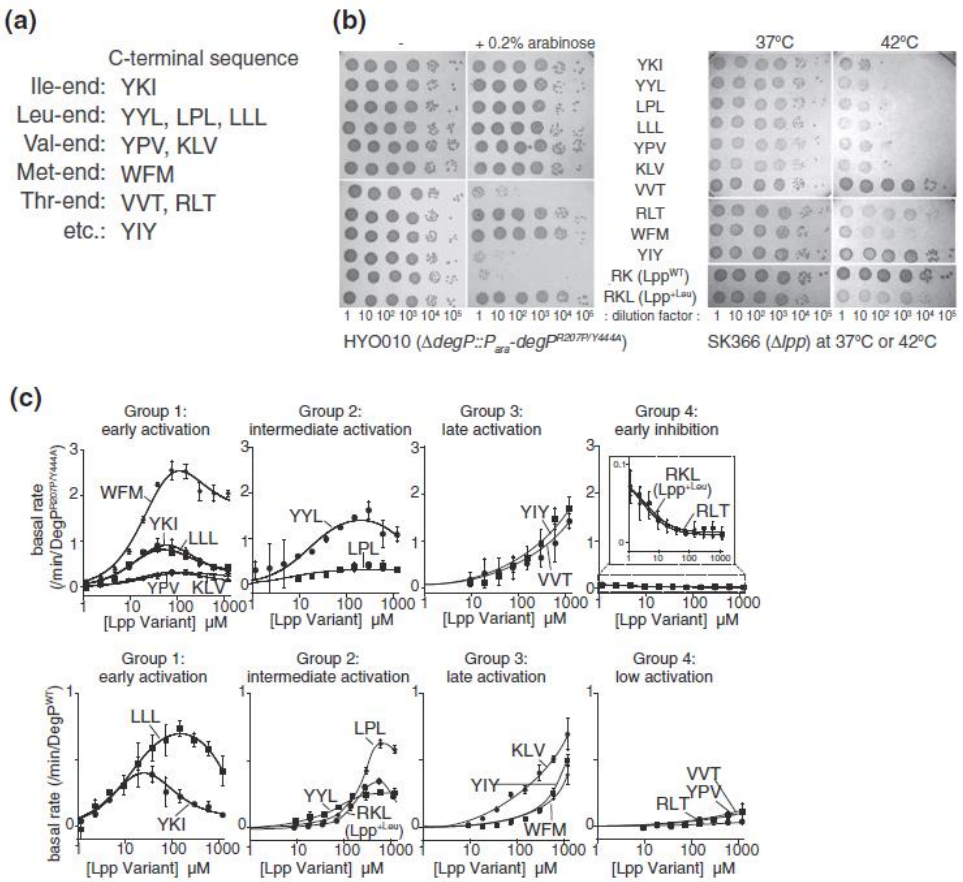


Figure 30. Lpp variants with diverse C-terminal sequences have distinct activity modulation effects.

(a) Ten Lpp variants with diverse C-terminal sequences were chosen from the growth competition experiments for further analysis (top). (b) Two strains expressing $DegP^{R207PY444A}$ under an arabinose-controlled promoter (left panel) or $DegP^{WT}$ under the native promoter (right panel) were transformed with plasmids that produce Lpp variants, Lpp^{WT}, or Lpp^{+Leu}. They were grown to log phase in LB broth, serially diluted, spotted on LB agar plates, and incubated overnight either in the absence or

in the presence of 0.2% arabinose for $DegP^{R207P/Y444A}$ expression at 37 °C (left panel) or at two temperatures (37 °C or 42 °C) for $DegP^{WT}$ expression (right panel). (c) Basal activities of $DegP^{R207P/Y444A}$ (top panel) and $DegP^{WT}$ (bottom panel) cleaving the reporter peptide (100 μM) were measured with different concentrations of Lpp variants. Lpp variants were grouped based on the patterns of activation and inhibition effects. Error bars are averages ± 1 SD (n = 3).

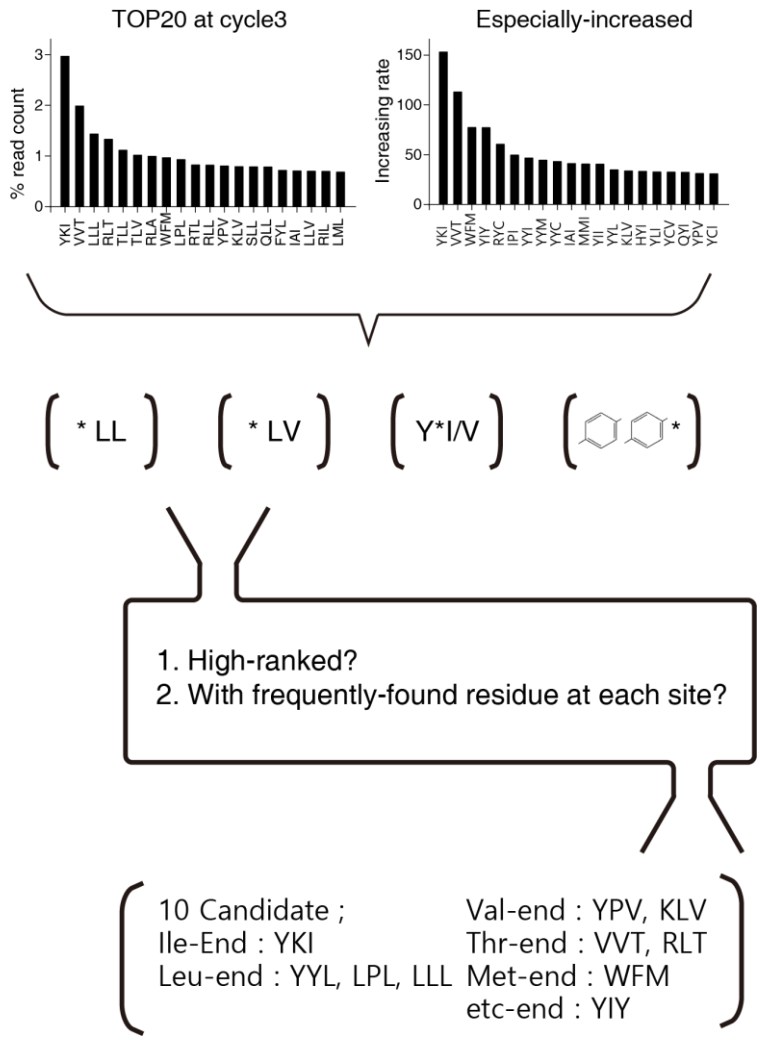


Figure 31. The logic behind the selection of the 10 Lpp variants used in this study is graphically presented.

Initially, 20 most-enriched variants after the final cycle and 20 most-increased variants through 3 cycles were chosen. Of them, 33 non-duplicates were classified into four groups with similar patterns, such as *LV, *LL, Y*I/V, etc. From each group, candidates with rarely-found residues were excluded. Then, we chose those

as high-ranked as possible, and tried to include at least one variant ending with I/L/V/M/T/Y, resulting in 10 Lpp variants: those with YKI, YYL, LPL, LLL, KLV, YPV, VVT, RLT, WFM, or YIY at the C-terminus.

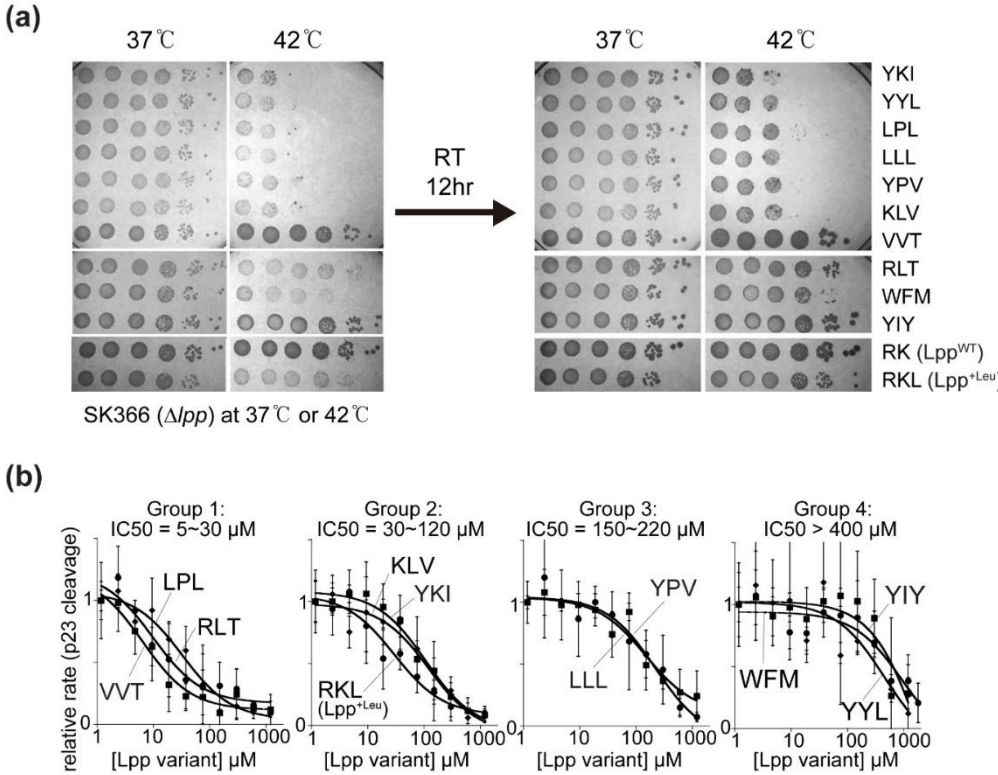


Figure 32. Lpp variants with diverse C-terminal sequences have distinct activity modulation effects.

(a) The *degP*^{WT} Δ*lpp s*trains expressing Lpp variants, Lpp^{WT}, or Lpp^{+Leu} were grown to log phase in LB broth, serially diluted, spotted on LB agar plates and grown at 37 ^{o}C or 42 ^{o}C. Overnight growth at high temperature (42 ^{o}C) showed a significantly slower colony formation except for those containing lpp variants with VVT and YIY at C-termini. The Lpp variants (RLT and WFM at C-termini) eventually grew to normal colonies when they were incubated for additional 12-hr at lower temperature (room temperature) as those with Lpp^{+Leu} did. However, the growth of cells containing lpp variants (YKI, YYL, LPL, LLL, YPV and KLV at C-termini) was

especially reduced 100 times. (b) Cleavage rates of the p23 substrate (10 μM) by $DegP^{R207P/Y444A}$ (0.2 μM), where $DegP^{R207P/Y444A}$ exists mostly in an active state, were measured in the presence of different quantities of Lpp variants at 37℃. Lpp variants were grouped based on the strength of inhibition effects. Error bars are averages ±1 SD (n = 3).

We tested how Lpp^{+Leu} affects the cell growth in the presence of $DegP^{WT}$ (Figure 28). Lpp^{+Leu} did not alter bacterial growth at 37 °C, where DegP is moderately expressed, but significantly hampered the normal colony formation on LB agar media at 42 °C, where the high level of misfolded proteins is produced by heat-shock and DegP expression is highly induced (Figures 28 and 29A). Lpp^{+Leu} does not kill bacteria because cells expressing Lpp^{+Leu} eventually grow into normal colonies when cultured at low temperatures (Figure 29B). These results suggest that Lpp^{+Leu}-mediated activity modulation of $DegP^{WT}$ alters the cellular fitness under misfolded protein stress.

We suspected that diverse Lpp variants that suppress $DegP^{R207P/Y444A}$ toxicity may have similar allosteric effects on DegP. To test this idea, we selected 10 Lpp variants with diverse C-terminal sequences that highly enriched in the growth competition experiment (Figures 30A and 31). Eight variants efficiently suppressed the toxicity of $DegP^{R207P/Y444A}$ on solid LB-agar media, whereas the two variants, VVT and YIY at the C-termini, did not, suggesting that other factors may contribute to the growth competition in liquid media (Figure 30B, left panel). Intriguingly, some variants reduced the viability of $DegP^{WT}$-expressing cells at 42 °C (Figures 30B, right panel and 32A). These results suggest that these variants alter DegP activity and reduce cellular fitness under heat stress more drastically than Lpp^{+Leu}.

Next , we tested the effect of modulation on basal activity with the reporter cleavage and on maximal activity with p23 cleavage (Figures 30C and 32B). All Lpp variants inhibited p23 cleavage by $DegP^{R207P/Y444A}$, but shoed a broad distribution of IC50 values, indicating that they bind differentially to and inhibit $DegP^{R207P/Y444A}$ active state (Figure 32B)

Their modulating effects differed more dramatically on basal activity (Figure 30C). The variant carrying RKL (Lpp^{+Leu}) and RLT at the C-terminus inhibited the basal activity of $DegP^{R207P/Y444A}$, while other variants showed an activating effect, which occurred at different concentrations of Lpp variant and were followed by inhibition effects at higher concentrations except for variants that ends with VVT and YIY (Figure 30C, top panel). The basal activity of $DegP^{WT}$ often have similar effects with that of $DegP^{R207P/Y444A}$, except for variants that ends with RKL, WFM, YPV, and KLV (Figure 30C, bottom panel). For example, variants with YKI and LLL showed strong activation and inhibition of two DegP variants at lower and higher concentrations, respectively, while those with YYL and LPL showed right-shift activation-inhibition patterns. The Lpp variant carrying YKI at the C-terminus showed the largest read number and the highest enrichment factor in growth competition experiments (Figure 31).

Overall, Lpp variants with diverse C-terminal sequences have very different modulation effects on DegP and can suppress $DegP^{R207P/Y444A}$ toxicity with strategies different from the one of Lpp^{+Leu}.

These results also suggest that we could find more potent allosteric modulators of DegP, which further reduce cellular fitness under heat stress by changing the C-terminal sequence.

Discussion

Many aspects of DegP allosterry are well characterized, such as substrate-dependent stabilization of active trimers and construction of cage-like polyhedrons; conformational changes in active sites, loops, and internal residues in allosteric activation clusters; and allosteric activity modulation by various mutations scattered throughout the DegP protease (46, 47, 50, 51, 53-62, 69). However, no allosteric effectors other than the DegP substrate or PDZ-binding peptide are known (70). Our results show that Lpp^{+Leu} and other Lpp strains with hydrophobic C-terminus can act as novel allosteric effectors for DegP protease. These are capable of DegP activation and inhibition, and the pattern varies greatly depending on the C-terminal sequence, concentration, and type of DegP variant interacting with. If the Lpp variant simply blocks the active site of DegP, then differential tampering of these activities is not possible. We recommend that each Lpp variant modulates DegP activity by altering the conformational equilibrium of active and inactive species with different affinities.

Our results suggest that Lpp^{+Leu} and other variants do not simply inhibit DegP activity, but inhibit $DegP^{R207P/Y444A}$ toxicity by altering the dynamic range of DegP activity (Figure 33). For example, Lpp^{+Leu} inhibits both active and inactive states, lowering the overall dynamic range of $DegP^{R207P/Y444A}$ activity.

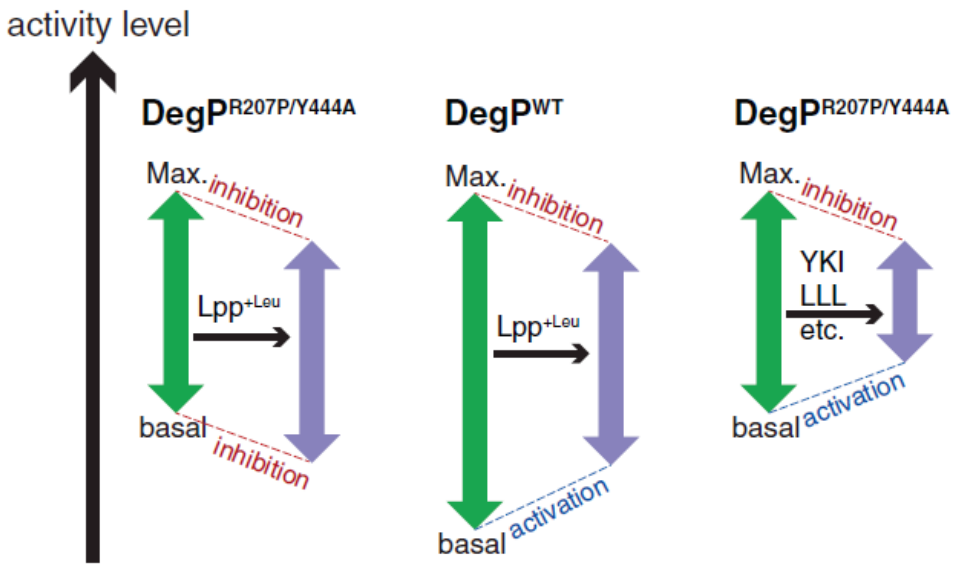


Figure 33. A model for the allosteric effects of Lpp^{+Leu} and other Lpp variants on DegP variants.

Lpp variants modulate the overall dynamic range of DegP activity. The C-terminal sequences of Lpp variants and mutations in DegP determine the direction and strength of modulation effects.

This activity-lowering effect of Lpp^{+Leu} is a possible explanation for the inhibitory effect against $DegP^{R207P/Y444A}$ toxicity in cells. Lpp^{+Leu}, however, reduces the dynamic range of $DegP^{WT}$ activity by inhibiting the active state and activating the inactive state of the protease. This can lead to poor growth suitability under misfolded protein stress, as shown in Figure 28. Other Lpp variants also alter the dynamic range in different patterns, suppress the toxicity of $DegP^{R207P/Y444A}$ in growth competition tests, and reduce the fitness of the cells that expresses $DegP^{WT}$ under misfolded protein stress.

We found that a tripodal structure with three hydrophobic ends is the minimal element to become an allosteric effector. Most of the coiled-coil area is not required for modulating activity, but the hydrophobic residues at the C-termini are essential. Residues in antepenultimate and penultimate positions also affect the strength and direction of activity regulation, suggesting that changes in these sites might allow to elaborately tune modulating activity of the allosteric effectors. Unexpectedly, the Lpp variant with C-terminal cysteines also suppresses $DegP^{R207P/Y444A}$ toxicity (Figures 19B, 20, and 21C). Considering the oxidative environment of the cytoplasm around the bacteria (71) C-terminal cysteines can form disulfide bonds with trimers or DegP or other proteins, which could be other mechanisms of toxicity inhibition. It is not yet clear where Lpp^{+Leu} or other variants bind to DegP. Althogh PDZ1 binds to the peptide with hydrophobic residues on C-termini (50, 54, 69), we believe that Lpp^{+Leu} does not solely bind to the peptide binding site of PDZ, since a PDZ1-peptide interaction, but not the Lpp^{+Leu} interaction, triggers cage assembly. One model for the DegP-Lpp^{+Leu} interaction is that, Lpp^{+Leu} binds to the active site region of DegP and causes an allosteric effect on the trimeric

DegP, the basic allosteric unit, as a substrate do. This model can explain the concentration-dependent inhibition and activation by Lpp^{+Leu}: The smaller the amount of Lpp^{+Leu} binds to one or two subunits of the DegP trimers and activates the remaining subunits allosterically, but the higher amounts saturate all three active sites, so inhibit DegP. This model differs from the model where Lpp^{+Leu} simply block the substrates access, because the latter does not contain any allosteric effect on DegP oligomers. Allosteric effect mainly stabilize active form of DegP (wild-type DegP, $DegP^{R207P}$ or $DegP^{\Delta linker}$), but stabilized form might have lower basal activity in some variants ($DegP^{R207P/Y444A}$).

Another model is that Lpp^{+Leu} binds to multiple sites of DegP with different affinity, at least one of which is an allosteric site distant from the active site of DegP. Since one mode of Lpp^{+Leu}-$DegP^{R207P/Y444A}$ interaction has a 1:1 stoichiometry, the central region of the trimer DegP might be one of the candidate allosteric sites. The L1 and L2 loops located in the central region undergo large conformational changes upon allosteric activation and contribute to the proper arrangement of the active site residues (51, 54, 57). Thus, Lpp^{+Leu} can interact with these loops and stabilize their distinct active or inactive form. Higher Lpp^{+Leu} concentrations or other variants may interact with active sites or other allosteric sites for differential activity modulation. Previous studies commonly suggest that the overall conformation of DegP trimer including active sites, loops, and internal allosteric cluster is thermodynamic coupled (42, 59).

Several other loops near the active site significantly change their conformations between the inactive and active states, and therefore, might function

as an allosteric hotspot where interacts with and is affected easily by allosteric effectors.

Allosteric inhibitors offer unique benefits for improving specificity or potency without directly blocking the active site of the enzymes. Particularly, targeting the allosteric sites of the proteases is emerging as a new promising drug discovery strategy, because traditional approaches to target active sites often fail due to their low specificity (72). Interestingly, acyldepsipeptide, a new class of antibiotics, allosterically activates the ClpP protease and induces excessive protein degradation in the cytoplasm (73). The cytotoxic effects of acyldepsipeptide have been shown to kill persistent bacteria that are very difficult to treat with traditional antibiotics (74). However, it is often challenging to identify molecules that regulate enzymatic activity allosterically. Our study shows that Lpp^{+Leu} and other variants can be a novel form of allosteric modulators for DegP. These molecules can reduce cellular fitness by disrupting delicate balance of regulated proteolysis in bacterial cytoplasm. Therefore, mimicking Lpp^{+Leu} or other variants with smaller tripodal molecules could be a new strategy for the antibiotic development against pathogenic bacteria.

Materials and Methods

Bacterial strains, plasmids, and primers used in this study are listed in Tables S1–S3.

Plasmid construction

Plasmids expressing the DegP variant (p7 for $DegP^{WT}$ and pSK576 for $DegP^{R207P/Y444A}$) or the Lpp variant (pSK709 for LppWT and pSK710 for Lpp^{+Leu}) have been described previously (46, 62). The plasmid expressing the His-tagged Lpp (His-Lpp) variant in the presence of IPTG (pSK709, pSK710, pHyo012-21, pYK1, pYK2, pYK15, pYK21, and pYK23) was subjected to reverse PCR mutagenesis using pSK710 as a template. Was built (75). The plasmid expressing the Lpp variant under the native promoter (pHyo022-31) was constructed from the low copy number plasmid pWSK29 (76).

First, the *lpp* gene with the native lpp promoter was amplified from W3110 strain using primers lpp-246_Not1_fw and lpp-246_EcoR1_rv, and inserted between the NotI and EcoRI sites of pWSK29. pHyo003 for lppWT and pHyo004 for Lpp^{+Leu} was created.

Protein purification and preparation of hybrid Lpp trimers

DegP and its variants were prepared as previously described (50). His-Lpp variants were expressed in the BL21 (DE3) strain and purified using Ni Sepharose 6 FastFlow (GE Healthcare). The eluted protein (50mM sodium phosphate (pH 8), 300mM NaCl, 500mM imidazole in elution buffer) was concentrated and exchanged for a buffer

containing 20mM MES (pH 6), 20mM sodium phosphate (pH 7 or 8).) Or 20mM Hepes (pH 8), further purified by ion exchange chromatography (Mono S 5/50 GL; GE Healthcare) according to the pI value of each protein using Amicon centrifugal filters (Millipore).

The His-tags of these proteins were digested with Bovine Thrombin (Millipore) for about 9 hours at 25 ℃ (room temperature). The cleaved protein was purified by ion exchange chromatography.

We adopted the previously reported procedure to obtain individual hybrid Lpp trimmers (26,37).

Briefly, equal amounts of purified His-Lpp^{WT} and His-Flag-Lpp^{+Leu} were denatured in 6M guanidinium, incubated for 10 minutes at room temperature, and then diluted 10-fold with buffer B (50 mM sodium phosphate (pH 8) and refolded with 100mM NaCl. After changing the buffer to 20 mM Tris (pH 8.8) using an Amicon centrifugal filter (Millipore), the hybrid trimers were separated by anion exchange chromatography (Mono Q 5/50 GL; GE Healthcare) and Bovine Thrombin (Millipore) to remove the His-tag and purified by anion exchange chromatography.

***In vitro* enzyme activity assay**

In vitro enzymatic assays were performed at 37 °C in 50 mM sodium phosphate (pH 8) and 100 mM NaCl.

The basal activity of the DegP variants (1–10 μM) in the presence of the Lpp variant was monitored by change in fluorescence after cleavage of reporter

peptide (100 μM, excitation, 320 nm, emission, 430 nm) (50, 78). Cleavage of peptide p23 (10 μM; excitation, 320 nm, emission, 430 nm (46)) by $DegP^{R207P/Y444A}$ (0.2 μM) was analyzed while adding the Lpp strain (1–1200 μM) using an Infinite F200Pro microplate reader (Tecan).

Assembly and binding assays

Assembly and fluorescence anisotropy experiments were performed as previously described (50). Briefly, FRET assays for assembly were conducted with 0.2 μM of donor- and acceptor-labeled DegP with a microplate reader (excitation, 520 nm; emission, 570 or 670 nm; Infinite M200pro, Tecan). Fluorescence anisotropy (FL-anisotropy) assay was performed with FL-labeled model substrate (flC18–58, 50 nM) mixing with $DegP^{S210A}$ or Lpp^{+Leu} using a microplate reader (excitation, 485 nm, emission, 535 nm, Infinite F200pro, Tecan).

BLI experiments were performed as described (68). $DegP^{S210A}$ is covalently attached to the surface of the AR2G (amine-reactive second generation) biosensor chip by 250 seconds immersion of N-hydroxysuccinimide/ethyl (dimethyl-aminopropyl) carbodiimide and 250 seconds in 1 methanolamine, and blocked by immersion. After immobilization ($DegP^{S210A}$ on a biosensor chip), incubate the chip with buffer (50mM phosphate, 100mM NaCl). The binding of Lpp^{+Leu} (1 to 500 μM) to $DegP^{S210A}$ was measured for 250 s and dissociation was measured for 250 s. Wavelength shift was measured using the BLItz system (ForteBio, MenloPark, CA).

Strain construction

For construction of the HYO010 strain expressing $DegP^{R207P/Y444A}$, the wild-type degP gene was replaced with an arabinose inducible mutant allele using overlap extension PCR and homologous recombination. Initially, the arabinose-inducible $degP^{R207P/Y444A}$ gene and the chloramphenicol resistance gene (CmR) were amplified in plasmid pSK636 (62). The two DNA fragments were combined by redundant expansion PCR for recombination into bacterial strains. The degP gene of E. coli ER2566 was replaced by the $CmR\text{-}P_{ara}\text{-}degP^{R207P/Y444A}$ allele by λ-Red mediated recombination with the pSIM6 plasmid (39). The pSIM6 plasmid was cured 3 times. To make the lpp deletion strain, the Δlpp-752::kan allele of the JW1667-5 (Keio collection; Coli Genetic Stock Center no. 9417) strain was transferred to W3110 by P1 transduction (62, 80) to create SK366. All the mutations were confirmed by Sanger sequencing of the amplified PCR fragment (Macrogen, South Korea) and possibly the phenotype of the mutant (e.g., low viability of degP deletions at elevated temperatures).

Lpp libraries expressing Lpp variants for selection and growth competition experiments (YRK*, YR**, Y*K*, *RK*, Y***, Y**, YRK**; * are random near the C-terminus Position) was constructed by reverse PCR of a plasmid expressing Lpp^{WT} (pHyo003) using primers carrying randomized codons (Table 3) (75). The positions chosen for mutagenesis were randomized to codon NNK (where N is A, T, G, or C, and K is T or G).

Purified PCR products were introduced into HYO010 by ligation, desalting and electroporation (MicroPulser, Bio-Rad, USA). These cells were grown on solid LB-agar plates supplemented with 0.1 mg/mL ampicillin and 0.2% arabinose to

perform direct selection experiments or in the absence of arabinose to prepare for growth competition experiments. For each library of the former, 10-20 colonies were selected for plasmid isolation and C-terminal sequence (Macrogen) determination.

We scraped approximately 6×10^4 colonies of the latter (before growth competition or cycle 0) and then diluted with 50 mL of supplemented LB broth and 0.1 mg/mL ampicillin. An estimated ~10^6 cells were transferred to a new flask and grown for 8 h at 37 °C plus 0.2% arabinose. Plasmid DNA was extracted after each growth competition cycle with Plasmid Miniprep Kit (Cosmo Genetech) and reintroduced into fresh HYO010 cells, where 10^6 colonies were scraped again for the next competition test. Samples before and after each cycle (three cycles in total) of the growth competition experiment were subjected to Illumina high-throughput sequencing (Macrogen) to analyze the sequence of the C-terminal region of the lpp gene.

Viability assay

To test the viability of strains, overnight cultures were diluted 100-fold in liquid LB broth, grown to log phase at 30 °C, and normalized to OD600 of ~ 0.2. Aliquots of 5 μL from a set of 10-fold serial dilutions were spotted on LB agar plates at 37 °C with or without 0.2% arabinose (46, 62).

Strain	Description	Reference
W3110	Wild-type	
SK324	W3110 degP::mPheS-kanR	(41)
SK339	W3110 degP(R207P/Y444A)	(42)
SK345	W3110 ΔdegP(orf)	(41)
SK366	W3110 Δlpp-752::kan	This study
SK367	W3110 Δlpp-752::kan degP(R207P/Y444A)	(42)
SK371	W3110 lpp(+Leu)	(42)
HYO010	ER2566 CmR-P_{ara}-degP(R207P/Y444A)	This study

Table 1. Bacterial strains used in this study

Plasmids	Description	Reference
pWSK29		(43)
pHyo003	pWSK29-plpp-lpp	This Study
pHyo004	pWSK29-plpp-lpp+Leu	This Study
p7	pET15b-DegP	(42)
pSK534	(6H)-ns_degP(d383-390) or (d357-364)	(41)
pSK617	(6H)-ns_DegP(R207P)	(42)
pSK473	(6H)-ns_degP(Y470A), Y444A	(41)
pSK576	(6H)-ns_degP(Y444A,R207P)	(42)
pSK709	pET28b-6His-(-ss)-lpp	(42)
pSK710	pET28b-6His-(-ss)-lpp+Leu	(42)
pYK2	pET28b-6His-lpp_Leu(ΔI27~D33)	This study
pYK23	pET28b-6His-lpp_Leu(ΔV41~D54)	This study
pYK21	pET28b-6His-lpp_Leu(ΔV55~R68)	This study
pYK1	pET28b-6His-Flag-lpp_Leu	This study
pYK15	pET28b-6His-lpp_Leu(Y56A)	This study
pHyo012	pET28b-6His-(-ss)-lpp_YKI	This study
pHyo013	pET28b-6His-(-ss)-lpp_YYL	This study
pHyo014	pET28b-6His-(-ss)-lpp_LPL	This study
pHyo015	pET28b-6His-(-ss)-lpp_LLL	This study
pHyo016	pET28b-6His-(-ss)-lpp_YPV	This study
pHyo017	pET28b-6His-(-ss)-lpp_KLV	This study
pHyo018	pET28b-6His-(-ss)-lpp_VVT	This study
pHyo019	pET28b-6His-(-ss)-lpp_RLT	This study
pHyo020	pET28b-6His-(-ss)-lpp_WFM	This study
pHyo021	pET28b-6His-(-ss)-lpp_YIY	This study
pHyo022	pWSK29-plpp-lpp_YKI	This study
pHyo023	pWSK29-plpp-lpp_YYL	This study
pHyo024	pWSK29-plpp-lpp_LPL	This study
pHyo025	pWSK29-plpp-lpp_LLL	This study
pHyo026	pWSK29-plpp-lpp_YPV	This study
pHyo027	pWSK29-plpp-lpp_KLV	This study
pHyo028	pWSK29-plpp-lpp_VVT	This study
pHyo029	pWSK29-plpp-lpp_RLT	This study
pHyo030	pWSK29-plpp-lpp_WFM	This study

pHyo031	pWSK29-plpp-lpp_YIY	This study

Table 2. Plasmids used in this study

Oligonucleotides	Sequence (5' to 3')	Description
araC-DegP-down_rv	gtgcccttaaacgGcaaaaagg ccatccgtcag	Amplification of arabinose-inducible DegP gene from pSK636 for recombination
R- Lpp N-cap	TTT AGC GTT GCT GGA CAT ATG G	Cloning truncated Lpp^{+Leu} variants ΔH1-H2 using pET28b (pSK710) or pYK2
F-Lpp-Trunc2	GTT GAC CAG CTG AGC AAC G	Cloning truncated Lpp^{+Leu} variants ΔH1-H2 using pET28b (pSK710) or pYK2
R-Lpp Trunc3B	TTT AGC GTT CAG AGT CTG AAC GTC	Cloning truncated Lpp^{+Leu} variants ΔH3-H4 using pET28b (pSK710) or pYK23
F-Lpp-Trunc4	GTT CAG GCT GCT AAA GAT GAC G	Cloning truncated Lpp^{+Leu} variant ΔH3-H4 using pET28b (pSK710) or pYK23
R-Lpp-Tunc5R	GTC GGA ACG CAT TGC GTT CAC 3'	Cloning truncated Lpp^{+Leu} variant ΔH5-H6 using pET28b (pSK710) or pYK21
F-Lpp-Trunc6	CTG GAC AAC ATG GCT ACT AAA TAC C	Cloning truncated Lpp^{+Leu} variants ΔH5-H6 using pET28b (pSK710) or pYK21
R-pET28 Flag/2	GTC TTT GTA GTC GCC GCT GTG ATG ATG ATG ATG ATG GCT G	Cloning of His-Flag tagged Lpp^{+Leu} using using pET28b (pSK710) or pYK1
F-pET28 Flag/2	GAT GAC GAC AAG GGC AGC GGC CTG GTG CCG CGC GGC AG	Cloning of His-Flag tagged Lpp^{+Leu} using using pET28b (pSK710) or pYK1
R-Lpp Y76A	CGC TTT AGT AGC CAT GTT GTC CAG ACG CTG	Cloning of Y56A variant of of Lpp^{+Leu} using pET28b (pSK710) or pYK15
F-LppLeu end	CGC AAG TTA TAA GCG GCC GCA C	Cloning of Y56A variant of of Lpp^{+Leu} using pET28b (pSK710) or pYK15

araC_up-DegP_Tfup-orf_rv	GGT TGA GGG AGA CCA ATT ATG ACA ACT TGA CGG C	Amplification of arabinose-inducible DegP gene from pSK636 for recombination
CmR_up-DegP_Tfup-orf_fw	CGA GAC TGA AAT ACG ACG TTG ATC GGC ACG TAA G	Amplification of chloramphenicol-resistance gene from pSK636 for recombination
CmR_down_overlap2	CTG ACG GAT GGC CTT TTT GCC GTT TAA GGG CAC	Amplification of chloramphenicol-resistance gene from pSK636 for recombination
lpp-246_Not1_fw	AAAT gcggccgc CTGATGGGCGCTTT TTTTATTTAATCG	Amplification of plpp-Lpp from W3110 chromosome and cloning plpp-lpp into pWSK29
lpp-246_EcoR1_rv	ATT T gaattc CC GCA GCC AGC AAT GC	Amplification of plpp-Lpp from W3110 chromosome and cloning plpp-lpp into pWSK29
lpp_RTH_fw	CTG TGA AGT GAA AAA TGG CGC ACA TTG	Construction of Lpp variant library on pHyo004
lpp-YRK(NNK)_RTH_rv	GTA CTA TTA MNN CTT GCG GTA TTT AGT AGC CAT GTT GTC C	Construction of Lpp variant (YRK*) library on pHyo004
lpp-YR(NNK)(NNK)_RTH_rv	GTA CTA TTA MNN MNN GCG GTA TTT AGT AGC CAT GTT GTCC	Construction of Lpp variant (YR**) library on pHyo004
lpp-YRK(NNK)(NNK)_RTH_rv	GTA CTA TTA MNN MNN CTT GCG GTA TTT AGT AGC CAT GTT GTCC	Construction of Lpp variant (YRK**) library on pHyo004
lpp-Y(NNK)K(NNK)_RTH_rv	GTA CTA TTA MNN CTT MNN GTA TTT AGT AGC CAT GTT GTC CAG ACG	Construction of Lpp variant (Y*K*) library on pHyo004

lpp-(NNK)RK(NNK)_RTH_rv	GTA CTA TTA MNN CTT GCG MNN TTT AGT AGC CAT GTT GTC C	Construction of Lpp variant (*RK*) library on pHyo004
lpp-Y(NNK)3_RTH_rv	GTA CTA TTA MNN MNN MNN GTA TTT AGT AGC CAT GTT GTC C	Construction of Lpp variant (Y***) library on pHyo004
lpp-Y(NNK)2_RTH_rv	GTA CTA TTA MNN MNN GTA TTT AGT AGC CAT GTT GTC CAG ACG	Construction of Lpp variant (Y**) library on pHyo004
pWSK29-lpp-(TAT)Y000_rv2	GTACTATTAMNNMN NMNNATATTTAGTA GCCAT	Construction of Lpp variant (Y***) library on pHyo004
NGS1-F-1	TCGTCGGCAGCGTC AGATGTGTATAAGA GACAGCGCAGCTCG TGCTAACCAGCG	Sample preparation for NGS
NGS1-R-1	GTCTCGTGGGCTCG GAGATGTGTATAAG AGACAGgattaagttgggta acgccagg	Sample preparation for NGS
NGS2-R-N701	CAAGCAGAAGACGG CATACGAGATTAAG GCGAGTCTCGTGGG CTCGG	Sample preparation for NGS
NGS2-F-S501	AATGATACGGCGAC CACCGAGATCTACA CTAGATCGCTCGTC GGCAGCGTC	Sample preparation for NGS
NGS2-F-S502	AATGATACGGCGAC CACCGAGATCTACA CCTCTCTATTCGTCG GCAGCGTC	Sample preparation for NGS

NGS2-F-S503	AATGATACGGCGAC CACCGAGATCTACA CTATCCTCTTCGTCG GCAGCGTC	Sample preparation for NGS
NGS2-F-S504	AATGATACGGCGAC CACCGAGATCTACA CAGAGTAGATCGTC GGCAGCGTC	Sample preparation for NGS
pWSK29_lpp_fw	TAA TAG TAC CTG TGA AGT GAA AAA TGG C	Cloning plpp-Lpp variant into pWSK29 (pHyo004) for individual viability assay
pET28b-Not1_fw	TAA GCG GCC GCA CTC GAG	Cloning His6-Lpp variant into pET28b (pSK710) for expression
Lpp_YKI_rv	GAT TTT GTA GTA TTT AGT AGC CAT GTT GTC CAG AC	Cloning Lpp variant (YKI) into pET28b (pSK710) or pWSK29 (pHyo004)
Lpp_YYL_rv	CAG GTA GTA GTA TTT AGT AGC CAT GTT GTC CAG AC	Cloning Lpp variant (YYL) into pET28b (pSK710) or pWSK29 (pHyo004)
Lpp_LPL_rv	CAG CGG CAG GTA TTT AGT AGC CAT GTT GTC CAG AC	Cloning Lpp variant (LPL) into pET28b (pSK710) or pWSK29 (pHyo004)
Lpp_LLL_rv	CAG CAG CAG GTA TTT AGT AGC CAT GTT GTC CAG AC	Cloning Lpp variant (LLL) into pET28b (pSK710) or pWSK29 (pHyo004)
Lpp_YPV_rv	CAC CGG GTA GTA TTT AGT AGC CAT GTT GTC CAG AC	Cloning Lpp variant (YPV) into pET28b (pSK710) or pWSK29 (pHyo004)

Lpp_KLV_rv	CAC CAG TTT GTA TTT AGT AGC CAT GTT GTC CAG AC	Cloning Lpp variant (KLV) into pET28b (pSK710) or pWSK29 (pHyo004)
Lpp_VVT_rv	GGT CAC CAC GTA TTT AGT AGC CAT GTT GTC CAG AC	Cloning Lpp variant (VVT) into pET28b (pSK710) or pWSK29 (pHyo004)
Lpp_RLT_rv	GGT CAG ACG GTA TTT AGT AGC CAT GTT GTC CAG AC	Cloning Lpp variant (RLT) into pET28b (pSK710) or pWSK29 (pHyo004)
Lpp_WFM_rv	CAT GAA CCA GTA TTT AGT AGC CAT GTT GTC CAG AC	Cloning Lpp variant (WFM) into pET28b (pSK710) or pWSK29 (pHyo004)
Lpp_YIY_rv	GTA GAT GTA GTA TTT AGT AGC CAT GTT GTC CAG AC	Cloning Lpp variant (YIY) into pET28b (pSK710) or pWSK29 (pHyo004)

Table 3. List of primers (oligonucleotides) used in this study.

References

(1) Zeymer, C.; Hilvert, D., Directed Evolution of Protein Catalysts. Annu Rev Biochem 2018, 87, 131-157.

(2) Morrison, M. S.; Podracky, C. J.; Liu, D. R., The developing toolkit of continuous directed evolution. Nat Chem Biol 2020, 16 (6), 610-619.

(3) Komor, A. C.; Kim, Y. B.; Packer, M. S.; Zuris, J. A.; Liu, D. R., Pro-grammable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 2016, 533 (7603), 420-4.

(4) Hess, G. T.; Fresard, L.; Han, K.; Lee, C. H.; Li, A.; Cimprich, K. A.; Montgomery, S. B.; Bassik, M. C., Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. Nat Methods 2016, 13 (12), 1036-1042.

(5) Halperin, S. O.; Tou, C. J.; Wong, E. B.; Modavi, C.; Schaffer, D. V.; Dueber, J. E., CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. Nature 2018, 560 (7717), 248-252.

(6) Ravikumar, A.; Arzumanyan, G. A.; Obadi, M. K. A.; Javanpour, A. A.; Liu, C. C., Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. Cell 2018, 175 (7), 1946-1957 e13.

(7) Camps, M.; Loeb, L. A., Use of Pol I-deficient *E. coli* for functional complementation of DNA polymerase. Methods Mol Biol 2003, 230, 11-8.

(8) Esvelt, K. M.; Carlson, J. C.; Liu, D. R., A system for the continuous directed

evolution of biomolecules. Nature 2011, 472 (7344), 499-503.

(9) Moore, C. L.; Papa, L. J., 3rd; Shoulders, M. D., A Processive Protein Chimera Introduces Mutations across Defined DNA Regions *In vivo*. J Am Chem Soc 2018, 140 (37), 11560-11564.

(10) Chen, H.; Liu, S.; Padula, S.; Lesman, D.; Griswold, K.; Lin, A.; Zhao, T.; Marshall, J. L.; Chen, F., Efficient, continuous mutagenesis in human cells using a pseudo-random DNA editor. Nat Biotechnol 2019.

(11) Lada, A. G.; Krick, C. F.; Kozmin, S. G.; Mayorov, V. I.; Karpova, T. S.; Rogozin, I. B.; Pavlov, Y. I., Mutator effects and mutation signatures of editing deaminases produced in bacteria and yeast. Biochemistry (Mosc) 2011, 76 (1), 131-46.

(12) Nishida, K.; Arazoe, T.; Yachie, N.; Banno, S.; Kakimoto, M.; Tabata, M.; Mochizuki, M.; Miyabe, A.; Araki, M.; Hara, K. Y.; Shimatani, Z.; Kondo, A., Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science 2016, 353 (6305).

(13) Durniak, K. J.; Bailey, S.; Steitz, T. A., The structure of a transcribing T7 RNA polymerase in transition from initiation to elongation. Science 2008, 322 (5901), 553-7.

(14) Salter, J. D.; Bennett, R. P.; Smith, H. C., The APOBEC Protein Family: United by Structure, Divergent in Function. Trends Biochem Sci 2016, 41 (7), 578-594.

(15) Mol, C. D.; Arvai, A. S.; Sanderson, R. J.; Slupphaug, G.; Kavli, B.; Krokan, H. E.; Mosbaugh, D. W.; Tainer, J. A., Crystal structure of human uracil-DNA

glycosylase in complex with a protein inhibitor: protein mimic-ry of DNA. Cell 1995, 82 (5), 701-8.

(16) Kast, P.; Hennecke, H., Amino acid substrate specificity of Escherichia coli phenylalanyl-tRNA synthetase altered by distinct mutations. J Mol Biol 1991, 222 (1), 99-124.

(17) Cirino, P. C.; Mayer, K. M.; Umeno, D., Generating mutant libraries using error-prone PCR. Methods Mol Biol 2003, 231, 3-9.

(18) Badran, A. H.; Liu, D. R., Development of potent *in vivo* mutagenesis plasmids with broad mutational spectra. Nat Commun 2015, 6, 8425.

(19) Bonner, G.; Lafer, E. M.; Sousa, R., Characterization of a set of T7 RNA polymerase active site mutants. J Biol Chem 1994, 269 (40), 25120-8.

(20) Radzicka, A.; Wolfenden, R., A proficient enzyme. Science 1995, 267 (5194), 90-3.

(21) Stemmer, W. P., Rapid evolution of a protein *in vitro* by DNA shuf-fling. Nature 1994, 370 (6488), 389-91.

(22) Zaccolo, M.; Gherardi, E., The effect of high-frequency random muta-genesis on *in vitro* protein evolution: a study on TEM-1 beta-lactamase. J Mol Biol 1999, 285 (2), 775-83.

(23) Barlow, M.; Hall, B. G., Experimental prediction of the natural evolu-tion of antibiotic resistance. Genetics 2003, 163 (4), 1237-41.

(24) Fujii, R.; Kitaoka, M.; Hayashi, K., RAISE: a simple and novel meth-od of

generating random insertion and deletion mutations. Nucleic Acids Res 2006, 34 (4), e30.

(25) Kim, S.; Sauer, R. T., Distinct regulatory mechanisms balance DegP proteolysis to maintain cellular fitness during heat stress. Genes Dev 2014, 28 (8), 902-11.

(26) Nishida, K.; Arazoe, T.; Yachie, N.; Banno, S.; Kakimoto, M.; Tabata, M.; Mochizuki, M.; Miyabe, A.; Araki, M.; Hara, K. Y.; Shimatani, Z.; Kondo, A., Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science 2016, 353 (6305).

(27) Ochman, H.; Gerber, A. S.; Hartl, D. L., Genetic applications of an inverse polymerase chain reaction. Genetics 1988, 120 (3), 621-3.

(28) Yang, S.; Kim, S.; Rim Lim, Y.; Kim, C.; An, H. J.; Kim, J. H.; Sung, J.; Lee, N. K., Contribution of RNA polymerase concentration variation to protein expression noise. Nat Commun 2014, 5, 4761.

(29) Komor, A. C.; Kim, Y. B.; Packer, M. S.; Zuris, J. A.; Liu, D. R., Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 2016, 533 (7603), 420-4.

(30) Kim, S.; Sauer, R. T., Distinct regulatory mechanisms balance DegP proteolysis to maintain cellular fitness during heat stress. Genes Dev 2014, 28 (8), 902-11.

(31) Kim, S.; Sauer, R. T., Cage assembly of DegP protease is not required for substrate-dependent regulation of proteolytic activity or high-temperature cell survival. Proc Natl Acad Sci U S A 2012, 109 (19), 7263-8.

(32) Kim, S.; Grant, R. A.; Sauer, R. T., Covalent linkage of distinct substrate

degrons controls assembly and disassembly of DegP proteolytic cages. Cell 2011, 145 (1), 67-78.

(33) Melancon, C. E., 3rd; Schultz, P. G., One plasmid selection system for the rapid evolution of aminoacyl-tRNA synthetases. Bioorg Med Chem Lett 2009, 19 (14), 3845-7.

(34) Davis, J. H.; Baker, T. A.; Sauer, R. T., Small-molecule control of protein degradation using split adaptors. ACS Chem Biol 2011, 6 (11), 1205-13.

(35) Bolger, A. M.; Lohse, M.; Usadel, B., Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014, 30 (15), 2114-20.

(36) Li, H.; Durbin, R., Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009, 25 (14), 1754-60.

(37) Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; Genome Project Data Processing, S., The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009, 25 (16), 2078-9.

(38) Moore, C. L.; Papa, L. J., 3rd; Shoulders, M. D., A Processive Protein Chimera Introduces Mutations across Defined DNA Regions In Vivo. J Am Chem Soc 2018, 140 (37), 11560-11564.

(39) Park, H.; Kim, Y. T.; Choi, C.; Kim, S., Tripodal Lipoprotein Variants with C-Terminal Hydrophobic Residues Allosterically Modulate Activity of the DegP Protease. J Mol Biol 2017, 429 (20), 3090-3101.

(40) Durniak, K. J.; Bailey, S.; Steitz, T. A., The structure of a transcribing T7 RNA polymerase in transition from initiation to elongation. Science 2008, 322 (5901),

553-7.

(41) A.R. Duguay, T.J. Silhavy, Quality control in the bacterial periplasm, Biochim. Biophys. Acta 1694 (2004) 121–134.

(42) T. Clausen, M. Kaiser, R. Huber, M. Ehrmann, HTRA proteases: regulated proteolysis in protein quality control, Nat. Rev. Mol. Cell Biol. 12 (2011) 152–162.

(43) B. Lipinska, O. Fayet, L. Baird, C. Georgopoulos, Identification, characterization, and mapping of the Escherichia coli htrA gene, whose product is essential for bacterial growth only at elevated temperatures, J. Bacteriol. 171 (1989) 1574–1584.

(44) B. Lipinska, M. Zylicz, C. Georgopoulos, The HtrA (DegP) protein, essential for Escherichia coli survival at high temperatures, is an endopeptidase, J. Bacteriol. 172 (1990) 1791–1797.

(45) K.L. Strauch, K. Johnson, J. Beckwith, Characterization of degP, a gene required for proteolysis in the cell envelope and essential for growth of Escherichia coli at high temperature, J. Bacteriol. 171 (1989) 2689–2696.

(46) S. Kim, R.T. Sauer, Cage assembly of DegP protease is not required for substrate-dependent regulation of proteolytic activity or high-temperature cell survival, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 7263–7268.

(47) T. Krojer, M. Garrido-Franco, R. Huber, M. Ehrmann, T. Clausen, Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine, Nature 416 (2002) 455–459.

(48) J. Skorko-Glonek, D. Zurawa-Janicka, T. Koper, M. Jarzab, D. Figaj, P. Glaza,

et al., HtrA protease family as therapeutic targets, Curr. Pharm. Des. 19 (2013) 977–1009.

(49) D. Zurawa-Janicka, T. Wenta, M. Jarzab, J. Skorko-Glonek, P. Glaza, A. Gieldon, et al., Structural insights into the activation mechanisms of human HtrA serine proteases, Arch. Biochem. Biophys. 621 (2017) 6–23.

(50) S. Kim, R.A. Grant, R.T. Sauer, Covalent linkage of distinct substrate degrons controls assembly and disassembly of DegP proteolytic cages, Cell 145 (2011) 67–78.

(51) T. Krojer, J. Sawa, E. Schafer, H.R. Saibil, M. Ehrmann, T. Clausen, Structural basis for the regulated protease and chaperone function of DegP, Nature 453 (2008) 885–890.

(52) H. Kolmar, P.R. Waller, R.T. Sauer, The DegP and DegQ periplasmic endoproteases of Escherichia coli: specificity for cleavage sites and substrate conformation, J. Bacteriol. 178 (1996) 5925–5929.

(53) M. Merdanovic, N. Mamant, M. Meltzer, S. Poepsel, A. Auckenthaler, R. Melgaard, et al., Determinants of structural and functional plasticity of a widely conserved protease chaperone complex, Nat. Struct. Mol. Biol. 17 (2010) 837–843.

(54) T. Krojer, J. Sawa, R. Huber, T. Clausen, HtrA proteases have a conserved activation mechanism that can be triggered by distinct molecular cues, Nat. Struct. Mol. Biol. 17 (2010) 844–852.

(55) D. Figaj, A. Gieldon, M. Bartczak, T. Koper, U. Zarzecka, A. Lesner, et al., The LD loop as an important structural element required for transmission of the allosteric

signal in the HtrA (DegP) protease from Escherichia coli, FEBS J. 283 (2016) 3471–3487.

(56) D. Figaj, A. Gieldon, A. Polit, A. Sobiecka-Szkatula, T. Koper, M. Denkiewicz, et al., The LA loop as an important regulatory element of the HtrA (DegP) protease from Escherichia coli: structural and functional studies, J. Biol. Chem. 289 (2014) 15880–15893.

(57) A. Sobiecka-Szkatula, A. Gieldon, A. Scire, F. Tanfani, D. Figaj, T. Koper, et al., The role of the L2 loop in the regulation and maintaining the proteolytic activity of HtrA (DegP) protein from Escherichia coli, Arch. Biochem. Biophys. 500 (2010) 123–130.

(58) J. Sohn, R.A. Grant, R.T. Sauer, Allosteric activation of DegS, a stress sensor PDZ protease, Cell 131 (2007) 572–583.

(59) A.K. de Regt, S. Kim, J. Sohn, R.A. Grant, T.A. Baker, R.T. Sauer, A conserved activation cluster is required for allosteric communication in HtrA-family proteases, Structure 23 (2015) 517–526.

(60) Q.T. Shen, X.C. Bai, L.F. Chang, Y. Wu, H.W. Wang, S.F. Sui, Bowl-shaped oligomeric structures on membranes as DegP's new functional forms in protein quality control, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 4858–4863.

(61) J. Jiang, X. Zhang, Y. Chen, Y. Wu, Z.H. Zhou, Z. Chang, et al., Activation of DegP chaperone-protease via formation of large cage-like oligomers upon binding to substrate proteins, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 11939–11944.

(62) S. Kim, R.T. Sauer, Distinct regulatory mechanisms balance DegP proteolysis

to maintain cellular fitness during heat stress, Genes Dev. 28 (2014) 902–911.

(63) V. Braun, Covalent lipoprotein from the outer membrane of Escherichia coli, Biochim. Biophys. Acta 415 (1975) 335–377.

(64) J.M. DiRienzo, K. Nakamura, M. Inouye, The outer membrane proteins of Gram-negative bacteria: biosynthesis, assembly, and functions, Annu. Rev. Biochem. 47 (1978) 481–532.

(65) W. Shu, J. Liu, H. Ji, M. Lu, Core structure of the outer membrane lipoprotein from Escherichia coli at 1.9 Å resolution, J. Mol. Biol. 299 (2000) 1101–1112.

(66) F.Schneider, P.Hammarstrom, J.W.Kelly, Transthyretin slowly exchanges subunits under physiological conditions: a convenient chromatographic method to study subunit exchange in oligomeric proteins, Protein Sci. 10 (2001) 1606–1613.

(67) J.L. Kimmel, G.D. Reinhart, Isolation of an individual allosteric interaction in tetrameric phosphofructokinase from Bacillus stearothermophilus, Biochemistry 40 (2001) 11623–11629.

(68) A. Sultana, J.E. Lee, Measuring protein–protein and protein–nucleic acid interactions by biolayer interferometry, Curr Protoc Protein Sci 79 (19 25) (2015) 11–26.

(69) T. Krojer, K. Pangerl, J. Kurt, J. Sawa, C. Stingl, K. Mechtler, et al., Interplay of PDZ and protease domain of DegP ensures efficient elimination of misfolded proteins, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 7702–7707.

(70) M. Meltzer, S. Hasenbein, P. Hauske, N. Kucz, M. Merdanovic, S. Grau, et al., Allosteric activation of HtrA protease DegP by stress signals during bacterial protein

quality control, Angew Chem Int Ed Engl 47 (2008) 1332–1334.

(71) H. Nakamoto, J.C. Bardwell, Catalysis of disulfide bond formation and isomerization in the Escherichia coli periplasm, Biochim. Biophys. Acta 1694 (2004) 111–119.

(72) M. Drag, G.S. Salvesen, Emerging principles in protease-based drug discovery, Nat. Rev. Drug Discov. 9 (2010) 690–701.

(73) H. Brotz-Oesterhelt, D. Beyer, H.P. Kroll, R. Endermann, C. Ladel, W. Schroeder, et al., Dysregulation of bacterial proteolytic machinery by a new class of antibiotics, Nat. Med. 11 (2005) 1082–1087.

(74) B.P. Conlon, E.S. Nakayasu, L.E. Fleck, M.D. LaFleur, V.M. Isabella, K. Coleman, et al., Activated ClpP kills persisters and eradicates a chronic biofilm infection, Nature 503 (2013) 365–370.

(75) H. Ochman, A.S. Gerber, D.L. Hartl, Genetic applications of an inverse polymerase chain reaction, Genetics 120 (1988) 621–623.

(76) R.F. Wang, S.R. Kushner, Construction of versatile low-copy number vectors for cloning, sequencing and gene expression in Escherichia coli, Gene 100 (1991) 195–199.

(77) R.L. Wiseman, N.S. Green, J.W. Kelly, Kinetic stabilization of an oligomeric protein under physiological conditions demonstrated by a lack of subunit exchange: implications for transthyretin amyloidosis, Biochemistry 44 (2005) 9265–9274.

(78) M.E. Lee, T.A. Baker, R.T. Sauer, Control of substrate gating and translocation into ClpP by channel residues and ClpX binding, J. Mol. Biol. 399 (2010) 707–718.

(79) J.H. Davis, T.A. Baker, R.T. Sauer, Small-molecule control of protein degradation using split adaptors, ACS Chem. Biol. 6 (2011) 1205–1213.

(80) L.C. Thomason, N. Costantino, D.L. Court, *E. coli* genome manipulation by P1 transduction, Curr Protoc Mol Biol Unitas 1 (2007) 17 (Chapter 1).

(81) S. Kim, R.T. Sauer, Cage assembly of DegP protease is not required for substrate-dependent regulation of proteolytic activity or high-temperature cell survival, Proc Natl Acad Sci U S A 109 (2012) 7263-7268.

(82) S. Kim, R.T. Sauer, Distinct regulatory mechanisms balance DegP proteolysis to maintain cellular fitness during heat stress, Genes Dev 28 (2014) 902-911.

(83) R.F. Wang, S.R. Kushner, Construction of versatile low-copy-number vectors for cloning, sequencing and gene expression in Escherichia coli, Gene 100 (1991) 195-199.

국문초록

지속적인 지향적 진화(Continuous Directed Evolution)는 세포에서 단백질의 빠르고 효율적인 진화를 가능하게 했다. 이를 위해 다양한 생체 내 돌연변이 유발 방법이 개발되었지만, 여전히 낮은 돌연변이율, 표적 불가, 좁은 편집 영역과 같은 한계가 있습니다. 우리는 높은 돌연변이 속도와 대장균에서 높은 유전자 특이성을 가진 eMutaT7 이라는 뮤테이터를 개발했다. T7 RNA 중합효소에 cytidine deaminase 를 융합한 eMutaT7 은 하루에 1kb 당 최대 4 개의 돌연변이를 도입 할 수 있고, 이는 기존의 시험관 내 돌연변이 유발 속도와 비슷한 수준이다. 이후, eMutaT7 으로 항생제 내성 단백질과 단백질 분해 효소의 알로스테릭 활성화를 했고, 이는 eMutaT7 이 신속한 연속 진화에 활용될 수 있다는 것이다.

Keywords: 지향적 진화, 유전자-특이적 돌연변이, 시티딘 디아미네이즈, T7 RNA 중합효소, 지속적인 지향적 진화, 세포 내 돌연변이

Student Number: 2014-30078