Title

Response and Adaptation of Escherichia coli to Suppression of the Amber Stop Codon

Authors

Wang, Qian
Sun, Tingting
Xu, Jianfeng
et al.

Peer reviewed

Response and Adaptation of *Escherichia coli* to Suppression of the Amber Stop Codon

Qian Wang,[a, b] Tingting Sun,[a, c] Jianfeng Xu,[a] Zhouxin Shen,[d] Steven P. Briggs,[d] Demin Zhou,[c] and Lei Wang*[a]

Some extant organisms reassign the amber stop codon to a sense codon through evolution, and suppression of the amber codon with engineered tRNAs has been exploited to expand the genetic code for incorporating non-canonical amino acids (ncAAs) in live systems. However, it is unclear how the host cells respond and adapt to such amber suppression. Herein we suppressed the amber codon in *Escherichia coli* with an orthogonal tRNA/synthetase pair and cultured the cells under such a pressure for about 500 generations. We discovered that *E. coli* quickly counteracted the suppression with transposon insertion to inactivate the orthogonal synthetase. Persistent amber suppression evading transposon inactivation led to global proteomic changes with a notable up-regulation of a previously uncharacterized protein (YdiI) for which we identified an unexpected function of expelling plasmids. These results should be valuable for understanding codon reassignment in genetic code evolution and for improving the efficiency of ncAA incorporation.

The canonical genetic code contains 61 sense codons specifying 20 amino acids and three nonsense codons (UAA, UAG and UGA) specifying the stop signal for protein translation. Sense codons are decoded by tRNAs, whereas the nonsense codons are recognized by proteins called class I release factors (RFs), which promote peptide release from the tRNA in the ribosome. The genetic code was once thought to be a "frozen accident" because it was universally preserved in all (then known) organisms and any changes would affect all proteins simultaneously and be deleterious to the host.[1] Small deviations from the

[a] Q. Wang, T. Sun, Dr. J. Xu, Prof. L. Wang
Jack H. Skirball Center for Chemical Biology and Proteomics
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037 (USA)
E-mail: lwang@salk.edu

[b] Q. Wang
College of Veterinary Medicine, China Agricultural University
2 Yuan Ming Yuan West Road, Beijing 100193 (China)

[c] T. Sun, Prof. D. Zhou
State Key Laboratory of Natural and Biomimetic Drugs
School of Pharmaceutical Sciences, Peking University
38 Xue Yuan Road, Beijing 100191 (China)

[d] Dr. Z. Shen, Prof. S. P. Briggs
Section of Cell and Development Biology, University of California at San Diego
9500 Gilman Drive, La Jolla, CA 92093 (USA)

canonical code were later discovered in the mitochondrial and nuclear genetic codes of a number of organisms.[2] These included the reassignment of sense codons from one amino acid to another and, more frequently, the reassignment of a non-sense codon to an amino acid. For instance, the eukaryotic release factor 1 (eRF1) of *Tetrahymena* restricts its recognition to UGA, and UAA/UAG are reassigned to Gln; the eRF1 of *Euplotes* recognizes UAA/UAG only as stop codons, and UGA is used to encode Cys.[3] These deviations in codon assignment suggest that the genetic code is flexible and might be still evolving in extant lineages. However, extant organisms harboring such changes are at the end-point of the evolution of the genetic code. Knowledge of the initial response by an organism, concurrent cellular adaptation, and eventual fixation of codon reassignments is lacking.[4]

Suppression of nonsense codons have been exploited for the incorporation of both canonical and non-canonical amino acids (ncAAs) into proteins.[5] Occurring only once per gene, the relative scarcity of nonsense codons may mitigate any damage caused by codon reassignment. Natural suppressor tRNAs decoding stop codons as common amino acids have been identified in *E. coli* and other organisms.[6] In addition, orthogonal tRNA/synthetase pairs have been generated to incorporate ncAAs into proteins in response to a stop codon directly in live cells.[5c, 7] Such an orthogonal tRNA/synthetase pair does not crosstalk with endogenous tRNA/synthetase pairs of the host cell, and functionally couples with the protein-translation machinery within the cell. The anticodon of the orthogonal tRNA is mutated to pair with a stop codon for specific recognition, and the orthogonal synthetase is engineered to use a desired ncAA as its substrate. This approach has enabled ncAAs with a variety of functional groups to be genetically incorporated into proteins in bacteria, eukaryotic cells, and even mammals.[5c, 8] Genetically encoding ncAAs has the attractive possibility of investigating proteins and biological processes in vivo.[9] However, because the stop codon chosen to encode the ncAA is also used by multiple endogenous genes for translation termination, it is important to understand how the host cell is affected when this stop codon is suppressed by the orthogonal tRNA genome-wide.

To begin to address these questions, we introduced an orthogonal tRNA/synthetase pair into *E. coli* to suppress the amber stop codon UAG, and maintained the suppression in *E. coli* for over 500 generations. We discovered that *E. coli* initially uses transposons to counteract the pressure, and that long term adaptation of *E. coli* to amber suppression involves a previously uncharacterized protein, YdiI, which we have

found to have the function of expelling plasmids from the cell. These results should facilitate the synthetic reassignment of the amber codon in *E. coli* and the improvement of ncAA incorporation efficiency by using amber codon suppression.

To introduce strong amber codon suppression in the *E. coli* genome, a suppressor tyrosyl tRNA derived from archaebacteria *Methanococcus jannaschii* (J17-tRNA$_{CUA}^{Tyr}$) and the cognate tyrosyl-tRNA synthetase (*Mj*TyrRS) were expressed in *E. coli*. The J17-tRNA$_{CUA}^{Tyr}$/*Mj*TyrRS pair is orthogonal to endogenous *E. coli* tRNA/synthetase pairs and incorporates a tyrosine residue in response to the amber stop codon, TAG, with high efficiency.[10] To monitor the amber codon suppression, the gene for enhanced green fluorescent protein (EGFP) containing a TAG codon at a permissive site (Tyr182) was coexpressed as a fluorescent reporter in a single plasmid, pBK-tYGT (Figure 1A). *E. coli* DH10β cells were transformed with the plasmid and found to be fluorescent when excited at a wavelength of 470 nm. Cells harboring pBK-tYGT (doubling time = 43 min)

grew significantly slower than cells harboring pBK-GT (doubling time = 27 min), a control plasmid without the tRNA and synthetase genes (Figure 1B), which suggests that strong amber codon suppression negatively impacts propagation of the cells.

To investigate how *E. coli* would react upon strong amber codon suppression, cells containing pBK-tYGT were continually passaged either in liquid media or on plates using kanamycin to maintain the plasmid. For liquid culture, cells were diluted 2.5×10^8-fold for subculture, and they could reach OD$_{600}$ = 2 in 24 hours. We found that non-fluorescent cells emerged and the percentage of green fluorescent cells decreased with growth time (Figure 1C). After three passages (one passage per day) the liquid culture quickly lost its fluorescence. Plasmids extracted from the non-fluorescent cells showed a shift in mobility on an agarose gel in comparison to those from the fluorescent cells (Figure 2). Separate PCR amplification of the
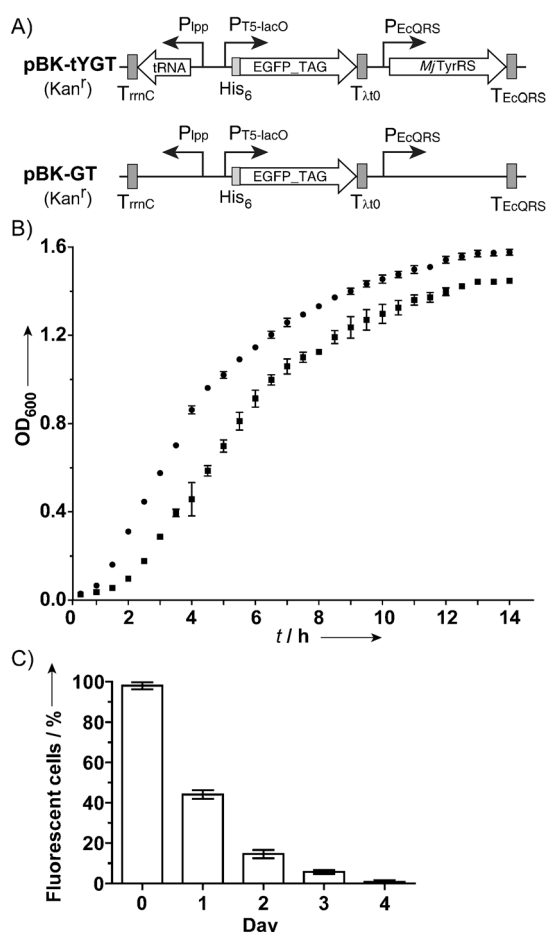

Figure 2. Non-fluorescent cells following long-term amber codon suppression had transposon insertions in the synthetase gene cassette of plasmid pBK-tYGT. Plasmids extracted from non-fluorescent colonies N1–N8 were amplified with primers specific for each gene cassette. Fluorescent colonies F1–F2 and untransformed plasmid pBK-tYGT were used as controls. The shift of plasmids was coordinated with the shift of the synthetase (aaRS) PCR products (bottom panel). M, molecular marker.

Figure 1. Strong amber codon suppression slowed down bacterial growth and led to a decrease in amber codon suppression over time. A) The orthogonal amber codon suppressor tRNA (J17-tRNA$_{CUA}^{Tyr}$), its cognate orthogonal synthetase (*Mj*TyrRS), and the gene for EGFP containing a TAG codon were assembled in the plasmid pBK-tYGT. Plasmid pBK-GT expresses EGFP-TAG only and was used as the control. P = promoter; T = terminator. B) Growth curves of DH10β cells harboring the plasmid pBK-tYGT (squares) or pBK-GT (circles). Error bars represent the s.d., $n = 3$. C) Percentage of fluorescent cells over time. Error bars represent the s.d., $n = 3$.
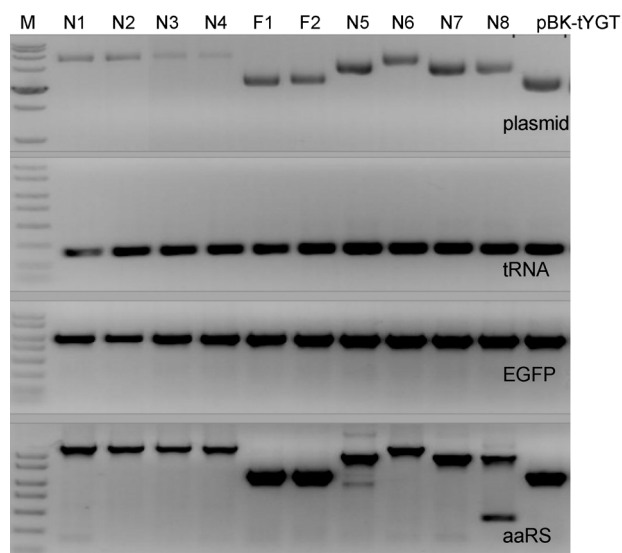
tRNA, GFP-TAG, and *Mj*TyrRS genes showed that the mobility shift of the plasmid resulted from a change in the size of the synthetase gene cassette. DNA sequencing of the isolated plasmids verified that there were insertions in the synthetase gene, and these insertions were all identified as *E. coli* endogenous transposons, IS1 or IS10. The five representative insertion sites and directions of insertion are listed in Table 1. The transposition occurred at a high frequency only in the pBK-tYGT plasmid, whereas the same plasmid construct harboring a tRNA only or a synthetase only was stable under the same growth conditions, indicating that the transposition is in response to the amber codon suppression.

Table 1. Transposon insertion in the synthetase gene cassette.			
Colony	Transposon type	Insertion site[a]	Insertion direction
N1	IS10	427	reverse
N3	IS10	1040	reverse
N5	IS1	404	reverse
N6	IS10	1040	forward
N7	IS1	748	forward
[a] Insertion sites were numbered from the start site of the *Mj*TyrRS gene.			

Transposition as an adaptation to stress has been reported in *E. coli* cells under UV light or continuous stationary-phase growth.[11] It is believed that insertion sequences function as a mutation generator to acquire growth fitness in evolution. They can dynamically regulate gene expression by abolishing, duplicating, or translocating target genes. We sequenced nine insertions in total, and they were all located within the synthetase gene cassette, abolishing the integrity of the gene.

To determine how *E. coli* cells would respond to long-term amber codon suppression, passaging was performed on plates to make the method more controllable. In each round, ten fluorescent colonies were randomly picked and suspended in media, and $1/10^5$ of the suspension was spread on the subsequent plate containing kanamycin. The continuous use of fluorescent cells ensured that the amber codon suppression had not been eliminated by transposons. Passaging was carried out for 22 rounds, which added up to approximately 450–500 generations. The initial and final cultures of cells were collected and subjected to quantitative whole-proteome profiling using mass spectrometry. Proteins whose amount differed more than twofold are listed in Table 2.

Among the proteins with large changes in expression, the elongation factor EF-Tu was the only protein directly related to translation, which was decreased to 0.39. Because EF-Tu facilitates acylated tRNA entering into the ribosome A site, decreased amounts of EF-Tu expression would slow down translation and cell growth. A decrease in EF-Tu would also decrease the delivery of the amber codon suppressor tRNA to the TAG site, whereas the competitor, release factor 1, was not affected. This could lead to more efficient termination than suppression at the TAG sites, which would miti-

gate the amber codon suppression pressure. The genes of four of the identified proteins end with a TAG stop codon, and their protein products were all decreased. A possible mechanism for the decrease of these proteins could be related to tmRNA-mediated protein degradation of no-stop mRNA products.[12] The family of outer membrane proteins had nine members with the highest amount of decrease, suggesting an impairment of the membrane translocation system.[13] Cold-shock proteins (Csp) were another family identified but, in contrast, show up-regulated expression. Csp family members normally show increased expression following cold shock, and are usually considered to be RNA or protein chaperones to maintain the necessary transcription and translation efficiency.[14] The most up-regulated protein identified was a hypothetical protein, YdiI, which showed a remarkable 16-fold increase (Table 2). We thus investigated its role in helping *E. coli* cells adapt to amber codon suppression.

YdiI contains a HotDog folding domain and was predicted as a putative thioesterase. Enzyme genomics confirmed its esterase activity and structural genomics showed high similarity to another CoA-conjugate thioesterase.[15] We cloned the *ydiI* gene into the pLEI plasmid and overexpressed it by inducing with IPTG. When pLEI-ydiI was co-transformed with pBK-tYGT into *E. coli* cells, we found that the overexpression of ydiI changed colony morphology (Figure 3 A). Colonies expressing YdiI were no longer uniformly fluorescent; less fluorescent, "white" spots appeared inside the colonies. When DH10β cells freshly transformed with pLEI-ydiI and pBK-tYGT grew for

| Table 2. Proteins up- or down-regulated more than twofold after persistent, long-term amber codon suppression in *E. coli* cells.					
Accession No.[a]	Protein	P22/P0[b]	Accession No.	Protein	P22/P0
NP_417479	ExbB	0.15	NP_415278	GalK	0.46
NP_416719	OmpC	0.19	NP_415280	GalE	0.47
NP_415772	OmpW	0.20	NP_418392	ArgE	0.47
NP_418232	WzzE[c]	0.21	NP_416818	FolC	0.48
NP_415895	OmpN	0.27	NP_417616	YraM	0.49
NP_415085	NmpC	0.27	NP_418258	UvrD	2.00
NP_415449	OmpF	0.27	NP_417573	YqjG	2.01
NP_417950	PitA	0.27	NP_417079	YfiQ	2.03
NP_418593	Hfq	0.32	NP_417348	YgeY	2.05
NP_418476	DnaB	0.32	NP_417432	AnsB	2.08
NP_417887	GlgA[c]	0.34	NP_414978	QueC	2.09
NP_415335	OmpX	0.34	NP_416476	HchA	2.11
NP_417875	MalQ[c]	0.34	NP_418222	IlvC	2.14
NP_416693	YejM	0.37	NP_414995	AcrB	2.36
NP_415710	LdcA	0.37	NP_416567	YegH	2.38
NP_414724	LpxB	0.38	NP_415581	YceB	2.43
NP_418407	EF-Tu	0.39	NP_415820	PspA	2.52
NP_416175	PurR	0.41	NP_414702	Dgt	2.68
NP_417244	CysJ	0.42	NP_415430	Cmk	2.76
NP_418142	IbpA	0.42	NP_416200	YdiH	3.12
NP_415348	MoeA	0.43	NP_417354	YgfK	3.93
NP_416930	AmiA	0.43	NP_415510	CspG	4.23
NP_417999	DppC	0.44	NP_415212	YbfF	5.15
NP_418141	IbpB	0.44	NP_418012	CspA	6.78
NP_416678	YeiR	0.46	NP_416075	CspB	8.34
NP_418193	AtpE[c]	0.46	NP_416201	YdiI	16.19
[a] NCBI accession number. [b] Ratios of proteins from final round (P22) and initial round (P0) cells are listed as P22/P0. [c] Proteins whose encoding genes end with a TAG codon.					
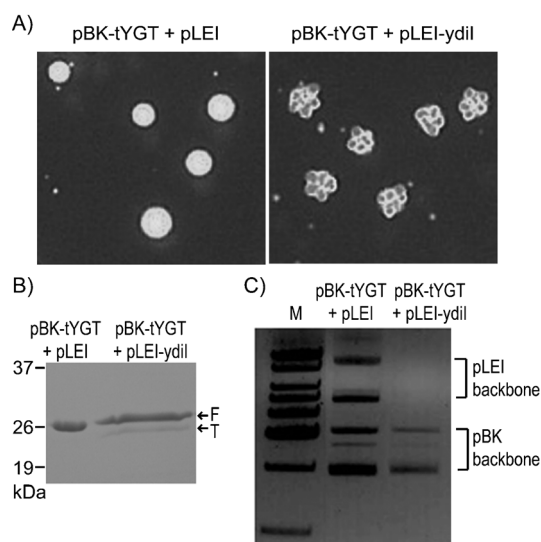
Figure 3. Ydil overexpression reduces EGFP reporter-gene expression by decreasing the plasmid copy number. A) Plasmid pBK-tYGT was co-transformed with pLEI-ydil or an empty vector pLEI into *E. coli* DH10β cells. Fluorescence images were taken of colonies grown on agarose plates. The colony morphology was altered in cells overexpressing Ydil. B) Full-length (F) and truncated (T) EGFP with an N-terminal His$_6$-tag were purified and separated by 12% SDS-PAGE. Note: 40-fold higher loading for Ydil overexpressing cells in a larger well. C) Plasmids extracted from the same number of cells were separated in an agarose gel and stained with ethidium bromide. The amount of plasmids in Ydil overexpressing cells was dramatically lower than the control. Note: plasmids were not linearized prior to electrophersis.

Figure 4. Expression of Ydil expels plasmids from *E. coli* cells. A) Method used to measure the percentage of cells that survived on plates containing different antibiotics. In cells transformed with pBK-tYGT and pLEI-ydil, Ydil was expressed by pLEI-ydil upon induction with IPTG. Control cells were transformed with pBK-tYGT and pLEI, a control plasmid without the *ydil* gene. Plasmid pBK-tYGT is kanamycin (Kan) resistant, and pLEI is chloramphenicol (Cm) resistant. B) Percentage of surviving cells transformed with pBK-tYGT and pLEI. C) Percentage of surviving cells transformed with pBK-tYGT and pLEI-ydil. Error bars represent the S.E.M., *n* = 3.

48 hours in liquid media in the presence of IPTG (0.1 mм), their green fluorescence intensity was dramatically reduced compared to that of cells transformed with pBK-tYGT and an empty vector (pLEI). A decrease in EGFP fluorescence could result from less efficient suppression of the TAG codon introduced at the Tyr182 site (which would increase the amount of truncated EGFP), or from overall lower expression of the plasmid pBK-tYGT. To examine the observed decrease and distinguish between the two possibilities, we purified the full-length and truncated EGFP proteins by Ni^{2+} affinity chromatography and separated them by SDS-PAGE (Figure 3B). The Ydil-overexpressing cells produced 20-fold less full-length EGFP, but didn't accumulate much more truncated EGFP. These results indicated that EGFP expression was reduced in the presence of Ydil but not because of increased premature termination at the introduced TAG codon.

To check if the gene for EGFP or the orthogonal tRNA/TyrRS pair were deactivated, we sequenced both plasmids, which had been isolated from the cells. However, no mutation, insertion, or deletion in the genes for the tRNA, synthetase, or EGFP was found. Surprisingly, plasmids extracted from Ydil-overexpressing cells showed a significant decrease in the amount of both the pBK-tYGT and the pLEI-ydil plasmid, suggesting that the decrease of EGFP expression was mainly due to a lower number of plasmid copies (Figure 3C).

To verify if Ydil overexpression indeed expels plasmids from *E. coli*, we grew *E. coli* cells with and without Ydil expression and quantified the number of colonies capable of surviving in
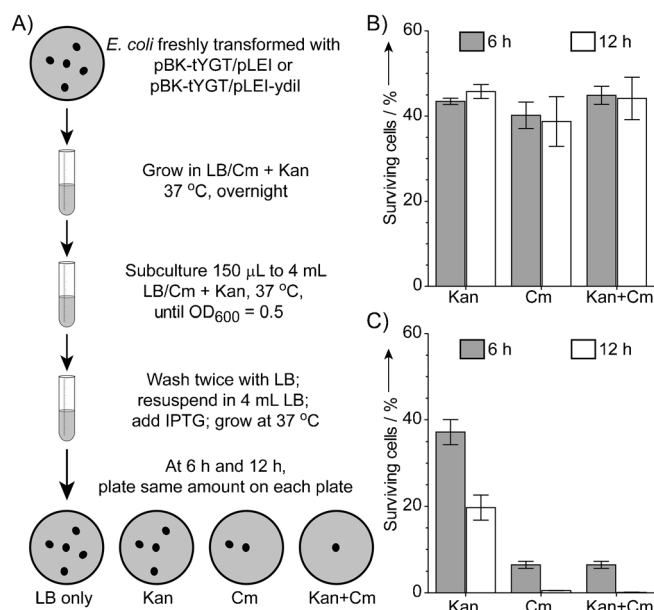
antibiotics corresponding to the resistance conferred by each plasmid (Figure 4A). The percentage of antibiotic-resistant cells relative to the total number of cells growing on an antibiotic-free plate was determined, as a measure of plasmid retention. In the absence of Ydil expression, no significant change in plasmid retention was observed for either plasmid pBK-tYGT containing a ColEI origin of replication (ori) or plasmid pLEI containing a p15A ori over time (Figure 4B). In contrast, when Ydil expression was induced by IPTG, it reduced kanamycin resistant cells (conferred by plasmid pBK-tYGT) from 46% to 20% in 12 hours. The decrease of the chloramphenicol resistant cells (conferred by plasmid pLEI-Ydil) was from 40 to 6.5% in six hours and further to 0.5% in 12 hours. When both antibiotics were present, only 6.5% of cells could survive after six hours and 0.1% after 12 hours. These results indicate that Ydil has the unexpected ability to expel plasmids from *E. coli*, the mechanism of which warrants further investigation.

In summary, upon strong amber codon suppression, *E. coli* DH10β cells immediately try to inactivate the responsible gene using transposon insertion. If transposons miss the target and amber codon suppression persists, the expression of Ydil is upregulated, which reduces the copy number of plasmids to attenuate the suppression effect. Amber codon suppression is currently the prevalent method to genetically incorporate ncAAs into proteins in live cells. The incorporation efficiency of ncAAs does not match that of canonical amino acids.[4, 16] Our results suggest that responses from host cells counteract the codon suppression and may set a limit for amber codon su-

pression. This limit might be overcome by using a transposon-free strain[17] as well as manipulating the proteins identified herein. In addition, codon reassignment has been hypothesized to occur during the evolution of the genetic code.[18] A detailed understanding of how cells respond to amber codon suppression will help to evaluate whether and how a stop codon can be reassigned to a sense codon, which could provide direct evidence for a challenging evolutionary question.

Experimental Section

Plasmid construction: The gene for EGFP with a TAG codon at the site of Tyr182 was first cloned into the plasmid pLEIG[19] to replace the gene for αGFP. The J17-tRNA$_{CUA}^{Tyr}$ and EGFP-TAG expression cassettes from the resultant plasmid were then subcloned into the plasmid pBK-JYRS[5c] to afford pBK-tYGT. Plasmid pBK-GT was derived from pBK-tYGT by deleting the genes for the tRNA and the *Mj*TyrRS. Plasmid pLEI was made from pLEIG by deleting the gene for αGFP and the tRNA expression cassette. The *ydiI* gene was amplified from *E. coli* genomic DNA using primers SpeI_ydiI_5: ccA CTA GTa tga tat gga aac gaa aaa tcc ccc t and BglII_ydiI_3: ccA GAT CTc aaa atg gcg gtc gtc aat cg. The PCR product was cloned into pLEI using *Spe*I and *Bgl*II sites to afford pLEI-ydiI, which expresses YdiI upon IPTG induction.

Liquid culture and fluorescent cell counting: Following 24h of liquid culture, cells were diluted 10^6-fold and 100 µL of the diluted culture was spread on an agarose plate with kanamycin (50 µg mL^{-1}). The fluorescence intensity of the colonies were recorded by a macro fluorescence-imaging system (Lightools Research; Encinitas, CA, USA) with an excitation wavelength of 470 nm (bandwidth 40 nm). The percentage of fluorescent colonies was calculated and plotted in Figure 1C.

Mass-spectrometric profiling of the proteomic change: P0 and P22 *E. coli* cells were harvested and washed three times using PBS buffer. The cells were lysed in RapiGest (2%; Waters Corp.) and HEPES buffer (10 mM HEPES, 150 mM NaCl, pH 7.2) by a Branson Sonifier 450 Ultrasonic Homogenizer. The proteins were reduced using tris-(2-carboxyethyl)-phosphine (2 mM, 95 °C, 5 min) and alkylated using iodoacetamide (5 mM, 37 °C, 30 min, in the dark). The proteins were then digested by using trypsin (1:50 dilution, overnight). The digested peptide samples of P0 and P22 (50 µg each) were labeled by using iTRAQ 115 and 117 reagents (Sigma–Aldrich), respectively. The labeled peptides were mixed and subjected to Nano-LC–MS/MS analysis. Automated 2D nanoflow LC–MS/MS analysis was performed on a LTQ tandem mass spectrometer (Thermo Electron Corporation), which employed automated data-dependent acquisition. An Agilent 1100 HPLC system was used to deliver a flow rate of 300 nL min^{-1} to the mass spectrometer through a splitter. Chromatographic separation was accomplished by using a three-phase capillary column. Using an in-house constructed pressure cell, Zorbax SB-C18 packing material (5 µm) was packed into fused-silica capillary tubing (200 µm ID, 360 µm OD, 20 cm long) to form the first-dimension reverse-phase (RP) column (RP1). A similar column (200 µm ID, 5 cm long) packed with PolySulfoethyl (5 µm, PolyLC) packing material was used as the ion-exchange (SCX) column. A 1 µm zero-dead-volume filter (Upchurch, M548) was attached to the exit of each column for column packing and connection. A fused-silica capillary (100 µm ID, 360 µm OD, 20 cm long) packed with Zorbax SB-C18 packing material (5 µm) was used as the analytical column (RP2). One end of the fused silica tubing was pulled into a sharp tip with an ID

less than 1 µm using a laser puller as the electrospray tip. The peptide mixtures were loaded onto the RP1 column using the same in-house pressure cell. Peptides were first eluted from the RP1 column to the SCX column by using a 0–80% acetonitrile gradient over 150 min. The peptides were then fractionated by using the SCX column through a series of salt gradients (20, 30, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, 120, 150, 180, 200, 1000 mM ammonium acetate for 20 min each), followed by high-resolution RP separation with an acetonitrile gradient of 0–80% over 120 min. The full MS scan range of 400–2000 *m/z* was divided into three smaller scan ranges (400–800, 800–1050, 1050–2000 *m/z*) to improve the dynamic range. Both collision-induced dissociation (CID) and pulsed-Q dissociation (PQD) scans of the same parent ion were collected for protein identification and quantitation. Each MS scan was followed by four pairs of CID-PQD MS/MS scans of the most intense ions from the parent MS scan. A dynamic exclusion of 1 min was used to improve the duty cycle of MS/MS scans. About 20 000 MS/MS spectra were collected for each sample. The raw data was extracted and searched using Spectrum Mill (Agilent, ver. A.03.02). The CID and PQD scans from the same parent ion were merged together. MS/MS spectra with a sequence tag length of one or less were considered to be poor spectra and discarded. The rest of the MS/MS spectra were searched against the NCBI RefSeq protein database limited to the *E. coli* proteome (16 324 sequences). The enzyme parameter was limited to full tryptic peptides with a maximum miscleavage of one. All other search parameters were set to the default setting for Spectrum Mill (carbamidomethylation of cysteines, iTRAQ modification, ±2.5 Da for precursor ions, ±0.7 Da for fragment ions, and a minimum matched peak intensity of 50%). A concatenated forward–reverse database was constructed to calculate the in-situ false discovery rate (FDR). Proteins with shared peptide(s) were grouped together into protein groups. A protein identification cut-off of 1% FDR at the protein-group level was used. A total of 1132 protein groups corresponding to 3130 RefSeq proteins were identified, among them 11 protein groups and 29 proteins were from the reverse database. Relative protein quantitation was performed by calculating the 117/115 (P22/P0) iTRAQ reported ion intensity ratios. Protein iTRAQ intensities were calculated by summing the peptide iTRAQ intensities from each protein group. Peptides shared among different protein groups were removed before quantitation. Among the 1132 identified protein groups, 1097 had iTRAQ reporter intensities strong enough to obtain a relative quantitation.

[1] F. H. Crick, *J. Mol. Biol.* **1968**, *38*, 367–379.
[2] a) S. Osawa, T. Jukes, K. Watanabe, A. Muto, *Microbiol. Rev.* **1992**, *56*, 229–264; b) R. D. Knight, S. J. Freeland, L. F. Landweber, *Nat. Rev. Genet.* **2001**, *2*, 49–58.
[3] a) C. Lozupone, R. Knight, L. Landweber, *Curr. Biol.* **2001**, *11*, 65–74; b) Y. Inagaki, W. F. Doolittle, *Nucleic Acids Res.* **2001**, *29*, 921–927.
[4] D. B. Johnson, C. Wang, J. Xu, M. D. Schultz, R. J. Schmitz, J. R. Ecker, L. Wang, *ACS Chem. Biol.* **2012**, *7*, 1337–1344.

[5] a) C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, *Science* **1989**, *244*, 182–188; b) M. W. Nowak, P. C. Kearney, J. R. Sampson, M. E. Saks, C. G. Labarca, S. K. Silverman, W. G. Zhong, J. Thorson, J. N. Abelson, N. Davidson, P. G. Schultz, D. A. Dougherty, H. A. Lester, *Science* **1995**, *268*, 439–442; c) L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498–500.

[6] a) S. Benzer, S. P. Champe, *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 1114–1121; b) H. Beier, M. Grimm, *Nucleic Acids Res.* **2001**, *29*, 4767–4782.

[7] L. Wang, P. G. Schultz, *Angew. Chem. Int. Ed.* **2005**, *44*, 34–66; *Angew. Chem.* **2005**, *117*, 34–68.

[8] a) K. Sakamoto, A. Hayashi, A. Sakamoto, D. Kiga, H. Nakayama, A. Soma, T. Kobayashi, M. Kitabatake, K. Takio, K. Saito, M. Shirouzu, I. Hirao, S. Yokoyama, *Nucleic Acids Res.* **2002**, *30*, 4692–4699; b) W. Liu, A. Brock, S. Chen, S. Chen, P. G. Schultz, *Nat. Methods* **2007**, *4*, 239–244; c) W. Wang, J. K. Takimoto, G. V. Louie, T. J. Baiga, J. P. Noel, K. F. Lee, P. A. Slesinger, L. Wang, *Nat. Neurosci.* **2007**, *10*, 1063–1072; d) B. Shen, Z. Xiang, B. Miller, G. Louie, W. Wang, J. P. Noel, F. H. Gage, L. Wang, *Stem Cells* **2011**, *29*, 1231–1240; e) J. Y. Kang, D. Kawaguchi, I. Coin, Z. Xiang, D. D. O'Leary, P. A. Slesinger, L. Wang, *Neuron* **2013**, *80*, 358–370.

[9] Reviews: a) L. Wang, J. Xie, P. G. Schultz, *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225–249; b) Q. Wang, A. R. Parrish, L. Wang, *Chem. Biol.* **2009**, *16*, 323–336; c) C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413–444; Recent examples: d) I. Nikić, T. Plass, O. Schraidt, J. Szymański, J. A. Briggs, C. Schultz, E. A. Lemke, *Angew. Chem. Int. Ed.* **2014**, *53*, 2245–2249; *Angew. Chem.* **2014**, *126*, 2278–2282; e) M. J. Schmidt, J. Borbas, M. Drescher, D. Summerer, *J. Am. Chem. Soc.* **2014**, *136*, 1238–1241; f) Y. J. Lee, B. Wu, J. E. Raymond, Y. Zeng, X. Fang, K. L. Wooley, W. R. Liu, *ACS Chem. Biol.* **2013**, *8*, 1664–1670; g) M. J. Gattner, M. Vrabel, T. Carell, *Chem. Commun.* **2013**, *49*, 379–381; h) F. Li, H. Zhang, Y. Sun, Y. Pan, J. Zhou, J. Wang, *Angew. Chem. Int. Ed.* **2013**, *52*, 9700–9704; *Angew. Chem.* **2013**, *125*, 9882–9886; i) Y. Yang, S. Lin, W. Lin, P. R. Chen, *ChemBioChem* **2014**; DOI: 10.1002/cbic.201400057 (this issue); j) I. Coin, V. Katritch, T. Sun, Z. Xiang, F. Y. Siu, M. Beyermann, R. C.

Stevens, L. Wang, *Cell* **2013**, *155*, 1258–1269; k) S. Zhu, M. Riou, C. A. Yao, S. Carvalho, P. C. Rodriguez, O. Bensaude, P. Paoletti, S. Ye, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 6081–6086.

[10] L. Wang, P. G. Schultz, *Chem. Biol.* **2001**, *8*, 883–890.

[11] a) R. Skaliter, Z. Eichenbaum, H. Shwartz, R. Ascarelli-Goell, Z. Livneh, *Mutat. Res.* **1992**, *267*, 139–151; b) Z. Eichenbaum, Z. Livneh, *Genetics* **1998**, *149*, 1173–1181.

[12] S. D. Moore, R. T. Sauer, *Annu. Rev. Biochem.* **2007**, *76*, 101–124.

[13] N. Ruiz, D. Kahne, T. J. Silhavy, *Nat. Rev. Microbiol.* **2006**, *4*, 57–66.

[14] S. Phadtare, *Curr. Issues Mol. Biol.* **2004**, *6*, 125–136.

[15] J. Badger, J. M. Sauder, J. M. Adams, S. Antonysamy, K. Bain, M. G. Bergseid, S. G. Buchanan, M. D. Buchanan, Y. Batiyenko, J. A. Christopher, S. Emtage, A. Eroshkina, I. Feil, E. B. Furlong, K. S. Gajiwala, X. Gao, D. He, J. Hendle, A. Huber, K. Hoda, et al., *Proteins Struct. Funct. Bioinf.* **2005**, *60*, 787–796.

[16] a) T. Mukai, A. Hayashi, F. Iraha, A. Sato, K. Ohtake, S. Yokoyama, K. Sakamoto, *Nucleic Acids Res.* **2010**, *38*, 8188–8195; b) D. B. Johnson, J. Xu, Z. Shen, J. K. Takimoto, M. D. Schultz, R. J. Schmitz, Z. Xiang, J. R. Ecker, S. P. Briggs, L. Wang, *Nat. Chem. Biol.* **2011**, *7*, 779–786; c) F. J. Isaacs, P. A. Carr, H. H. Wang, M. J. Lajoie, B. Sterling, L. Kraal, A. C. Tolonen, T. A. Gianoulis, D. B. Goodman, N. B. Reppas, C. J. Emig, D. Bang, S. J. Hwang, M. C. Jewett, J. M. Jacobson, G. M. Church, *Science* **2011**, *333*, 348–353.

[17] G. Posfai, G. Plunkett III, T. Feher, D. Frisch, G. M. Keil, K. Umenhoffer, V. Kolisnychenko, B. Stahl, S. S. Sharma, M. de Arruda, V. Burland, S. W. Harcum, F. R. Blattner, *Science* **2006**, *312*, 1044–1046.

[18] D. W. Schultz, M. Yarus, *J. Mol. Biol.* **1994**, *235*, 1377–1380.

[19] L. Wang, J. Xie, A. A. Deniz, P. G. Schultz, *J. Org. Chem.* **2003**, *68*, 174–176.