Kawahara-Kobayashi et al. Supplementary Data

Supplementary Methods

DNA constructs and *in vitro* transcription

Mutations were introduced into MBP, LexA, CAT, GFP and tRNA constructs by site directed mutagenesis. The mutations were confirmed by DNA sequencing. MBP-W_all_A was chemically synthesized (GeneArt) and cloned into the pK7 vector. The N-terminal His tag was introduced into the MBP, LexA, and CAT genes. The C-terminal His tag was introduced into the GFP genes. A PreScission Protease (GE Healthcare) recognition site was introduced between the His tag and the target protein in the GFP constructs used for crystallization, and in the LexA constructs.

Cell-free protein expression

For protein synthesis with the $tRNA^{Ser}$ variant that reassigns Cys to the UGU/UGC codons, Cys-SA was added to both the internal and external solutions of the dialysis mode reaction mixture at a final concentration of 5.0 μM every hour, to maintain the inhibition of cysteinyl-tRNA synthetase.

Purification

For the GFP purification under native conditions, the translation products were centrifuged at 12,000 g and 4 ℃ for 10 min. Aliquots (3 ml) of the supernatants were mixed with 7 ml of buffer A (40 mM Tris-HCl, pH 7.6, 300 mM NaCl, 10 mM imidazole, 1 mM DTT) and filtered through Millex-GV 0.45 μm filters (Millipore). The products were incubated with TALON resin (Clontech) at 4 ℃ for 1 hr. The beads were washed with buffer A. The proteins were eluted with elution buffer (same as the buffer A, but including 300 mM imidazole).

For the purification of MBP and CAT under denaturing conditions, 1 ml portions of the translation products were mixed with 9 ml of denaturing buffer A (8 M urea, 50 mM phosphate buffer, pH 7.8, 300 mM NaCl, 10 mM imidazole) and vortexed at 37 ℃ for 1 hr. The products were then incubated with TALON resin at room temperature for 1 hr. The beads were washed with denaturing buffer B (same as denaturing buffer, but including 40 mM imidazole). The proteins were eluted with denaturing-elution buffer (same as denaturing buffer A, but including 150 mM imidazole).

Amino acid composition analysis

To detect Cys, the products were reduced with 1% DTT and alkylated with 2.5% iodoacetic acid. After the reduction and alkylation, the samples were fractionated by SDS-PAGE, electroblotted onto PVDF membranes and stained with Coomassie Brilliant Blue. The band on the membrane was excised with a clean razor blade. The samples were hydrolyzed for 24, 48, and 72 hr. Derivatization with aminoquinolyl-N-hydroxysuccinimidyl carbamate and chromatographic analyses of the AQC-amino acids were performed as described (53), except an Agilent Technologies 1200 series SL

HPLC system was used. The data thus obtained, except for the WT/Univ-24 hr data, were used for linear extrapolation to 0 time.

The amino acid content for amino acid X per protein molecule, C_{X1}, represented in Table S1, was calculated using M_X, the measured value of the amino acid of interest, by the following equation:

$$C_{X1} = \frac{M_X}{\frac{1}{2}\left(\frac{M_F}{C_{FT}} + \frac{M_L}{C_{LT}}\right)}$$

where M_F and M_L denote the measured values of Phe and Leu, respectively. C_{FT} and C_{LT} denote the theoretical amino acid contents per protein molecule for Phe and Leu, respectively.

The amino acid content for amino acid X per protein molecule, C_{X3}, represented in Table S3, was calculated using M_X, the value of the measured amount of the amino acid of interest, by the equation below:

$$C_{X3} = \frac{M_X}{\frac{1}{3}\left(\frac{M_F}{C_{FT}} + \frac{M_L}{C_{LT}} + \frac{M_A}{C_{AT}}\right)}$$

where M_A denotes the measured value of Ala. C_{AT} denotes the theoretical amino acid content per protein molecule for Ala.

Phe, Leu, Ala were employed in the normalization because they are relatively stable in the acid hydrolysis.

Tryptophan fluorescence measurement

Translation of MBP was performed by using the small-scale (60 µl internal/600 µl external) dialysis mode cell-free reaction (54) at 37 ℃ for 8 hr. The products containing the N-terminal polyhistidine tag were purified under denaturing conditions. The tryptophan fluorescence spectra were measured at room temperature with an FP-6500 spectrofluorometer (JASCO). The excitation wavelength was set at 280 nm, and the emission was recorded from 290 to 500 nm (band-width = 1 nm for excitation and 1 nm for emission). For the measurements, 15.0 µg of each purified sample were diluted in 500 µl of denaturing buffer A.

Crystallization data collection

For data collection, the drop solution was slowly equilibrated against the same reservoir solution used for crystallization. The crystals thus obtained were briefly transferred to 1.2 × reservoir solution and were flash-cooled in a cryo-stream of nitrogen gas at 100 K. The data sets of the crystals were collected at station BL41XU of SPring-8 (Harima, Japan). The data sets of ΔUGG-A110X(UGG)/Sim and ΔUGG/Univ were collected at 100 K in a cryo-stream of nitrogen gas. The collected data sets were processed with HKL2000 (55).

Structure determination and refinement

Molecular replacement was performed with MOLREP (56). The resulting initial model was manually improved to conform to the electron density, using the program COOT (57). Eventually, the atomic models of gfpΔUGG-A110X(UGG)/Sim and ΔUGG/Univ were refined up to 2.10 Å and 1.85 Å resolutions, respectively, using the programs PHENIX (58) and REFMAC5 (59). The crystallographic data collection and refinement statistics are presented in Supplementary Table S2. Molecular graphics were illustrated with CueMol 2 (http://www.cuemol.org/).

LexA cleavage assay

The plasmid containing the LexA substrate gene was transformed into Rosetta2 (DE3) pLysS cells (Merck), which were grown at 37 ℃. At an optical density of 0.6 at 590 nm, protein expression was induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside. After four hours of induction, the bacterial cells were collected by centrifugation. The harvested cells were suspended in buffer B (same as buffer A, but including 40 mM imidazole) before sonication. The disrupted cells were centrifuged at 12,000 g and 4 ℃ for 10 min. Cell-free translation of LexA enzymes was performed by using the middle-scale (1 ml internal/10 ml external) dialysis mode reaction (60) at 30 ℃ for 8 hr. The cell-free products were centrifuged as above. Each supernatant was purified under native conditions, as described above for GFP. The LexA cleavage assay was performed at 37℃ for 36 hr, in 20 mM CAPS-NaOH (pH 10.0), 200 mM NaCl, 3 μM LexA enzyme, and 30 μM LexA substrate. At the end of the incubation time, the samples were mixed with 0.2 vol of 6 × protein stop buffer, containing 10% SDS, 36% glycerol, 0.5 M DTT, 0.175 M Tris-HCl (pH 6.8) and 0.012% bromophenol blue, and were analyzed by electrophoresis on a 15% polyacrylamide gel containing 6 M urea, 0.1 M sodium phosphate (pH 7.2), and 0.1% SDS, with a 3.5% stacking gel in the same buffer; the running buffer was 0.1 M sodium phosphate (pH 7.2), 0.1% SDS. This gel system afforded better separation of these low molecular weight proteins than the Laemmli gel system. The gel was stained with Coomassie Brilliant Blue.

Supplementary References

53. Masuda, A. and Dohmae, N. (2010) Automated Protein Hydrolysis Delivering Sample to a Solid Acid Catalyst for Amino Acid Analysis. *Analytical Chemistry*, **82**, 8939-8945.
54. Matsuda, T., Kigawa, T., Koshiba, S., Inoue, M., Aoki, M., Yamasaki, K., Seki, M., Shinozaki, K. and Yokoyama, S. (2006) Cell-free synthesis of zinc-binding proteins. *J Struct Funct Genomics*, **7**, 93-100.
55. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology*, **276**, 307-326.
56. Vagin, A. and Teplyakov, A. (1997) MOLREP: an automated program for molecular

replacement. *J. Appl. Crystallogr.*, **30**, 1022-1025.

57. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, **60**, 2126-2132.

58. Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K. and Terwilliger, T.C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr*, **58**, 1948-1954.

59. Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*, **53**, 240-255.

60. Kigawa, T., Inoue, M., Aoki, M., Matsuda, T., Yabuki, T., Seki, E., Harada, T., Watanabe, S. and Yokoyama, S. (2007) In Spirin, A. S. and Swartz, J. R. (eds.), *CELL-FREE PROTEIN SYNTHESIS : METHODS AND PROTOCOLS.* Wiley-VCH, Weinheim, Germany, pp. 99-109.

Figure Legends

Supplementary Figure S1. Complete chromatograms of MBP hydrolysates with methanesulfonic acid. The noise peaks between 5 to 25 min arise from the salt produced by the NaOH added for the methanesulfonic acid neutralization before the chromatography.

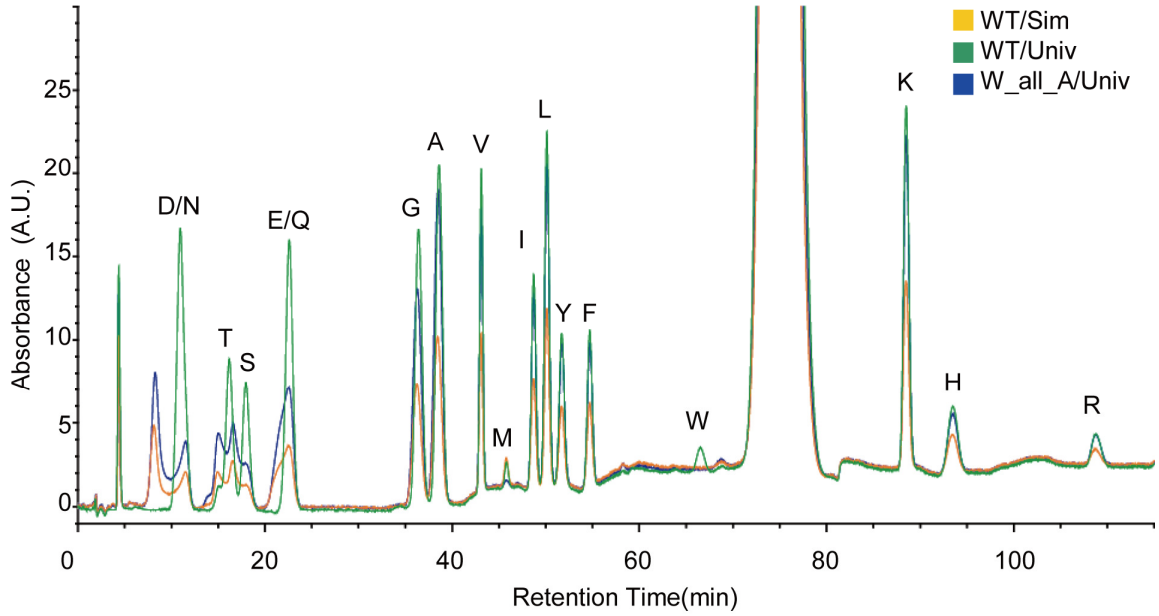
Supplementary Figure S2. Fluorescence spectra of MBPs translated by either the simplified genetic code or the universal code. The samples were excited at 280 nm in denaturing buffer. The apparent peak near 350nm was seen only in WT/Univ. The small peaks near 300 nm arise from the Tyr fluorescence in each protein.

Supplementary Figure S3. To confirm that the amount of proteins translated from the GFP mutant mRNAs are almost equal, the products were detected by western blotting, using an anti-His tag antibody. Lane 1, gfpΔUGG/Sim. Lane 2, gfpΔUGG/Univ. Lane 3, gfpΔUGG-A110X(UGG)/Sim. Lane 4, gfpΔUGG-A110X(UGG)/Univ. Lane 5, S1/Sim. Lane 6, S1/Univ.

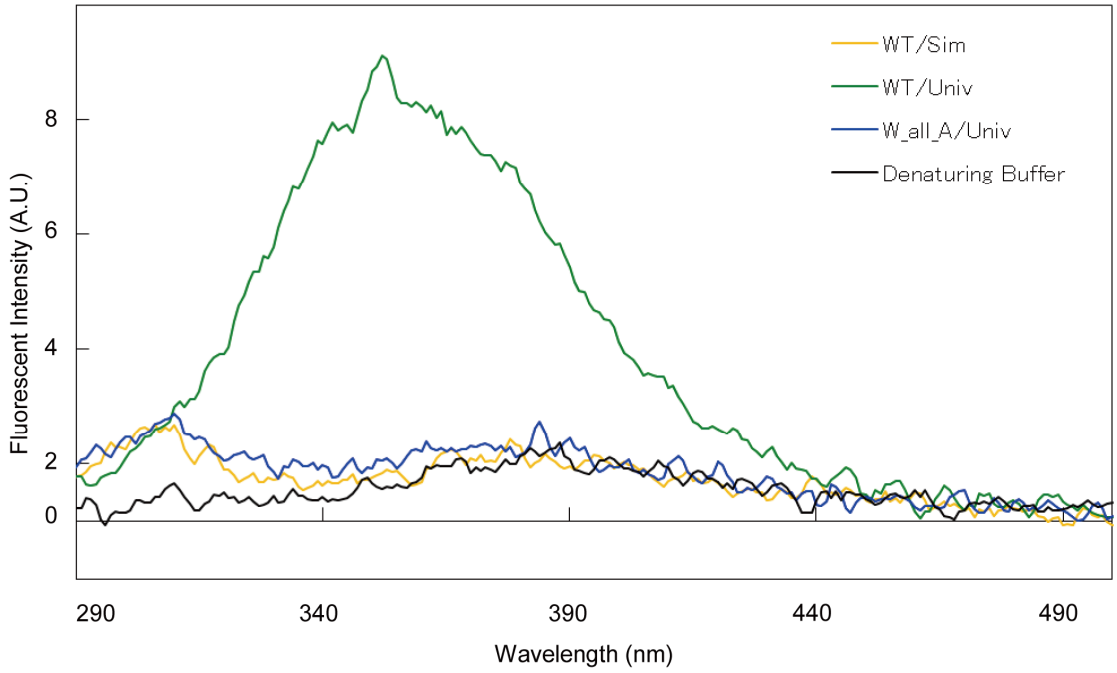
Supplementary Figure S4. Reassignment of UGU/UGC codons by the $tRNA^{Ser}$ variant. (A) The nucleotide sequences of the anticodon stem loop of the $tRNA^{Ser}$ variant and the UGC codon on the mRNA. The anticodon loop of $tRNA^{Ser}$ was substituted with that of $tRNA^{Cys}$. Positions 32 and 38 in the anticodon loop are numbered. (B) The CAT mRNA, which contains 4 UGU and 1 UGC codons, was translated under the conditions noted at the top of each lane. An autoradiogram of a polyacrylamide gel, with the products labeled with [^{14}C] Leu, is shown. (C) The difference in the

amino acid compositions between the proteins translated from the wild-type CAT mRNA by the simplified genetic code and from the same mRNA by the universal genetic code. WT/Sim: the protein synthesized by the simplified genetic code, using the wild-type CAT mRNA. C_all_S/Univ: the protein synthesized by the universal genetic code, using a mutant CAT mRNA in which all of the UGU/UGC codons were altered to UCU. WT/Univ: a protein synthesized by the universal genetic code, using the wild-type CAT mRNA. Before the hydrolysis step, the Cys residues were converted to carboxymethyl cysteine (CMC) to prevent decay. Refer to Table S3 for each value.

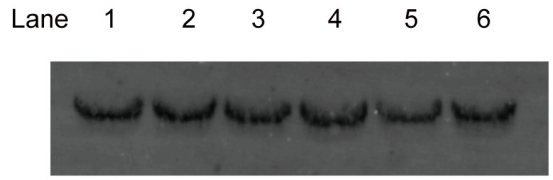
Supplementary Figure 1



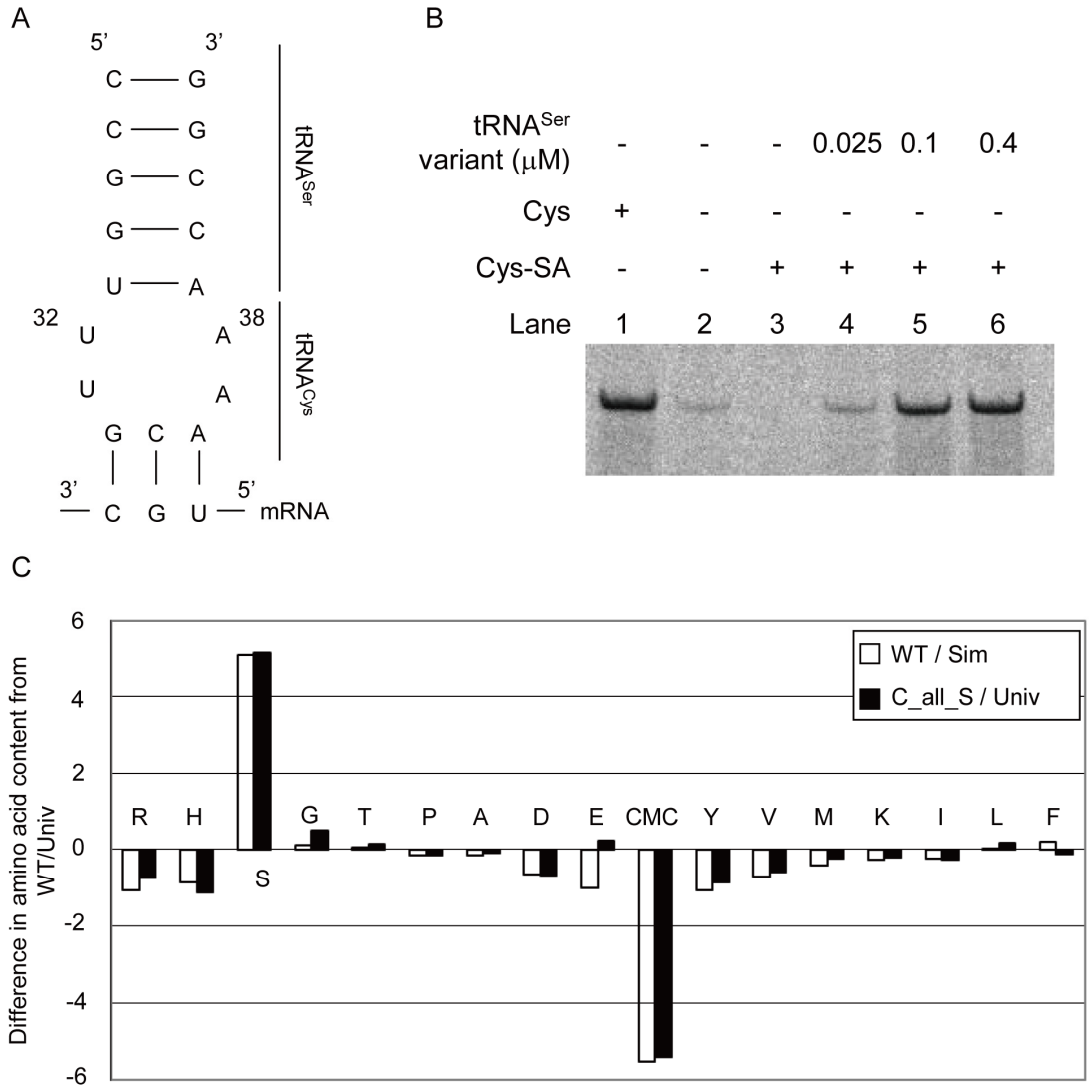
Supplementary Figure 2



Supplementary Figure 3



Supplementary Figure 4



Supplementary Tables

Supplementary Table. S1

Amino acid compositions measured after HCl hydrolysis of MBP

Amino acid content (residues / mol)

Residue	Theoretical	WT/Univ	W_all_A /Univ	WT/Sim
Ala (A)	45	45.7	53.3	53.8
Arg (R)	6	5.9	5.7	6.0
His (H)	13	9.8	11.2	9.4
Ser (S)	20	17.7	17.8	19.5
Gly (G)	35	34.5	34.2	35.7
Thr (T)	19	17.9	18.2	18.4
Pro (P)	23	19.3	19.5	20.4
Asp/Asn (D/N)	46	45.0	45.0	44.7
Glu/Gln (E/Q)	38	40.5	41.1	41.4
Tyr (Y)	16	13.8	14.1	13.5
Val (V)	21	20.0	20.0	19.9
Lys (K)	36	35.0	34.6	34.0
Ile (I)	22	19.7	19.9	19.9
Leu(L)	32	32.2	32.3	32.0
Phe (F)	16	15.9	15.9	16.0

Asp/Asn (D/N) is the sum of the Asp (D) and Asn (N) contents

Glu/Gln (E/Q) is the sum of the Glu (E) and Gln (Q) contents

Supplementary Table. S2

Data collection and refinement statistics

	gfpΔUGG-A110X(UGG)/Sim	gfpΔUGG/Univ
Data collection		
Beamline	SPring-8 BL41XU	SPring-8 BL41XU
Wavelength (Å)	1.000	1.000
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell parameters	a = 51.78 b = 63.05 c = 67.40 $\alpha = \beta = \gamma = 90°$	a = 51.84 b = 62.96 c = 67.37 $\alpha = \beta = \gamma = 90°$
Resolution (Å)	50.00-2.10 (2.14-2.10)	50.00-1.85 (1.88-1.85)
Unique reflections	13,473	19,367
Redundancy	9.5 (5.7)	10.5 (6.8)
Completeness (%)	99.6 (97.7)	99.6 (99.1)
$I/\sigma(I)$	30.6 (3.7)	51.8 (5.5)
R_{sym}	0.110 (0.387)	0.095 (0.479)
Refinement		
Resolution (Å)	33.70-2.09	46.00-1.85
R_{work}/R_{free}	18.8/24.0	18.8/23.8
No. of atoms		
Protein	1796	1807
Water	74	126
Average B-factor (Å^2)		
Protein	31.5	29.2
Water	44.1	42.9
R.m.s. deviation		
Bond length (Å)	0.022	0.022
Bond angle (°)	2.36	2.30
Ramachandran plot (%)		
Most favored	96.30	96.35
Allowed	3.70	3.65
Generously allowed	0.00	0.00
PDB ID code	3UG0	3UFZ

Values in parentheses are for the highest resolution shell

Supplementary Table. S3

Amino acid compositions measured after HCl hydrolysis of CAT

Residue	Amino acid content (residues / mol)				Difference in amino acid content from WT/Univ (residues / mol)	
	Theoretical	WT/Univ	C_all_S/Univ	WT/Sim	(C_all_S/Univ) -(WT/Univ)	(WT/Sim) -(WT/Univ)
CMC	5	6.0	0.6	0.5	-5.4	-5.5
Ser (S)	15	12.3	17.5	17.4	5.1	5.1
Arg (R)	6	6.6	5.8	5.5	-0.8	-1.1
His (H)	19	18.8	17.7	18.0	-1.1	-0.9
Gly (G)	13	12.2	12.7	12.3	0.5	0.1
Thr (T)	13	12.8	13.0	12.9	0.1	0.1
Pro (P)	8	7.9	7.8	7.8	-0.1	-0.1
Ala (A)	15	15.3	15.2	15.2	-0.1	-0.1
Asp/Asn (D/N)	22	22.3	21.6	21.6	-0.7	-0.7
Glu/Gln (E/Q)	25	26.7	26.9	25.7	0.2	-1.0
Tyr (Y)	11	10.7	9.9	9.7	-0.9	-1.0
Val (V)	17	16.9	16.3	16.2	-0.6	-0.7
Lys (K)	12	11.8	11.6	11.5	-0.2	-0.3
Ile (I)	9	8.5	8.2	8.3	-0.3	-0.2
Leu (L)	14	14.1	14.2	14.1	0.2	0.0
Phe (F)	20	19.5	19.4	19.7	-0.1	0.2

Asp/Asn (D/N) is the sum of the Asp (D) and Asn (N) contents

Glu/Gln (E/Q) is the sum of the Glu (E) and Gln (Q) contents