





## Reconstituted cell-free protein synthesis using in vitro transcribed tRNAs

Keita Hibi<sup>1</sup>, Kazuaki Amikura <sup>1,2</sup>, Naoki Sugiura <sup>1</sup>, Keiko Masuda<sup>3</sup>, Satoshi Ohno<sup>4</sup>, Takashi Yokogawa <sup>4</sup>, Takuya Ueda<sup>1,5</sup> & Yoshihiro Shimizu <sup>3✉</sup>

Entire reconstitution of tRNAs for active protein production in a cell-free system brings flexibility into the genetic code engineering. It can also contribute to the field of cell-free synthetic biology, which aims to construct self-replicable artificial cells. Herein, we developed a system equipped only with in vitro transcribed tRNA (iVTtRNA) based on a reconstituted cell-free protein synthesis (PURE) system. The developed system, consisting of 21 iVTtRNAs without nucleotide modifications, is able to synthesize active proteins according to the redesigned genetic code. Manipulation of iVTtRNA composition in the system enabled genetic code rewriting. Introduction of modified nucleotides into specific iVTtRNAs demonstrated to be effective for both protein yield and decoding fidelity, where the production yield of DHFR reached about 40% of the reaction with native tRNA at 30°C. The developed system will prove useful for studying decoding processes, and may be employed in genetic code and protein engineering applications.

<sup>1</sup>Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba 277-8562, Japan. <sup>2</sup>Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA. <sup>3</sup>Laboratory for Cell-Free Protein Synthesis, RIKEN Center for Biosystems Dynamics Research (BDR), Suita, Osaka 565-0874, Japan. <sup>4</sup>Department of Chemistry and Biomolecular Science, Faculty of Engineering, Gifu University, Gifu 501-1193, Japan. <sup>5</sup>Department of Integrative Bioscience and Biomedical Engineering, Graduate School of Science and Engineering, Waseda University, Tokyo, Shinjuku 162-8480, Japan. ✉email: [yshimizu@riken.jp](mailto:yshimizu@riken.jp)

Transfer RNAs (tRNAs) are key molecules that carry amino acids to ribosomes and decode the genetic code using anticodons to translate genetic information into the primary sequence of a polypeptide chain. Since the first pivotal study<sup>1</sup>, tRNAs have been engineered to alter connections between the genetic code and amino acids and reassign the genetic code for site-specific incorporation of noncanonical amino acids into proteins in a cell-free protein synthesis system<sup>2,3</sup>. Along with the development of aminoacyl-tRNA synthetase (aaRS) engineering and ribozyme utilization via flexizymes<sup>3</sup>, the creation of novel functional peptides or proteins with noncanonical amino acids has been extensively examined. Such genetic code engineering is expected to be highly beneficial for drug development research<sup>4</sup>.

Despite considerable effort expended on genetic code engineering, complete reconstitution of tRNAs in a cell-free system, which could prove game-changing for codon reassignment, has not been fully realized to our knowledge. The major obstacle to achieving this goal is the presence of a variety of modified nucleotides in tRNA structures. In contrast with many non-essential tRNA modifications outside anticodons<sup>5</sup>, those in the anticodon loop region directly interact with the mRNA codon on the ribosome, and become essential in some cases. Among the most widely-studied examples are modifications in tRNA<sup>Lys</sup><sub>UUU</sub>. The anticodon loop of tRNA<sup>Lys</sup><sub>UUU</sub> is modified at position 34 (5-methylaminomethyl-2-thiouridine; mnm<sup>5s2</sup>U34) and 37 (N6-threonylcarbamoyladenine; t<sup>6</sup>A37). These modifications are essential in decoding processes by both biochemical and structural analyses<sup>6,7</sup>. Nevertheless, several researchers have tried to entirely reconstitute the tRNA in a cell-free protein expression system utilizing a set of in vitro transcribed tRNAs (iVTtRNAs). Synthesizing active proteins using a set of three native tRNAs and 48 iVTtRNAs<sup>8</sup> or redesigning the genetic code using a set of 32 iVTtRNAs to incorporate noncanonical amino acids into synthesized polypeptides<sup>9</sup> have been studied so far. Chemically synthesized 21 tRNAs have been used for the in vitro protein synthesis experiment though detailed analysis of individual effect of each added tRNA was not examined<sup>10</sup>.

Notably, such a research direction can also contribute to the field of cell-free synthetic biology, which aims to construct artificial cells based on encapsulation of cell-free systems in a synthetic compartment such as phospholipid vesicles or microfluidic chips<sup>11,12</sup>. One of the major goals of this field is the construction of a self-replicable artificial cell by organizing essential purified biological macromolecules<sup>13,14</sup>. Such bottom-up approaches can be used for the construction of minimal cells, whereas top-down approaches utilize natural living cells possessing entirely synthesized genomes<sup>15,16</sup>. This may appear to be a formidable task requiring the identification, preparation, and integration of essential molecules for reconstituting self-replication. Nevertheless, it could be useful to explore the boundary between living and nonliving systems because self-replication is one of the most fundamental features of life.

The PURE system is a reconstituted cell-free protein synthesis system composed of individually prepared components required for gene expression in *Escherichia coli*<sup>17</sup>. This system is suitable for accomplishing self-replicable artificial cells via a bottom-up approach<sup>18</sup>. As mentioned in a previous review<sup>13</sup>, exploring a set of macromolecules that can create a new automaton composed of the same molecules, the basis of von Neumann's theory of self-replication, is crucial for the construction of self-reproducing systems. Thus, in the present study, we focused on the reconstitution of transfer RNA (tRNA) in a controllable manner to explore a minimal set of tRNAs that are functional for protein expression within the PURE system, and that can be synthesized by the PURE system.

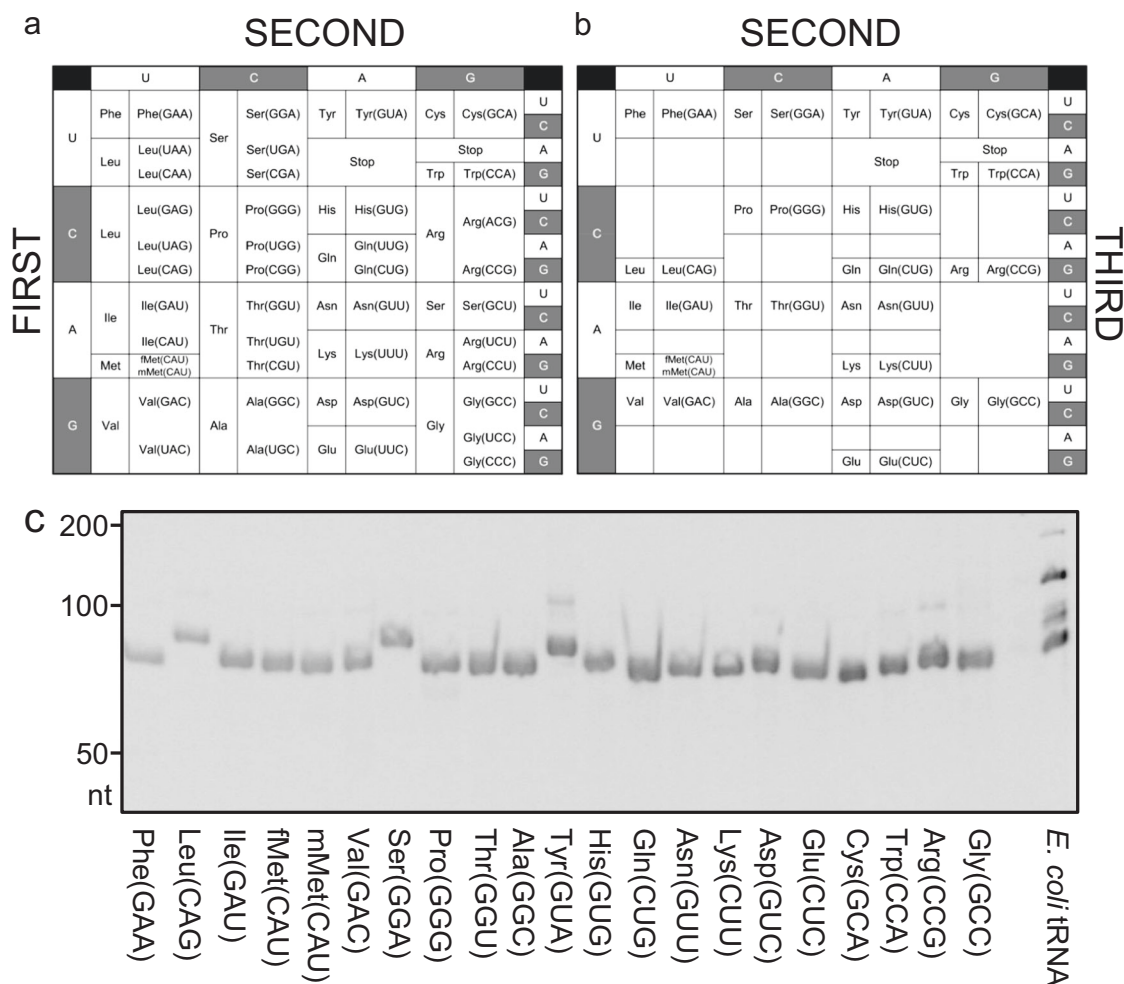
Herein, we present a cell-free protein synthesis system, based on the PURE system, equipped only with iVTtRNAs. The system, consisting of a set of 21 iVTtRNAs without any modifications, is able to synthesize active proteins according to the redesigned simplified genetic code. We succeeded in producing active dihydrofolate reductase (DHFR) and super-folder green fluorescent protein (sfGFP) variants with comparable activity to proteins synthesized using native tRNA mixtures. We also successfully rewrote the genetic code by assigning Ala to the Ser codon (UCG) through engineering tRNA<sup>Ala</sup><sub>GCC</sub>. The developed system displayed orthogonality in terms of decoding and produced active DHFR according to the redesigned genetic code, whereas inactive polypeptides were synthesized using a native tRNA mixture based on the universal genetic code. Introduction of modified nucleotides into specific iVTtRNAs further demonstrated to be effective for both protein yield and decoding fidelity.

## Results

**Selection and preparation of iVTtRNAs.** *E. coli* tRNAs contain a number of modified nucleotides, among which modified nucleotides 34 and 37 are essential for decoding specific codons<sup>19</sup>. In addition, in *E. coli*, some genes responsible for the modification of these nucleotides are essential genes<sup>20,21</sup>. Therefore, reconstitution of protein synthesis using a full set of iVTtRNAs may not be practical. However, codons are degenerate, and most amino acids are encoded by multiple codons. Therefore, by using a restricted set of tRNAs that may not require these modified nucleotides, reconstitution of tRNAs in the PURE system may be possible.

Nucleotide 34 is directly involved in base pairing at the wobble position, and can be modified in various ways (Fig. 1a and Supplementary Fig. 1). In particular, when the nucleotide U is present, it tends to be modified. Therefore, we preferentially selected tRNAs with a G or C at position 34, similar to a previous study<sup>9</sup>. However, Lys and Glu codons are decoded only by tRNA<sup>Lys</sup><sub>UUU</sub> and tRNA<sup>Glu</sup><sub>UUC</sub>, respectively, in *E. coli*, in which nucleotide 34 is U and modified to mnm<sup>5s2</sup>U. Because these modifications are reportedly required for efficient aminoacylation<sup>22,23</sup>, and are also essential for codon-anticodon interactions and subsequent translocation on the ribosome<sup>6,7</sup>, it is predictable that the corresponding iVTtRNA may not be functional in the translation machinery. Therefore, we artificially prepared tRNA<sup>Lys</sup><sub>CUU</sub> and tRNA<sup>Glu</sup><sub>CUC</sub> in an attempt to improve codon-anticodon interactions on the ribosome, and the aminoacylation efficiency, as reported in a previous study<sup>24</sup>. When the anticodon sequence (positions 34–36) of tRNA is GUN (tRNA<sup>Tyr</sup><sub>GUA</sub>, tRNA<sup>His</sup><sub>GUG</sub>, tRNA<sup>Asn</sup><sub>GUU</sub>, and tRNA<sup>Asp</sup><sub>GUC</sub>), where N is any nucleotide, G at position 34 is modified to queuosine (Q). In such cases, we decided to use an unmodified version. It was reported that these iVTtRNAs can be aminoacylated without the modification<sup>25–27</sup>. Consequently, we selected 21 tRNAs for the reconstitution of tRNAs with iVTtRNAs in the PURE system from *E. coli* genomic sequence (Fig. 1b, colored yellow and blue in Supplementary Fig. 1, and Supplementary Data 1). Except for Met, in which both initiator and elongator tRNAs were selected, a single tRNA was selected for the decoding of each amino acid.

Preparation of iVTtRNAs was performed by separate runoff transcription reactions with T7 RNA polymerase. To prevent addition of an extra nucleotide to the 3'-terminus of tRNA, 2'-methoxy modification was introduced at the second nucleotide from the 5'-terminus of the antisense strand in PCR-amplified DNA templates using modified DNA primers<sup>28</sup>. We also followed a previous protocol using RNase P for the efficient transcription<sup>29</sup>. Precursor tRNA with 27 extra nucleotides universally



**Fig. 1 Selection and preparation of iVTtRNAs for tRNA reconstitution.** Genetic code table showing 41 tRNA species in *E. coli* (a) and the table showing 21 tRNAs used in this study (b) are shown. c PAGE analysis of the prepared 21 iVTtRNAs. Purified iVTtRNAs (0.01  $A_{260}$  units) and native tRNA mixtures (0.04  $A_{260}$  units) were analyzed by urea-PAGE.

introduced at the 5'-terminus was transcribed and processed using RNase P to generate mature tRNA, followed by purification with anion-exchange chromatography. All iVTtRNAs were successfully prepared and found to be uniform (Fig. 1c).

**Aminoacylation of iVTtRNAs.** The ability of the prepared iVTtRNAs to accept amino acids was investigated by aminoacylation assays using each aaRS and radioisotope-labeled amino acids. All possible combinations of iVTtRNAs and aaRSs were tested to investigate orthogonality between aaRSs and iVTtRNAs (Supplementary Fig. 2). The measurements essentially demonstrated orthogonality, even in the absence of any modifications (i.e., only the amino acid corresponding to each iVTtRNA was specifically aminoacylated). The ratios of amino acid acceptance for each tRNA were also sufficient to support translation (15–50%), except for iVTtRNA<sup>Ile</sup><sub>GAU</sub> (4.2%) and iVTtRNA<sup>Glu</sup><sub>CUC</sub> (0.5%). The acceptance ratio for iVTtRNA<sup>Glu</sup><sub>CUC</sub> was particularly low, and this made it difficult to confirm orthogonality (Supplementary Fig. 2p).

These observations are consistent with previous studies in which threonylcarbamoyladenosine ( $t^6A$ ) at position 37 in tRNA<sup>Ile</sup><sub>GAU</sub> and  $mnm^5s^2U$  at position 34 in tRNA<sup>Glu</sup><sub>UUC</sub> were shown to be important for aminoacylation efficiency<sup>30,31</sup>. Therefore, we tried to compensate for the lack of modification

in these tRNAs by increasing the aaRS concentration. By increasing the isoleucyl-tRNA synthetase and glutamyl-tRNA synthetase concentration 30-fold (from 50 nM to 1.5  $\mu$ M), the amino acid acceptance ratios were increased to 15% (isoleucylation) and 13% (glutamylation), with sufficient orthogonality (Supplementary Fig. 3). The results indicate that all of the prepared iVTtRNAs could be sufficiently aminoacylated by each aaRS and were therefore suitable for construction of an iVTtRNA-based cell-free protein synthesis system.

**Decoding of iVTtRNAs.** We next investigated which codons could be decoded by the prepared iVTtRNAs in the PURE system. Hypothetically, decoding fidelity is maintained by canonical Watson-Crick base pairing, while the third base pair in the codon-anticodon interaction is less stringent and can be “wobble.” Although there are some exceptions such as the stop codon (UAG and UAA) decoded by Gln-tRNA<sup>Gln</sup><sub>CUG</sub><sup>32</sup> and the GGC codon (Gly) decoded by Ser-tRNA<sup>Ser</sup><sub>GCU</sub><sup>33</sup>, the efficiency of such unusual decoding is very low, hence we investigated all four codons in which the first and second base pairs engage in canonical Watson-Crick base pairing for each iVTtRNA.

Assays were performed by using template DNA including test codons that encode octapeptide (Supplementary Data 2), and peptide synthesis in the PURE system containing reconstituted

iVTtRNAs was evaluated by measuring radioisotope-labeled amino acid incorporation into the trichloroacetic acid (TCA)-insoluble fraction. Hydrophobic amino acids such as Leu, Phe, Ile, and Val were used as components of the synthesized peptides to ensure the insolubility. At first, the peptide sequence MFFLFXLF, where X is the test amino acid to be evaluated, was used for evaluation of iVTtRNA<sup>Ile</sup><sub>GAU</sub>, iVTtRNA<sup>Val</sup><sub>GAC</sub>, iVTtRNA<sup>mMet</sup><sub>CAU</sub>, and iVTtRNA<sup>fMet</sup><sub>CAU</sub>. This sequence was derived from our previous manuscript<sup>34</sup> where decoding abilities of iVTtRNA<sup>Phe</sup><sub>GAA</sub> and iVTtRNA<sup>Leu</sup><sub>CAG</sub> have already been confirmed. Next, we switched the sequence from MFFLFXLF to MIIIIXLF to increase the hydrophobicity of synthesized peptides because testing of some hydrophilic amino acids were found to be difficult by using the sequence MFFLFXLF. After examining other iVTtRNAs except for iVTtRNA<sup>Phe</sup><sub>GAA</sub> and iVTtRNA<sup>Leu</sup><sub>CAG</sub>, the sequence MIIIIXVV was used to evaluate iVTtRNA<sup>Phe</sup><sub>GAA</sub> and iVTtRNA<sup>Leu</sup><sub>CAG</sub>. To investigate the decoding ability of each iVTtRNA in a highly sensitive manner, each iVTtRNA was added in excess relative to the ribosome (6 μM of each tRNA, 0.2 μM ribosome).

The results were both similar and different to those of previous studies<sup>8,9</sup> (Supplementary Figs. 4–8). Each iVTtRNA with an anticodon starting with the nucleotide G (iVTtRNA<sup>Phe</sup><sub>GAA</sub>, iVTtRNA<sup>Ile</sup><sub>GAU</sub>, iVTtRNA<sup>Val</sup><sub>GAC</sub>, iVTtRNA<sup>Ser</sup><sub>GGA</sub>, iVTtRNA<sup>Pro</sup><sub>GGG</sub>, iVTtRNA<sup>Thr</sup><sub>GGU</sub>, iVTtRNA<sup>Ala</sup><sub>GGC</sub>, iVTtRNA<sup>Tyr</sup><sub>Tyr</sub>, iVTtRNA<sup>His</sup><sub>GUG</sub>, iVTtRNA<sup>Asn</sup><sub>GUU</sub>, iVTtRNA<sup>Asp</sup><sub>GUC</sub>, iVTtRNA<sup>Cys</sup><sub>GCA</sub>, and iVTtRNA<sup>Gly</sup><sub>GCC</sub>) effectively decoded target codons ending with the nucleotides U or C. Presumably due to the high concentration of iVTtRNA, iVTtRNA<sup>Ile</sup><sub>GAU</sub>, and iVTtRNA<sup>Asn</sup><sub>GUU</sub>, which have been demonstrated not to decode target codons in previous studies, were shown to be functional. In addition, iVTtRNA<sup>Tyr</sup><sub>GUA</sub>, iVTtRNA<sup>His</sup><sub>GUG</sub>, iVTtRNA<sup>Asn</sup><sub>GUU</sub>, and iVTtRNA<sup>Asp</sup><sub>GUC</sub>, that have Q at 34 position in their native state (Q series iVTtRNAs), decoded both target codons, while they tended to decode codons ending in C better than those ending in U.

All iVTtRNAs for which their anticodons started with the nucleotide C (iVTtRNA<sup>Leu</sup><sub>CAG</sub>, iVTtRNA<sup>mMet</sup><sub>CAU</sub> and/or iVTtRNA<sup>fMet</sup><sub>CAU</sub>, iVTtRNA<sup>Gln</sup><sub>CUG</sub>, iVTtRNA<sup>Lys</sup><sub>CUU</sub>, iVTtRNA<sup>Glu</sup><sub>CUC</sub>, iVTtRNA<sup>Trp</sup><sub>CCA</sub>, and iVTtRNA<sup>Arg</sup><sub>CCG</sub>) effectively decoded target codons ending with the nucleotide G. These iVTtRNAs were divided into two types; iVTtRNA<sup>Gln</sup><sub>CUG</sub>, iVTtRNA<sup>Lys</sup><sub>CUU</sub>, iVTtRNA<sup>Trp</sup><sub>CCA</sub>, and iVTtRNA<sup>Arg</sup><sub>CCG</sub> that only decoded codons ending with G, and iVTtRNA<sup>Leu</sup><sub>CAG</sub> and iVTtRNA<sup>Glu</sup><sub>CUC</sub> that also decoded codons ending in A. The former group is consistent with a previous study<sup>8</sup>, and the latter might reflect the addition of excess iVTtRNAs in the cell-free expression system. We note that the buffer composition was based on potassium glutamate, which is known to be an efficient ingredient for the cell-free protein synthesis reaction<sup>35</sup>, and glutamic acid was present in the system in submolar quantities. This may cause efficient decoding with unmodified iVTtRNA<sup>Glu</sup><sub>CUC</sub> in addition to excess addition of iVTtRNA<sup>Glu</sup><sub>CUC</sub>. Further information on the detailed differences between decoding properties among iVTtRNAs and inconsistencies with the genetic code table can be found in Supplementary Note 1, but the results presented here indicate the possibility of synthesizing proteins using a set of iVTtRNAs without any native tRNA.

**Cell-free protein synthesis using iVTtRNAs.** To evaluate the activity of iVTtRNAs for the synthesis of long polypeptides (i.e., protein synthesis), DHFR and sfGFP synthesis were examined using the PURE system containing iVTtRNAs. DNA templates for protein expression (Supplementary Data 3) were prepared according to the genetic code table where each amino acid corresponds to only a single codon (Fig. 2a). To decrease the GC content in order to avoid internal secondary structure formation

within mRNAs, the use of wobble base pairings were prioritized for iVTtRNAs whose anticodons start with the nucleotide G, while they were not used for Q series iVTtRNAs that may affect the decoding efficiency.

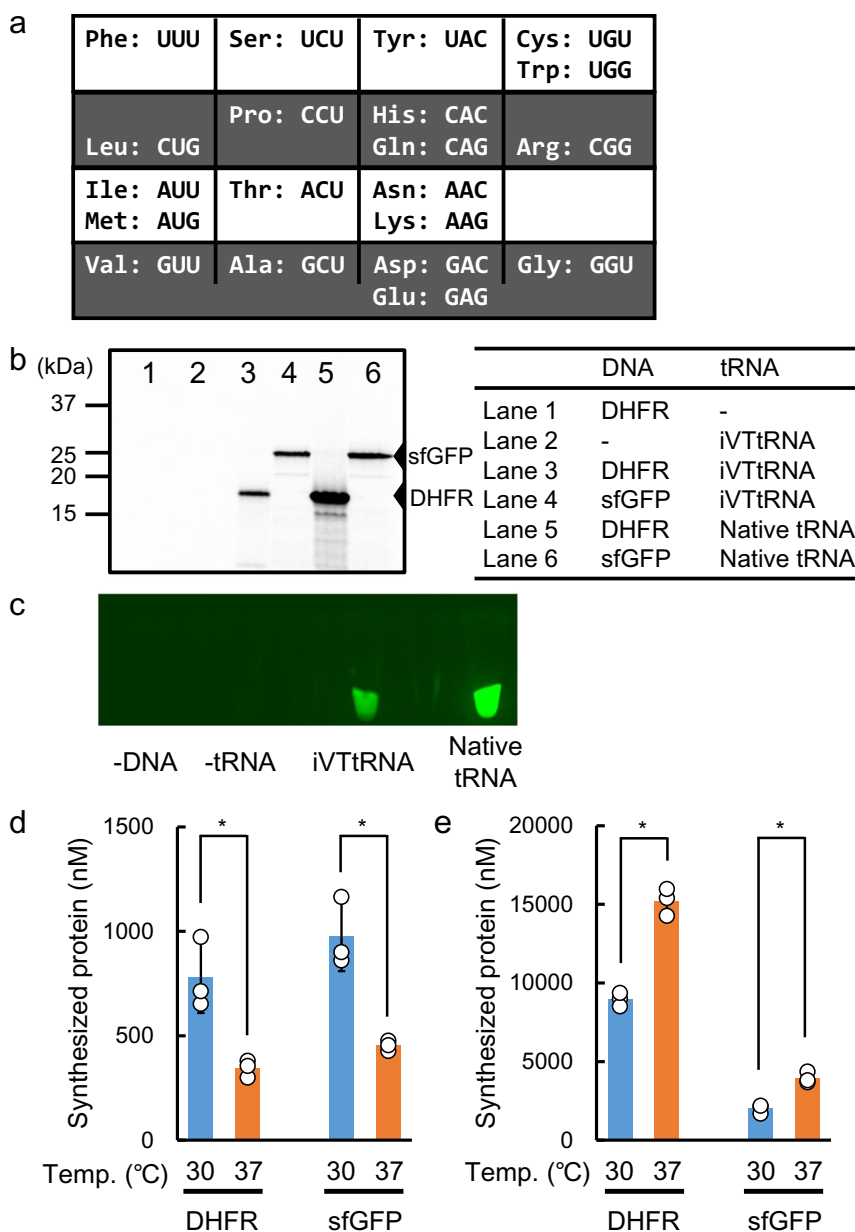
Despite lacking any modifications, iVTtRNAs were shown to support the synthesis of both proteins, and products synthesized using iVTtRNAs had the same molecular weight as those synthesized using native tRNA mixtures (Fig. 2b, c). Two temperatures were tested, and the yields of proteins synthesized using iVTtRNAs were higher when incubated at 30 °C than 37 °C, whereas yields with native tRNA mixtures were higher at 37 °C than at 30 °C (Fig. 2d, e, and Supplementary Fig. 9). We note that sfGFP synthesis was also successful without using any wobble base pairings (Supplementary Fig. 10 and Supplementary Data 3).

The relationship between the productivity of DHFR synthesis and tRNA concentration was monitored, and it increased linearly with increasing iVTtRNAs added, while it was saturated when 60 A<sub>260</sub> unit/mL native tRNA mixtures were added (Supplementary Fig. 11). These observations suggest the possibility that some reactions involving inefficient iVTtRNAs, such as aminoacylation and decoding, may become rate-limiting reactions, and do not reach a plateau when iVTtRNAs are used. Therefore, we fixed the iVTtRNA concentration at 20 A<sub>260</sub> unit/mL and excessively added only four iVTtRNA species (iVTtRNA<sup>Ile</sup><sub>GAU</sub>, iVTtRNA<sup>Pro</sup><sub>GGG</sub>, iVTtRNA<sup>Asn</sup><sub>GUU</sub>, and iVTtRNA<sup>Glu</sup><sub>CUC</sub>), which appeared to be inefficient from aminoacylation experiments in this study (Supplementary Figs. 2 and 3) and previous in vitro tRNA reconstitution studies<sup>8,9</sup>. The results demonstrated that higher yields of DHFR were obtained when specific iVTtRNAs were added, compared with a mixture of all iVTtRNAs, suggesting that reactions involving these specific iVTtRNAs are rate-limiting (Fig. 3a).

The addition of specific iVTtRNAs was also shown to affect the decoding fidelity. The specific activity of synthesized DHFR was measured, and when the protein was synthesized using a mixture of equally abundant iVTtRNAs, the activity was almost half of the protein synthesized using native tRNA mixtures (Fig. 3b). By contrast, activity was recovered to levels comparable with the protein synthesized using native tRNAs when specific iVTtRNAs were added (Fig. 3b). A previous study showed that tRNAs without modifications in anticodon loops can mistakenly decode near-cognate codons<sup>36</sup>. Similarly, our results suggest that a decrease in decoding fidelity may be recovered by manipulating the tRNA composition, specifically by increasing the concentrations of inefficient tRNAs.

**Genetic code redesigning using iVTtRNAs.** We explored the possibility of redesigning the genetic code table using iVTtRNAs. It is known that alanyl-tRNA synthetase (AlaRS) specifically recognizes the G2:U71 wobble base pair in the acceptor stem region of tRNA<sup>Ala</sup>, and this base pair is the major identity element that allows AlaRS to discriminate tRNA<sup>Ala</sup> from other tRNAs<sup>37</sup>. By contrast, tRNA<sup>Ser</sup> has a long variable loop as its identity element for discrimination by seryl-tRNA synthetase<sup>27</sup>. Both aaRS do not recognize anticodon regions for the discrimination and thus, we tried to reassign Ala to the Ser codon by transplanting the CGA anticodon into tRNA<sup>Ala</sup><sub>GGC</sub> to generate tRNA<sup>Ala</sup> with a CGA anticodon (iVTtRNA<sup>Ala</sup><sub>CGA</sub>), corresponding to a Ser UCG codon (Supplementary Fig. 12). The scheme utilizing the identity element for AlaRS has been adopted for genetic code engineering studies<sup>38</sup>.

After preparing an appropriate DNA template (Supplementary Data 3) for DHFR based on the newly designed genetic code (Fig. 4a), in which all GCU codons used in experiments to generate the results shown in Fig. 2 were overwritten by the UCG

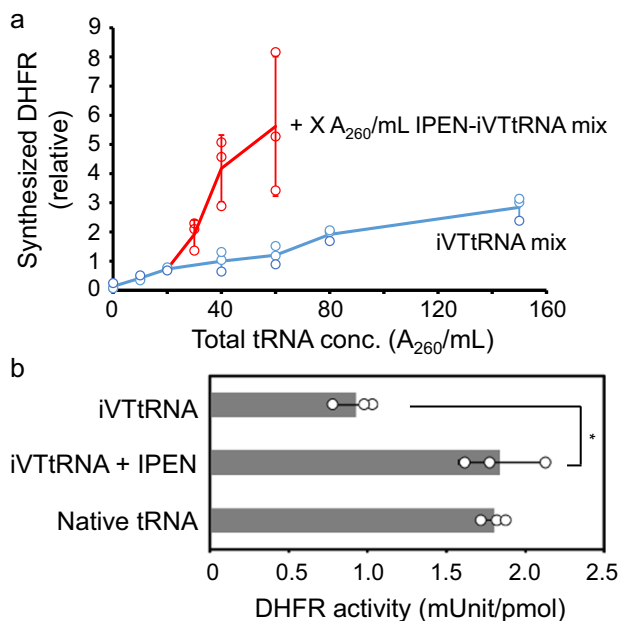


**Fig. 2 Protein synthesis with iVTtRNAs.** **a** Genetic code table used for protein synthesis with iVTtRNAs. DNA sequences for protein expression were designed according to this table. **b** SDS-PAGE analysis of synthesized proteins labeled with [ $^{35}$ S]Met. **c** Fluorescence images of synthesized sfGFP in reaction mixtures. Temperature dependence of the synthesized protein yield. Yields with iVTtRNAs (**d**) and yields with native tRNA mixtures (**e**) are shown. All experiments were performed with 60  $A_{260}$  unit/mL iVTtRNA mixtures or 40  $A_{260}$  unit/mL native tRNA mixtures. Error bars indicate standard deviation of independent repeats of triplicate measurements. Each dot represents individual observed value. An asterisks indicate that  $P$  values  $< 0.05$ . Welch's  $t$  test was applied between yields at 30 and 37 °C.

codon, DHFR synthesis was performed with iVTtRNAs complemented with iVTtRNA<sup>Ala<sub>CGA</sub></sup> or native tRNA mixtures. Of note, excessive addition of the four specific iVTtRNAs used in experiments that generated the results shown in Fig. 3 were also performed in this experiment. Amino acid sequences of the synthesized DHFR were hypothetically different between two conditions; the correct amino acid sequence should only be synthesized using the new iVTtRNA mixtures, whereas all Ala residues should be changed to Ser when native tRNA mixtures are used, according to the canonical genetic code. SDS-PAGE analysis of the synthesized DHFR revealed a slight band shift, indicating that different polypeptides were produced in the two conditions (Fig. 4b). Liquid chromatography-mass spectrometry (LC-MS)

analysis confirmed the successful codon reassignment (Supplementary Fig. 13 and Supplementary Data 4). Specific activity measurements of synthesized DHFR clearly demonstrated that only DHFR synthesized using iVTtRNAs possessed activity comparable with DHFR harboring the correct amino acid sequence (compare Fig. 3b with Fig. 4c), whereas DHFR synthesized using native tRNA mixtures was inactive. These results indicate that Ala was not incorporated into DHFR when native tRNA mixtures were used, whereas the UCG codon was successfully reassigned to Ala when iVTtRNAs were employed.

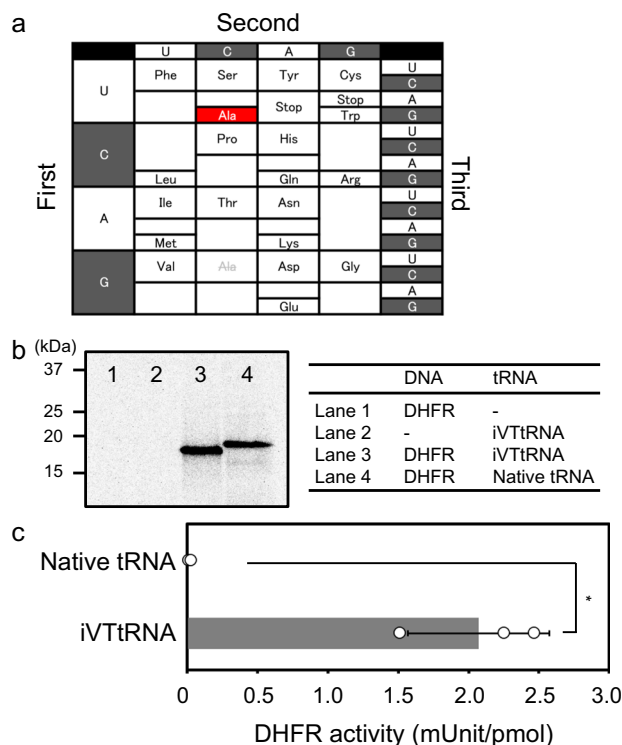
**Introduction of modified nucleotides into iVTtRNAs.** Excessive addition of four iVTtRNA species (iVTtRNA<sup>Ile<sub>GAU</sub></sup>,



**Fig. 3** Effects of addition of four specific iVTtRNAs. **a** Dependency of tRNA concentration on the yield of synthesized DHFR. The red line shows the results of added all iVTtRNAs, while the blue line shows the results of adding 20 A<sub>260</sub> unit/mL iVTtRNAs with a specified amount of the IPEN-iVTtRNA mixture composed of four specific iVTtRNAs (iVTtRNA<sup>Ile</sup><sub>GAU</sub>, iVTtRNA<sup>Pro</sup><sub>GGG</sub>, iVTtRNA<sup>Glu</sup><sub>CUC</sub>, and iVTtRNA<sup>Asn</sup><sub>GUU</sub>). Experiments were performed with 0, 10, 20, and 40 A<sub>260</sub> unit/mL IPEN-iVTtRNA mix. **b** Specific activity of synthesized DHFR. Activities are shown for synthesized DHFR prepared with 60 A<sub>260</sub> unit/mL iVTtRNA mixtures, 20 A<sub>260</sub> unit/mL iVTtRNA mixtures, and 40 A<sub>260</sub> unit/mL native tRNA mixtures. Error bars indicate standard deviation of independent repeats of triplicate measurements. Each dot represents individual observed value. An asterisk indicates that *P* value < 0.05. Welch's *t* test was applied between DHFR activities synthesized with iVTtRNA mixtures and iVTtRNA and IPEN-iVTtRNA mixtures.

iVTtRNA<sup>Pro</sup><sub>GGG</sub>, iVTtRNA<sup>Asn</sup><sub>GUU</sub>, and iVTtRNA<sup>Glu</sup><sub>CUC</sub> were shown to be effective for both yields and fidelity (Fig. 3). Therefore, we addressed this matter by introducing modification on anticodon loop regions of these iVTtRNA. t<sup>6</sup>A modification at position 37 of iVTtRNA<sup>Ile</sup><sub>GAU</sub> and iVTtRNA<sup>Asn</sup><sub>GUU</sub>, 1-methylguanosine (m<sup>1</sup>G) modification at position 37 of iVTtRNA<sup>Pro</sup><sub>GGG</sub>, and 5-methylaminomethyluridine (mnm<sup>5</sup>U) modification at position 34 of iVTtRNA<sup>Glu</sup><sub>UUC</sub> were successfully introduced using corresponding modification enzymes (Supplementary Fig. 14). We note that tRNA<sup>Glu</sup><sub>UUC</sub> that has a native anticodon sequence was used in this experiment instead of tRNA<sup>Glu</sup><sub>CUC</sub>.

Aminoacylation experiments showed that the modification on iVTtRNA<sup>Ile</sup><sub>GAU</sub>, iVTtRNA<sup>Pro</sup><sub>GGG</sub>, and iVTtRNA<sup>Glu</sup><sub>UUC</sub> were effective for increasing the aminoacylation efficiency whereas that for iVTtRNA<sup>Asn</sup><sub>GUU</sub> did not show improvement (Supplementary Fig. 15). Particularly, t<sup>6</sup>A37 modification on iVTtRNA<sup>Ile</sup><sub>GAU</sub> showed about sevenfold increase in the efficiency while mnm<sup>5</sup>U34 modification on iVTtRNA<sup>Glu</sup><sub>UUC</sub> and m<sup>1</sup>G37 modification on iVTtRNA<sup>Pro</sup><sub>GGG</sub> showed about 2- and 1.4-fold increase, respectively. We next examined the DHFR synthesis using modified or unmodified iVTtRNA<sup>Ile</sup><sub>GAU</sub>, iVTtRNA<sup>Pro</sup><sub>GGG</sub>, and iVTtRNA<sup>Glu</sup><sub>UUC</sub> with all possible combinations. The result demonstrated the addition of modified iVTtRNA<sup>Ile</sup><sub>GAU</sub> increased the yield and the addition of modified iVTtRNA<sup>Glu</sup><sub>UUC</sub> further increased the

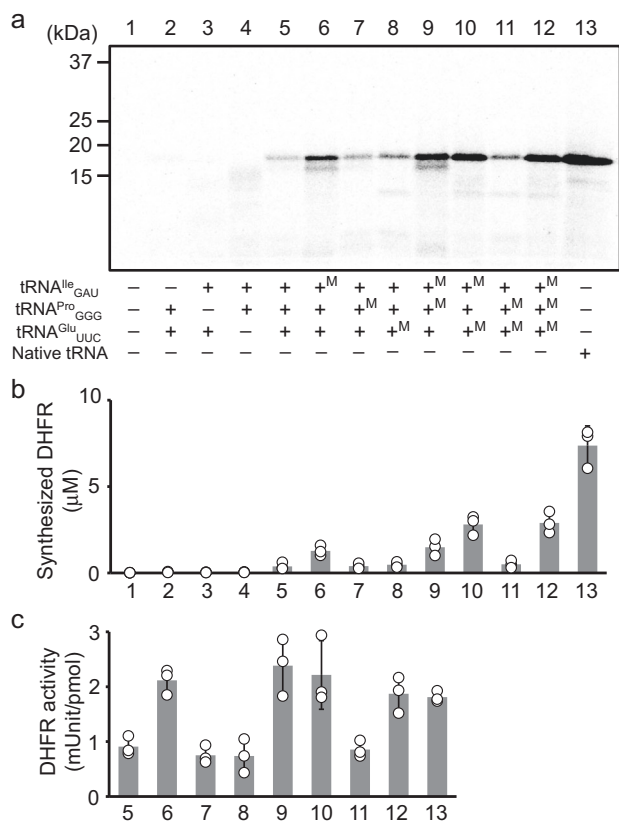


**Fig. 4** Genetic code redesigning with iVTtRNAs. **a** Redesigned genetic code table using newly prepared iVTtRNA<sup>Ala</sup><sub>CGA</sub> but without iVTtRNA<sup>Ala</sup><sub>GCC</sub>. The newly assigned UCG codon for Ala is highlighted in red, and original Ala codons not assigned to any amino acids in the redesigned genetic code table are shaded with gray. **b** SDS-PAGE analysis of synthesized protein labeled with [<sup>35</sup>S]Met. **c** Specific activity of synthesized DHFR. All experiments were performed with 40 A<sub>260</sub> unit/mL native tRNA mixtures or 20 A<sub>260</sub> unit/mL iVTtRNA mixtures and 40 A<sub>260</sub> unit/mL IPEN-iVTtRNA mixtures. Error bars indicate standard deviation of independent repeats of triplicate measurements. Each dot represents individual observed value. An asterisk indicates that *P* value < 0.05. Welch's *t* test was applied between DHFR activities synthesized with native tRNA mixtures and iVTtRNA mixtures.

yield where it reached about 40% compared with the native tRNA mixtures (Fig. 5a, b). Specific activities of synthesized DHFR showed that the addition of modified iVTtRNA<sup>Ile</sup><sub>GAU</sub> showed the level comparable with DHFR synthesized with native tRNAs, suggesting t<sup>6</sup>A modification on iVTtRNA<sup>Ile</sup><sub>GAU</sub> is crucial for the decoding fidelity (Fig. 5c).

## Discussion

Herein, we demonstrated that the PURE system containing iVTtRNAs without any modified nucleotides is capable of synthesizing active proteins. The set of 21 iVTtRNAs (Fig. 1 and Supplementary Fig. 1) used in this study can be regarded as hypothetically minimal, and the resulting protein expression system based on a simplified genetic code table in which only a single tRNA corresponds to a single amino acid, including formyl-methionine, proved effective (Fig. 2a). It is interesting to note that exploring better set of the tRNA sequences which can improve the translation efficiency, which should be performed with the developed system in the future. Natural genomic sequences cannot be used for the gene expression in this system and the gene synthesis technologies are required. We consider that such a feature does not become a limitation because of the



**Fig. 5** Effects of modifications on specific iVTtRNAs on protein synthesis yield and fidelity. **a** SDS-PAGE analysis of synthesized proteins labeled with [<sup>35</sup>S]Met. **b** Yields of synthesized DHFR. **c** Specific activity of synthesized DHFR. All experiments were performed with 40 A<sub>260</sub> unit/mL native tRNA mixtures or 60 A<sub>260</sub> unit/mL iVTtRNA mixtures. Error bars indicate standard deviation of independent repeats of triplicate measurements. Each dot represents individual observed value.

recent development of these technologies<sup>39</sup>, which enables an easy access to the custom-designed DNA templates. When the composition of the iVTtRNAs was optimized, the activity of the synthesized protein was comparable to that synthesized with native tRNA mixtures (Fig. 3b), suggesting this set of 21 iVTtRNAs is intrinsically capable of maintaining translational fidelity. This may imply that primordial features of tRNAs, which may have functioned without any modifications when the genetic code first evolved, are still preserved in these selected tRNAs.

Extant tRNAs have a variety of modified nucleotides, and they are added through complicated enzymatic networks. For example, to introduce 2-thiouridine, the intermediate of mnm<sup>5</sup>s<sup>2</sup>U, an enzymatic network for sulfur transfer composed of seven proteins is necessary<sup>40</sup>. However, upon the appearance of the first translation apparatus, tRNAs may have performed their functions without any modifications, which require such complicated enzymatic networks. Thus, the present results might reflect the translation system during the early stages of life, and may provide clues to the characteristics of the primitive translation system and how the translation apparatus evolved. For example, it would be interesting to examine prebiotic metals<sup>41</sup> in the iVTtRNA-based system. Elucidating how such metals affect tRNA function and the behavior of the overall translation machinery might be insightful.

In addition to the modifications present in the anticodon loop of tRNAs, which are directly involved in codon-anticodon interactions in the decoding process, many modified nucleotides are added to tRNAs, mainly for stabilization of the tertiary structure<sup>42</sup>. Therefore, a lack of all modification in iVTtRNAs might cause overall structural instability in reaction mixtures, and a shift in the optimal temperature may be needed to maximize the yield of synthesized proteins (Fig. 2d, e). As shown by our experiments using native tRNA mixtures, the *E. coli* translation apparatus is optimal at 37°C, but the optimal temperature was 30 °C when iVTtRNAs were employed, suggesting that a lack of modified nucleotides outside the anticodon loop may be one of the factors responsible for the low translation efficiency with iVTtRNAs, and this may reflect structural instability.

Our results further indicate the importance of modifications in the anticodon loop region for both translation efficiency and fidelity. The overall efficiency of protein synthesis was much lower when using iVTtRNAs than when using native tRNA mixtures (Supplementary Fig. 11), and this was not improved greatly by the addition of all iVTtRNAs, but it was improved by the addition of four specific iVTtRNAs (iVTtRNA<sup>Ile</sup><sub>GAU</sub>, iVTtRNA<sup>Pro</sup><sub>GGG</sub>, iVTtRNA<sup>Asn</sup><sub>GUU</sub>, and iVTtRNA<sup>Glu</sup><sub>CUC</sub>) (Fig. 3a). Surprisingly, addition of these iVTtRNAs also resulted in the recovery of translational fidelity, as shown by analysis of the specific activities of the synthesized proteins (Fig. 3b), suggesting that these four iVTtRNAs are more inefficient during aminoacylation, and presumably during decoding, than the other iVTtRNAs.

This was further validated by the introduction of modified nucleotides into these four iVTtRNAs. The introduction of t<sup>6</sup>A37 into iVTtRNA<sup>Ile</sup><sub>GAU</sub> was most effective for the aminoacylation, the protein yield, and the fidelity (Fig. 5). The introduction of mnm<sup>5</sup>U34 into iVTtRNA<sup>Glu</sup><sub>UUC</sub> was also shown to increase the aminoacylation and the protein yield (Fig. 5). The yield reached almost 40% of those with native tRNA mixtures (Fig. 5). Further attempts to modify the anticodon loop region, especially for those with U at 34 position may increase the number of iVTtRNAs and variety of codon sets that can be used for the reconstitution, which may facilitate the study of the effects on horizontal gene transfer and evolution<sup>10</sup>.

Also, the improvement of the yield may require universal modifications on TΨC-loop and D-loop of tRNA, which stabilizes the tertiary structure of tRNA<sup>42</sup>. In this way, development of in vitro reconstitution systems for tRNA modification processes and their integration with the iVTtRNA-based system is an important future perspective for improving the productivity and decoding accuracy, both of which are crucial for the reconstitution of self-replicable artificial cells using a bottom-up approach. For this goal, development of the tRNA transcription-coupled and/or tRNA modification-coupled protein synthesis system is also important perspective. This approach may require the in vitro synthesis of sufficient amount of the tRNAs for efficient protein synthesis, requiring high amount of transcription substrates. Therefore, integration of nonequilibrium system through the construction of artificial cells<sup>11,43</sup> might be effective for this direction.

We also demonstrated that the iVTtRNA-based system can easily redesign the genetic code table, by switching the Ala codon from GCN to the Ser UCG codon (Fig. 4). Although the redesigning of the genetic code table has already been studied, e.g., flexizyme studies for peptide production, the protein synthesis with comparable activity to those synthesized with native tRNA mixtures based on the canonical genetic code table (Figs. 3b and 4c), was demonstrated in this study (Fig. 4a). LC-MS analysis (Supplementary Fig. 13) clearly showed that the codon reassigning from

Ser to Ala was performed efficiently as intended. Thus, the iVTtRNA-based system could contribute to the field of protein engineering by facilitating genetic code expansion. Since the system sustains translational activity for long peptide synthesis, and because many codons lack their cognate tRNA, there are no potential competitor tRNAs when extra tRNAs are incorporated in the system. Combined with aaRS engineering or flexizyme technology<sup>2,3</sup>, as well as ribosome engineering<sup>44</sup>, introduction of multiple noncanonical amino acids into long polypeptides and proteins could be made possible by reassignment of sense codons in combination with amber codon suppression techniques. The system may also prove to be a useful platform for developing orthogonal aaRS and tRNA pairs for sense codon suppression via directed evolution.

It is also interesting to note that natural tRNA mixtures failed to synthesize active proteins when the redesigned genetic code table was applied (Fig. 4c). The products obtained using natural tRNA mixtures were slightly shifted in the SDS-PAGE analysis, suggesting that Ser was incorporated in Ala positions (Fig. 4b). This indicates that template DNA based on the redesigned genetic code table has dual coding sequences; active proteins can be expressed only using the iVTtRNA-based system, and products are inactive when the canonical genetic code is applied. This feature can be applied to the study of dangerous proteins, such as virulence factors from pathogenic cells or viruses. The release of strains carrying dangerous genes into the environment can be a risk through horizontal gene transfer to living cells with unknown mechanisms. However, if the genetic materials for such dangerous proteins are based on the redesigned genetic code table, they would only be expressed as inactive proteins based on the canonical genetic code, even if strains are released, thereby reducing the risk to the environment. Therefore, the present system could also serve as a platform for studying dangerous proteins without biosecurity issues<sup>45</sup>.

## Methods

**Preparation of protein components for reconstitution.** Components of the *E. coli* translation apparatus, including translation factors and aaRS, were prepared as previously described<sup>146</sup>. Preparation of RNaseP components, M1 RNA, and C5 protein were also prepared as described previously<sup>29</sup>. We modified the protocol for the C5 protein by precipitating it with 80% saturated ammonium sulfate, and dissolving it by dialyzing against buffer A (50 mM sodium acetate pH 6.5, 5 mM EDTA, 0.25 M NaCl, and 7 mM 2-mercaptoethanol). The resulting precipitate was recovered by centrifugation and dissolved by dialyzing against buffer B comprising 50 mM HEPES-KOH pH 7.6, 100 mM NH<sub>4</sub>Cl, 6 M urea, and 10 mM dithiothreitol (DTT). Dissolved proteins were applied onto a 5 mL HiTrap SP HP column (#17115101, GE Healthcare, USA), washed with buffer B containing 7 mM 2-mercaptoethanol as a substitute for 10 mM DTT, and eluted with a linear gradient from 100 mM to 2 M NH<sub>4</sub>Cl in buffer B. Fractions containing C5 protein were analyzed by SDS-PAGE, recovered, dialyzed against buffer D (50 mM HEPES-KOH pH 7.6, 0.8 M NH<sub>4</sub>Cl, 10 mM MgCl<sub>2</sub>, 2 M urea, 7 mM 2-mercaptoethanol), then further dialyzed against buffer D without urea. Resultant solutions were concentrated using an Amicon Ultra 3 kDa (#UFC800324, Merck Millipore, USA) and dialyzed against buffer D without urea containing 50% glycerol. The purified C5 protein was stored at -30 °C. We note that we can share all of the plasmids on requests.

**Preparation of modification enzymes for iVTtRNAs.** Modification enzymes for iVTtRNAs were prepared as follows. *E. coli* genes of TsaB, TsaC, TsaD, TsaE, TrmD, GlyA, MnmC, MnmE, and GidA were amplified from *E. coli* A19 genome using appropriate primers (Supplementary Data 5). Amplified genes for TsaC, TsaD, TsaE, GlyA, MnmC, MnmE, and GidA were cloned into pET15b (#69661, Merck Millipore, USA) as small ubiquitin-like modifier (SUMO) protein-fusion proteins where His-tag, SUMO protein, and modification enzymes were tandemly arranged. Genes for TsaB and TrmD were cloned into pET15b as His-tag fusion protein. All genes were cloned with Gibson assembly technique (#E2611, New England Biolabs, USA). Resultant plasmids were transformed into an *E. coli* BL21 (DE3) strain and grown in 1 L LB medium to an OD<sub>660</sub> of 0.6–1.0 at 37 °C. Overexpression was induced by the addition of IPTG to a final concentration of 1 mM for TsaB, TsaC, TsaD, TsaE, and TrmD, or 0.1 mM for GlyA, MnmC, MnmE, and GidA. After 3 h of cultivation at 37 °C, cells were harvested. TsaB, TsaC, TsaE, TrmD, and MnmE-overexpressed cells were resuspended in 40 mL of

Lysis buffer (50 mM Hepes-KOH, pH 7.6, 1 M NH<sub>4</sub>Cl, 10 mM MgCl<sub>2</sub>, and 7 mM 2-mercaptoethanol) and disrupted by sonication. Resultant lysate was centrifuged and the supernatant was recovered and mixed with 5 mL of cComplete His-tag Purification Resin (#05893801001, Roche, Switzerland) for 1 h with rotator. The resin was washed with 100 mL of Lysis buffer and then the protein was eluted with 25 mL of Elution buffer (50 mM Hepes-KOH, pH 7.6, 400 mM KCl, 10 mM MgCl<sub>2</sub>, 400 mM imidazole, and 7 mM 2-mercaptoethanol). TsaB and TrmD were concentrated by Amicon Ultra 3 kDa (#UFC800324, Merck Millipore, USA) and then dialyzed against Stock buffer (50 mM Hepes-KOH, pH 7.6, 500 mM KCl, 10 mM MgCl<sub>2</sub>, 7 mM 2-mercaptoethanol, and 30% glycerol). They were flash frozen with liquid nitrogen and stored at -80 °C. For TsaC, TsaE, and MnmE, Ulp1 (#12588018, Thermo Fischer Scientific, USA) was added to the recovered fractions to a final concentration of 23 µg/ml to remove the His-tagged SUMO protein. The fractions were dialyzed against Cleavage buffer (50 mM Hepes-KOH, pH 7.6, 100 mM KCl, and 7 mM 2-mercaptoethanol) overnight while the Cleavage buffer contained 500 mM KCl for MnmE preparation. The dialyzed samples were again mixed with 5 mL of cComplete His-tag Purification Resin with rotator. Then the flow-through fractions that contain modification enzymes were collected. Recovered samples were concentrated by Amicon Ultra 3 kDa for TsaE or Amicon Ultra 10 kDa (#UFC901024, Merck Millipore, USA) for TsaC and MnmE. They were dialyzed against Stock buffer, flash frozen with liquid nitrogen, and stored at -80 °C. For GlyA preparation, all buffer contained 10% Glycerol and the other procedures were same as MnmE preparation. TsaD was found to be insoluble after sonication. Therefore, the protein was pelleted by centrifugation at 20,400 × g at 4 °C for 45 min and it was resuspended in Lysis buffer supplemented with 4% Triton X-100. The suspension was again pelleted by centrifugation at 20,400 × g for 45 min. The pellet was dissolved with Lysis buffer supplemented with 6 M urea. The other procedure was same as MnmE preparation except that all buffers contained 2 M urea. For MnmC preparation, all steps until removal of His-tagged SUMO protein were same as GlyA preparation. Because MnmC nonspecifically bound to the His-tag purification resin, the purification with anion-exchange chromatography was selected. The solution after Ulp1 treatment was dialyzed against IEX buffer (50 mM Hepes-KOH, pH 7.6, 100 mM KCl, 10 mM MgCl<sub>2</sub>, and 7 mM 2-mercaptoethanol) and then they were applied onto a 5 mL HiTrap Q HP column (#17115401, GE Healthcare, USA). The column was washed with 25 mL of IEX buffer and then MnmC was eluted with a linear gradient from 100 mM to 1 M KCl in IEX buffer. Fractions containing MnmC were concentrated by Amicon Ultra 30 kDa (#UFC903024, Merck Millipore, USA), dialyzed against Stock buffer, flash frozen with liquid nitrogen, and stored at -80 °C. We note that we can share all of the plasmids on requests.

**Preparation of iVTtRNA.** Genes for each iVTtRNA (Supplementary Data 1) were cloned into the pGEMEX-1 vector (#P2211, Promega, USA) between *Xba*I and *Bam*HI restriction sites. Using the resultant plasmids as templates, DNA templates for in vitro transcription were PCR-amplified using a T7 promoter primer as forward primer and appropriate reverse primers for each iVTtRNA (Supplementary Data 1). Each reverse primer contained a 2'-methoxy modification at the second nucleotide from the 5'-terminus to prevent template-independent addition of an extra nucleotide<sup>28</sup>. Products were purified by phenol/chloroform/isoamyl alcohol (25:24:1) extraction, followed by ethanol precipitation, and precipitants were dissolved in water. Run-off transcription of precursor iVTtRNAs with 27 extra nucleotides introduced at the 5'-terminus (5'-GGGAGACCA-CAACGGTTTCCCTCTAGA-3') was performed using the resultant DNA templates for 3 h at 37 °C in 20 mL reaction mixtures containing 30 nM T7 RNA polymerase, 1 mM each ATP, GTP, CTP, and UTP, 40 mM HEPES-KOH pH 7.6, 20 mM MgCl<sub>2</sub>, 1.5 mM spermidine, 5 mM DTT, 20 µg PCR products, and 0.2 U/mL inorganic pyrophosphatase (#10108987001, Roche, Switzerland). RNase P components composed of C5 protein and M1 RNA were subsequently added to reaction mixtures at 150 nM for removal of the 27 extra nucleotides, and reactions were incubated for 1 h at 37 °C. Transcribed iVTtRNAs were then processed with acidic phenol extraction followed by chloroform/isoamyl alcohol (10:1) extraction, then purification by anion-exchange chromatography. Samples containing iVTtRNAs were loaded onto 10 mL HiTrap Q HP columns (#17115401, GE Healthcare, USA) and washed with Q buffer (20 mM HEPES-KOH pH 7.6 and 200 mM KCl). iVTtRNAs were eluted with a linear gradient from 200 mM to 1 M KCl in Q buffer. Fractions containing target iVTtRNAs, determined by urea-PAGE, were recovered by isopropanol precipitation, and precipitated iVTtRNAs were dissolved in water and stored at -80 °C until use. We note that we can share all of the plasmids on requests.

**Modification of iVTtRNA.** t<sup>6</sup>A37 modification was performed in the reaction mixture containing 50 mM Hepes-KOH, pH 7.6, 300 mM KCl, 20 mM MgCl<sub>2</sub>, 5 mM DTT, 50 mM NaHCO<sub>3</sub>, 1 µM CaCl<sub>2</sub>, 1 mM ATP, 1 mM Thr, 5 µM TsaC, 5 µM TsaB, 5 µM TsaD, 5 µM TsaE, 4 A<sub>260</sub> unit/mL tRNA<sup>Leu</sup><sub>GAU</sub> or tRNA<sup>Asp</sup><sub>GUU</sub> or tRNA<sup>Phe</sup><sub>GAA</sub>. m<sup>1</sup>G37 modification was performed in the reaction mixture containing 50 mM Hepes-KOH, pH 7.6, 200 mM KCl, 10 mM MgCl<sub>2</sub>, 36.4 µM S-adenosyl methionine (SAM), 1.5 µM TrmD, 10 A<sub>260</sub> unit/mL tRNA<sup>Pro</sup><sub>GCG</sub> or tRNA<sup>Phe</sup><sub>GAA</sub>. mnm<sup>5</sup>U34 was performed in the reaction mixture containing 50 mM Hepes-KOH, pH 7.6, 150 mM KCl, 12.5 mM MgCl<sub>2</sub>, 5 mM DTT, 500 µM FAD, 1 mM tetrahydrofolate, 4 mM GTP, 2.5 mM NADH, 0.2 mM pyridoxal-5'-



phosphate, 1 mM Ser, 1 mM Gly, 36.4  $\mu$ M SAM, 0.1 unit/ $\mu$ L recombinant RNase inhibitor (#2313 A, TaKaRa, Japan), 10  $\mu$ M GlyA, 3  $\mu$ M GidA, 3  $\mu$ M MnmE, 2.5  $\mu$ M MnmC, 6  $A_{260}$  unit/mL tRNA<sup>Glu</sup><sub>UUC</sub> or tRNA<sup>Glu</sup><sub>CUC</sub>. After incubation at 37°C for 2 h for t<sup>6</sup>A37 and m<sup>1</sup>G37 modification or 4 h for mnm<sup>5</sup>U34 modification, tRNAs were processed with acidic phenol extraction followed by chloroform/isoamyl alcohol (10:1) extraction and recovered by isopropanol precipitation. Precipitated tRNAs were dissolved in water and stored at -80°C. For mnm<sup>5</sup>U34 modification, the reaction was again performed with recovered tRNAs. They were processed with acidic phenol extraction followed by chloroform/isoamyl alcohol (10:1) extraction, desalted by MicroSpin G-25 Columns (#27532501, GE Healthcare, USA), and recovered by isopropanol precipitation. Precipitated tRNAs were dissolved in water and stored at -80°C. To quantify the modification efficiency, 57.1  $\mu$ M [<sup>14</sup>C]Thr was used instead of 1 mM Thr and the mixture without TsaD was used as control for t<sup>6</sup>A37 modification. 36.4  $\mu$ M S-[methyl-<sup>14</sup>C]-adenosyl-Methionine was used and the mixture without TrmD was used as control for m<sup>1</sup>G modification and without GidA was used as control for mnm<sup>5</sup>U34 modification. After incubation at 37°C, aliquots (10  $\mu$ L) were withdrawn and spotted on Whatman 25 mm GF/C filter disks (#1822-025, GE Healthcare, USA). Filter disks were washed twice with 10% TCA, then ethanol, followed by measurement of radioactivity by a liquid scintillation counter. For the quantification of mnm<sup>5</sup>U34 modification, tRNAs were purified as above and then 0.02  $A_{260}$  unit was spotted on the filter disks instead.

**Aminoacylation.** Aminoacylation experiments were performed according to a previous report<sup>47</sup>. Reaction mixtures (10  $\mu$ L) contained 100 mM HEPES-KOH pH 7.6, 15 mM MgCl<sub>2</sub>, 40 mM KCl, 1 mM DTT, 4 mM ATP, 1 unit/ $\mu$ L recombinant RNase inhibitor (#2313 A, TaKaRa, Japan), 50 nM or 1.5  $\mu$ M aaRS corresponding to each amino acid, 20  $\mu$ M [<sup>14</sup>C]-amino acid or 68  $\mu$ M [<sup>14</sup>C]Asn for asparagine aminoacylation, and 2  $A_{260}$  unit/mL tRNA. For measuring methylation, a mixture of 0.6  $\mu$ M [<sup>35</sup>S]Met and 3.4  $\mu$ M cold L-methionine was used instead. [<sup>14</sup>C]Cys was obtained by reducing [1,2,1',2'-<sup>14</sup>C]-cystine with 50 mM DTT at 37°C for 15 min. For measuring cysteinylating, the DTT concentration was 5 mM. When native tRNA mixtures were used, 40  $A_{260}$  unit/mL tRNA mixtures (#10109541001, Sigma-Aldrich, USA) were added instead. After incubation at 37°C for 30 min, aliquots (8  $\mu$ L) were withdrawn and spotted on Whatman 25 mm GF/C filter disks (#1822-025, GE Healthcare, USA). Filter disks were washed twice with 10% TCA, then with ethanol, followed by measurement of radioactivity by a liquid scintillation counter.

**Octapeptide synthesis with the PURE system.** DNA templates encoding octapeptides were PCR-amplified with appropriate primers (Supplementary Data 2) using the pURE1 vector (#PUREV001, BioComber, Japan) containing a T7 promoter sequence and the ribosome binding site (Shine-Dalgarno sequence) as a template. Amplified DNA templates were purified using a QIAquick PCR purification kit (#28104, QIAGEN, Germany). Octapeptide synthesis reactions contained 50 mM HEPES-KOH pH 7.6, 100 mM potassium glutamate, 13 mM magnesium acetate, 2 mM spermidine, 1 mM DTT, 2 mM ATP, 2 mM GTP, 1 mM UTP, 1 mM CTP, 20 mM creatine phosphate, 10  $\mu$ g/mL 10-formyl-5,6,7,8-tetrahydrofolic acid, 0.2  $\mu$ M ribosome, 4 nM DNA template, proteinaceous PURE system components including translation factors and enzymes, amino acids, and tRNAs. The concentrations of proteinaceous PURE system components were as described in a previous protocol<sup>46</sup>. We note that all 20 aaRSs were included in this experiment, regardless of DNA templates and test codons. The composition and concentration of amino acids, which depend on the test codons, are listed in Supplementary Data 2. The composition of iVTtRNAs depends on the template, and reactions were performed using basal iVTtRNAs (Supplementary Data 2) in the presence or absence of test iVTtRNAs (Supplementary Data 2). The concentration of each iVTtRNA was fixed at 6  $\mu$ M. When native tRNA mixtures were used, 56  $A_{260}$  unit/mL tRNA mixtures (#10109541001, Sigma-Aldrich, USA) were added instead. Reactions were performed for 60 min at 37°C and aliquots were withdrawn, spotted onto Whatman 3 MM filter papers (#1822-025, GE Healthcare, USA), boiled in 10% TCA at 90°C for 30 min to deacylate aminoacyl-RNAs. The radioactivity in the 10% TCA-insoluble fraction was measured with a liquid scintillation counter.

**Protein synthesis with the PURE system.** Protein synthesis experiments were performed with a PUREflex 2.0 kit (#PF201, GeneFrontier Corporation, Japan) without tRNA mixtures in Solution I (Buffer mix). We additionally added 0.2  $\mu$ M [<sup>35</sup>S]Met, 4 nM PCR-amplified DNA template for DHFR or sGFP expression (Supplementary Data 3), and a specified amount of iVTtRNA mixture or native tRNA mixture. Reactions were performed at 30 or 37°C for 12 h, and synthesized proteins were analyzed.

**Analysis of synthesized proteins.** Synthesized DHFR and sGFP containing radioactive [<sup>35</sup>S]Met were analyzed by 15% SDS-PAGE, and the gel image was visualized using a BAS-5000 bio-imaging analyzer (GE Healthcare, USA). Time-course analysis of sGFP expression was performed by measuring sGFP fluorescence every 3 min using a StepOne qRT-PCR system (#4376373, Applied Biosystems, USA). Fluorescence images of synthesized sGFP in reaction mixtures were

obtained using an LAS-4000 instrument (GE Healthcare, USA). The activity of synthesized DHFR was measured as described previously<sup>44</sup>. Reaction mixtures (2 mL) contained 50 mM MES-KOH pH 7.0, 25 mM TRIS-HCl pH 7.0, 25 mM ethanolamine, 100 mM NaCl, 10 mM 2-mercaptoethanol, 0.1 mM EDTA, 100  $\mu$ M dihydrofolic acid, and a 10  $\mu$ L aliquot of the PURE reaction mixture, and they were incubated at 37°C for 15 min. Next, 20 mM NADPH was added to a final concentration of 200  $\mu$ M, and the decrease in absorbance at 340 nm was measured over a period of 10 min with a V-550 spectrophotometer (Jasco, Japan). One unit of DHFR was defined as the amount of enzyme required to process 1  $\mu$ mol of dihydrofolic acid in 1 min at 37°C.

**LC-MS analysis of synthesized proteins.** DHFR with FLAG sequence at its terminus (Supplementary Data 3) was synthesized with a PUREflex 2.0 kit (#PF201, GeneFrontier Corporation, Japan) without tRNA mixtures in Solution I (Buffer mix). We additionally added 4 nM PCR-amplified DNA template for DHFR tagged with FLAG sequence at its C-terminus (Supplementary Data 3) and a specified amount of iVTtRNA mixture or native tRNA mixture. Reactions were performed at 30°C for 12 h. Synthesized DHFR was purified with Anti-FLAG M2 Magnetic Beads (#M8823, Sigma-Aldrich, USA). Aliquots (55  $\mu$ L) of the reaction mixtures were mixed with 245  $\mu$ L of the FLAG wash buffer (50 mM Hepes-KOH, pH 7.6, 150 mM NaCl, 20 mM Mg(OAc)<sub>2</sub>), and further mixed with 30  $\mu$ L beads for 1 h. The beads were washed with 210  $\mu$ L of the FLAG wash buffer followed by washing with the buffer without Mg(OAc)<sub>2</sub> twice (210  $\mu$ L and 180  $\mu$ L). Then, DHFR was eluted with 50  $\mu$ L of the FLAG wash buffer without Mg(OAc)<sub>2</sub> but supplemented with 100  $\mu$ g/mL 3xFLAG peptide (#F4799, Sigma-aldrich, USA) after gently mixing for 1 h. Preparation of samples for LC-MS analysis was performed according to a phase transfer surfactant (PTS)-aided protocol<sup>48</sup>. The 4  $\mu$ L of a dense PTS buffer (100 mM sodium deoxycholate, 100 mM sodium N-laurylsarcosinate, and 500 mM NH<sub>4</sub>HCO<sub>3</sub>) was added to 40  $\mu$ L of the eluted samples. They were reduced with 10 mM TCEP at 37°C for 30 min, alkylated with 20 mM iodoacetamide at 37°C for 30 min, and quenched with 20 mM Cys. Each reaction solution was divided into two portions. One was digested by 10 ng/ $\mu$ L Lys-C (#90051, Thermo Fisher Scientific, USA) and 10 ng/ $\mu$ L trypsin (#90057, Thermo Fisher Scientific, USA) at 37°C overnight. The other was digested by 10 ng/ $\mu$ L Asp-N (#90053, Thermo Fisher Scientific, USA) at 37°C overnight. After digestion, 10% trifluoroacetic acid (TFA) was added to a final concentration of 1%, and the reaction solutions were centrifuged at 15,000  $\times$  g for 5 min to precipitate the detergents. Supernatants were desalted using self-prepared stage tips<sup>49</sup> and dried with SpeedVac. LC-MS analysis was performed using an Orbitrap mass spectrometer (LTQ Orbitrap Velos Pro, Thermo Fisher Scientific, USA) equipped with a nanospray ion source (Nanospray Flex, Thermo Fisher Scientific, USA) and a nano-LC system (UltiMate 3000, Thermo Fisher Scientific, USA). The dried peptide mixtures were dissolved in a solution containing 5% acetonitrile and 0.1% TFA, and each sample was applied to the nano-LC system. Peptides were concentrated using a trap column (#164535, 0.075  $\times$  20 mm, 3  $\mu$ m, Acclaim PepMap 100 C18, Thermo Fisher Scientific, USA) and then separated using a nano capillary column (#NTCC-360/100-3-153, 0.1  $\times$  150 mm, 3  $\mu$ m, C18, Nikkyo Technos, Japan) using two mobile phases A (0.1% formic acid) and B (acetonitrile and 0.1% formic acid) with a gradient (5% B for 5 min, 5–45% B in 45 min, 45–90% B in 1 min, and 90% B in 4 min) at a flow rate of 500 nL/min. Elution was directly electrosprayed (2.2 kV) into the MS (positive mode, scan range of 200–1500 m/z, 60,000 FWHM resolution)<sup>50</sup>.

## Statistics and reproducibility

Error bars indicate standard deviation of triplicate measurements. Welch's *t* test was applied to show statistical significance using values of triplicate measurements. Exact number of replicates and measurement values are shown in Supplementary Data 6.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

DNA or amino acid sequences of the plasmids and DNA primers used in preparation of DNA templates for *in vitro* transcription are available in Supplementary Data 1, 2, 3, and 5. LC-MS data is available in Supplementary Data 4. Source data for main figures are presented in Supplementary Data 6.

Received: 30 April 2020; Accepted: 12 June 2020;

Published online: 03 July 2020

## References

1. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C. & Schultz, P. G. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244**, 182–188 (1989).

2. Reynolds, N. M., Vargas-Rodriguez, O., Söll, D. & Crnković, A. The central role of tRNA in genetic code expansion. *Biochim. Biophys. Acta Gen. Subj.* **1861**, 3001–3008 (2017).
3. Tajima, K., Katoh, T. & Suga, H. Genetic code expansion via integration of redundant amino acid assignment by finely tuning tRNA pools. *Curr. Opin. Chem. Biol.* **46**, 212–218 (2018).
4. Krall, N., da Cruz, F. P., Boutureira, O. & Bernardes, G. J. Site-selective protein-modification chemistry for basic biology and drug development. *Nat. Chem.* **8**, 103–113 (2016).
5. Alexandrov, A. et al. Rapid tRNA decay can result from lack of nonessential modifications. *Mol. Cell* **21**, 87–96 (2006).
6. Phelps, S. S., Malkiewicz, A., Agris, P. F. & Joseph, S. Modified nucleotides in tRNA<sup>Lys</sup> and tRNA<sup>Val</sup> are important for translocation. *J. Mol. Biol.* **338**, 439–444 (2004).
7. Murphy, F. V. 4th, Ramakrishnan, V., Malkiewicz, A. & Agris, P. F. The role of modifications in codon discrimination by tRNA<sup>Lys</sup><sub>UUU</sub>. *Nat. Struct. Mol. Biol.* **11**, 1186–1191 (2004).
8. Cui, Z., Stein, V., Tnimov, Z., Mureev, S. & Alexandrov, K. Semisynthetic tRNA complement mediates *in vitro* protein synthesis. *J. Am. Chem. Soc.* **137**, 4404–4413 (2015).
9. Iwane, Y. et al. Expanding the amino acid repertoire of ribosomal polypeptide synthesis via the artificial division of codon boxes. *Nat. Chem.* **8**, 317–325 (2016).
10. Calles, J., Justice, I., Brinkley, D., Garcia, A. & Endy, D. Fail-safe genetic codes designed to intrinsically contain engineered organisms. *Nucleic Acids Res.* **47**, 10439–10451 (2019).
11. Karzbrun, E., Tayar, A. M., Noireaux, V. & Bar-Ziv, R. H. Synthetic biology. Programmable on-chip DNA compartments as artificial cells. *Science* **345**, 829–832 (2014).
12. Berhanu, S., Ueda, T. & Kuruma, Y. Artificial photosynthetic cell producing energy for protein synthesis. *Nat. Commun.* **10**, 1325 (2019).
13. Noireaux, V., Maeda, Y. T. & Libchaber, A. Development of an artificial cell, from self-organization to computation and self-reproduction. *Proc. Natl Acad. Sci. USA.* **108**, 3473–3480 (2011).
14. Stano, P. & Luisi, P. L. Semi-synthetic minimal cells: origin and recent developments. *Curr. Opin. Biotechnol.* **24**, 633–638 (2013).
15. Hutchison, C. A. 3rd et al. Design and synthesis of a minimal bacterial genome. *Science* **351**, aad6253 (2016).
16. Fredens, J. et al. Total synthesis of *Escherichia coli* with a recoded genome. *Nature* **569**, 514–518 (2019).
17. Shimizu, Y. et al. Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19**, 751–755 (2001).
18. Libicher, K., Hornberger, R., Heymann, M. & Mutschler, H. *In vitro* self-replication and multicistronic expression of large synthetic genomes. *Nat. Commun.* **11**, 904 (2020).
19. Agris, P. F. et al. Celebrating wobble decoding: half a century and still much is new. *RNA Biol.* **15**, 537–553 (2018).
20. Arigoni, F. et al. A genome-based approach for the identification of essential bacterial genes. *Nat. Biotechnol.* **16**, 851–856 (1998).
21. Freiberg, C. et al. Identification of novel essential *Escherichia coli* genes conserved among pathogenic bacteria. *J. Mol. Microbiol. Biotechnol.* **3**, 483–489 (2001).
22. Tamura, K., Himeno, H., Asahara, H., Hasegawa, T. & Shimizu, M. *In vitro* study of *E. coli* tRNA<sup>Arg</sup> and tRNA<sup>Lys</sup> identity elements. *Nucleic Acids Res.* **20**, 2335–2339 (1992).
23. Madore, E. et al. Effect of modified nucleotides on *Escherichia coli* tRNA<sup>Glu</sup> structure and on its aminoacylation by glutamyl-tRNA synthetase. Predominant and distinct roles of the mnm<sup>5</sup> and s<sup>2</sup> modifications of U34. *Eur. J. Biochem.* **266**, 1128–1135 (1999).
24. Iwane, Y., Katoh, T., Goto, Y. & Suga, H. Artificial division of codon boxes for expansion of the amino acid repertoire of ribosomal polypeptide synthesis. *Methods Mol. Biol.* **1728**, 17–47 (2018).
25. Himeno, H. et al. Role of the extra G-C pair at the end of the acceptor stem of tRNA<sup>His</sup> in aminoacylation. *Nucleic Acids Res.* **17**, 7855–7863 (1989).
26. Hasegawa, T., Himeno, H., Ishikura, H. & Shimizu, M. Discriminator base of tRNA<sup>Asp</sup> is involved in amino acid acceptor activity. *Biochem. Biophys. Res. Commun.* **163**, 1534–1538 (1989).
27. Shimizu, M., Asahara, H., Tamura, K., Hasegawa, T. & Himeno, H. The role of anticodon bases and the discriminator nucleotide in the recognition of some *E. coli* tRNAs by their aminoacyl-tRNA synthetases. *J. Mol. Evol.* **35**, 436–443 (1992).
28. Kao, C., Zheng, M. & Rüdiger, S. A simple and efficient method to reduce nontemplated nucleotide addition at the 3' terminus of RNAs transcribed by T7 RNA polymerase. *RNA* **5**, 1268–1272 (1999).
29. Fukunaga, J. et al. Use of RNase P for efficient preparation of yeast tRNA<sup>Tyr</sup> transcript and its mutants. *J. Biochem.* **139**, 123–127 (2006).
30. Thiaville, P. C. et al. Essentiality of threonylcarbamoyladenine (t<sup>6</sup>A), a universal tRNA modification, in bacteria. *Mol. Microbiol.* **98**, 1199–1221 (2015).
31. Sylvers, L. A., Rogers, K. C., Shimizu, M., Ohtsuka, E. & Söll, D. A 2-thiouridine derivative in tRNA<sup>Glu</sup> is a positive determinant for aminoacylation by *Escherichia coli* glutamyl-tRNA synthetase. *Biochemistry* **32**, 3836–3841 (1993).
32. Nilsson, M. & Rydén-Aulin, M. Glutamine is incorporated at the nonsense codons UAG and UAA in a suppressor-free *Escherichia coli* strain. *Biochim. Biophys. Acta* **1627**, 1–6 (2003).
33. Toth, M. J., Murgola, E. J. & Schimmel, P. Evidence for a unique first position codon-anticodon mismatch *in vivo*. *J. Mol. Biol.* **201**, 451–454 (1988).
34. Shimizu, Y. ArfA recruits RF2 into stalled ribosomes. *J. Mol. Biol.* **423**, 624–631 (2012).
35. Kim, D. M., Kigawa, T., Choi, C. Y. & Yokoyama, S. A highly efficient cell-free protein synthesis system from *Escherichia coli*. *Eur. J. Biochem.* **239**, 881–886 (1996).
36. Manickam, N., Joshi, K., Bhatt, M. J. & Farabaugh, P. J. Effects of tRNA modification on translational accuracy depend on intrinsic codon-anticodon strength. *Nucleic Acids Res.* **44**, 1871–1881 (2016).
37. Giegé, R., Sissler, M. & Florentz, C. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.* **26**, 5017–5035 (1998).
38. Ma, C., Kudlicki, W., Odom, O. W., Kramer, G. & Hardesty, B. *In vitro* protein engineering using synthetic tRNA<sup>Ala</sup> with different anticodons. *Biochemistry* **32**, 7939–7945 (1993).
39. Hughes, R. A. & Ellington, A. D. Synthetic DNA Synthesis and Assembly: Putting the Synthetic in Synthetic Biology. *Cold Spring Harb. Perspect. Biol.* **9**, a023812 (2017).
40. Ikeuchi, Y., Shigi, N., Kato, J., Nishimura, A. & Suzuki, T. Mechanistic insights into sulfur relay by multiple sulfur mediators involved in thiouridine biosynthesis at tRNA wobble positions. *Mol. Cell* **21**, 97–108 (2006).
41. Bray, M. S. et al. Multiple prebiotic metals mediate translation. *Proc. Natl Acad. Sci. USA.* **115**, 12164–12169 (2018).
42. Hori, H. et al. Transfer RNA modification enzymes from thermophiles and their modified nucleosides in tRNA. *Microorganisms* **6**, E110 (2018).
43. Noireaux, V., Bar-Ziv, R., Godefroy, J., Salman, H. & Libchaber, A. Toward an artificial cell based on gene expression in vesicles. *Phys. Biol.* **2**, 1–8 (2005).
44. Maini, R. et al. Ribosome-mediated incorporation of dipeptides and dipeptide analogues into proteins *in vitro*. *J. Am. Chem. Soc.* **137**, 11206–11209 (2015).
45. Lajoie, M. J., Söll, D. & Church, G. M. Overcoming challenges in engineering the genetic code. *J. Mol. Biol.* **428**, 1004–1021 (2016).
46. Shimizu, Y. & Ueda, T. PURE technology. *Methods Mol. Biol.* **607**, 11–21 (2010).
47. Kiga, D. et al. An engineered *Escherichia coli* tyrosyl-tRNA synthetase for site-specific incorporation of an unnatural amino acid into proteins in eukaryotic translation and its application in a wheat germ cell-free system. *Proc. Natl Acad. Sci. USA.* **99**, 9715–9720 (2002).
48. Masuda, T., Tomita, M. & Ishihama, Y. Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *J. Proteome Res.* **7**, 731–740 (2008).
49. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906 (2007).
50. Shimojo, M. et al. *In vitro* reconstitution of functional small ribosomal subunit assembly for comprehensive analysis of ribosomal elements in *E. coli*. *Commun. Biol.* **3**, 142 (2020).

## Acknowledgements

We thank Nono Takeuchi-Tomita for helpful discussions. KA was partly supported by an NIGMS grant (R35GM122560 to Prof. Dieter Söll in Yale University). This work was supported by a Grant-in-Aid to KH (19J12955), KA (15K16083), and YS (17H05680) from the Japan Society for the Promotion of Science (JSPS), the Human Frontier Science Program (RGP0043/2017 to YS), the Astrobiology Center Project of the National Institutes of Natural Sciences (AB271004 and AB281007 to KA, and AB311005 to YS and KA), and an intramural Grant-in-Aid from the RIKEN Center for Biosystems Dynamics Research (to YS).

## Author contributions

K.H., K.A., T.U., and Y.S. designed the study. K.H., K.A., and Y.S. wrote the manuscript. K.H., N.S., K.M., and Y.S. performed the experiments. S.A. and T.Y. prepared plasmids for tRNA transcription. All authors discussed the results and commented on the manuscript.

**Competing interests**

The authors declare no competing interests.

**Additional information**

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s42003-020-1074-2>.

**Correspondence** and requests for materials should be addressed to Y.S.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020