REGENERATION

Sustainable regeneration of 20 aminoacyl-tRNA synthetases in a reconstituted system toward self-synthesizing artificial systems

Katsumi Hagino[1], Keiko Masuda[2], Yoshihiro Shimizu[2], Norikazu Ichihashi[1,3,4]*

In vitro construction of self-reproducible artificial systems is a major challenge in bottom-up synthetic biology. Here, we developed a reconstituted system capable of sustainably regenerating all 20 aminoacyl-transfer RNA synthetases (AARS), which are major components of the translation system. To achieve this, we needed five types of improvements: (i) optimization of AARS sequences for efficient translation, (ii) optimization of the composition of the translation system to enhance translation, (iii) employment of another bacterial AlaRS and SerRS to improve each aminoacylation activity, (iv) diminishing the translational inhibition caused by certain AARS sequences by codon optimization and EF-P addition, and (v) balancing the DNA concentrations of 20 AARS to match each requirement. After these improvements, we succeeded in the sustainable regeneration of all 20 AARS for up to 20 cycles of 2.5-fold serial dilutions. These methodologies and results provide a substantial advancement toward the realization of self-reproducible artificial systems.

INTRODUCTION

Self-reproduction is a fundamental characteristic of living organisms, and its implementation in artificial systems remains a major challenge (1–17). Although notable progress has been made in reconstituting biological functions such as DNA replication, metabolism, cellular proliferation, and cytoskeletal function in vitro, the ability to continuously reproduce essential biomolecules such as DNA, RNA, and proteins has not been achieved in synthetic systems (18–34). In living organisms, self-replication relies on the synthesis of polymers from small molecules via processes such as DNA replication, transcription, and translation.

To construct a self-reproducible artificial system, the reconstituted transcription/translation system of *Escherichia coli* (PURE system) is a reasonable starting point, because it consists of minimal elements of transcription and translation (35). The PURE system consists of macromolecules, including ribosomes, tRNAs, 20 aminoacyl-tRNA synthetases (AARS), translational initiation/elongation/releasing factors, and T7 RNA polymerase. If all these RNAs and proteins in the PURE system could be sustainably regenerated by the PURE system itself, it would be much closer to a self-reproducible system.

Several studies have successfully synthesized translation proteins (TPs) using the PURE system. Awai *et al.* expressed each of 20 AARS in the PURE system and detected activity except for PheRS (36). Wei and Endy expressed each of 36 TPs in the PURE system and detected each activity of 19 of 23 testable TPs after purification (37). Li *et al.* coexpressed most of the ribosomal proteins in one pot and detected reconstitution of the 30S subunit (38). Libicher *et al.* coexpressed multiple TPs in the PURE system from three large plasmids encoding most of the translation factors (39), although the expression amounts were insufficient to regenerate the original amount of TPs, and the activities of the expressed TPs were not

verified. Doerr *et al.* also reported that a significant portion of the coexpressed proteins was truncated owing to inefficient ribosome processivity (40). However, most of these studies did not demonstrate regeneration, because the expressed translation factors were not used for further translation.

As a study demonstrating the regeneration of TPs during transcription/translation reactions, Libicher and Mutschler performed partial regeneration of T7 RNA polymerase and adenylate kinase, as well as 12 AARS and RF1, to the second generation by serial dilution experiments using the PURE system in which the target proteins were removed (41). In addition, Lavickova *et al.* demonstrated the sustainable self-regeneration of T7 RNA polymerase or up to seven AARS using a microfluidic reactor, which allowed continuous expression of green fluorescent protein (GFP) for more than 24 hours (42). Inspired by these studies, we recently constructed a system that couples the expression of 20 AARS with replication of DNA encoding AARS. Although DNA replication was sustained over four generations in this system, translation activity rapidly declined over generations, indicating that the 20 AARS were not sufficiently regenerated (43). In summary, sustainable regeneration in PURE systems is still limited to seven AARS (42), and increasing this number is the next most important issue.

In this study, we attempted to increase the number of regenerative TPs up to all 20 AARS, which account for ~5% (0.2 mg/ml) of TPs in our customized PURE system excluding ribosomes. For this purpose, we first quantified the translation efficiency and activity per protein of the AARS expressed in the PURE system. Second, we improved the AARS sequences to increase translation efficiency. Third, we found and used highly active AlaRS and SerRS from *Geobacillus stearothermophilus*. Fourth, we found that some AARS sequences had inhibitory effects on the translation activity of the PURE system and mitigated this effect by removing rare codons and adding EF-P. Fifth, we further optimized the DNA concentration ratio to achieve ideal AARS expression balance. After these improvements, we succeeded in maintaining the translation activity in 20 AARS-depleted PURE systems by supplying 20 AARS from DNA through a 2.5-fold serial dilution for 20 cycles.

[1]Department of Life Science, Graduate School of Arts and Science, University of Tokyo, Meguro, Tokyo 153-8902, Japan. [2]Laboratory for Cell-Free Protein Synthesis, RIKEN Center for Biosystems Dynamics Research (BDR), Suita, Osaka 565-0874, Japan. [3]Komaba Institute for Sciences, University of Tokyo, Meguro, Tokyo 153-8902, Japan. [4]Universal Biology Institute, University of Tokyo, Meguro, Tokyo 153-8902, Japan.
*Corresponding author. Email: ichihashi@g.ecc.u-tokyo.ac.jp

RESULTS

Assay of translation efficiency and activities of 20 AARS

To systematically achieve regeneration of 20 AARS, we first evaluated the translation efficiency and activity of the 20 AARS expressed in the PURE system. In contrast to translation, which has been measured previously (*43*), the activity per protein of each AARS has not yet been measured. The assay procedure is outlined in Fig. 1A.

For the translation efficiency assay [Fig. 1A(i)], the DNA encoding each AARS (AARS-DNA) was incubated with a customized PURE system (version 1) containing stable isotope–labeled amino acids (Lys and Arg) to label the newly synthesized AARS. All the proteins, including the expressed AARS, were purified and digested with trypsin. The resulting peptides were quantified using liquid chromatography–mass spectrometry (LC-MS). The expressed AARS, containing isotope-labeled amino acids, had a higher molecular weight (i.e., heavy) than the endogenous (i.e., light) AARS in the PURE system, allowing for calculation of the abundance ratio [heavy-to-light (H/L) ratio] between the expressed and endogenous AARS (fig. S1). Translation efficiency was determined by multiplying the H/L ratio by the known concentration of endogenous AARS (Fig. 1B). The translation efficiency varies greatly from 0.05 to 3.75 μM with the average of ~1 μM. Some AARS (LeuRS, MetRS, PheRS, SerRS, and ValRS) exhibited significantly lower translations than average, suggesting that there is room for improvement in these AARS translations. In addition, it is notable that this translation efficiency may be underestimated if the tRNA mixture we used here, which is purified from *E. coli*, contains precharged tRNAs for lysine and arginine.

For the assay of the activity per protein of the expressed AARS, each AARS-DNA or no DNA (as a control) was incubated in the customized PURE system (version 1) using the same method as the translation assay. The mixture was then diluted with the customized PURE system lacking the AARS of interest at various ratios, and GFP expression was monitored [Fig. 1A(ii)]. The increase in GFP fluorescence was plotted against dilution rate (fig. S2), and the slopes
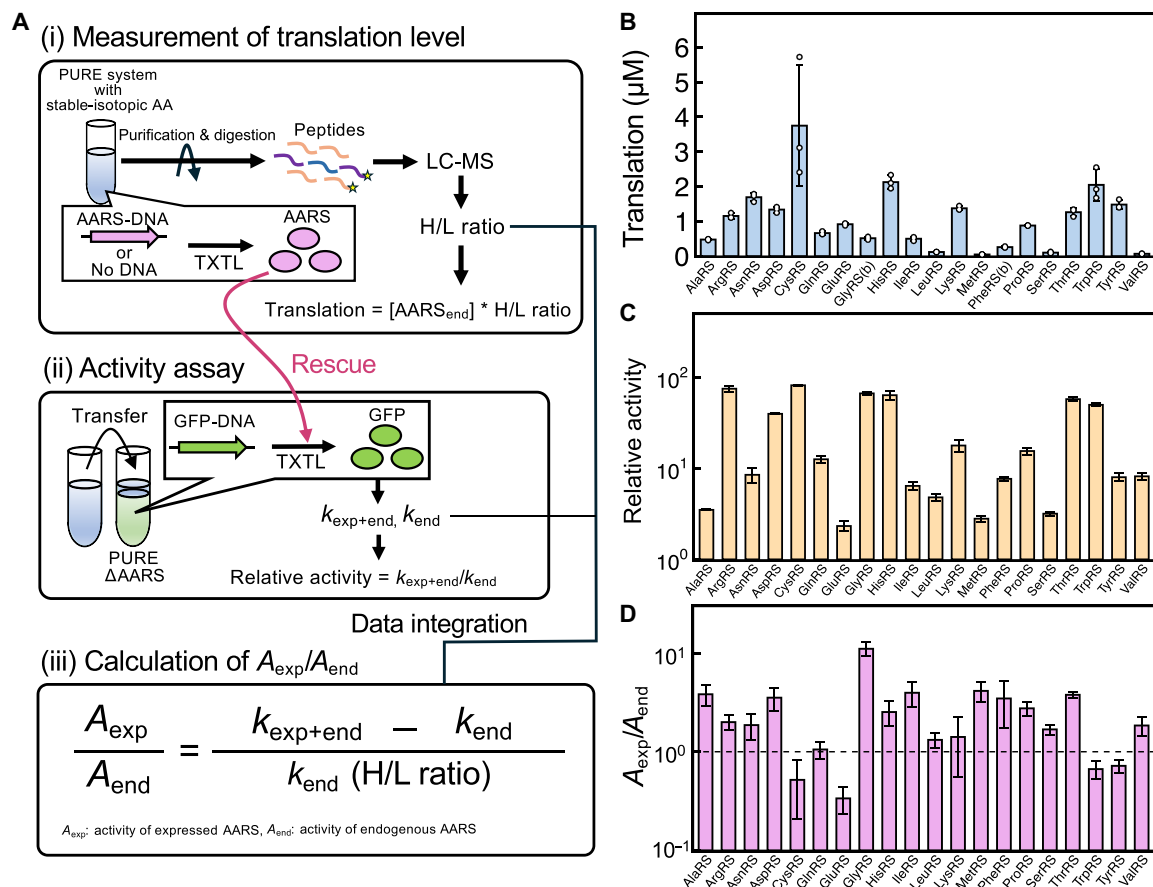
Fig. 1. Characterization of 20 AARS expressed in the PURE system. (**A**) Characterization scheme for 20 AARS. Each AARS was expressed in a customized 1st PURE system (version 1) containing stable isotope–labeled amino acids at 30°C for 8 hours. The expressed AARS were purified together with endogenous AARS, treated with trypsin, and analyzed using LC-MS to determine the H/L ratio. The translation efficiencies were obtained by multiplying the H/L ratio by the known AARS concentrations in the PURE system. AARS activity was measured by expressing GFP in the 2nd PURE system lacking the AARS of interest and by assessing the GFP synthesis rate. The activity per protein ratio (A_{exp}/A_{end}) was calculated from the H/L ratio and AARS activity according to the equation. (**B**) Translation efficiency of each AARS. Each point represents an individual detected peptide, with error bars indicating standard deviation. The H/L ratios are presented in fig. S1. (**C**) Total activity of each AARS. k_{end} represents the total activity of each endogenous AARS, and $k_{exp+end}$ represents the total activity of both endogenous and expressed AARS. The error bars represent the standard errors of the slope of the linear regression shown in fig. S2. The raw data for each AARS are presented in figs. S2 and S3. (**D**) Activity per protein ratio of expressed to endogenous AARS. The error bars indicate the propagated standard errors calculated using Eq. 1.

of the regression curves were obtained. The slope of the data with AARS-DNA represents the sum of the expressed and endogenous AARS activity ($k_{exp+end}$). The slope of the data with no DNA represents endogenous AARS activity (k_{end}). From these data, we obtained the relative activities ($k_{exp+end}/k_{end}$) of the total expressed AARS to the total endogenous AARS (Fig. 1C), which also exhibited varying amounts. For TrpRS, we used phi29 DNA polymerase instead of GFP as a reporter gene because GFP has only one tryptophan residue (fig. S3) (43).

To estimate the activity per protein of each expressed AARS, we divided the relative activity [$(k_{exp+end} - k_{end})/k_{end}$] by the relative translation amount (H/L ratio) obtained previously [Fig. 1A(iii)]. The resultant value (A_{exp}/A_{end}) represents the relative activity per protein of the AARS expressed in the PURE system compared to that of the endogenous purified AARS protein. The relative activities per protein of most AARS were more than one (Fig. 1D), indicating that the expressed proteins of most AARS had similar or higher activities per protein than the endogenous ones. This result suggests that sustainable regeneration can be achieved with lower protein expression for these AARS. However, some AARS (CysRS, GluRS, TrpRS, and TyrRS) exhibit lower activity per protein than endogenous AARS, indicating room for improvement.

Improvement of the 5′-terminal sequences

First, we sought to enhance the translation efficiency. The AARS sequences used in this study originated from plasmids used for expression in *E. coli* (44). By comparing the 5′ untranslated region (5′UTR), we found that some AARS (LeuRS, MetRS, SerRS, and ValRS), which show relatively smaller translation in Fig. 1B, have a shorter 5′UTR that lacks a spacer sequence and an epsilon sequence (table S1), which are important regions for translation efficiency (45). We also found that for some AARS (LysRS, ProRS, GlnRS, LeuRS, and SerRS), the GC ratio up to the seventh residue, which should be as low as possible for translation (46), could be reduced without changing the amino acid sequences (table S1). Accordingly, we added the 5′UTR sequences, reduced the GC ratio in the N-terminal regions of AARS, and observed some improvements in translation efficiency (fig. S4A).

Improvement of the composition of the PURE system

Two AARS (CysRS and GluRS), which exhibit low activities per protein, contain Zn ions in their active sites (47, 48). To address the possibility of zinc deficiency in these AARS, we added 1 mM Zn(OAc)$_2$ to the PURE system (version 1). The activities per protein (A_{exp}/A_{end}) of CysRS and GluRS increased two to four times and reached one (fig. S4B), indicating that the resultant activities per protein were comparable to those of the endogenous AARS.

Subsequently, we changed the composition of the PURE system to improve its translation capacity. In a previous study (43) and in the experiments described so far, we used a customized PURE system tailored for DNA replication (termed "version 1" in this study), in which some components, such as T7 RNA polymerase, nucleotide triphosphates, and tRNA, were reduced (table S2, version 1). By restoring the concentrations of these components to their original levels (49), the translation level increased by ~14-fold (fig. S4C and table S2). The PURE system of this composition was termed "version 2" in this study.
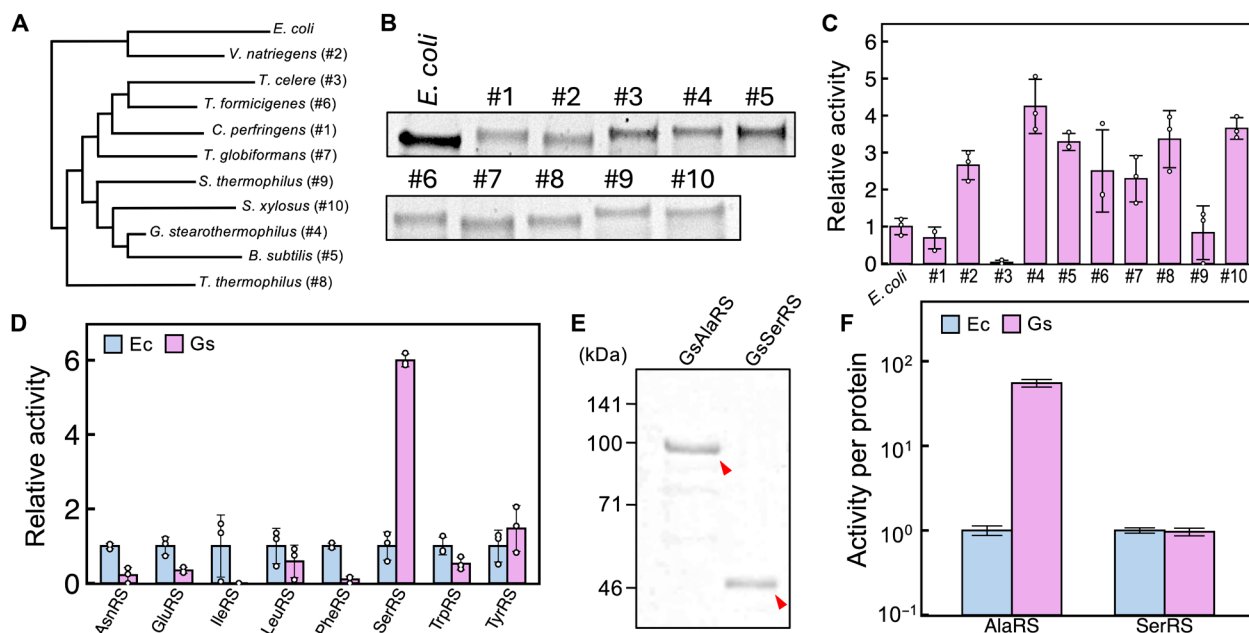
Highly active AARS derived from other bacteria

One possible strategy for AARS regeneration is to minimize the required translation level by enhancing the activities per protein of the expressed AARS. We first focused on AlaRS, which is the most abundant AARS in PURE systems. We selected AlaRSs from 10 bacteria with growth rates similar to that of *E. coli* (Fig. 2A) (50). They were expressed in the PURE system, and their activities were assessed using the GFP expression assay shown in Fig. 1A(ii). Although all 10 variants exhibited lower translation efficiencies than *E. coli* AlaRS (Fig. 2B), seven of these showed higher total activities (Fig. 2C). Among these, we selected the most active AlaRS (no. 4) from *G. stearothermophilus* (GsAlaRS) for subsequent regeneration experiments. Next, we investigated activities of other AARS isolated from *G. stearothermophilus*. Of the nine tested AARS, SerRS (GsSerRS) was approximately six times more active than *E. coli* SerRS (EcSerRS) when expressed in the PURE system (Fig. 2D). We also used GsSerRS for the subsequent regeneration experiments.

To compare the activities per protein of GsAlaRS and GsSerRS to *E. coli* ones, we purified recombinant GsAlaRS and GsSerRS from *E. coli* (Fig. 2E) and assayed for their activities. We found that GsAlaRS was ~50 times more active than its *E. coli* counterpart, whereas GsSerRS showed no significant difference (Fig. 2F). These results indicate that GsAlaRS is more active than EcAlaRS even when expressed in *E. coli*, whereas GsSerRS is more active only when expressed in the PURE system.

Serial dilution experiments for 1 to 10 AARS regeneration

We attempted to regenerate 1 to 10 AARS by serial dilution experiments using the data and the improved AARS obtained above. The experimental procedure is illustrated in Fig. 3A. In the first reaction (round 1), 1 to 10 AARS-DNAs of interest and firefly luciferase–encoding DNA (Fluc-DNA) were incubated in the customized PURE system (version 2), containing all 20 AARS proteins, for 8 hours at 30°C. Firefly luciferase was used as a translation reporter in this experiment because its activity was sufficiently detectable at a DNA concentration that did not compete with AARS translation. We also confirmed that luciferase expression was dependent on any of the 20 AARS, similar to GFP expression (fig. S7). The concentration of each AARS-DNA was optimized on the basis of the translation efficiency and the relative activity per protein (A_{exp}/A_{end}) according to Eq. 3 (table S3) regenerating the original AARS ratios in the PURE system. During incubation, AARS of interest and reporter luciferase were expressed. Thereafter, an aliquot of the reaction mixture was diluted fivefold with another customized PURE system (PUREΔAARS) that lacks each AARS of interest and contains AARS-DNA and Fluc-DNA at the same concentration and incubated again. In this second reaction, both luciferase and the target AARS were continuously expressed if the lacking AARS protein was sufficiently supplied (i.e., rescued) from the previous round of reaction. This serial dilution process was repeated for up to 10 rounds, and at the end of each round, luciferase activity was assayed as a measure of translation activity. If each AARS is regenerated continuously in sufficient amounts, translation activity should remain constant. As a negative control, we conducted the same experiment using DNA encoding GFP (GFP-DNA) instead of AARS-DNA. In addition, as a positive control, we conducted the same serial dilution experiment using GFP-DNA and a customized PURE system containing all 20 AARS proteins (Full PURE).

Fig. 2. Translation and activities of AARS derived from other bacteria. (**A**) Phylogenetic tree of AlaRSs obtained from *E. coli* and 10 other bacteria. A tree was constructed using the neighbor-joining method. *V.* natriegens, *Vibrio natriegens*; *T. celere, Thermobrachium celere*; *T. formicigenes, Tepidibacter formicigenes*; *C. perfringes, Clostridium perfringens*; *T. globiformans, Thermosipho globiformans*; *S. thermophilus, Streptococcus thermophilus*; *S. xylosus, Staphylococcus xylosus*; *B. subtilis, Bacillus subtilis*; *T. thermophilus, Thermus thermophilus*. (**B**) Translation of the AlaRSs. Each AARS was expressed in the PURE system (version 1) with fluorescently labeled lysyl-tRNA at 30°C for 8 hours. The mixture was analyzed by SDS-PAGE, and fluorescence was detected. The entire gel image is shown in fig. S5A. (**C**) Activity of AlaRSs. AlaRSs were expressed in the PURE system (version 1) and diluted with the PURE system lacking AlaRS, and GFP fluorescence was measured as the AlaRS activity. Values relative to *E. coli* AlaRS are shown. Error bars show the standard deviations from three experiments. (**D**) Activity of eight AARS of *G. stearothermophilus* (Gs). Eight *G. stearothermophilus*–derived AARS were expressed and measured as described in (C). Error bars show standard deviations from three experiments. The results of SDS-PAGE analysis are shown in fig. S5B. Ec, *E. coli*. (**E**) SDS-PAGE analysis of recombinant GsAlaRS and GsSerRS purified from *E. coli*. (**F**) Activity per protein of purified AARS. Each purified AARS was added to a PURE system lacking the target AARS, and GFP expression was used to calculate activity per protein. The error bars represent the standard errors of the slope of the linear regression shown in fig. S6. The raw data are shown in fig. S6.

We observed that for a single AARS (IleRS), five AARS (AsnRS, TyrRS, PheRS, IleRS, and GluRS), and the positive control (Full PURE), the translational activities at each round were maintained at a level similar to the initial level, at least until round 10 (Fig. 3B, upper and lower right panels). In contrast, the translational activity of the 10 AARS-DNAs decreased continuously in each round (Fig. 3B, lower left panel). We conducted the same experiments using 10 AARS-DNAs in different combinations. All translation activities eventually decreased (Fig. 3C). These results demonstrate that the current reaction conditions are insufficient for the regeneration of 10 AARS. This result was unexpected because the predicted translation amount based on each AARS translation and activity data (Fig. 1) was at least 12 times higher than the required level for regeneration. This implies that unknown factors diminish AARS expressions or their activities.

Inhibitory effect of some AARS for translation

One of the possible factors that has not been considered in the previous experiment is the interaction among the 20 AARS-DNAs, because we measured each AARS translation independently. To examine the effect of each AARS-DNA on the expression of other genes, we incubated each AARS-DNA with Fluc-DNA in the PURE system and compared the luminescence produced by the expressed luciferase. The results were then plotted against the expected translation of each AARS used in Fig. 1B. If there is simple competition between AARS-DNA and Fluc-DNA for translation resources, luminescence should

linearly decrease as the expected translation increases. As expected, approximately half of the AARS followed this trend (Fig. 4A, indicated by the wide gray line). In contrast, other eight AARS-DNAs (GsAlaRS, AspRS, GlnRS, GluRS, GlyRS, IleRS, LeuRS, and ValRS) are accumulated in the lower left part the figure (yellow), implying that these AARS-DNAs inhibit luciferase expression more than expected from simple competition.

Next, we attempted to eliminate the inhibitory effects of four AARS-DNAs (LeuRS, ValRS, GlnRS, and GluRS) that require relatively large amounts for sustainable regeneration. One possible mechanism for translational inhibition is ribosome sequestration caused by ribosome stall. Inhibitory AARS may have a sequence that induces the stalling of ribosomes, which could reduce active free ribosomes and inhibit the translation of other genes. LeuRS contains four closely located rare codons (AGA, ATA, CGA, and CTA codons at positions 6, 9, 46, and 47, respectively) that may induce ribosome stalling at this position. Therefore, we created a modified LeuRS (LeuRS_v2) by replacing all rare codons with a common one and compared its inhibitory effect when coexpressed with Fluc-DNA. Coexpression with the original LeuRS reduced luciferase activity by 40% compared to GFP-DNA, but this reduction was mitigated with LeuRS_v2 (Fig. 4B). GlnRS also contains a rare codon (AGA at position 6), while the substitution of this rare codon (GlnRS_v2) did not mitigate the inhibition effect (Fig. 4B). These results indicated the importance of avoiding multiple or clustered rare codons.
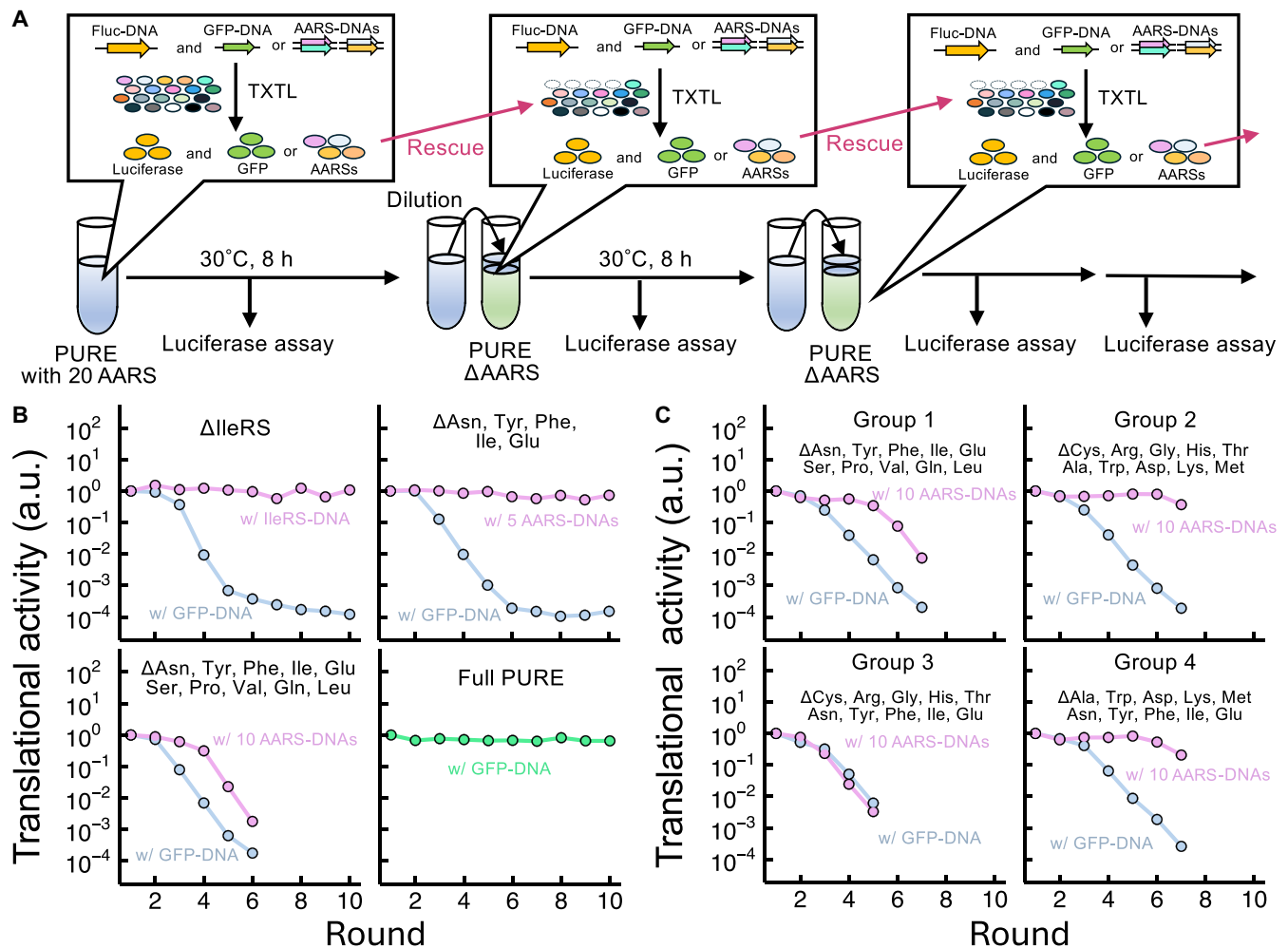
Fig. 3. Serial dilution experiments for 1 to 10 AARS regenerations. (**A**) Scheme of serial dilution experiment. The first reaction mixture contains 1 to 10 AARS-DNAs (5 nM in total) and 0.01 nM Fluc-DNA encoding luciferase in the customized PURE system (version 2) containing all 20 AARS proteins. In the control experiment, GFP-DNA (5 nM) was used instead of 20 AARS-DNAs. Each DNA concentration is presented in table S3. In the first reaction (round 1), AARS, GFP, and luciferase were expressed at 30°C for 8 hours (h). From round 2 onward, the reaction solution of the previous round was diluted fivefold with the customized PURE system (version 2), which lacks each AARS of interest (PUREΔAARS) and contains the same DNA sets as round 1. The diluted mixture was then incubated at 30°C for 8 hours. If AARS are sufficiently synthesized in the previous round, the lack of AARS in the next round is rescued, and luciferase and AARS are expressed again. Luciferase activity (luminescence) was measured after each round to assess the translation activity. (**B** and **C**) Trajectories of the translation activity in the serial dilution experiment. The translational activity in each round was plotted after normalization on the basis of the value at round 1. a.u., arbitrary units.

The three AARS (ValRS, GlnRS, and GluRS) do not contain a region with multiple rare codons but commonly contain proline repeats. GlnRS and GluRS have a proline doublet (PP), while ValRS has a proline triplet (PPP), which is known to cause ribosome stalling (*51–53*). Ribosome stalling at the polyproline site can be resolved by addition of EF-P (*52, 53*). We tested whether EF-P could eliminate the inhibition caused by these three AARS. When GlnRS, GluRS, and control GFP (no polyproline) were coexpressed with Fluc-DNA, EF-P addition slightly increased luciferase luminescence, even for GFP, whereas when ValRS was coexpressed, EF-P significantly enhanced luminescence (Fig. 4C). This suggests that the inhibitory effect of ValRS, which could be caused by its proline triplet, can be alleviated by the addition of EF-P.

In these experiments (Fig. 4, B and C), we identified two methods (using LeuRS_v2 and adding EF-P) to mitigate the inhibitory effects of LeuRS and ValRS-DNAs. To verify the total effect of these methods

under conditions that can be used for 20 AARS regeneration experiments, we performed a coexpression experiment of Fluc-DNA with all four inhibitory AARS-DNAs at realistic concentrations. We observed ~40% inhibition of luciferase activity, which was reduced to the original level by LeuRS_v2 and the addition of EF-P (fig. S8).

Another interesting result shown in Fig. 4A is the higher luminescence when coexpressed with AsnRS-DNA (N) than the expected competition line. This result suggests that AsnRS expression increases the translation of other genes. This result implies that AsnRS protein is insufficient in the PURE system. This point was not addressed in this study but should be improved in the future.

Calibration of DNA concentration ratio

Another possible factor that could cause the failure of AARS regeneration is the deviation from the expected expression level when multiple AARS are coexpressed owing to the competition of transcription/

translation resources among AARS-DNAs. In the previous experiment in Fig. 3, we prepared an AARS-DNA mixture that was expected to express 20 AARS at the optimum balance based on the individually measured data (table S4, left). In this experiment, we prepared a 20 AARS-DNA mixture using the same method and measured its expression in a PURE system (version 2) using isotope labeling and LC-MS as shown in Fig. 1A(i). The actual translation of each AARS (Fig. 5A) and

the ratio of the actual to predicted translation (Fig. 5B) were calculated. If each AARS is expressed as predicted, the ratio should be one for all AARS. However, Fig. 5B shows that the ratio [TL (translation)/predict] varies from 0.04 to 1.2 (coefficient of variation is 0.92), and also the average ratio (0.37) was smaller than 1, indicating that when coexpressed, both the balance and averaged translation level of 20 AARS are different from our prediction.
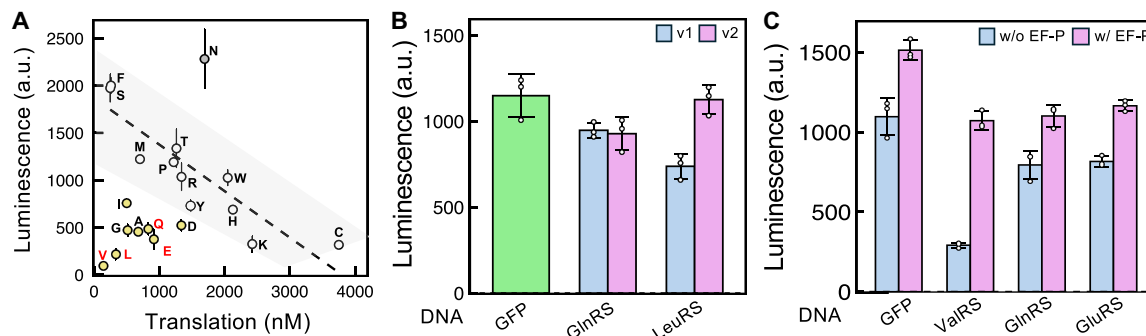
Fig. 4. Inhibitory effect of each AARS-DNA on translation of another gene. (**A**) Relationship between the translations of each AARS (Fig. 1B) and coexpressed luciferase (luminescence). Each AARS-DNA (5 nM) and Fluc-DNA (0.01 nM) were incubated in the PURE system (version 2) at 30°C for 8 hours, and luminescence was measured. Each point represents an AARS. The points that follow a linearly decreasing trend determined by the random sample consensus method are shown in white. Yellow points represent candidates for inhibitory AARS. (**B**) Effect of rare codons in GlnRS-DNA or LeuRS-DNA on the inhibitory effect. One and four rare codons in GlnRS and LeuRS were substituted with the common codons in GlnRS_v2 and LeuRS_v2, respectively. Each AARS-DNA (1 nM) was coexpressed with GFP-DNA (4 nM) and Fluc-DNA (0.01 nM) in the PURE system (version 2) at 30°C for 8 hours, and luminescence was measured. Error bars represent the standard deviation of three independent experiments. (**C**) Effect of EF-P on inhibitory effects. As in (B), each of the three AARS-DNAs was coexpressed with GFP-DNA and Fluc-DNA with or without the addition of EF-P (1 μM), and luminescence was measured. Error bars represent the standard deviations of three independent experiments.
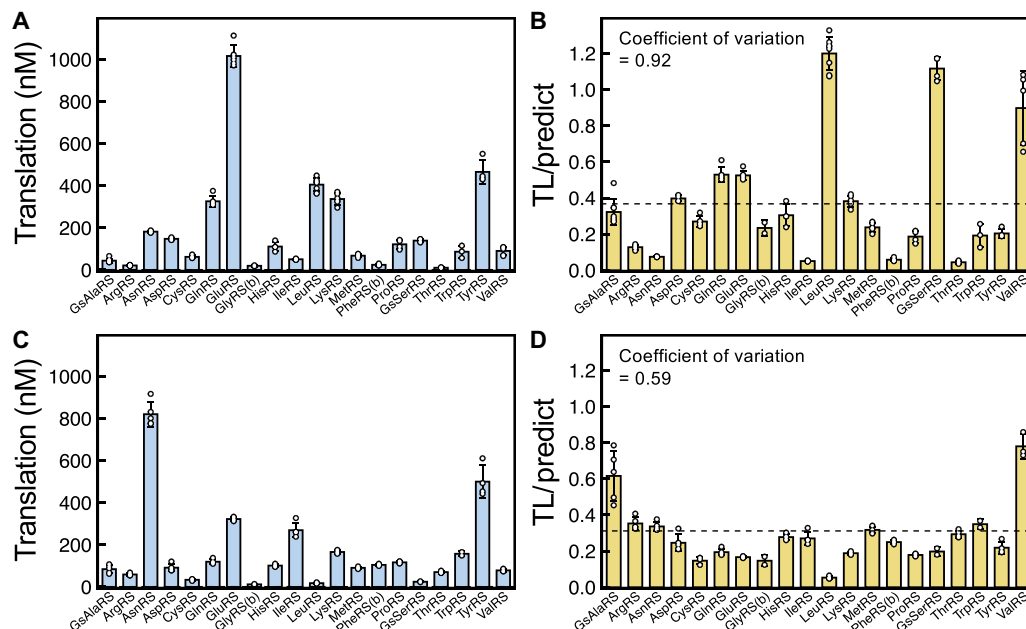

Fig. 5. Quantification and calibration of the coexpressed 20 AARS. All 20 AARS were incubated in a customized PURE system (version 2) containing stable isotope amino acids at 30°C for 8 hours. After LC-MS analysis using the method shown in Fig. 1A(ii), the translation of each AARS-DNA was calculated. (**A** and **B**) Results with the AARS-DNA composition (table S4, left, 5 nM in total), which was determined on the basis of the individually measured translation efficiency in Fig. 1B. The translation amount of each AARS is shown in (A), and the ratio of the translation to the prediction from the individually measured data is shown in (B). (**C** and **D**) Results with the AARS-DNA composition (table S4, right, 2 nM in total), which was calibrated on the basis of the deviation from the prediction shown in (B). The translation amount of each AARS is shown in (C), and the ratio of the translation to the prediction after the calibration is shown in (D). Each point represents an individual detected peptide, with error bars indicating standard deviation. H/L ratio plots are shown in fig. S9. The dotted line represents the average translation prediction.

To calibrate the expression balance, we adjusted the DNA concentration ratios by increasing the DNA for insufficient AARS and decreasing it for sufficient AARS based on the data in Fig. 5B (table S4, right). In addition, to increase the total expression level, we used LeuRS_v2-DNA, which omitted rare codons, and added EF-P to mitigate the inhibitory effects of LeuRS and ValRS-DNA. After this calibration, the translation level (Fig. 5C) and the ratio to the predicted level (Fig. 5D) were measured by the same method. The coefficient of variation (0.59) became smaller than the value before calibration, indicating that the expression balance was improved. Although the average ratio (0.31) still did not approach 1, we decided to perform a regeneration experiment because the expected value was determined on the basis of the expected AARS activities in the original PURE system, which are usually in excess (54).

Serial dilution experiments for 20 AARS regeneration

Last, we attempted to regenerate 20 AARS using all the improvements we had shown above. The experimental procedure is shown in Fig. 6A, which is almost the same as Fig. 3A, but the reaction mixture contains all 20 AARS-DNAs, and the PURE system used for dilution (the composition is shown in table S5) lacks all 20 AARS. First, we performed serial dilution experiments at four total DNA concentrations (2, 3, 5, and 7.5 nM) to determine the optimal total DNA concentration. Translation activity was maintained at a similar level up to 10 rounds at 2 and 3 nM, whereas it gradually declined at 5 and 7.5 nM (fig. S10), indicating that higher DNA concentrations were unsuitable for AARS regeneration. We further conducted serial dilution experiments under four conditions (2 or 3 nM AARS-DNA mixture and 2- or 2.5-fold dilution) for up to 20 rounds and found that the translation activities (luciferase activity) were
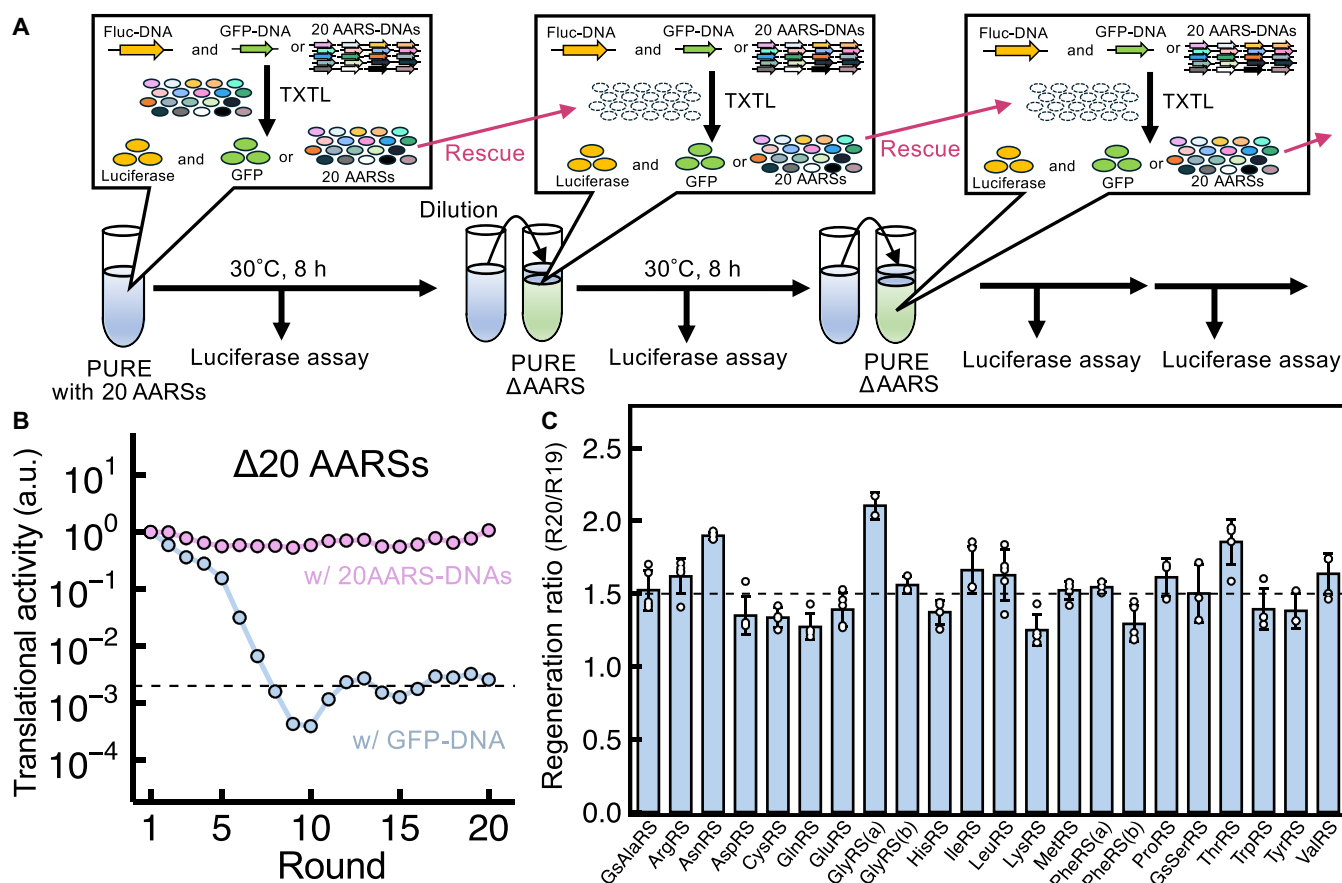
Fig. 6. Serial dilution experiments for 20 AARS regeneration. (**A**) Scheme of the serial dilution experiment. The first reaction mixture contained 20 AARS-DNAs (2 nM in total) and 0.01 nM Fluc-DNA encoding luciferase in the customized PURE system (version 2) containing all 20 AARS proteins. In the control experiment, GFP-DNA (2 nM) was used instead of the AARS-DNA. At round 1, AARS, GFP, and luciferase were expressed. From rounds 2 to 20, the reaction solution of the previous round was diluted 2.5-fold with a customized PURE system that lacked all 20 AARS proteins (PUREΔAARS) but contained 20 AARS-DNA (or GFP-DNA) and Fluc-DNA at the same concentrations. Then, the reaction mixture was incubated. If AARS is sufficiently synthesized in the previous round, the lack of each of the 20 AARS is rescued, and thus, luciferase and AARS are expressed again. Luciferase activity (luminescence) was measured after each round to assess translation activity. In round 20, stable isotope–labeled amino acids were added to the mixture to measure the AARS translation. (**B**) Trajectories of the translation activity. The translation activity (luciferase activity) was normalized on the basis of luminescence after round 1. (**C**) Regeneration ratio of each AARS at round 20. The ratio of isotope-labeled AARS (heavy) to unlabeled AARS (light) was measured using LC-MS. The heavy and light peptide concentrations corresponded to AARS synthesized at round 20 and those synthesized until round 19, respectively. If each AARS concentration is maintained, the ratio should be near 1.5 at this dilution rate. Each point represents an individual detected peptide, with error bars showing standard deviation of them.

maintained at a certain level in the final rounds under all four conditions (Fig. 6B and fig. S11). These data suggested that all 20 AARS were sustainably regenerated during the serial dilution cycle at a dilution rate ($\mu = -0.002$) and a doubling time ($t_d = 347$ min).

As additional control experiments, we performed serial dilution experiments without one of the 20 AARS-DNAs to examine the requirement of each AARS-DNA. We focused on five AARS (AlaRS, CysRS, IleRS, ThrRS, and TrpRS) that were previously suggested to be contaminated with other protein components included in the PURE system (*54*). When one of these five AARS-DNAs was omitted, the translation activity decreased to <10% of the original level after nine rounds of dilution (fig. S12), indicating that contamination did not explain the maintenance of translation activity until 20 rounds with 20 AARS-DNAs (Fig. 6B) at least for these five AARS.

To further verify the sustainable regeneration of 20 AARS, we compared AARS expressed at round 20 to those expressed before round 20. We first replaced the amino acids in PUREΔAARS with stable isotopes at round 20 and quantified the synthesized 20 AARS using LC-MS. We then calculated "the regeneration ratio of each AARS (R20/R19)," which represents the ratio of AARS expressed at round 20 to AARS transferred from round 19, from the H/L ratio by normalizing the isotope ratio. If 20 AARS were expressed at round 20 to the same level as in round 19 (i.e., sustainably regenerated), the regeneration ratio should be 1.5, because the reaction mixture was diluted 2.5-fold in each round (1 for the transferred AARS and 1.5 for the newly synthesized AARS). The regeneration ratios of all 20 AARS were ~1.5 (Fig. 6C), indicating that all 20 AARS were sustainably regenerated even at round 20, where the purified AARS proteins included in the first reaction in round 1 must be diluted to $\sim 10^{-8}$.

In this study, we used six improvements: (i) sequence modification, (ii) zinc addition, (iii) PURE system composition change, (iv) GsAlaRS and GsSerRS utilization, (v) diminishing inhibitory effects, and (vi) adjustment of DNA concentration for coexpression. To evaluate the contribution of each improvement to regeneration, we conducted serial dilution experiments again by restoring each improvement yet under the same condition as in Fig. 6. When restoring (i) sequence modification, the AARS-DNA sequences before modification were used ("ΔSequence modification"). When restoring (ii) zinc addition, zinc acetate was omitted ("ΔAdding Zn"). When restoring (iii) the PURE system composition change, version 1 composition (table S2) was used ("ΔPURE composition improvement"). When restoring (iv) GsAlaRS and GsSerRS utilization, *E. coli* AlaRS and SerRS were used instead of GsAlaRS and GsSerRS ("ΔGsAlaRS, GsSerRS"). When restoring (v) diminishing inhibitory effects, the original version of LeuRS-DNA was used instead of version 2, and EF-P was removed ("ΔMitigation of inhibitory AARS"). When restoring (vi) adjustment of DNA concentration for coexpression, the DNA concentration ratio before calibration (table S4) was used ("ΔCalibration of DNA conc."). We observed an immediate decline in translation activity when any one of the five improvements was restored, except for (ii) zinc addition (Fig. 7). These results demonstrate that all five improvements, excluding zinc addition, were crucial for the sustainable regeneration of 20 AARS.

DISCUSSION

Self-reproduction is one of the most unique functions of living organisms, and its artificial realization is a significant challenge in bottom-up synthetic biology (*2*, *14*, *15*, *17*). To advance this challenge, we focused on 20 AARS, which comprise approximately half of the TPs included in the PURE system, except for ribosomes. Unlike previous studies, we measured the activities per protein of all AARS expressed in the PURE system, together with translation, and systematically improved both translation efficiencies and activities
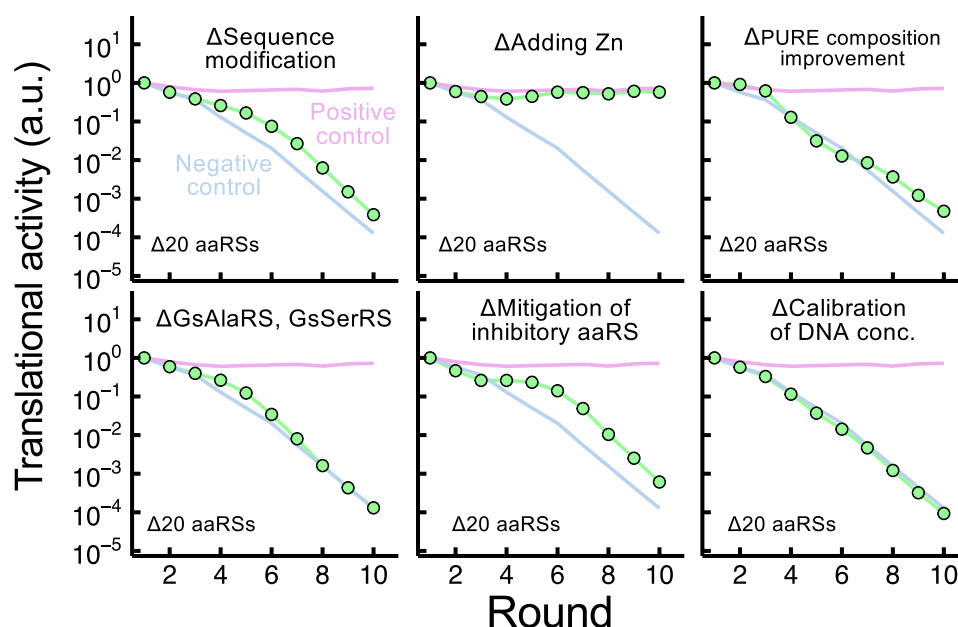
Fig. 7. Contribution of each improvement to 20 AARS regeneration. Serial dilution experiments, shown in Fig. 6A, were performed under conditions that reverted each of the six improvements. The trajectories of the translation activity are shown (green points). The red line indicates the condition in which all improvements were applied (positive control, a new experiment under the same condition as Fig. 6B), and the blue line indicates the condition in which GFP-DNA was used instead of 20 AARS-DNAs (negative control).

of AARS with insufficient activities using six methodologies: (i) sequence modification, (ii) zinc addition, (iii) PURE system composition change, (iv) GsAlaRS and GsSerRS utilization, (v) diminishing inhibitory effects, and (vi) calibration of the DNA concentration ratio. Because of these improvements, we succeeded in the sustainable regeneration of all 20 AARS twice under a condition (2 nM DNA and 2.5-fold dilution) for up to 20 rounds (Fig. 6) and up to 10 rounds (Fig. 7, positive control). We also succeeded in the sustainable regeneration under other three conditions (2 or 3 nM DNA and 2- or 2.5-fold dilution, fig. S11). These results represent a significant advancement toward the realization of self-synthesizing artificial systems. In addition, the methodologies and knowledge obtained in this study will be useful for further increasing the regeneratable proteins in the future.

One of the new strategies used in this study is to adjust the expression balance for 20 AARS based on translation efficiency as well as the activity per protein of each AARS. Our previous approach relied solely on translation efficiency and endogenous AARS concentration to determine DNA ratios, resulting in an insufficient regeneration performance (43). In this study, we introduced a more refined strategy by integrating activities per protein into the calculations. Accordingly, we found that some AARS (e.g., GlyRS, IleRS, and MetRS) were more active when expressed in the PURE system than the purified proteins, which means that the required expression level of AARS for regeneration is lower than previously thought. Another important insight obtained is that the DNA concentrations adjusted on the basis of the individually measured data provided discrepancies between the predicted and actual translation amounts (Fig. 5B), probably due to the transcription/translation competition between genes, and calibration with the coexpression data was required for balanced expression. Even under the final condition, some AARS expression levels (e.g., LeuRS) were still lower than predicted (Fig. 5D). Understanding the reason for this discrepancy will allow for more precise control of gene expression in the PURE system.

A new finding of this study was the highly active AlaRS and SerRS from *G. stearothermophilus* when expressed in the PURE system. The increases in the translation activity were approximately four- and sixfold for AlaRS (Fig. 2C, no. 4) and SerRS (Fig. 2D, SerRS), respectively. In general, more active translation factors require fewer translation resources to achieve the same level of activity and are thus useful for self-regeneration. Although the exact mechanism by which these two AARS exhibit high activity in the PURE system remains unclear, the approach of selecting highly active translation factors from other bacteria can be a useful strategy to improve translation factors.

Another finding is the translation inhibition caused by some AARS-DNAs, such as LeuRS, ValRS, GlnRS, and GluRS (Fig. 4A). Considering that the inhibitory effects were mitigated by removing rare codons or adding EF-P, a possible mechanism of the inhibitory effect is ribosome stalling, which decreases the availability of ribosomes. Several ribosome-stalling sequences, other than rare codons or polyproline, have been reported (55, 56). These stalling sequences may have caused the remaining inhibitory effects of GlnRS and GluRS. This problem may be serious when the number of coexpressed genes is increased because the inhibitory effect is expected to accumulate. Further identification of inhibitory sequences is required to realize self-reproducible systems. In addition, knowledge of ribosome stalling sequences will contribute to a deeper understanding of translational mechanisms in natural living systems.

Although this study achieved stable regeneration of 20 AARS, some challenges remain. The first is the regeneration of additional genes, including other translational proteins, ribosomes, and tRNAs. The major obstacle is the translation capacity of the current PURE system, which can express up to 3.8 mg/ml of proteins under dialysis conditions (49), which is still lower than the protein concentration in our customized PURE system (7.1 mg/ml, PURE version 7). However, the activities per protein of many of the newly expressed AARS in the PURE system were higher than those of purified AARS (Fig. 1D). It is also known that proteins in the PURE system are usually in excess, as shown by Ganesh and Maerkl (54), and that most translation factors can be coexpressed in the PURE system at levels higher than those of endogenous proteins, as shown by Doerr et al. (40). On the basis of these results, we think that the challenge of whole PURE system regeneration may be more easily overcome than expected. Another challenge for self-reproduction is coupling with DNA replication. As shown previously, the buffer composition optimized for translation is significantly different from that optimized for DNA replication (57). Further studies are needed to develop a system that combines high translational activity with DNA replication.

MATERIALS AND METHODS
DNA preparation
The DNA encoding each AARS was prepared as follows. First, we polymerase chain reaction (PCR) amplified the DNA fragments encoding each AARS using primers 1 (GCGAAATTAATACGACT-CACTATAGGG) and 2 (CCGCTGAGCAATAACTAGCATAACC) and the plasmid (pET-AARS) as templates, prepared in our previous study (43). The PCR product was purified using the QIAquick PCR Purification Kit (QIAGEN), which was used for all DNA purification procedures in this study. Nine sequence-modified pET-AARS were prepared by PCR amplification followed by self-ligation using a mutated and In-Fusion cloning kit (TaKaRa). The circular DNA encoding the phi29 DNA polymerase used in the AARS activity assay (fig. S3) has been described previously (20, 43). The plasmid that encodes GFP under the T7 promoter (pET-GFP) was previously constructed as pETG5tag (58). A linear DNA fragment encoding GFP-DNA was prepared by PCR using pET-GFP as a template and primers 1 and 2. A plasmid encoding firefly luciferase under the T7 promoter (Fluc) was previously constructed (59). A linear DNA fragment encoding firefly luciferase (Fluc-DNA) was prepared by PCR using pET-Fluc as a template and primers 3 (TTCTCTGCCCAATACGCAAAC) and 4 (GAGCAGACAAGCCC-GTCAG). Linear DNAs encoding AARS under the T7 promoter derived from other bacteria were prepared as follows. First, the amino acid sequences of the AARS genes of the 10 bacteria were obtained from the National Center for Biotechnology Information database. After optimizing the nucleotide sequence for *E. coli* translation using three company sites or algorithms [Eurofin (Fig. 2, B and C), Twist Bio Sciences (Fig. 2D), and CodHonEditor (60) (Figs. 3 to 7)], the gene sequence and upstream T7 promoter were synthesized and inserted into the pET vector using Twist Bio Science's artificial gene synthesis service. The linear DNA fragment encoding each AARS was prepared by PCR using each pET-AARS as a template and primers 1 and 2. All plasmid sequences used in this study are shown in the Supplementary Materials.

PURE system preparation

The customized PURE system used in this study was prepared from independently purified components in our laboratory. Protein purification was conducted using affinity chromatography with a histidine tag, followed by gel-filtration chromatography as described previously (*61*). The compositions of the versions 1 and 2 are listed in table S2. The final compositions used in Fig. 6 are listed in table S5. The total tRNA mixture (*E. coli* MRE 600) was purchased from Roche.

Stable isotope labeling and quantification of PURE-expressed AARS

To measure the translation efficiency, AARS-DNA (5 nM) was incubated at 30°C for 8 hours in a customized PURE system in which Lys and Arg were replaced with stable isotope amino acids (0.36 mM $^{13}C_6$ L-lysine-2HCl and 0.36 mM $^{13}C_6$, $^{15}N_4$ L-arginine-HCl) (Thermo Fisher Scientific) to express AARS. Then, 10 μl of the reaction mixture was used to purify the proteins contained in the PURE system using a His-tag spin column (New England Biolabs) according to the manufacturer's instructions.

The eluate was concentrated by methanol-chloroform precipitation and digested according to a phase-transfer surfactant-aided protocol (*62*). The pellets were dissolved in 20 μl of buffer (10 mM sodium deoxycholate, 10 mM sodium N-lauroylsarcosinate, and 50 mM NH_4HCO_3), reduced with 10 mM tris(2-carboxyethyl)phosphine at 37°C for 30 min, alkylated with 20 mM iodoacetamide at 37°C for 30 min, quenched with 20 mM L-cysteine, and digested with 100 ng of Lys-C/Trypsin Protease Mix (Thermo Fisher Scientific) at 37°C overnight. After digestion, the detergents were precipitated by adding final 1% trifluoroacetic acid (TFA), followed by centrifugation at 15,000*g* for 5 min at 4°C. The supernatant was desalted using GL-Tip SDB (GL Sciences) according to the manufacturer's instructions and dried under reduced pressure.

For the AARS sample at round 20 (Fig. 6C), we used GL-Tip SDB-SCX (GL Sciences) instead of GL-Tip SDB for further fractionation to improve detection efficiency (*63*). Briefly, peptides were concentrated on SDB-SCX and then eluted stepwise in the following buffers: 0.1% TFA, 0.1% TFA, 80% acetonitrile, 0.5% TFA, 30% acetonitrile, 1% TFA, 30% acetonitrile, 3% TFA, 30% acetonitrile, 3% TFA, 30% acetonitrile, 100 mM ammonium acetate, 4% TFA, 30% acetonitrile, 500 mM ammonium acetate, 30% acetonitrile, and 500 mM ammonium acetate. The eluates were dried under reduced pressure.

Dried peptides were dissolved in 0.1% TFA. The peptides were concentrated and separated using a nano-LC system (UltiMate 3000, Thermo Fisher Scientific, Bremen, Germany) equipped with a trap column (C18, 0.075 by 20 mm, 3 μm, Acclaim PepMap 100, Thermo Fisher Scientific) and nanocapillary analytical column (C18, 0.075 by 150 mm, 3 μm, Nikkyo Technos, Tokyo, Japan) at a flow rate of 300 nl/min. Mobile phases A (0.1% formic acid) and B (acetonitrile and 0.1% formic acid) were combined in the following gradient: 5% B for 5 min, 5 to 40% B for 75 min, 40 to 90% B for 1 min, and 90% B for 4 min. MS analysis was performed using an Orbitrap mass spectrometer (Q Exactive, Thermo Fisher Scientific) equipped with a nanospray ion source (Nanospray Flex, Thermo Fisher Scientific) with the following parameters: spray voltage, 2.2 kV, positive mode; scan range, *m/z* 310 to 2000; and 70,000 resolution. The 10 most intense multiply charged ions (*z* = 2 to 4) were fragmented in the collision cell by high-energy collisional dissociation with a normalized collision energy of 30%. Calculation of peak areas and H/L ratio was performed using Skyline software (MacCoss group at the University of Washington, WA).

In the quantification of AARS regenerated in round 20 in Fig. 6C, we assumed that 40% of the regenerated AARS were not labeled with stable isotope amino acids because of the influence of the normal amino acids carried over from round 19, and the AARS regenerated at round 20 were calculated by dividing the H/L ratio by 0.6.

Activity assay of the AARS synthesized in the PURE system

First, the DNA encoding each AARS (5 nM) was incubated at 30°C for 8 hours using a customized PURE system (version 1), in which Lys and Arg were replaced with stable isotope amino acids (Thermo Fisher Scientific) to express AARS. Thereafter, an aliquot of the reaction solution was diluted with dilution buffer [50 mM Hepes-KOH (pH 7.6), 100 mM KCl, 10 mM $MgCl_2$, 7 mM 2-mercaptoethanol, bovine serum albumin (5 mg/ml), 10 mM dithiothreitol, and 30% glycerol] at different rates and added to the second customized PURE system (version 1), which lacks AARS of interest and contains GFP-DNA (3 nM). For AsnRS, CysRS, GluRS, HisRS, and TyrRS, Fluc-DNA (5 nM) was included as a translation load in the 2nd PURE system. The mixture was then incubated at 30°C for 12 hours. GFP fluorescence was measured every 15 min for up to 4 hours and every 30 min thereafter (Mx3005P; Agilent Technologies). The slope of the GFP fluorescence from 0 to 4 hours was evaluated as the GFP synthesis rate. The volume ratio of 1st PURE to 2nd PURE system was then plotted against the GFP synthesis rate, and the activities $k_{exp+end}$ and k_{end} were calculated using linear regression. For TrpRS, the DNA replication rate was measured instead of the GFP synthesis rate, as previously described (*43*).

The ratio of the activity per protein of expressed AARS to that of endogenous AARS (A_{exp}/A_{end}) was calculated using the two obtained activities ($k_{exp+end}$ and k_{end} obtained above) according to the following equation (Eq. 1)

$$\frac{A_{exp}}{A_{end}} = \frac{k_{exp+end} - k_{end}}{k_{end} \times (\text{H/L ratio})} \quad (1)$$

The standard error of the relative activity and A_{exp}/A_{end} was calculated on the basis of the error propagation theory according to a previous report (*64*).

SDS–polyacrylamide gel electrophoresis for quantifying AARS expression amount

DNA (5 nM) encoding each AARS was incubated at 30°C for 8 hours in a customized PURE system (version 1) containing fluorescently labeled lysyl-tRNA (FluoroTect GreenLys, Promega). After expression, an aliquot (5 μl) was treated with 0.5 μl of ribonuclease A (5 mg/ml; QIAGEN, Hilden, Germany) at 37°C for 30 min, incubated at 95°C for 5 min in SDS sample buffer [17 mM tris-HCl (pH 7.4), 0.7% SDS, 0.3 M 2-mercaptoethanol, and 3% glycerol], and subjected to 10% SDS–polyacrylamide gel electrophoresis (SDS-PAGE). The synthesized fluorescently labeled proteins were detected using FUSION-SL4 (Vilber-Lourmat), and band intensities were analyzed using ImageJ.

Activity assay of AARS derived from other bacteria

As shown in Fig. 1, each AARS was expressed in the 1st PURE system, diluted with buffer, and added to the 2nd PURE system lacking the AARS of interest to monitor the GFP expression. GFP fluorescence was measured every 15 min up to 4 hours and every 30 min thereafter (Mx3005P, Agilent Technologies). The slope of GFP fluorescence from 0 to 4 hours was evaluated as the activity. To quantify the background activity of endogenous AARS, we used the PURE system without incubating any DNA, and the activity of each AARS was evaluated by subtracting the background activity from the quantified activity of AARS.

Expression and purification of AARS derived from *G. stearothermophilus*

Expression plasmids encoding GsAlaRS (pET-GsAlaRS) and GsSerRS (pET-GsSerRS_v2) with histidine tags at the N terminus and C terminus, respectively, were synthesized by Twist Biotech (sequences are shown in the Supplementary Materials). The plasmids were introduced into Rosetta (DE3) pLysS (Novagen) and cultured at 37°C with shaking at 120 rpm in 1 liter of LB medium. In the late log phase, isopropyl β-D-thiogalactopyranoside (final concentration of 1 mM) was added and further incubated for 3 hours. After harvesting, each AARS was purified by affinity chromatography with a histidine tag, followed by gel-filtration chromatography using the same method as that used for the other protein factors of the PURE system, as described above (*61*). The purified *G. stearothermophilus*–derived AARS band was subjected to SDS-PAGE and stained with Coomassie brilliant blue (Fig. 2E).

Activity per protein measurements of the purified GsAARS

Purified AlaRS or SerRS derived from *E. coli* and *G. stearothermophilus* was diluted with dilution buffer [50 mM Hepes-KOH (pH 7.6), 100 mM KCl, 10 mM MgCl_2, 7 mM 2-mercaptoethanol, bovine serum albumin (5 mg/ml), 10 mM dithiothreitol, and 30% glycerol] at some different rates and added to the customized PURE system that lacks AlaRS or SerRS and contains GFP-DNA (3 nM). The mixture was incubated at 30°C for 12 hours. GFP fluorescence was measured every 15 min for up to 4 hours and every 30 min thereafter (Mx3005P; Agilent Technologies). The slope of the GFP fluorescence from 0 to 4 hours was evaluated for each AARS activity. The concentration of each AARS was plotted against AARS activity, and the activity per protein was calculated using linear regression.

Serial dilution experiment

In round 1, a single or multiple DNA encoding each AARS (AARS-DNA) and firefly luciferase (Fluc-DNA) was incubated in the customized PURE system (version 2) containing all 20 AARS proteins (the complete composition is shown in table S2) at 30°C for 8 hours. The total DNA concentrations were 5 nM (Fig. 3), 2 nM (Fig. 6), 2 to 7.5 nM (fig. S10), and 2 to 3 nM (fig. S11). The concentration ratio of each AARS-DNA is presented in tables S3 and S4. In the subsequent rounds, the reaction solution after incubation in the previous round was diluted 2-fold, 2.5-fold (Figs. 6 and 7 and figs. S10 and S11), or 5-fold (Fig. 3) with another customized PURE system (version 2, PUREΔAARS) that lacks AARS of interest and contains AARS-DNA and Fluc-DNA at the same concentrations as round 1 and incubated again at 30°C for 8 hours. The serial dilution process was repeated during the indicated rounds. At the end of each round,

1 μl of an aliquot of the reaction mixture was added to 30 μl of the luciferase assay reagent (Promega), and luminescence was measured as translational activity using a luminometer (GloMax, Promega). For negative control experiments, GFP-DNA was used instead of AARS-DNA. In addition, as a positive control, we conducted the same serial dilution experiment using GFP-DNA and a customized PURE system (version 2) containing 20 AARS proteins (Full PURE in Fig. 3).

Assay of the inhibitory effect of AARS-DNA on translation

To examine the inhibitory effect (Fig. 4A), AARS-DNA (5 nM) and Fluc-DNA (0.01 nM) were incubated at 30°C for 8 hours in a customized PURE system (version 2) containing all AARS proteins. For the assay in Fig. 4 (B and C), AARS-DNA (1 nM), GFP-DNA (4 nM), and Fluc-DNA (0.01 nM) were incubated in the PURE system (version 2). As controls, GFP-DNA (5 nM) and Fluc-DNA (0.01 nM) were incubated in the PURE system (version 2). EF-P (1 μM, GeneFrontier) was added in Fig. 4C. After incubation, 1 μl of an aliquot of the reaction mixture was added to 30 μl of the luciferase assay reagent (Promega), and luminescence was measured as translational activity using a luminometer (GloMax, Promega).

Determination of DNA concentration ratio

The concentration ratio of each AARS in the AARS-DNA mix used in serial dilution experiments was determined as follows. First, we defined a requirement index, R, which reflects the requirement of each AARS-DNA for regeneration. The requirement index should be proportional to each concentration and the activity per protein of the endogenous AARS included in the PURE system but inversely proportional to each concentration and the activity per protein of the AARS expressed in the PURE system. Therefore, R can be defined using the H/L ratio (concentration ratio of expressed to endogenous AARS) and activity per protein (A_{exp}/A_{end}) obtained in Fig. 1 as follows (Eq. 2)

$$R = \frac{A_{end}}{A_{exp}(\text{H/L ratio})} \tag{2}$$

The concentration of each DNA was determined as follows (Eq. 3). The total DNA concentration differed depending on the experimental conditions used.

$$[\text{AARS} - \text{DNA}_i] = [\text{Total DNA}] \times \frac{R_i}{\sum\limits_{i}^{20} R_i} \tag{3}$$

In Figs. 5C and 6, AARS-DNA concentrations were calibrated by multiplying a calibration value (C_i) as follows (Eq. 4)

$$[\text{AARS} - \text{DNA}_i] = [\text{Total DNA}] \times \frac{R_i}{\sum\limits_{i}^{20} R_i} \times C_i \tag{4}$$

The calibration value (C_i) was applied to correct discrepancies between the predicted and experimental translation levels as shown in Fig. 5B. Each calibration value is equal to the average of the *y*-axis values (dotted line) in Fig. 5B divided by each *y*-axis value. After each

DNA concentration ([AARS-DNA$_i$]) was calculated, the DNA concentrations were normalized to ensure that the sum of the calculated DNA concentrations was equal to the target total DNA concentration.

Prediction of the translation amount in coexpressing 20 AARS
The translation amount was predicted as follows on the basis of the translation efficiency (T) measured in Fig. 1B, the DNA concentration ratio, and the activity improvement rate of the PURE system obtained in fig. S4C (14.4) (Eq. 5).

$$\left[\mathrm{AARS}_i\right] = T \times \frac{\left[\mathrm{AARS-DNA}_i\right]}{\left[\mathrm{Total\ DNA}\right]} \times 14.4 \tag{5}$$

Supplementary Materials
The PDF file includes:
Figs. S1 to S12
Tables S1 to S5
Legend for data S1
Supplementary Methods
References

Other Supplementary Material for this manuscript includes the following:
Data S1

REFERENCES AND NOTES

1. J. W. Szostak, D. P. Bartel, P. L. Luisi, Synthesizing life. *Nature* **409**, 387–390 (2001).
2. A. C. Forster, G. M. Church, Towards synthesis of a minimal cell. *Mol. Syst. Biol.* **2**, 45 (2006).
3. N. Ichihashi, What can we learn from the construction of in vitro replication systems? *Ann. N. Y. Acad. Sci.* **1447**, 144–156 (2019).
4. M. Forlin, R. Lentini, S. S. Mansy, Cellular imitations. *Curr. Opin. Chem. Biol.* **16**, 586–592 (2012).
5. B. C. Buddingh', J. C. M. van Hest, Artificial cells: Synthetic compartments with life-like functionality and adaptivity. *Acc. Chem. Res.* **50**, 769–777 (2017).
6. N. A. Yewdall, A. F. Mason, J. C. M. Van Hest, The hallmarks of living systems: Towards creating artificial cells. *Interface Focus* **8**, 20180023 (2018).
7. A. D. Silverman, A. S. Karim, M. C. Jewett, Cell-free gene expression: An expanded repertoire of applications. *Nat. Rev. Genet.* **21**, 151–170 (2020).
8. V. Noireaux, A. P. Liu, The new age of cell-free biology. *Annu. Rev. Biomed. Eng.* **22**, 51–77 (2020).
9. L. Damiano, P. Stano, On the "life-likeness" of synthetic cells. *Front. Bioeng. Biotechnol.* **8**, 953 (2020).
10. E. Cho, Y. Lu, Compartmentalizing cell-free systems: Toward creating life-like artificial cells and beyond. *ACS Synth. Biol.* **9**, 2881–2901 (2020).
11. I. Ivanov, S. L. Castellanos, S. Balasbas, L. Otrin, N. Maruscaroniccaron, T. Vidakovicacute-Koch, K. Sundmacher, Bottom-up synthesis of artificial cells: Recent highlights and future challenges. *Annu. Rev. Chem. Biomol. Eng.* **12**, 287–308 (2021).
12. C. Wang, J. Yang, Y. Lu, Modularize and unite: Toward creating a functional artificial cell. *Front. Mol. Biosci.* **8**, 781986 (2021).
13. Y. Lyu, R. Peng, H. Liu, H. Kuai, L. Mo, D. Han, J. Li, W. Tan, Protocells programmed through artificial reaction networks. *Chem. Sci.* **11**, 631–642 (2020).
14. L. Olivi, M. Berger, R. N. P. Creyghton, N. De Franceschi, C. Dekker, B. M. Mulder, N. J. Claassens, P. R. ten Wolde, J. van der Oost, Towards a synthetic cell cycle. *Nat. Commun.* **12**, 4531 (2015).
15. J. C. Blain, J. W. Szostak, Progress toward synthetic cells. *Annu. Rev. Biochem.* **83**, 615–640 (2014).
16. N. J. Gaut, K. P. Adamala, Reconstituting natural cell elements in synthetic cells. *Adv. Biol.* **5**, e2000188 (2021).
17. J. De Capitani, H. Mutschler, The long road to a synthetic self-replicating central dogma. *Biochemistry* **62**, 1221–1232 (2023).
18. M. Su'etsugu, H. Takada, T. Katayama, H. Tsujimoto, Exponential propagation of large circular DNA by reconstitution of a chromosome-replication cycle. *Nucleic Acids Res.* **45**, 11525–11534 (2017).
19. P. van Nies, I. Westerlaken, D. Blanken, M. Salas, M. Mencía, C. Danelon, Self-replication of DNA by its encoded proteins in liposome-based synthetic cells. *Nat. Commun.* **9**, 1583 (2018).
20. Y. Sakatani, T. Yomo, N. Ichihashi, Self-replication of circular DNA by a self-encoded DNA polymerase through rolling-circle replication and recombination. *Sci. Rep.* **8**, 13089 (2018).
21. H. Okauchi, Y. Sakatani, K. Otsuka, N. Ichihashi, Minimization of elements for isothermal DNA replication by an evolutionary approach. *ACS Synth. Biol.* **9**, 1771–1780 (2020).
22. S. Berhanu, T. Ueda, Y. Kuruma, Artificial photosynthetic cell producing energy for protein synthesis. *Nat. Commun.* **10**, 1325 (2019).
23. K. Y. Lee, S.-J. Park, K. A. Lee, S.-H. Kim, H. Kim, Y. Meroz, L. Mahadevan, K.-H. Jung, T. K. Ahn, K. K. Parker, K. Shin, Photosynthetic artificial organelles sustain and control ATP-dependent reactions in a protocellular system. *Nat. Biotechnol.* **36**, 530–535 (2018).
24. S. Giaveri, N. Bohra, C. Diehl, H. Y. Yang, M. Ballinger, N. Paczia, T. Glatter, T. J. Erb, Integrated translation and metabolism in a partially self-synthesizing biochemical network. *Science* **385**, 174–178 (2024).
25. S. Sugii, K. Hagino, R. Mizuuchi, N. Ichihashi, Cell-free expression of RuBisCO for ATP production in the synthetic cells. *Synth. Biol.* **8**, ysad016 (2023).
26. D. Blanken, D. Foschepoth, A. C. Serrão, C. Danelon, Genetically controlled membrane synthesis in liposomes. *Nat. Commun.* **11**, 4317 (2020).
27. S. Eto, R. Matsumura, Y. Shimane, M. Fujimi, S. Berhanu, T. Kasama, Y. Kuruma, Phospholipid synthesis inside phospholipid membrane vesicles. *Commun. Biol.* **5**, 1016 (2022).
28. S. Kohyama, A. Merino-Salomón, P. Schwille, In vitro assembly, positioning and contraction of a division ring in minimal cells. *Nat. Commun.* **13**, 6098 (2022).
29. G. Sato, S. Miyazawa, N. Doi, K. Fujiwara, Cell-free protein expression by a reconstituted transcription–translation system energized by sugar catabolism. *Molecules* **29**, 2956 (2024).
30. E. Karzbrun, A. M. Tayar, V. Noireaux, R. H. Bar-Ziv, Programmable on-chip DNA compartments as artificial cells. *Science* **345**, 829–832 (2014).
31. M. A. Boyd, W. Thavarajah, J. B. Lucks, N. P. Kamat, Robust and tunable performance of a cell-free biosensor encapsulated in lipid vesicles. *Sci. Adv.* **9**, eadd6605 (2024).
32. Z.-H. Luo, C. Chen, Q.-H. Zhao, N.-N. Deng, Functional metal-phenolic cortical cytoskeleton for artificial cells. *Sci. Adv.* **10**, eadj4047 (2024).
33. S. Takada, N. Yoshinaga, N. Doi, K. Fujiwara, Mode selection mechanism in traveling and standing waves revealed by Min wave reconstituted in artificial cells. *Sci. Adv.* **8**, eabm8460 (2024).
34. M. Levy, R. Falkovich, S. S. Daube, R. H. Bar-Ziv, Autonomous synthesis and assembly of a ribosomal subunit on a chip. *Sci. Adv.* **6**, eaaz6020 (2024).
35. Y. Shimizu, A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa, T. Ueda, Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19**, 751–755 (2001).
36. T. Awai, N. Ichihashi, T. Yomo, Activities of 20 aminoacyl-tRNA synthetases expressed in a reconstituted translation system in *Escherichia coli*. *Biochem. Biophys. Rep.* **3**, 140–143 (2015).
37. E. Wei, D. Endy, Experimental tests of functional molecular regeneration via a standard framework for coordinating synthetic cell building. bioRxiv 433818 [Preprint] (2021). https://doi.org/10.1101/2021.03.03.433818.
38. J. Li, W. Haas, K. Jackson, E. Kuru, M. C. Jewett, Z. H. Fan, S. Gygi, G. M. Church, Cogenerating synthetic parts toward a self-replicating system. *ACS Synth. Biol.* **6**, 1327–1336 (2017).
39. K. Libicher, R. Hornberger, M. Heymann, H. Mutschler, In vitro self-replication and multicistronic expression of large synthetic genomes. *Nat. Commun.* **11**, 904 (2020).
40. A. Doerr, D. Foschepoth, A. C. Forster, C. Danelon, In vitro synthesis of 32 translation-factor proteins from a single template reveals impaired ribosomal processivity. *Sci. Rep.* **11**, 1898 (2021).
41. K. Libicher, H. Mutschler, Probing self-regeneration of essential protein factors required for in vitro translation activity by serial transfer. *Chem. Commun.* **56**, 15426–15429 (2020).
42. B. Lavickova, N. Laohakunakorn, S. J. Maerkl, A partially self-regenerating synthetic cell. *Nat. Commun.* **11**, 6340 (2020).
43. K. Hagino, N. Ichihashi, In vitro transcription/translation-coupled DNA replication through partial regeneration of 20 aminoacyl-tRNA synthetases. *ACS Synth. Biol.* **12**, 1252–1263 (2023).
44. Y. Shimizu, T. Kanamori, T. Ueda, Protein synthesis by pure translation systems. *Methods* **36**, 299–304 (2005).
45. P. O. Olins, C. S. Devine, S. H. Rangwala, K. S. Kavka, The T7 phage gene 10 leader RNA, a ribosome-binding site that dramatically enhances the expression of foreign genes in *Escherichia coli*. *Gene* **73**, 227–235 (1988).
46. T. Fuse-Murakami, R. Matsumoto, T. Kanamori, N-terminal amino acid affects the translation efficiency at lower temperatures in a reconstituted protein synthesis system. *Int. J. Mol. Sci.* **25**, 5264 (2024).
47. C.-M. Zhang, T. Christian, K. J. Newberry, J. J. Perona, Y.-M. Hou, Zinc-mediated amino acid discrimination in cysteinyl-tRNA synthetase. *J. Mol. Biol.* **327**, 911–917 (2003).
48. J. Liu, Y. Gagnon, J. Gauthier, L. Furenlid, P.-J. L'Heureux, M. Auger, O. Nureki, S. Yokoyama, J. Lapointe, The zinc-binding site of *Escherichia coli* glutamyl-tRNA synthetase is located in the acceptor-binding domain. *J. Biol. Chem.* **270**, 15162–15169 (1995).

49. Y. Kazuta, T. Matsuura, N. Ichihashi, T. Yomo, Synthesis of milligram quantities of proteins using a reconstituted in vitro protein synthesis system. *J. Biosci. Bioeng.* **118**, 554–557 (2014).

50. J. S. Madin, D. A. Nielsen, M. Brbic, R. Corkrey, D. Danko, K. Edwards, M. K. M. Engqvist, N. Fierer, J. L. Geoghegan, M. Gillings, N. C. Kyrpides, E. Litchman, C. E. Mason, L. Moore, S. L. Nielsen, I. T. Paulsen, N. D. Price, T. B. K. Reddy, M. A. Richards, E. P. C. Rocha, T. M. Schmidt, H. Shaaban, M. Shukla, F. Supek, S. G. Tetu, S. Vieira-Silva, A. R. Wattam, D. A. Westfall, M. Westoby, A synthesis of bacterial and archaeal phenotypic trait data. *Sci. Data* **7**, 170 (2020).

51. R. Krafczyk, F. Qi, A. Sieber, J. Mehler, K. Jung, D. Frishman, J. Lassak, Proline codon pair selection determines ribosome pausing strength and translation efficiency in bacteria. *Commun. Biol.* **4**, 589 (2021).

52. K. R. Hummels, D. B. Kearns, Translation elongation factor P (EF-P). *FEMS Microbiol. Rev.* **44**, 208–218 (2020).

53. S. Ude, J. Lassak, A. L. Starosta, T. Kraxenberger, D. N. Wilson, K. Jung, Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. *Science* **339**, 82–85 (2013).

54. R. B. Ganesh, S. J. Maerkl, Towards self-regeneration: Exploring the limits of protein synthesis in the protein synthesis using recombinant elements (PURE) cell-free transcription–translation system. *ACS Synth. Biol.* **13**, 2555–2566 (2024).

55. M. C. J. Yip, S. Shao, Detecting and rescuing stalled ribosomes. *Trends Biochem. Sci.* **46**, 731–743 (2021).

56. T. Inada, R. Beckmann, Mechanisms of translation-coupled quality control. *J. Mol. Biol.* **436**, 168496 (2024).

57. K. Seo, N. Ichihashi, Investigation of compatibility between DNA replication, transcription, and translation for in vitro central dogma. *ACS Synth. Biol.* **12**, 1813–1822 (2023).

58. Y. Ito, T. Kawama, I. Urabe, T. Yomo, Evolution of an arbitrary sequence in solubility. *J. Mol. Evol.* **58**, 196–202 (2004).

59. Y. Murase, H. Nakanishi, G. Tsuji, T. Sunami, N. Ichihashi, In vitro evolution of unmodified 16S rRNA for simple ribosome reconstitution. *ACS Synth. Biol.* **7**, 576–583 (2018).

60. K. Takai, CodHonEditor: Spreadsheets for codon optimization and editing of protein coding sequences. *Nucleosides Nucleotides Nucleic Acids* **35**, 223–232 (2016).

61. S. Fujii, T. Matsuura, T. Sunami, T. Nishikawa, Y. Kazuta, T. Yomo, Liposome display for in vitro selection and evolution of membrane proteins. *Nat. Protoc.* **9**, 1578–1591 (2014).

62. T. Masuda, M. Tomita, Y. Ishihama, Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *J. Proteome Res.* **7**, 731–740 (2008).

63. J. Adachi, K. Hashiguchi, M. Nagano, M. Sato, A. Sato, K. Fukamizu, Y. Ishihama, T. Tomonaga, Improved proteome and phosphoproteome analysis on a cation exchanger by a combined acid and salt gradient. *Anal. Chem.* **88**, 7899–7903 (2016).

64. P. Fornasini, *The Uncertainty in Physical Measurements* (Springer New York, 2008).

65. T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. Datsenko, M. Tomita, B. L. Wanner, H. Mori, Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).

66. E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. Wilkins, R. Appel, A. Bairoch, "Protein identification and analysis tools on the Expasy server," in *The Proteomics Protocols Handbook*, J. M. Walker, Ed. (Humana Press, 2005), pp. 571–607.