ACS Synthetic Biology

Subscriber access provided by HARVARD UNIVERSITY

Article

Co-generating synthetic parts toward a self-replicating system

Jun Li, Wilhelm Haas, Kirsten Jackson, Erkin Kuru, Michael C. Jewett, Z. Hugh Fan, Steven P. Gygi, and George M. Church

ACS Synth. Biol., **Just Accepted Manuscript** • DOI: 10.1021/acssynbio.6b00342 • Publication Date (Web): 23 Mar 2017

Downloaded from http://pubs.acs.org on March 24, 2017

Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.

ACS Publications

ACS Synthetic Biology is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Published by American Chemical Society. Copyright © American Chemical Society. However, no copyright claim is made to original U.S. Government works, or works produced by employees of any Commonwealth realm Crown government in the course of their duties.

Co-generating synthetic parts toward a self-replicating system

Jun Li[1,2], Wilhelm Haas[3,4], Kirsten Jackson[5], Erkin Kuru[1,2], Michael C. Jewett[6], Z Hugh Fan[5,7], Steven Gygi[4], and George M. Church[1,2,*]

1. Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, USA.
2. Wyss Harvard Institute of Biologically Inspired Engineering, 3 Blackfan Circle, Boston, MA 02115, USA.
3. Massachusetts General Hospital Cancer Center, Charlestown, Boston, MA; Department of Medicine, Harvard Medical School, Boston, MA, USA.
4. Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA.
5. J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, P.O. Box 116131, Gainesville, FL 32611, USA
6. Department of Chemical and Biological Engineering, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208, United States
7. Department of Mechanical and Aerospace Engineering, University of Florida, P.O. Box 116250, Gainesville, FL, 32611, USA

Correspondences should be addressed to G. M. Church at Email:

gchurch@genetics.med.harvard.edu; Tel: +1. 617.432.7562; Fax: +1. 617.432.6513.

ABSTRACT

To build replicating systems with new functions, the engineering of existing biological machineries requires a sensible strategy. Protein synthesis Using Recombinant Elements (PURE) system consists of the desired components for transcription, translation, aminoacylation and energy regeneration. PURE, might be the basis for a radically alterable, lifelike system after optimization. Here, we regenerated 54 *E. coli* ribosomal (r-) proteins individually from DNA templates in the PURE system. We show that using stable isotope labeling with amino acids, mass spectrometry based quantitative proteomics could detect 26 of the 33 50S and 20 of the 21 30S subunit r-proteins when co-expressed in batch format PURE system. By optimizing DNA template concentrations and adapting a miniaturized Fluid Array Device with optimized feeding solution, we were able to cogenerate and detect at least 29 of the 33 50S and all of the 21 30S subunit r-proteins in one pot. The boost on yield of a single r-protein in co-expression pool varied from ~1.5 to 5-fold compared to the batch mode, with up to ~ 2.4 μM yield for a single r-protein. Reconstituted ribosomes under physiological condition from PURE system synthesized 30S r-proteins and native 16S rRNA showed ~13% activity of native 70S ribosomes, which increased to 21% when supplemented with GroEL/ES. This work also points to what is still needed to obtain self-replicating synthetic ribosomes *in-situ* in the PURE system.

Key words: cell-free protein synthesis; continuous-exchange; self-replicating system; proteomics; ribosome

INTRODUCTION

Construction of biochemical system capable of autonomous replication and Darwinian evolution, i.e. synthetic life, with radically new functions like novel amino acids or nucleotides represent ambitious projects[1-9]. There are two basic approaches for achieving a lifelike, self-replicating system. The "top-down" approach seeks to manipulate existing genomes from 'natural' cells through genome editing and evolution[4, 5]. The "bottom-up" approach seeks to assemble a synthetic cell from simple standard parts *in vitro,* such as a self-replicating system made from RNA[1] and the proposed 151 gene minimal genomic components sufficient to enable autonomous replication *in vitro*[10] currently with limited success. This has advantages over top-down approaches for enabling system-level to molecular-level debugging, replacing any or many components (e.g. RFs, r-proteins, tRNAs and synthetases) with radically altered versions (possibly even overall change in chirality). A suite of components (all of known function, minimizing unwanted or unknown side reactions), from the three main polymer systems of "the central dogma": DNA, RNA, and protein syntheses, is necessary.

Because the PURE system is a cell-free protein synthesis system that is constructed in a bottom-up approach based on well-characterized *E. coli* biochemistry[11, 12], it is the most prominent technology for a comprehensive self-renewing cell-like system. In the PURE system, various components (Table S1), responsible for transcription, translation, aminoacylation and energy regeneration, are individually prepared and reconstituted[13]. This potentially allows expression of cellular genes with known factors and avoids cellular contaminants such as proteases and nucleases inhibitory to protein synthesis.

A synthetic cell needs to be able to duplicate all components of itself to sustain its system. As the number and variety of synthesized components increase, balancing and regulating the protein synthesis become crucial. Previous models propose that ~42,000 peptide bonds per ribosome need to be synthesized in order to produce all the components of a self-replicating system[10]. Even though classical PURE system contains wild-type ribosomes, it is still

~1-2 orders of magnitude less efficient than this predicted necessary efficiency[14]. Therefore, it is necessary to improve protein expression efficiency in order to fulfill the self-reproduction. Different approaches have been applied to improve PURE system productivity. Although increasing the amount of input materials such as ribosomes and other translation enzymes can boost the yield by various fold[15, 16], it is not considered as a true improvement of the efficiency because the increase of input material is much more than the increase of output (reporter proteins synthesized) when counted by peptide bonds. Recent studies have improved the translation yield of the PURE system at **c**ontinuous-**e**xchange (CE) mode[17, 18]. Compared to the traditional batch format, CE mode increases protein expression by enabling continuous supply of nutrients and energy molecules from a feeding solution and removal of accumulating inorganic phosphate byproducts from the reaction solution. For example, Jackson K. *et al.* developed a **m**iniaturized **F**luid **A**rray **D**evice (mFAD) with an optimized feeding solution to the PURE system and yielded 0.18 mg/mL of GFP, which was around 72.5 times higher than the GFP production with the conventional PURE system[17] significantly approaching to the proposed peptide bond formation efficiency necessary for a synthetic life.

Yet, to construct a lifelike entity with PURE system, the biggest challenge remains to enable the self-replication of ribosomes, the key component of protein synthesis. Construction of *E. coli* ribosomes requires the synthesis and assembly of 21 small subunit r-proteins (designated S1–S21, Table 1), 33 large subunit r-proteins (designated L1–L36, Table 2), and 3 rRNAs (23S, 16S, and 5S rRNA). As a step towards this long-term goal, in this study, we reconstructed the 54 r-protein genes into circular protein expression vectors and successfully expressed them individually in the PURE system reconstituted in-house. We adopted mFAD with the optimized feeding solution described in previous study [17] to individual expression of r-proteins in the PURE system and found the yields increased by ~2 to 5-fold. For batch and CE mode PURE system, increasing the DNA template concentration to 20 ng/μl boosted the yield of S1 by ~1.5-fold compared to 10 ng/μl. Further increase in DNA template concentration did not

increase the yield. Toward reconstituting a self-replicating ribosome, we also co-expressed the 21 30S r-proteins or 33 50S r-proteins in the PURE system with ^{13}C-labelled lysine and arginine and probed the co-expression by Mass Spectrometry with a method similar to **s**table **i**sotope **l**abeling with **a**mino acids in **c**ell culture (SILAC)[19], termed **s**table **i**sotope **l**abeling with **a**mino acids in **c**ell-**f**ree **p**rotein **s**ynthesis (SILACFPS) (Figure 1). In CE format PURE system, we detected all of the 21 30S r-proteins and at least 29 of the 33 50S r-proteins indicating successful co-expression. The co-expression efficiency in CE format PURE system was significantly higher than the batch format base on the paired t-test of the ratios of newly synthesized protein to initially added protein of PURE system ribosomes. To test the functional of these proteins, we reconstituted 30S subunits from purified PURE system synthesized 30S r-proteins and native 16S rRNA under physiological condition and measured reconstituted ribosome activity by synthesizing firefly luciferase. This study extends the technical capabilities of the PURE system toward achieving a lifelike system by showing that 30S and 50S r-proteins can be co-synthesized in one pot, that was previously limited to co-synthesis and assembly of the five *E. coli* RNA polymerase subunits[20].

RESULTS

Individual expression of 30S and 50S r-proteins in batch format PURE system

PURE system contains T7 RNA polymerase to couple transcription to translation from a double-stranded DNA template containing a T7 promoter. Therefore, we cloned a total of 54 genes from *E. coli* MG1655 strain genome, the 21 30S r-protein and 33 50S r-protein genes, each to a pET-24b vector (except for rplB, see methods) between upstream T7 promoter and downstream T7 terminator. We reconstituted PURE system in-house based on previous study[12] with 10 ng/μl plasmid template concentrations encoding each of the r-protein. All of the r-proteins, namely 21 30S (Figure S1) and 33 50S (Figure S2) were successfully expressed in the PURE system individually, when the reactions were incubated at 37 ℃ for 2 h.

Individual expression of 30S and 50S r-proteins in CE format PURE system

Previous studies developed and optimized a feeding solution specific to the PURE system using a CE device mFAD[17, 18] in a 96-well format (Figure 1). The GFP expression yield in the mFAD with non-optimized feeding solution, which was based on the existing buffer and energy reagent concentrations, increased 20.9 times over the traditional batch format, while with optimized feeding solution (recipe listed in methods) increased 72.5-fold compared to the batch format[17]. Further optimization on experimental parameters, such as solution volume ratio, shaking and temperature, GFP expression increased 77.8-fold compared to the batch format in a standard microplate[18]. Here, to increase the yield of r-proteins, we applied the mFAD and the optimized feeding solution to the expression reaction. Specifically, 30S subunit r-protein S1, S2, S3, S4, S5, and S19 with various protein sizes from 10 kD to 61 kD were expressed in the batch format in the conventional microplate and the mFAD with optimized feeding solutions. Compared to the batch format, mFAD with the optimized feeding solution enhanced protein expression yield for all proteins (Figure 2): The CE mode increased protein expression yields of S1, S2, S3, S4, S5, and S19 ~3, ~3, ~3, ~2, ~2, and ~5-fold increase, respectively, over the batch format.

Previous studies have showed that the amount of synthesized protein in the PURE system is largely correlated to the amount of the encoding mRNA (within a certain concentration range) transcribed by T7 RNA polymerase[20, 21]. This suggests that the amount of DNA template could also influence the amount of synthesized proteins. Indeed, increasing the DNA template concentration from 10 ng/μl to 20 ng/μl slightly increased the yield of S1 in the batch format, while further increase from 20 ng/μl to 120 ng/μl did not further enhance the yield (Figure 3). We observed a similar trend for the CE format: Yield reached maximum at a DNA template

concentration of 20 ng/μl and stayed steady when template concentration exceeded 20 ng/μl (Figure 3).

Co-expression of 30S and 50S r-proteins in batch and CE format PURE system

To achieve a self-replicating of ribosome, synthesis of adequate amounts of 54 the r-proteins in one pot of PURE system are a curial first step. Therefore, we tried to co-express these proteins in PURE system in batch and CE format. In order to distinguish ribosomes, i.e. r-proteins, already present in the PURE system from the newly synthesized r-proteins, we employed a quantitative mass spectrometry-based proteomics approach SILACFPS[19] . In order to label the newly synthesized r-proteins, we replaced lysine and arginine amino acids of the PURE co-expression reaction mixture and feeding solution with ^{13}C-labelled lysine and arginine, so that after trypsin digestion, each digested newly synthesized peptide fragment has at least one ^{13}C-labelled lysine or arginine. These peptides therefore can be distinguished from peptides coming from ribosomes already present in the PURE system, allowing us to quantify their relative expression levels (Table 1 and 2). Other protein components in the PURE system can be taken as negative control: Only light form peptides were supposed to be detected (Table S2 and S3).

Specifically, we co-expressed 30S and 50S r-proteins in the PURE system in the batch (DNA template concentration: each r-protein 10 ng/μl) and CE format (DNA template concentration: each r-protein 80 ng/μl or 120 ng/μl) containing ^{13}C-labelled arginine and lysine. When using the batch format, we found that 20 of the 21 30S r-proteins (Table 1) as well as 26 of the 33 50S r-proteins (Table 2) were identified as carrying ^{13}C-labelled lysine or arginine, showing that most of these proteins were successfully co-expressed in our system. The ratio of detected peptides representing newly synthesized protein (Heavy form) to initially added ribosomes (Light form), H/L ratio for short, was calculated to measure the efficiency of

expression (Table 1 and 2). Some H/L ratios were found to as high as 1.7 (S11) for 30S r-proteins and 1 (L16) for 50S r-proteins showing high efficiency of expression of these proteins. When using the CE format, all of the 30S r-proteins (Table 1) and 29 of the 33 50S r-proteins (Table 2) were identified as carrying ^{13}C-labelled lysine or arginine for both of the two DNA template concentration sets. The highest H/L ratio 2 was observed for S18 (30S, CE format, 120 ng/μl template concentrations) and L32 (50S, CE format, 80 ng/μl template concentrations). Paired t-test of H/L ratios showed that CE mode with 80 ng/μl and 120 ng/μl template concentrations produced significant higher amount of r-proteins than batch mode with a p-value of 0.005 and 0.012, respectively, for the 30S, and a p-value of 0.0026 and <0.0001, respectively, for the 50S (Figure 4). According to the H/L ratios, the yield increase of a single r-protein in co-expression pool in CE mode varied from ~1.5 to 5-fold compared to the batch mode. Paired t-test of H/L ratios between 80 ng/μl and 120 ng/μl template concentrations in CE mode showed that the difference is not significant (Figure 4). In optimal conditions (120 ng/μl template and CE mode), 10 of the 21 30S r-proteins and 9 of the 33 50S r-proteins had an H/L ratio ≥ 1 (Figure 4), indicating that we could generate at least as much proteins as we initially added for these proteins. Thus we could estimate the yield of each r-protein based on the initial ribosome concentration added to the system. As the initial ribosome concentration is 1.2 μM, the yields of the proteins with H/L ratio ≥ 1 were estimated to be at least 1.2 μM.

30S subunit reconstitution with PURE system synthesized 30S r-proteins

To our knowledge 30S subunit reconstitution has never been done with *in vitro* synthesized 30S r-proteins instead of TP30. Thus, in an effort to make these semi-synthetic 30S subunits, we purified PURE system synthesized 30S r-proteins using Reverse His-tag Purification method described in PURE system handbook. S7, S9, S11, S12, S13 and S18

failed to pass the Amicon Ultracel 0.5 ml-100 K spin concentrator and could not be purified using this method. Therefore a strep-tagged version of S7, S9, S11, S12, S13 and S18 was constructed. Step-tagged proteins were synthesized in PURE system and purified as described in the methods. Purified proteins were analyzed by SDS-PAGE (Figure S3) and concentrations were determined by Bradford assay.

30S subunit reconstitution was conducted in PURE system with no ribosomes (PURE Δ ribosome system), serving as a platform to integrate ribosome reconstitution and protein synthesis in one compartment. Reconstituted ribosome activity was measured by synthesizing a reporter protein firefly luciferase (Fluc) here. Specifically, 30S subunit reconstitution, its coupling with native 50S subunit, Fluc transcription and translation occurred simultaneously in a 2-h 37℃ 15 μl batch reaction and Fluc was quantified by luciferase assay. Coupling native 50S and 30S subunits (50S + 30S; ratio = 1:1) exhibited ~87% activity of intact native 70S ribosomes (70S) (Figure 5), indicating that the recoupling process is not a bottleneck for ribosome reconstitution in our system. On the other hand, when we reconstituted 30S subunits from **nat**ive **16S** rRNA (nat16S) and **t**otal r-**p**roteins of **30**S subunit (TP30) extracted from *E. coli* cells and coupled them with native 50S (nat16S + TP30 + 50S; ratio = 1:3:1), they exhibited ~46% activity of 70S (Figure 5). This decrease in activity could be explained by the lack of assembly cofactors and/or rRNA processing functionalities. Finally, we replaced TP30 with PURE system synthesized 30S r-proteins in our integrated reconstitution reaction. Reconstituted 30S subunits from native 16S rRNA and PURE system synthesized 30S r-proteins (nat16S + ivt 30S r-pro + 50S; ratio = 1:3:1) showed ~13% activity of native 70S ribosomes, which increased to ~21% when supplemented with chaperone system GroEL/ES (Figure 5). The activity drop from 46% to 13% caused by replacing TP30 with PURE system synthesized proteins suggested that a fraction of r-proteins synthesized in PURE was not active or properly folded, possibly due to the missing of protein factors facilitating protein folding in PURE[22].

DISCUSSION

A cell-free protein synthesis environment that consists of the optimized concentrations of proteins, energy source, monomer substrates, cofactors and DNA with desired genes, in theory, could constitute a self-replicating dynamic system. The ribosome, as the key component for protein translation, needs to regenerate itself to ensure the self-replication of the entire system. Previously, Tian J *et al.* reported that 30S r-proteins S2, S4, S5, S6, S9, S10, S12, S13, S15, S16, S17, and S21 could individually be expressed in the *E. coli* cell lysate based cell-free protein synthesis system using PCR product templates, but small subunit proteins S1, S3, S7, S8, S11, S14, S18, S19, S20 could not[23]. This might have been due to the degradation of linear DNA templates, mRNAs or proteins caused by the nucleases and proteases present in the cell lysate based system. PURE system is a defined reconstituted cell-free system with minimal components for protein transcription and translation. With no nucleases and proteases contamination present and being entirely modular in terms of content and concentrations of system components, PURE system appears to be a more attractive platform toward constructing a self-replicating entity.

In this study, we demonstrated that PURE system can be utilized to express all of the 54 r-proteins individually in batch format with a ~2 to 5-fold additional increase in yield when a CE device mFAD was applied. We also tried combinations of different DNA template concentrations of r-proteins to find the optimal template concentration for the highest expression levels. Increasing the DNA template concentration to 20 ng/µl boosted the yield of 30S subunit r-protein S1 in both batch and CE mode, but the yield did not increase further beyond this template concentration. By labelling the newly synthesized proteins with ^{13}C lysine and arginine, we established a Mass Spectrometry-based detection method which can distinguish the newly co-synthesized r-proteins from the original ones present in the PURE system. The mass spectrometry data showed 20 of the 21 30S r-proteins and 26 of the 33 50S r-proteins were co-

expressed in batch format PURE system; while in CE format PURE system, the de novo co-expression was significantly higher and heavy forms of all the 30S r-proteins and 29 of the 33 50S r-proteins could be detected, for both of the two DNA template concentration sets. Notably, the four 50S subunit r-proteins (L30, L34, L35 and L36) that could not be detected are all smaller than 8 kD and light form peptides coming from two of those r-proteins (L34 and L36) already present in the PURE system could not be detected either. The underrepresentation of these small proteins may be due to a low efficiency of precipitating these proteins by applying the TCA precipitation protocol or due to relatively small number of peptides being generated by trypsin digestion[24].

A robust, self-replicating entity would rely on *in situ* reconstitution of the protein translation machinery, from individual components (most notably r-proteins) synthesized *de novo*. Therefore balancing the amount of proteins synthesized for each target protein individually, so that they can stoichiometrically be assembled into ribosome complex, is our ultimate goal. The optimal ratio for the relative amounts of the templates encoding r-proteins was not determined in this study. A systematic optimization may be necessary to further balance the amount of the synthesized r-proteins. For the proteins which were not synthesized efficiently by the PURE system, several fine-tuning strategies, such as altering promoter strengths or codon optimization, will be pursued. For example, gene codons can be altered to be A + U rich for destabilizing mRNA secondary structure[25]. Genes can also be changed to have only the most frequently used codon (e.g. only the triplet CAG for glutamine - present 69% of the time in *E. coli* genes – would be used). Nevertheless, our results suggest that co-expression of r-proteins in the PURE system approaches to (or exceeds) the level of r-proteins that are initially added to boot the system up and that assembling a synthetic lifelike system is not a far-fetched dream. Moreover, these *in vitro* translated 30S r-proteins, after purification, can reconstitute 70S ribosomes with native 50S subunits and native 16S rRNA with ~13% activity of native 70S ribosomes under physiological condition in PURE. This activity

increased to 21% when chaperone system GroEL/ES was supplemented in the reconstitution reaction. Future improvements of PURE system productivity and ribosome reconstitution efficiency can be possibly achieved by adding translation facilitating and protein folding factors to PURE system protein synthesis[15, 22, 26] or by adding ribosome assembly cofactors to reconstitution reaction[27]. Our findings also provide hope that, fully synthetic ribosomes capable of self-replicating can be achieved in the PURE system based platform by integrating rRNA transcription, r-protein co-translation and ribosome assembly together. More ambitiously, further improvements of this system will enable a foundation to construct a DNA, RNA and protein based replicating system[10].

METHODS

Media, chemicals, and reagents

Unless specified, all chemicals were obtained from Sigma-Aldrich. Tryptone and yeast extract were obtained from BD Difco. GroEL/ES were obtained from Takara.

PURE system reconstitution

PURE system component plasmids were obtained from the Ueda group. Home-made PURE system was prepared according to [12, 17]. Reagent concentrations: 1.2 μM ribosomes, 2.7 μM IF1, 0.4 μM IF2, 1.5 μM IF3, 0.26 μM EF-G, 0.92 μM EF-Tu, 0.66 μM EF-Ts, 0.25 μM RF1, 0.24 μM RF2, 0.17 μM RF3, 0.5 μM RRF, 1900 U/ml AlaRS, 2500 U/ml ArgRS, 20 mg/ml AsnRs, 2500 U/ml AspRs, 630 U/ml CysRs, 1300 U/ml GlnRs, 1900 U/ml GluRs, 5000 U/ml GlyRs, 630 U/ml HisRs, 2500 U/ml IleRS, 3800 U/ml LeuRS, 3800 U/ml LysRS, 6300 U/ml MetRS, 1300 U/ml PheRS, 1300 U/ml ProRS, 1900 U/ml SerRS, 1300 U/ml ThrRS, 630 U/ml

TrpRS, 630 U/ml TyrRS, 3100 U/ml ValRS, 4500 U/ml MTF, 4 μg/ml creatine kinase, 3 μg/ml myokinase, 1.1 μg/ml nucleoside-diphosphate kinase, 2.0 units/ml pyrophosphatase, 10 μg/ml T7 RNA polymerase, 2 mM ATP, 2 mM GTP, 1 mM CTP, 1 mM UTP, 20 mM creatine phosphate, 50 mM HEPES-KOH (pH 7.6), 100 mM potassium glutamate, 13 mM magnesium acetate, 2 mM spermidine, 1 mM dithiothreitol (DTT), 0.3 mM 20 amino acids, 10 μg/ml 10-formyl-5,6,7,8-tetrahydrofolic acid and 56 A260 per ml tRNA mix.

Device fabrication

Fabrication of the mFAD was described in [17].

Optimized feeding solution

Optimized feeding solution contains 50 mM HEPES-KOH (pH 7.6), 100 mM potassium glutamate, 17 mM magnesium acetate, 2 mM spermidine, 1 mM DTT, 4 mM ATP, 4 mM GTP, 1 mM CTP, 1 mM UTP, 20 mM creatine phosphate, 0.5 mM amino acids, 10 μg/ml 10-formyl-5,6,7,8-tetrahydrofolic acid as listed in [17].

Molecular cloning

50S and 30S r-protein genes were PCR amplified from *E.coli* MG1655 genome (ATCC) with primers listed in Table S4 and S5 and cloned to pET-24b (Novagen) using restriction enzyme NdeI and XhoI except rplB. RplB was cloned to pIVEX 2.3d vector (Roche) with restriction enzyme NcoI and XhoI due to its internal cleavage site of NdeI. Genes were cloned in its natural format with no additional amino acid or affinity tags. PCR and standard Sanger sequencing primers were obtained from Integrated DNA Technologies with standard purification. Sanger sequencing verification of each clone was performed by Genewiz.

Individual expression of r-proteins in batch format PURE system

Each protein was expressed in a 25 µl PURE system reaction with 0.8 U/µl Murine RNase Inhibitor (New England Biolabs) at 37℃ for 2 h with a plasmid template concentration of 10 ng/µl or other concentrations as indicated in the text. Reactions were later analyzed on 4-12% Bis-Tris Protein Gels (Life Technologies) and stained by Coomassie Blue G-250.

Individual expression of r-proteins in CE format PURE system

For the CE format in the mFAD, 10 µl PURE system reaction solution with 0.8 U/µl Murine RNase Inhibitor and a plasmid template concentration as indicated in the text was dispensed into the reaction chamber, while a total of 200 µl of the feeding solution was pipetted into the feeding chambers (100 µl in each of the two feeding chambers). Negative controls were prepared by replacing the DNA in the reaction solution with an equal volume of nuclease-free water. Batch format controls were carried in 10 µl PURE system reaction with 0.8 U/µl Murine RNase Inhibitor and a plasmid template concentration as indicated in the text and dispensed into a conventional microplate. To prevent evaporation, the conventional microplate and the mFAD were covered with a tape. Then the microplate and mFAD were incubated at 37℃ (optimized temperature in [18]) for 4 h with shaking. Samples were later analyzed on 4-12% Bis-Tris Protein Gels (Life Technologies) and stained by Coomassie Blue G-250.

Co-expression of 30S and 50S r-proteins

For batch format controls, 10 µl PURE system reaction was reconstituted with 10 ng/µl of each plasmid encoding 30S or 50S r-proteins, 0.8 U/µl Murine RNase Inhibitor, 0.3 mM ^{13}C-labelled

lysine, arginine and 0.3 mM the other 18 amino acids. Reactions were dispensed into a conventional microplate. For CE format, 10 µl PURE system reaction was reconstituted with 80 ng/µl or 120 ng/µl of each plasmid encoding 30S or 50S r-proteins, 0.8 U/µl Murine RNase Inhibitor, 0.3 mM ^{13}C-labelled lysine, arginine and 0.3 mM the other 18 amino acids. Reactions were dispensed into mFAD. Negative controls were prepared by replacing the DNA in the reaction solution with an equal volume of nuclease-free water. The microplate and mFAD were sealed and incubated at 37℃ for 4 h with shaking.

Mass spectrometry sample preparation

10 µl of co-expression reaction was taken and diluted to 1 mg/ml with Alkylation buffer (8 M Urea, 25 mM Tris-HCl pH 8.0, 10 mM DTT) and incubated at 56℃ for 30 min, then cool down to room temperature. Iodoacetamide was added to the protein solution to a final concentration of 30 mM. The tube was wrapped with foil and incubated at room temperature for 30 min. DTT was added to 20 mM and incubated at 37℃ for 30 min to quench the reaction. 1/4 volume of 100% TCA was added drop wise to each sample and incubated for 10 min on ice. Each sample was then spun at ~14K rpm for 5 min at 4℃. Supernatant was discarded. Pellet was washed with 200 µl ice cold acetone per tube by vortexing and spun again at ~14K rpm for 5 min at 4℃. The washing step was repeated twice. Pellet was air-dried on bench. 500 µl digestion buffer (8 M Urea, 50 mM Tris-HCl pH 8.0) was added to protein pellet and incubated at 56℃ for 60 min to denature the proteins. Each sample was then cooled. 100 mM Tris-HCl (pH 8.0) was added to each sample until urea concentration was less than 1M. 20 µg trypsin was added to protein samples and incubated at 37℃ for 6 h. 1/2 volume of 5% acetonitrile/5% formic acid was added to samples to reach a final pH <4. Water's tc18 column was pre-wet with 1 ml 100% acetonitrile and 1 ml 90% acetonitrile/5% formic acid and then equilibrated with 1 ml 5% acetonitrile/5%

formic acid. Acidified samples were loaded to the column slowly (< 1ml/min), washed with 1 ml 5% acetonitrile/5% formic acid and eluted with 500 µl 50% acetonitrile/ 5% formic acid. Samples were then dried using Speedvac.

Mass spectrometry data analysis

The generated peptides were analyzed using liquid-chromatography tandem mass spectrometry (LC-MS/MS) essentially as described previously[28]. Briefly, the analysis was performed using an Orbitrap Elite mass spectrometer (Thermo Scientific, San Jose, CA) equipped with an Accela 600 binary HPLC pump (Thermo Scientific) and a Famos autosampler (LC Packings). Peptides were fractionated over a 100 µm I.D. in-house-made microcapillary column packed first with approximately 0.5 cm of Magic C_4 resin (5 µm, 100 Å, Michrom Bioresources) followed by 20 cm of Maccel C_{18}AQ resin (3 µm, 200 Å, Nest Group). Fractionation was achieved by applying a gradient from 10 % to 32 % acetonitrile in 0.125 % formic acid over 75 min at a flow rate of approximately 300 nl min^{-1}. The mass spectrometer was operated in a data-dependent mode collecting survey MS spectra in the Orbitrap over an m/z range of 300-1500, followed by the collection of MS2 spectra acquired in the dual pressure linear ion trap on the up to 20 most abundant ions detected in the survey MS spectrum. The settings for collecting survey MS spectra were: AGC target, $1x10^6$; maximum ion time, 50 ms; resolution $6x10^3$. The settings for the acquisition of MS2 spectra were: isolation width, 2 m/z; AGC target, $2x10^3$; max. ion time, 100 ms; normalized collision energy, 35; dynamic exclusion, 40 s at 10 ppm.

Peptides were identified from the MS2 data using the Sequest algorithm[23] operated in an in-house-developed software suite environment that was also applied for filtering the search results and extracting the quantitative data. The data was searched against a protein sequence database comprised of the sequences of *E. coli* ribosomal proteins, of known contaminants

such as porcine trypsin, and of 500 nonsense protein sequences derived from all *S. cerevisiae* protein sequences using a fourth order Markov chain model[29]. To this forward (target) database component we added a reversed (decoy) component including all listed protein sequences in reversed order[30]. The Markov chain model derived nonsense protein sequences were added to generate a protein sequence database size allowing an accurate estimation of the false discovery rate of assigned peptides and proteins. Searches were performed using a 50 ppm precursor ion mass tolerance and we required that both termini of the assigned peptide sequences were consistent with trypsin specificity allowing two missed cleavages. Carbamidomethylation of cysteines (+57.02146) was set as static information and full ^{13}C labeling on arginine and lysine (+6.020129) as well as oxidation on methionine (+15.99492) were set as variable modifications. A false discovery rate of less than 1 % for the assignments of both peptides and proteins was achieved using the target-decoy search strategy [30]. Assignments of MS2 spectra were filtered using linear discriminant analysis to define one composite score from the following spectral and peptide sequence specific properties: mass accuracy, XCorr, ΔCn, peptide length, and the number of missed cleavages[31]. Protein identifications were filtered based on the combined probabilities of being correctly assigned for all peptides assembled into each protein[31]. Both peptide and protein assignments were filtered to a false-discovery rate of less than 1 %. Relative peptide quantification was done in an automated manner by producing extracted ion chromatograms (XIC) for the light (^{12}C arginine or lysine) and heavy (^{13}C arginine or lysine) forms of a peptide ion followed by measuring the area under the chromatographic peaks[32]. A peptide ion was considered to be quantified when the sum of the signal-to-noise ratio of both light and heavy form was larger or equal to 10. The median of the log2 (heavy/light) ratios measured for all peptides assembled into a protein was defined as the protein log2 ratio. If only the ion signal of the light peptide form was observed above the noise level the log2 heavy-to-light ratio was defined as being smaller than log2 of the signal-to-noise for the light peptide ion. As orthogonal approach to determine the occurrence of

the light and heavy forms of a protein we counted the number of MS2 spectra assigned to each forms of a protein.

Isolation of tightly coupled 70S ribosomes and ribosomal subunits

Tightly coupled 70S ribosomes, 30S and 50S subunits were prepared as described by Michael C Jewett[33]. Results from three independent ribosome preparations and subsequent rRNA and total protein preparations were used and averaged to generate the final results shown in the manuscript.

Isolation of native rRNA and r-proteins

TP30 and native 16S rRNA were prepared as described by Nierhaus[34]. TP30 was purified using acetone precipitation.

Preparation of PURE system synthesized 30S r-proteins

30S r-proteins, except S7, S9, S11, S12, S13 and S18, were expressed in PURE system reactions and purified using reverse his-tag purification method. Specifically, after expression, 200 μl reaction was collected for each protein and then diluted with 200 μl dilution buffer (50 mM Hepes-KOH, 10 mM Magnesium acetate, 0.8 M NaCl, 0.1% Tween 20). The diluted reaction mixture was applied to Amicon Ultracel 0.5 ml-100 K spin concentrator and centrifuged for 50 min at 15000 × g at 4℃. The permeate was transferred to a new tube and mixed with 0.25 volumes Ni-NTA magnetic beads (Qiagen) for 40 min at 4℃. The reaction mixture slurry was applied to an empty Bio-Rad micro-spin column and centrifuged for 2 min at 1500 × g at 4℃. The elute containing purified protein was concentrated by Amicon Ultracel 0.5 ml-3 K spin concentrator and exchanged to a storage buffer to a final concentration of 50 mM Tris-HCl pH 7.6, 100 mM KCl, 10 mM Magnesium acetate, 1 mM DTT and 30% glycerol. A strep-tag was inserted to S7 at C-terminal, S9 at N-terminal, S11 at N-terminal, S12 at C-terminal, S13 at N-

terminal and S18 at N-terminal on the protein expression plasmids using QuikChange Lightening Site-Directed Mutagenesis Kits (Agilent Genomics) (Table S6). These strep-tagged proteins were expressed in PURE system. 200 µl Strep-Tactin magnetic beads (Qiagen) were washed three times with 1 ml wash buffer (50 mM Tris HCl pH 7.6, 0.5 M NaCl, 6 mM β-mercaptoethanol) and resuspended in 200 µl wash buffer. 200 µl PURE reaction for each strep-tagged protein was collected and mixed with the beads mixture on a rotator at 4℃ for 3 h. The beads were then immobilized with a magnet. Supernatant was discarded and beads were washed twice with 200 µl wash buffer. Strep-tagged r-protein was eluted by 150 µl elution buffer (50 mM Tris HCl pH 7.6, 0.5 M NaCl, 10 mM biotin, 6 mM β-mercaptoethanol), concentrated by EMD Amicon Ultracel 0.5 ml-3 K spin concentrator and exchanged to a storage buffer to a final concentration of 50 mM Tris-HCl pH 7.6, 100 mM KCl, 10 mM Magnesium acetate, 1 mM DTT and 30% glycerol. All purified r-proteins were analyzed on 4-12% Bis-Tris Gel (Life Technologies) and concentrations were determined by Bradford Assay (Bio-Rad). After concentration determination, all of the 21 30S r-proteins were mixed at an equal molar ratio and buffer exchanged to 50 mM Tris-HCl pH 7.6, 100 mM KCl, 10 mM Magnesium acetate, 1 mM DTT and concentrated by Amicon Ultracel 0.5 ml-3 K spin concentrator to a certain fold. The final concentration of the r-protein mixture stock was the original concentration times the fold number. The stock was used for the reconstitution assay.

Integrated ribosome assembly, in vitro transcription and translation

Integrated assay was set up to 15 µl with PURE system factors, 0.8 U/µl Murine RNase Inhibitor, 10 ng/µl pIVEX 2.3d Fluc plasmid (expressing firefly luciferase under a T7 promoter control, obtained from [15]), 0.3 µM 50S subunit, 0.9 µM PURE system synthesized 30S r-proteins or 0.9 µM TP30 and 0.3 µM native 16S rRNA with or without 0.6 µM GroEL/ES. When TP30 or PURE system synthesized 30S r-proteins mixture were added, an equal volume of the same

buffer was added to other reconstitution reactions to rule out the influence from buffers. Reactions were incubated at 37℃ for 2 h. After incubation, 7 µl reaction was mixed with 40 µl luciferase assay substrate (Promega) and incubated at room temperature for 10 min. Luminescence was measured by SpectraMaxM5 plate reader (Molecular Devices, Sunnyvale, CA). 5 replicates were conducted for each condition.

SUPPORTING INFORMATION

Contains 6 supporting tables and 3 supporting figures.

AUTHOR CONTRIBUTIONS

JL, MCJ and GMC designed the study. JL performed the experiments. WH and SG contributed to the mass-spec data acquisition and analysis. KJ and ZHF manufactured the CE device mFAD. JL, WH and EK wrote the manuscript. All authors helped edit the final manuscript.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

ACKNOWLEDGEMENTS

This work was funded by programs from the Department of Energy (Genomes to Life Center) [Grant #DE-FG02-02ER63445].

REFERENCES

1. Szostak, J. W.; Bartel, D. P.; Luisi, P. L., Synthesizing life. *Nature* **2001,** 409, (6818), 387-90.

2. Chen, I. A.; Roberts, R. W.; Szostak, J. W., The emergence of competition between model protocells. *Science* **2004,** 305, (5689), 1474-6.

3. Posfai, G.; Plunkett, G., 3rd; Feher, T.; Frisch, D.; Keil, G. M.; Umenhoffer, K.; Kolisnychenko, V.; Stahl, B.; Sharma, S. S.; de Arruda, M.; Burland, V.; Harcum, S. W.; Blattner, F. R., Emergent properties of reduced-genome Escherichia coli. *Science* **2006,** 312, (5776), 1044-6.

4. Smith, H. O.; Hutchison, C. A., 3rd; Pfannkoch, C.; Venter, J. C., Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proc Natl Acad Sci U S A* **2003,** 100, (26), 15440-5.

5. Chan, L. Y.; Kosuri, S.; Endy, D., Refactoring bacteriophage T7. *Mol Syst Biol* **2005,** 1, 2005 0018.

6. Gibson, D. G.; Glass, J. I.; Lartigue, C.; Noskov, V. N.; Chuang, R. Y.; Algire, M. A.; Benders, G. A.; Montague, M. G.; Ma, L.; Moodie, M. M.; Merryman, C.; Vashee, S.; Krishnakumar, R.; Assad-Garcia, N.; Andrews-Pfannkoch, C.; Denisova, E. A.; Young, L.; Qi, Z. Q.; Segall-Shapiro, T. H.; Calvey, C. H.; Parmar, P. P.; Hutchison, C. A., 3rd; Smith, H. O.; Venter, J. C., Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **2010,** 329, (5987), 52-6.

7. Mansy, S. S.; Schrum, J. P.; Krishnamurthy, M.; Tobe, S.; Treco, D. A.; Szostak, J. W., Template-directed synthesis of a genetic polymer in a model protocell. *Nature* **2008,** 454, (7200), 122-5.

8. Mansy, S. S.; Szostak, J. W., Reconstructing the emergence of cellular life through the synthesis of model protocells. *Cold Spring Harb Symp Quant Biol* **2009,** 74, 47-54.

9. Schrum, J. P.; Zhu, T. F.; Szostak, J. W., The origins of cellular life. *Cold Spring Harb Perspect Biol* **2010,** 2, (9), a002212.

10. Forster, A. C.; Church, G. M., Towards synthesis of a minimal cell. *Mol Syst Biol* **2006,** 2, 45.

11. Shimizu, Y.; Inoue, A.; Tomari, Y.; Suzuki, T.; Yokogawa, T.; Nishikawa, K.; Ueda, T., Cell-free translation reconstituted with purified components. *Nat Biotechnol* **2001,** 19, (8), 751-5.

12. Shimizu, Y.; Kanamori, T.; Ueda, T., Protein synthesis by pure translation systems. *Methods* **2005,** 36, (3), 299-304.

13. Shimizu, Y.; Ueda, T., PURE technology. *Methods Mol Biol* **2010,** 607, 11-21.

14. Jewett, M. C.; Forster, A. C., Update on designing and building minimal cells. *Curr Opin Biotechnol* **2010,** 21, (5), 697-703.

15. Li, J.; Gu, L.; Aach, J.; Church, G. M., Improved cell-free RNA and protein synthesis system. *PLoS One* **2014,** 9, (9), e106232.

16. Kazuta, Y.; Matsuura, T.; Ichihashi, N.; Yomo, T., Synthesis of milligram quantities of proteins using a reconstituted in vitro protein synthesis system. *J Biosci Bioeng* **2014,** 118, (5), 554-7.

17. Jackson, K.; Kanamori, T.; Ueda, T.; Fan, Z. H., Protein synthesis yield increased 72 times in the cell-free PURE system. *Integr Biol (Camb)* **2014,** 6, (8), 781-8.

18. Jackson, K.; Jin, S.; Fan, Z. H., Optimization of a miniaturized fluid array device for cell-free protein synthesis. *Biotechnol Bioeng* **2015,** 112, (12), 2459-67.

19. Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **2002,** 1, (5), 376-86.

20. Asahara, H.; Chong, S., In vitro genetic reconstruction of bacterial transcription initiation by coupled synthesis and detection of RNA polymerase holoenzyme. *Nucleic Acids Res* **2010,** 38, (13), e141.

21. Maki, K.; Uno, K.; Morita, T.; Aiba, H., RNA, but not protein partners, is directly responsible for translational silencing by a bacterial Hfq-binding small RNA. *Proc Natl Acad Sci U S A* **2008,** 105, (30), 10332-7.

22. Li, J.; Zhang, C.; Huang, P.; Kuru, E.; Forster-Benson, E. T. C.; Li, T.; Church, G. M., Dissecting limiting factors of the Protein synthesis Using Recombinant Elements (PURE) system. *bioRxiv 099838* **2017,** https://doi.org/10.1101/099838.

23. Tian, J.; Gong, H.; Sheng, N.; Zhou, X.; Gulari, E.; Gao, X.; Church, G., Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* **2004,** 432, (7020), 1050-4.

24. Feist, P.; Hummon, A. B., Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples. *Int J Mol Sci* **2015,** 16, (2), 3537-63.

25. Shabalina, S. A.; Spiridonov, N. A.; Kashina, A., Sounds of silence: synonymous nucleotides as a key to biological regulation and complexity. *Nucleic Acids Res* **2013,** 41, (4), 2073-94.

26. Wang, H. H.; Huang, P. Y.; Xu, G.; Haas, W.; Marblestone, A.; Li, J.; Gygi, S. P.; Forster, A. C.; Jewett, M. C.; Church, G. M., Multiplexed in vivo His-tagging of enzyme pathways for in vitro single-pot multienzyme catalysis. *ACS Synth Biol* **2012,** 1, (2), 43-52.

27. Connolly, K.; Culver, G., Deconstructing ribosome construction. *Trends Biochem Sci* **2009,** 34, (5), 256-63.

28. Tolonen, A. C.; Haas, W.; Chilaka, A. C.; Aach, J.; Gygi, S. P.; Church, G. M., Proteome-wide systems analysis of a cellulosic biofuel-producing microbe. *Mol Syst Biol* **2011,** 7, 461.

29. Haas, W.; Faherty, B. K.; Gerber, S. A.; Elias, J. E.; Beausoleil, S. A.; Bakalarski, C. E.; Li, X.; Villen, J.; Gygi, S. P., Optimization and use of peptide mass measurement accuracy in shotgun proteomics. *Mol Cell Proteomics* **2006,** 5, (7), 1326-37.

30. Elias, J. E.; Gygi, S. P., Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **2007,** 4, (3), 207-14.

31. Huttlin, E. L.; Jedrychowski, M. P.; Elias, J. E.; Goswami, T.; Rad, R.; Beausoleil, S. A.; Villen, J.; Haas, W.; Sowa, M. E.; Gygi, S. P., A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **2010,** 143, (7), 1174-89.

32. Bakalarski, C. E.; Elias, J. E.; Villen, J.; Haas, W.; Gerber, S. A.; Everley, P. A.; Gygi, S. P., The impact of peptide abundance and dynamic range on stable-isotope-based quantitative proteomic analyses. *J Proteome Res* **2008,** 7, (11), 4756-65.

33. Jewett, M. C.; Fritz, B. R.; Timmerman, L. E.; Church, G. M., In vitro integration of ribosomal RNA synthesis, ribosome assembly, and translation. *Mol Syst Biol* **2013,** 9, 678.

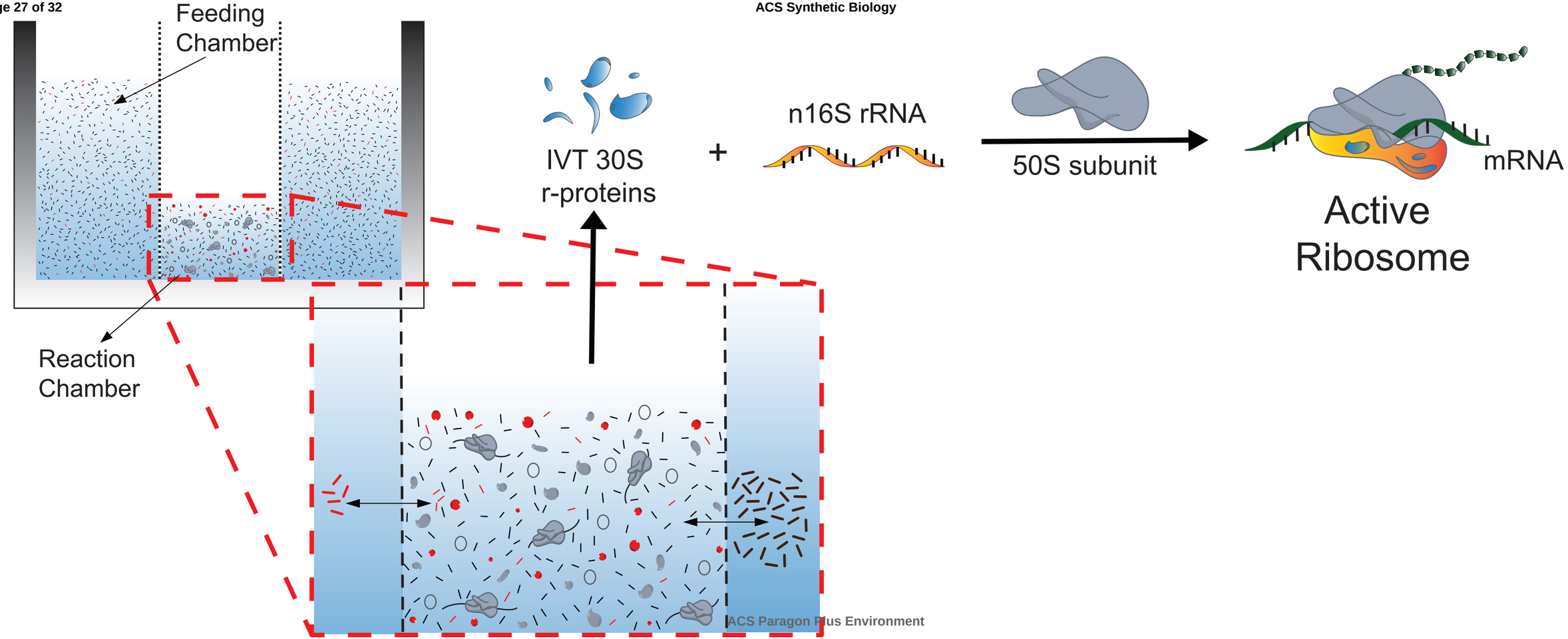
34. Nierhaus, K. H., Reconstitution of ribosomes, in Ribosomes and Protein Synthesis: A Practical Approach. *Spedding, G., Editor Oxford University Press, Oxford* **1990**, 161-189.

Table 1. Mass Spectrum analysis of PURE system co-synthesized 30S subunit r-proteins. Newly synthesized 30S r-proteins are labelled with ^{13}C on lysine and arginine (heavy form), while the other protein components in the PURE system are in the ^{12}C light form. Samples are first alkylated by iodoacetamide, precipitated by trichloroacetic acid, and finally digested by trypsin before subjected to mass spectrum. Spectral count data of detected peptides are presented by their encoded genes. The values for spectral counts indicate the numbers of spectra correctly assigned to the listed and grouped into L (light) and H (heavy) form.

			CE 120 ng/μl			CE 80 ng/μl			Batch 10 ng/μl		
Protein	Gene name	Protein size	Spectral Counts (L)	Spectral Counts (H)	H/L ratio	Spectral Counts (L)	Spectral Counts (H)	H/L ratio	Spectral Counts (L)	Spectral Counts (H)	H/L ratio
S1	rpsA	61.2kD	107	64	0.60	56	37	0.66	87	46	0.53
S2	rpsB	27kD	60	28	0.47	45	17	0.38	64	12	0.19
S3	rpsC	25.6kD	28	32	1.14	20	28	1.40	22	22	1.00
S4	rpsD	23.5kD	22	25	1.14	17	14	0.82	21	11	0.52
S5	rpsE	17.6kD	23	7	0.30	16	6	0.38	27	7	0.26
S6	rpsF	15.2kD	13	12	0.92	11	5	0.45	13	5	0.38
S7	rpsG	20kD	26	13	0.50	21	10	0.48	22	9	0.41
S8	rpsH	14.1kD	30	18	0.60	17	6	0.35	18	9	0.50
S9	rpsI	15.1kD	11	11	1.00	8	8	1.00	9	6	0.67
S10	rpsJ	11.7kD	25	13	0.52	13	7	0.54	19	11	0.58
S11	rpsK	13.8kD	10	12	1.20	9	10	1.11	7	6	0.86
S12	rpsL	13.7kD	5	6	1.20	6	6	1.00	5	3	0.60
S13	rpsM	13,1kD	23	15	0.65	17	11	0.65	18	7	0.39
S14	rpsN	11.6kD	1	1	1.00	2	2	1.00	1	0	0.00
S15	rpsO	10.3kD	7	1	0.14	7	1	0.14	9	1	0.11
S16	rpsP	9.2kD	10	12	1.20	10	10	1.00	9	4	0.44
S17	rpsQ	9.3kD	5	4	0.80	3	4	1.33	6	2	0.33
S18	rpsR	9kD	2	4	2.00	3	2	0.67	3	3	1.00
S19	rpsS	10.1kD	4	4	1.00	3	4	1.33	3	5	1.67
S20	rpsT	9.7kD	4	5	1.25	4	4	1.00	4	2	0.50
S21	rpsU	8.5kD	6	3	0.50	6	4	0.67	5	2	0.40

Table 2. Mass Spectrum analysis of PURE system co-synthesized 50S subunit r-proteins. Newly synthesized 50S r-proteins are labelled with ^{13}C on lysine and arginine (heavy form), while the other protein components in the PURE system are in the ^{12}C light form. Samples are first alkylated by iodoacetamide, precipitated by trichloroacetic acid, and finally digested by trypsin before subjected to mass spectrum. Spectral count data of detected peptides are presented by their encoded genes. The values for spectral counts indicate the numbers of spectra correctly assigned to the listed and grouped into L (light) and H (heavy) form.

			CE 120 ng/µl			CE 80 ng/µl			Batch 10 ng/µl		
Protein	Gene name	Protein size	Spectral Counts (L)	Spectral Counts (H)	H/L ratio	Spectral Counts (L)	Spectral Counts (H)	H/L ratio	Spectral Counts (L)	Spectral Counts (H)	H/L ratio
L1	rplA	24.7kD	45	21	0.47	34	22	0.65	35	13	0.37
L2	rplB	30kD	18	20	1.11	19	17	0.89	18	11	0.61
L3	rplC	22.3kD	13	15	1.15	12	13	1.08	15	9	0.60
L4	rplD	22.1kD	27	10	0.37	24	11	0.46	22	6	0.27
L5	rplE	20.3kD	39	5	0.13	39	6	0.15	38	4	0.11
L6	rplF	18.9kD	19	4	0.21	14	3	0.21	19	3	0.16
L9	rplI	15.8kD	40	18	0.45	34	13	0.38	33	11	0.33
L10	rplJ	17.7kD	14	7	0.50	11	5	0.45	12	2	0.17
L11	rplK	14.9kD	10	4	0.40	10	2	0.20	12	2	0.17
L12	rplL	12.3kD	25	3	0.12	26	3	0.12	19	3	0.16
L13	rplM	16kD	10	10	1.00	8	11	1.38	9	3	0.33
L14	rplN	13.5kD	12	7	0.58	11	7	0.64	12	3	0.25
L15	rplO	15kD	16	9	0.56	16	8	0.50	17	8	0.47
L16	rplP	15kD	6	10	1.67	4	6	1.50	3	3	1.00
L17	rplQ	14.4kD	14	8	0.57	9	5	0.56	11	5	0.45
L18	rplR	12.8kD	5	7	1.40	6	4	0.67	3	3	1.00
L19	rplS	13.1kD	20	5	0.25	25	3	0.12	20	0	0.00
L20	rplT	13.5kD	10	4	0.40	9	3	0.33	6	3	0.50
L21	rplU	11.6kD	6	1	0.17	7	1	0.14	3	1	0.33
L22	rplV	12.1kD	21	13	0.62	22	9	0.41	16	5	0.31
L23	rplW	11.2kD	3	3	1.00	3	1	0.33	3	1	0.33
L24	rplX	11.3kD	20	15	0.75	13	7	0.54	19	5	0.26
L25	rplY	10.7kD	13	6	0.46	10	3	0.30	10	1	0.10
L27	rpmA	9.1kD	7	2	0.29	5	1	0.20	4	0	0.00
L28	rpmB	9kD	3	3	1.00	5	1	0.20	2	1	0.50
L29	rpmC	7kD	5	2	0.40	2	2	1.00	4	1	0.25
L30	rpmD	6.5kD	3	0	0.00	4	0	0.00	3	0	0.00
L31	rpmE	7.9kD	5	5	1.00	3	4	1.33	5	2	0.40
L32	rpmF	6.4kD	4	7	1.75	2	4	2.00	0	1	-------
L33	rpmG	6.4kD	2	1	0.50	2	1	0.50	2	0	0.00
L34	rpmH	5.4kD	0	0	0.00	0	0	0.00	0	0	0.00
L35	rpmI	7.3kD	1	0	0.00	1	0	0.00	0	0	0.00
L36	rpmJ	4.4kD	0	0	0.00	0	0	0.00	0	0	0.00


Feeding
Chamber
Reaction
Chamber
IVT 30S
r-proteins
+
n16S rRNA
50S subunit
mRNA
Active
Ribosome

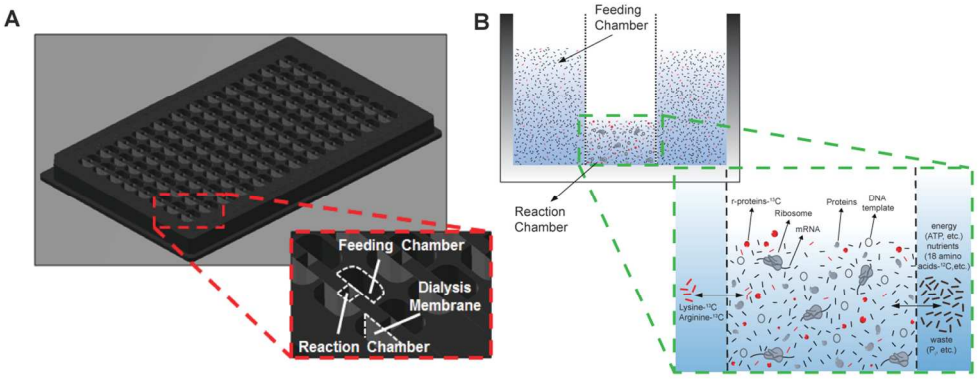


Figure 1. Illustration of the CE device mFAD and the SILACFPS r-protein co-expression strategy. (A) Three dimensional rendering of the 96-well mFAD. An expanded view of several protein expression units. As labeled, the reaction chamber is isolated from the two feeding chambers by vertically oriented dialysis membranes. (B) Cross-sectional illustration of one protein expression unit of the mFAD. The reaction chamber (central chamber) contains the PURE system capable of protein synthesis with 13C-labelled arginine and lysine, r-protein DNA templates along with the synthesized r-proteins in heavy form. The feeding solution consists of the energy and nutrient molecules including 13C-labelled arginine and lysine needed to sustain protein expression. An expanded view shows the different components for the reaction and feeding solutions.

165x65mm (220 x 220 DPI)

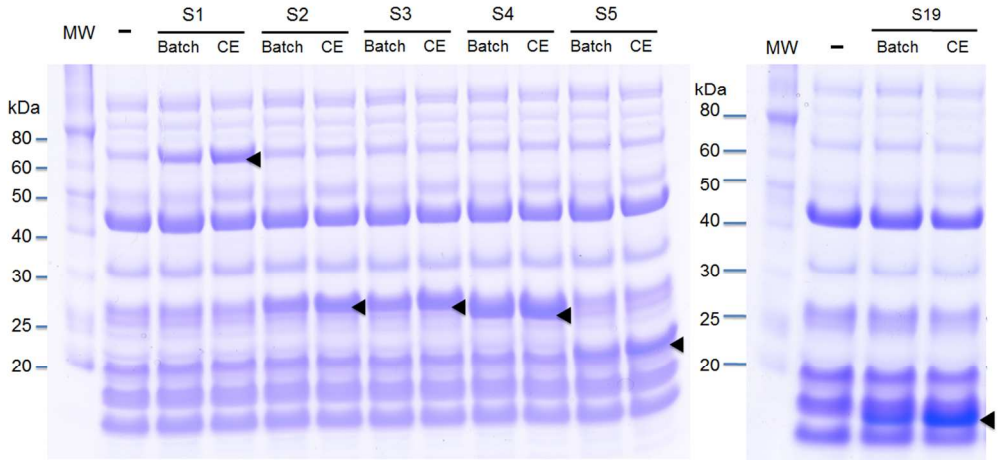


Figure 2. SDS PAGE analysis of sample 30S r-proteins S1, S2, S3, S4, S5 and S19 expressed in batch and CE mode PURE system by Coomassie Blue staining. The expected migration of each protein is marked on the gel. DNA template concentration of 10 ng/µl was used for the expression reaction for both batch and CE format. The resulting protein expression yields were quantified by the band intensity acquired by a typhoon scanner on SDS-PAGE.

341x158mm (96 x 96 DPI)

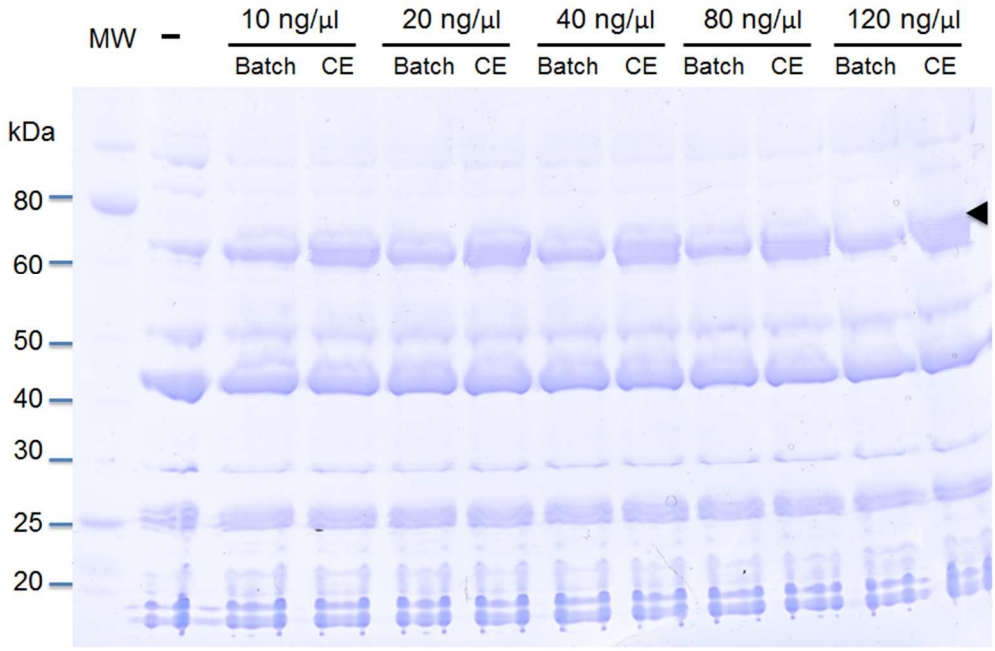


Figure 3. SDS-PAGE analysis of S1 synthesis in batch and CE format PURE system with different DNA template concentrations. The gel was stained by Coomassie Blue. The expected migration of each protein was marked on the gel. The CE format produced larger amount of S1 compared to the batch format, no matter which template concentration was used.

232x152mm (96 x 96 DPI)

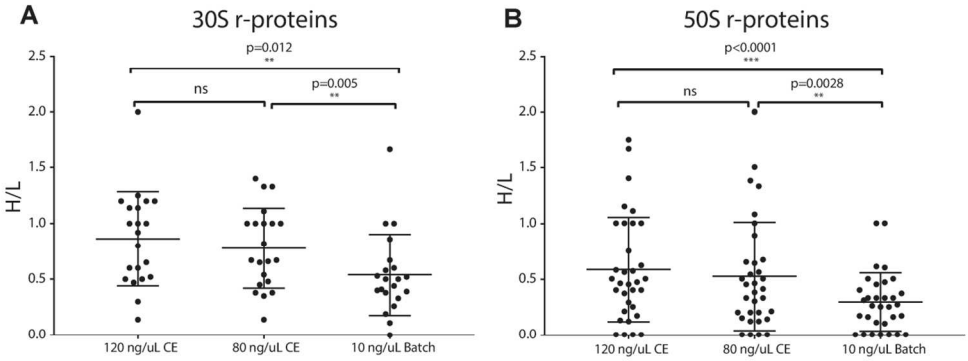


Figure 4. H/L ratio plots of r-proteins co-expressed in batch and CE format PURE system. (A) H/L ratio plots of 30S subunit r-proteins co-expressed in batch and CE format PURE system containing 13C-labelled arginine and lysine. (B) H/L ratio plots of 50S subunit r-proteins co-expressed in batch and CE format PURE system containing 13C-labelled arginine and lysine. DNA template concentrations of each protein in the reactions are indicated as above. P-values were calculated by a paired t-test.

164x73mm (220 x 220 DPI)

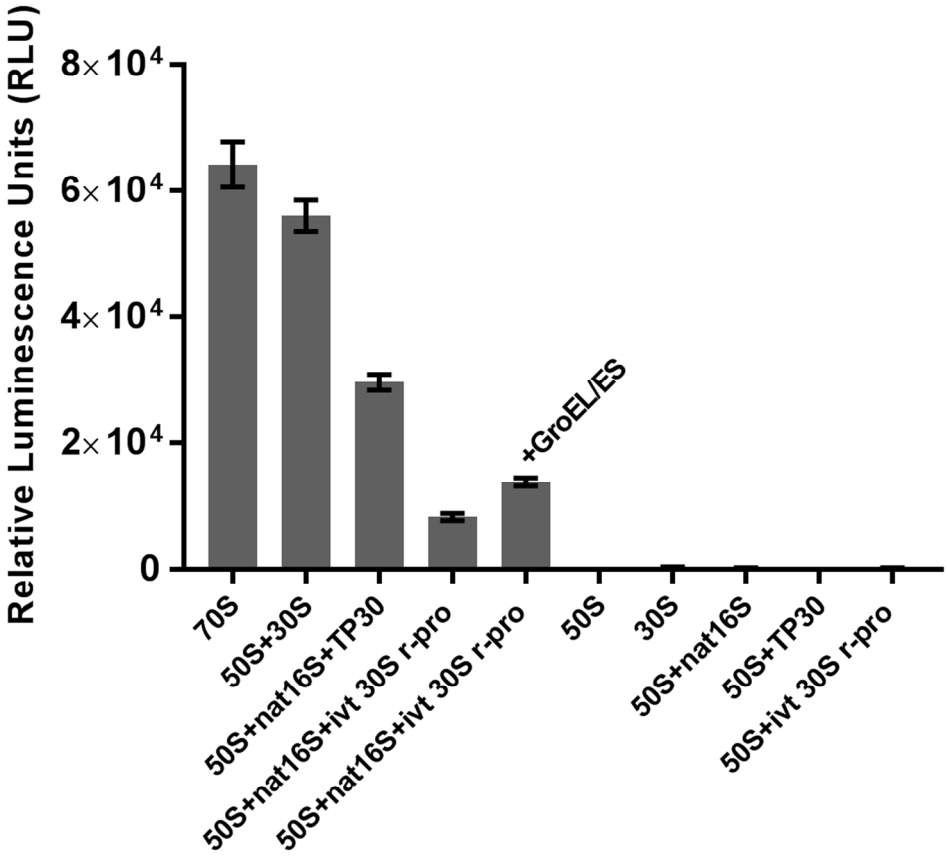


Figure 5. Reconstitution of 30S subunits in PURE Δ ribosome system under physiological condition. 70S ribosomes isolated from E. coli was taken as a positive control. 30S ribosome reconstitutions from components as labelled in the figure were measured by luciferase production. The last five columns were negative control reactions. For reconstitution reactions, 0.3 μM 70S, 0.3 μM 50S, 0.3 μM 30S, 0.3 μM native 16S rRNA (nat16S), 0.9 μM TP30, 0.9 μM PURE system synthesized r-proteins (ivt 30S r-pro) and 0.6 μM of GroEL/ES were added as indicated. Reactions were incubated at 37°C for 2 h. Luciferase activities were measured in relative luminescence unit. Error bars are ± standard deviations, with n=5.

314x287mm (96 x 96 DPI)