Holistic engineering of cell-free systems through proteome-reprogramming synthetic circuits.

Permalink
https://escholarship.org/uc/item/6b5317b1

Journal
Nature communications, 11(1)

ISSN
2041-1723

Authors
Contreras-Llano, Luis E
Meyer, Conary
Liu, Yao
et al.

Publication Date
2020-06-19

DOI
10.1038/s41467-020-16900-7

Peer reviewed
Holistic engineering of cell-free systems through proteome-reprogramming synthetic circuits

Luis E. Contreras-Llano1,4, Conary Meyer1,4, Yao Liu1, Mridul Sarker2, Sierin Lim2, Marjorie L. Longo3 & Cheemeng Tan1

Synthetic biology has focused on engineering genetic modules that operate orthogonally from the host cells. A synthetic biological module, however, can be designed to reprogram the host proteome, which in turn enhances the function of the synthetic module. Here, we apply this holistic synthetic biology concept to the engineering of cell-free systems by exploiting the crosstalk between metabolic networks in cells, leading to a protein environment more favorable for protein synthesis. Specifically, we show that local modules expressing translation machinery can reprogram the bacterial proteome, changing the expression levels of more than 700 proteins. The resultant feedback generates a cell-free system that can synthesize fluorescent reporters, protein nanocages, and the gene-editing nuclease Cas9, with up to 5-fold higher expression level than classical cell-free systems. Our work demonstrates a holistic approach that integrates synthetic and systems biology concepts to achieve outcomes not possible by only local, orthogonal circuits.
The mantra of synthetic biology advocates for the use of orthogonal genetic modules to engineer and control cellular behavior. However, the use of orthogonal genetic modules often faces the challenges of varying cellular context, such as growth rate, crosstalk, and noise. These challenges highlight the necessity to complement orthogonal genetic module design with a system-based approach that functions in conjunction with cell physiology. Such systems-synthetic biology approaches have been applied in two major ways. First, systems-level properties can be considered for the control of local synthetic modules. For instance, previous studies have investigated the impact of global physiology on cell-free protein synthesis (CFPS). Second, the global host circuits are modified before the insertion of local synthetic modules. One classic example is the gene knockout and knockin of the BL21 Escherichia coli strain for subsequent conversion into BL21 (DE3) using lacUV5-T7 RNAP-based synthetic modules.

A powerful alternative, referred to as the holistic synthetic biology approach in this work, is to use local synthetic modules to reprogram the global host physiology, which in turn becomes beneficial to the function of the local synthetic modules.

Here, we apply the holistic synthetic biology approach to the engineering of CFPS systems. CFPS decouples cellular growth from protein production, allowing for applications such as synthesis of toxic or metabolically interfering proteins. Paper-based diagnostics, a priori prediction of metabolic burden, and function of genetic circuits have been used to modify the proteome of source bacteria through gene knockins or knockouts. While these approaches have been added to CFPS reactions, in addition, genome recoding methods, Sections M2, M3, and M6, have also been added to CFPS reactions. In addition, genome recoding approaches have been used to modify the proteome of source bacteria through gene knockins or knockouts. While these approaches can precisely change the concentration of a few proteins, they are challenging to scale up for targeting multiple pathways that can impact CFPS. Furthermore, manipulating the expression levels of many essential genes while maintaining cell viability is often inhibitory complex.

A holistic synthetic biology approach could overcome this challenge.

**Results**

**Enhanced CFPS via overexpression of translation machinery.** To implement the holistic approach, we used synthetic modules to express all or a subset of the 34 proteins of the core E. coli translation machinery within multiple strains of E. coli BL21 (DE3) that were lysed and used in CFPS. Distribution of protein overexpression across multiple strains was chosen to decrease the metabolic burden caused by protein expression and plasmid maintenance. The burden imposed by plasmid maintenance manifests in the form of decreased growth rates, which in turn generates lower concentrations of ribosomes and other translation machinery proteins. This has been shown to be a limiting factor for efficient CFPS.

We hypothesized that the overexpression of translation machinery should benefit CFPS in two ways. First, it should compensate for the increased metabolic burden by virtue of being supplied with translation factors. And second, it should shift the global proteome to a high-growth-rate-like state where translation factors are enriched, and the cell reaches peak protein synthesis efficiency. We produced two different microbial consortia, one with 18 strains (BL-18S) and the other with 7 strains (BL-7S) to obtain cell lysates enriched in translation machinery without the need to purify and supplement individual proteins. BL-18S expressed 11 initiation, elongation, and termination factors (IETFs), as well as 23 aminoacyl-tRNA transferases (AAT). BL-7S expressed 11 IETFs and 1 AAT (Fig. 1a and Supplementary Table 1). Throughout this study, we used the expression level of deGFP, a truncated version of eGFP with the same fluorescence properties to quantify the absolute yield of the CFPS (Supplementary Fig. 6, “Methods, Section M4”, and Supplementary Note 1). In addition, we optimized the reaction buffer and lysate preparation for the new CFPS (Supplementary Figs. 1–4 and Supplementary Note 2).

To compare our modified extracts to existing systems, we ran several experiments to quantify the differences. The whole-cell lysate of BL-18S (BL-18SWCE) and BL-7S (BL-7SWCE) had comparable expression activities (Supplementary Fig. 5). Thus, we proceeded with BL-7SWCE due to its simpler preparation procedure. To assess the influence of translation machinery overexpression in CFPS, we compared the protein yield against a cell lysate produced using E. coli BL21 (DE3) without any plasmids (BL-EWCE), a cell lysate from the same strain carrying the original plasmid vectors (BL-PWCE), and the commercial S30 T7 High-Yield Expression System (Promega Corporation) (S30) (“Methods, Section M2”). In batch reaction mode, BL-7SWCE produced a maximum of 1.51 mg mL\(^{-1}\) of deGFP (Fig. 1b), and 4.8 mg mL\(^{-1}\) in a semicontinuous exchange mode (Fig. 1c). The S30 expression system performed poorly when adapted to the semicontinuous exchange mode. Therefore, the data were not included because the protocol for this setup was not defined by the manufacturer. The yields of BL-7SWCE were two- to threefold higher than the controls in both formats. When examining the expression dynamics (Fig. 1d), CFPS reactions assembled using our in-cell lab lysates show a 20 min lag-period before the production of deGFP can be detected. During this initial lag-period, the transcription machinery (T7 RNAP) likely ramped up mRNA synthesis until the mRNA reached the concentration necessary for starting protein synthesis. Once protein synthesis was started, reactions assembled using BL-7SWCE expressed deGFP at a higher rate than other cell lysates. Altogether, the data show that BL-7SWCE can achieve higher expression levels than conventional systems.

**Effect of translation machinery concentration on CFPS.** Our next set of experiments are intended to investigate the cause of the improved CFPS efficiency. Specifically, we sought to decouple the direct effect of increasing the translation factor concentrations in the CFPS reaction from the indirect effects of protein overexpression and feedback from the overexpressed proteins. To study the effects of the increased translation factors in a standard CFPS reaction, we purified the translation machinery proteins overexpressed in BL-18S and supplemented it to BL-EWCE (Fig. 2a, “Methods, Sections M3 and M6”). The expression level of deGFP increased proportionally with the addition of translation machinery (Fig. 2b). These results demonstrate that the increased concentration of protein machinery is not the only factor responsible for the increased protein expression of our multi-strain CFPS systems. Furthermore, we intended to rule out any additional effects that plasmid maintenance or protein overexpression could be causing in our multi-strain CFPS systems. To rule out the plausible effects, we purified translation machinery proteins (overexpressed in BL-7S) and supplemented them to an extract that was generated from BL21 (DE3) overexpressing cyan fluorescent protein (CFP) (BL-CFPPWCE) (“Methods, Sections M2, M3, and M6”). The expression level once again increased proportionally with the amount of protein added but plateaued at a twofold increase (Fig. 2c). Our results show that...
The concentrations of translation machinery are comparable between single-strain preparations supplemented with purified translation machinery and multi-strain preparations (Supplementary Note 3 and Supplementary Fig. 7), the yields obtained in CFPS are not equivalent. These results are consistent with our hypothesis that the overexpression of translation machinery causes an auxiliary effect on the host circuits that create an environment more favorable for CFPS.

Influence of protein overexpression on the host proteome. The above results suggest that the protein profile in our BL-7S WCE is more amenable to CFPS than any of the controls. Specifically, our data indicate that this proteome reprogramming occurs directly as a result of the overexpression of translation machinery by our synthetic modules. To further understand the favorable changes occurring in BL-7S, we analyzed the protein composition of five overexpressed proteins (EF-Tu and EF-Ts) occurring in BL-7S, we analyzed the protein composition of several whole-cell lysates through mass spectrometry ("Methods, Section M7"). For this experiment, an additional extract using one of the IET strains was created, specifically Strain-1 that overexpresses the elongation factors EF-Tu and EF-Ts (BL-1S WCE). We decided to analyze Strain-1 of our 7-strain consortium due to the major roles of EF-Tu and EF-Ts for increasing elongation rates in CFPS and because it represents 50% of the inoculation mixture (Supplementary Table 1). All four
were upregulated. Even though BL-CFP showed a proteome shift, the proteome of BL-7S and BL-1S showed a decrease in the levels of up- and downregulated proteins (changes >25%, mostly unchanged with a nearly even split between the number of upregulated and downregulated proteins). These results show that the proteome of BL-7S was affected by the expression of translation machinery. This change in protein profile and content likely results in the generation of an environment more favorable for CFPS.

We further characterized the proteome change uncovered by our mass spectrometry results. To this end, we categorized each protein based on their assigned gene ontological function (“Methods, Section M7”). We then summed the intensity of each protein in each category for protein content comparison (Fig. 3d). Again, the proteome of BL-CFP and BL-E exhibited no significant difference. However, BL-7S and BL-1S exhibited a 17% increase in the gene expression category (e.g., translation factors, aminoacyl tRNA synthetases, ribosomes). They also showed a decrease of 14% in the metabolism (e.g., TCA cycle and amino acid catabolism) and 3% in the homeostasis (e.g., iron homeostasis, proteases, and cell cycle regulators) categories. To better understand the specific proteome changes, more detailed functions were assigned to the proteins. The fold changes between the means of each protein in each category were compared with the BL-E control. The proteins were then grouped by their function, and the average fold change of each protein was calculated. On the other hand, the proteome of BL-7S and BL-1S showed a decrease in over a third of all observed proteins, while <5% of all proteins were upregulated. Even though BL-CFP showed a proteome shift, it is likely caused by the metabolic burden of protein overexpression, and the proteome change did not boost the yield of BL-CFPWCE.

These results show that the proteome of BL-7S was affected by the expression of the translational machinery. This change in protein profile and content likely results in the generation of an environment more favorable for CFPS.
ribosomal proteins); and downregulation of metabolic proteins that compete with nutrients in CFPS (e.g., tryptophanase and pyruvate kinase) (Fig. 3f). However, we note some exceptions to this general expectation, such as increases in a few metabolic proteins, including glycerol-3-phosphate acyltransferase and 2,3-dihydroxybenzoate-AMP ligase (Supplementary Note 4).

**Demonstrating the versatility of enhanced CFPS.** To explore the potential of our BL-7S_WCE beyond the enhanced deGFP expression (Fig. 1b), we decided to test its versatility through different applications. For our first trial, we produced ferritin from *Archaeoglobus fulgidus* (AfFtn), an archaeal iron storage protein capable of self-assembly forming nanocages. AfFtn has been...
shown to encapsulate and release molecular cargo. As AfFtn requires the assembly of precisely 24 subunits of 22 kDa to form nanocages, it is a good test case for the CFPS system to produce large protein assemblies while maintaining its function. Reactions assembled with BL-7SWCE expressed 50% more ferritin than our controls assembled with BL-EWCE (Fig. 4a and Supplementary Fig. 10, A, B). TEM images demonstrate the AfFtn nanocages of 12 nm (Fig. 4b and Supplementary Fig. 10C). The iron core formation in the unstained TEM images confirms the function of the produced AfFtn.

One of the major challenges for E. coli-based CFPS systems is their limited ability to efficiently synthesize large proteins. This problem becomes particularly pronounced in the expression of proteins larger than 70 kDa. We decided to test if our multi-strain system offers an advantage over traditional approaches in this task. Thus, we expressed the biotechnologically relevant protein Cas9 (159 kDa) and compared its expression against BL-EWCE and S30 (Fig. 4c and Supplementary Fig. 11). Our BL-7SWCE produced 0.52 mg mL$^{-1}$ of Cas9. This is approximately threefold higher than BL-EWCE and approximately fivefold higher than S30. These results show that our system can synthesize broad sizes of proteins between 20 and 160 kDa in higher quantities than conventional systems. The modularity of the bacterial consortium enables the incorporation of additional strains in our system as a plug-and-play feature. By exploiting this feature, we could confer a new function to our cell-free system, such as expressing proteins from linear templates. To implement this, we added a strain expressing the Gam protein (a nuclease inhibitor), resulting in an 8-strain WCE (BL-8S-GamWCE; "Methods, Section M2"). We used rolling circle amplification (RCA) to generate the deGFP encoding template ("Methods, Section M3"). The resulting double-stranded linear DNA was added directly (21% V/V) into three different cell lysates: two BL-8S-GamWCE with different inoculation ratios of the Gam-expressing strain (1:1 and 1:5), and controls without Gam (BL-7SWCE). Using the amplified linear DNA, the two BL-8S-GamWCE synthesized approximately twofold more deGFP than the BL-7SWCE controls (Fig. 4d). The deGFP expression levels increased proportionally with the amount of Gam-expressing strain. The maximum yield achieved using the linear template in reactions assembled with BL-8S-GamWCE is ~75% of the yield achieved using BL-7SWCE and plasmid template (Fig. 4c).

The results of this experiment show how the modularity of our bacterial consortium approach can be exploited to custom modify cell lysates to match the requirements of a given experiment. However, further optimization and benchmarking may be necessary.
necessary to make a fair comparison between our and commercial CFPS systems that are designed for specific applications. Altogether, these applications demonstrate the power of our holistic synthetic biology approach in generating versatile high-yield CFPS systems.

Discussion

Our work highlights both the utility and the potential of holistic synthetic biology approaches in boosting the performance of local synthetic modules. We demonstrate that the protome reprogramming described in our study is the direct result of the overexpression of translation machinery in the host cells. Furthermore, we show that the use of a plasmid system in the source strain does not result in a decrease in the activity of the CFPS system. This illustrates how plasmid-based approaches can be implemented to functionalize cell lysates without sacrificing CFPS efficiency. This study opens a new research direction in cell-free synthetic biology, showcasing how the integration of orthogonal circuits, cell physiology, and systems biology can become a powerful tool that maximizes the output of a given cell-free system. Similar approaches have been used for transcriptional rewiring with the aim of increasing the production of proteins free system. Similar approaches have been used for transcriptional rewiring with the aim of increasing the production of proteins free system. Similar approaches have been used for transcriptional rewiring with the aim of increasing the production of proteins free system. Similar approaches have been used for transcriptional rewiring with the aim of increasing the production of proteins free system. Similar approaches have been used for transcriptional rewiring with the aim of increasing the production of proteins free system.

Methods

M1: Construction of plasmids and strains. We used the plasmids pJVE2x.3d (Roche), pET15b (Novagen), pLysS (Novagen), and pSC10139 as the backbones for all our constructions. The backbones of pET15b, pLysS, and pSC101 were used to create the plasmids pUURAH, pUURCM, and pUURKL, respectively. Briefly, the three plasmids have compatible replication origins, distinct copy number, carry a NsiI/PalCl cloning site downstream of a PT7–lacI hybrid promoter, and have a T7 RNAP terminator sequence. pUURAH contains the ampicillin resistance gene/ColEI replication origin and expresses lacI, pUURCM contains the chloramphenicol resistance gene/p15A replication origin and expresses T7 lysozyme, and pUURKL contains kanamycin resistance gene/pSC101 replication origin. The plasmids pUURAH and pUURKL were used as backbones to generate all 34 vectors encoding translation machinery expressing plasmids were made by Villareal et al. and are available through Addgene.

Methods

M2: Preparation of whole-cell extract. For our whole-cell extract preparations, we variate the specific strain or consortium used and the inoculation ratios (ratio represent % of the strain in the total volume of the mix). Culture and induction times and all subsequent steps were made generic among all preparations.

BL-7wC and BL-18wC were prepared using the following protocol: each strain comprising the 7 or the 18-strain consortium was individually grown in 3 mL of 2YTP media supplemented with carbenicillin/chloramphenicol/kanamycin at 37 °C with shaking at 200 rpm overnight. The overnight cultures were used to establish the BL-7w or BL-18w consortia by mixing strains at the indicated ratios (See Supplementary Table 1). The mixtures were then used to inoculate 300 mL of 2YTP media supplemented with carbenicillin and kanamycin at a 1/250 dilution.

BL-EwC was prepared using the following protocol: the strain BL21 (DE3) was grown in 3 mL of 2YTP media at 37 °C with shaking at 200 rpm overnight. The saturated overnight culture was then used to inoculate 300 mL of 2YTP media supplemented with carbenicillin and kanamycin at a 1/250 dilution.

BL-PwC was prepared using the following protocol: the strain BL21 (DE3) transformed with the plasmids pUURAH, pUURCM, and pUURKL was grown in 3 mL of 2YTP media supplemented with carbenicillin/chloramphenicol/kanamycin at 37 °C with shaking at 200 rpm overnight. The saturated overnight culture was then used to inoculate 300 mL of 2YTP media supplemented with carbenicillin and kanamycin at a 1/250 dilution.

BL-SwC was prepared using the following protocol: the strain BL21 (DE3) transformed with the plasmids pET15bL-CFP, pUURCM, and pUURKL was grown in 3 mL of 2YTP media supplemented with carbenicillin/chloramphenicol/kanamycin at 37 °C with shaking at 200 rpm overnight. The saturated overnight culture was then used to inoculate 300 mL of 2YTP media supplemented with carbenicillin and kanamycin at a 1/250 dilution.

M3: Assembly of CFPS reactions. The assembly of CFPS reactions for batch experiments was carried out as follows: CFPS reactions (10 μL) were assembled in 1.5 mL low protein binding microcentrifuge tubes (Thermo Scientific) by mixing the following components: 1.2 mM each of ATP and GTP; 0.85 mM each of UTP and CTP (Promega); 34 μg mL⁻¹ folinic acid (Sigma–Aldrich); 170 μg mL⁻¹ E. coli tRNA mixture from E. coli MR600 (Roche); 2 mM each of the 20 standard amino acids (Sigma–Aldrich); 0.33 mM NAD (Roche); 0.27 mM CoA (Sigma–Aldrich); 4 mM spermidine (Sigma–Aldrich); 180 mM potassium glutamate (Sigma–Aldrich); 12 mM magnesium glutamate (Sigma–Aldrich); 50 mM HEPS pH 7.6 (Sigma); 67 mM creatine phosphate (Roche); 80 μg mL⁻¹ creatine kinase (Roche); 0.64 mM cAMP (Sigma–Aldrich); 2% PEG8k (Sigma–Aldrich); 0.2 mg mL⁻¹ BSA;
2.7 μL (27% v/v) of cell extract, and 100 ng plasmid DNA. Each CPS reaction was collected on ice and incubated overnight at 30 °C with shaking at 300 rpm unless noted otherwise. As individual reagent concentrations were optimized, their optimal value listed above were used for all reactions from that point onward. The assembly of reactions supplemented with purified translation machinery mixtures was carried out as follows: reactions were assembled as described above and supplemented with varying amounts of purified translation machinery mixtures such as the translation machinery proteins overexpressed in BL-18S (Methods, Section M6) to a CPS reaction assembled with BL-EWCE. For the experiments in Fig. 2c, we supplemented the 11 translation machinery proteins overexpressed in BL-7S (Methods, Section M6) to the CPS reaction assembled with BL-EWCE. Some authors of proteins did not affect the final concentration of any of the components in the CPS reactions. Negative controls were assembled using the same volume of Buffer A than the volume of supplemented translation machinery mixtures (Methods, Section M6).

The assembly of reactions under semicontinuous agitation was carried out as follows: reactions were scaled up to 15 μL, assembled into 1.5 mL low protein binding microcentrifuge tubes, and transferred to a 384-well plate (Corning). Once all the reactions were loaded into the plate, the wells were sealed with film and the plate was loaded into an m1000Pro Infinite plate reader to measure fluorescence. Reactions were incubated at 30 °C with semicontinuous sharking at 30 rpm (30 s ON, 30 s OFF) for 12 h. Fluorescence was measured every 10 min and followed for 12 h. Note: the yield of all reactions carried out in 384-well plate format under semicontinuous agitation was considerably lower compared with control batch reactions carried out in parallel. This decrease in yield is consistent across all our in-lab cell lysates and points out to agitation as a crucial parameter for achieving high-yield protein expression.

The assembly of CPS reactions using a linear template was carried out as follows: we amplified 1 ng of the plasmid pIVEX-Eps-deGFP using the commercial kit TempliPhi for RCA according to the manufacturer’s instructions (GE Healthcare, UK). The resulting double-stranded linear DNA template was directly added (21% V/V) to CFPS reactions assembled using BL-7SWCE or BL-8S-GamWCE. This is the maximum percentage by volume that we could add to the CPS reactions without perturbing the concentrations of the rest of the components. The precise concentration of DNA obtained through RCA could not be quantified using a Nanodrop spectrophotometer. This issue is because even in the absence of input DNA, the RCA reaction yields nonspecific products. However, according to the manufacturer’s indications, we estimate that the amount of double-stranded linear DNA added to each CPS reaction is between 150 and 500 ng.

M4: Quantification of deGFP expression. Fluorescent measurements were taken of CPS reactions diluted 1:50 in dilution buffer (50 mM HEPES pH 7.6, 4 mM spermidine, 2% PEG8k, 12 mM magnesium glutamate, 180 mM potassium glutamate, and 0.4 mg mL^-1 BSA). Active deGFP protein yields were quantified by measuring fluorescence using a NanoQuant plate (Tecan) and an m1000Pro finite plate reader. Excitation and emission wavelength used to measure the fluorescence of deGFP were 488 and 507 nm, respectively. deGFP fluorescence units were converted to concentration using a standard curve.

M5: Semicontinuous exchange reaction. The semicontinuous reactions were conducted using the 96-Well Equilibrium dialyzer (MWCO 10 kDa) purchased from Harvard Apparatus (Holliston, MA). Reactions were set up with 20 μL cell-free reactions loaded on one side of the dialyzer with 200 μL of feeding solution on the other. The feed solution has the same composition as cell-free reaction buffer. Finally, 50 mM TEAB with 6 μg of trypsin (1:25 ratio) is added and the sample is incubated overnight with one addition of 50 mM TEAB with trypsin after 2 h. The following day the digested peptides were released from the S-trap solid phase support by vacuum drying at 30 °C for 1 min with a series of solutions starting with 50 mM TEAB which is placed on top of the digestion solution, then 5% formic acid followed by 50% acetonitrile with 0.1% formic acid. The solution is then vacuum centrifuged to almost dryness and resuspended in 2% acetonitrile 0.1% trifluoroacetic acid and subjected to fluorescent peptide quantification (Thermo Scientific).

M7: Mass spectrometry. The following protocol was used for peptide sample preparation: the proteins in the whole-cell extract preparations were quantified using a NanoDrop (Thermo Scientific) by measuring the absorbance at 280 nm. The equivalent amount of protein was used for S-Trap (PROTIFIT) digestion. Digestion followed the S-trap protocol: briefly, the proteins were reduced and alkylated, the buffer concentrations were adjusted to a final concentration of 5% SDS 50 mM TEAB, 12% phosphoric acid was added at 1:1 ratio with a final concentration of 1.2% and S-trap buffer (100 mM TEAB in 99% water at 1:10 V/V ratio) was added at a 1:1 ratio (V/V ratio). The protein mixture was then spun through the S-trap column and washed 3 times with S-trap buffer. Finally, 50 mM TEAB with 6 μg of trypsin (1:25 ratio) is added and all labeled samples were combined with each TMT-10plex was pooled together to create a reference to account for bias between the two TMT runs. Each sample was labeled with the TMT-10plex Mass Tag Labeling Kit (Thermo Scientific). Briefly, 20 μL of each TMT label (126-131) was added to each digested peptide sample and incubated for an hour. The reaction was quenched with 1 μL of 5% hydroxyamine and incubated for 15 min. All labeled samples were then mixed and lyophilized to almost dryness. The TMT labeled sample was reconstituted, desalted, and separated into eight fractions by high pH fractionation (Thermo Scientific). One-third of each fraction (~800 ng) was used for LC-MS/MS analysis.

The following protocol was used for liquid chromatography and mass spectrometry of the samples: liquid chromatographic separation was conducted on a Dionex nano Ultimate 3000 (Thermo Scientific) with a Thermos Easy-Spray source. The digested peptides were reconstituted in 2% acetonitrile/0.1% trifluoroacetic acid and 1 μg in 5 μL of each sample was loaded onto a PepMap 100 A 3U 75 μm × 20 mm reverse-phase trap where they were desalted online before being separated on a 100 Å 2U 50 μm × 150 mm PepMap Easy-Spray reverse-phase column. Peptides were eluted using a 120-min gradient of 0.1% formic acid (A) and 80% acetonitrile (B) with a flow rate of 200 nL/min. The separation gradient was run with 2–5% B over 1 min, 5–50% B over 89 min, 50–99% B over 2 min, a 4-min hold at 99% B, and finally 99% B to 2% B held at 2% B for 18 min.

The following protocol was used for mass spectra acquisition: mass spectra were collected on a Fusion Lumos mass spectrometer (Thermo Fisher Scientific) in a data-dependent MS3 synchronous precursor selection method. MS1 spectra were acquired in the Orbitrap, 120 °C declustering, 50 ms max injection time, 5 × 105 max injection time. MS2 spectra were acquired in the linear ion trap with a 0.7 Da isolation window, CID fragmentation energy of 35%, turbo scan speed, 50 ms max injection time, 1 × 104 AGC, and maximum parallelizable time turned on. MS2 ions were isolated in the ion trap and fragmented with an HCD energy of 65%. MS3 spectra were acquired in the orbitrap with a resolution of 50 K and a scan range of 100–500 Da, 105 ms max injection time, and 1 × 105 AGC.

The following process was followed for peptide and protein identification: identification of peptides and proteins was conducted using the PAW pipeline. In brief, the ProteoWizard toolkit is used to convert the MS scans into intensity values and protein TMT reporter for each file. The Comet database-searching engine is then used to identify peptides. The E. coli BL21 (DE3) proteome UP00002032 and a list of known contaminants and expressed protein sequences were used for protein identification. Results are filtered based on a desired false discovery rate using the target decoy method. Identified proteins with sequence coverage of <5% were excluded from the downstream analysis.
The following process was used to scale the two TMT results: the protein intensities from the pooled samples in each 10plex were used to calculate scaling factors that can be applied to the intensity values from each sample in each TMT-10plex, eliminating the bias that results from independent MS runs.46 The data are presented before and after normalization in Supplementary Fig. 8.

The following process was used for the assignment of gene ontological function: identified proteins were assigned gene ontological functions based on the gene ontology identifiers provided in the E. coli BL21 (DE3) proteomes UP000002022. The gene ontology identifiers were grouped based on the general functional categories of interest. The specific identifiers used for this assignment are detailed in Supplementary Data 1. Proteins that lacked identifiers or only possessed broad identifiers were classified as Unknown. Several proteins with known functions that lacked identifiers were manually assigned the correct functional group. These manual assignments are also detailed in Supplementary Data 1.

M8: Protein quantification and SDS-PAGE analysis. Analysis of proteins by SDS-Polyacrylamide Gel Electrophoresis (PAGE) was carried out by separating proteins from whole-cell lysates and CPRS reactions using 4–20% Mini-PROTEAN TGX precast gels (Bio-Rad). We used Precision Plus Protein Dual Color Standards (10–250 kDa) as a reference standard for molecular weight verification. Protein gels were endpoint stained using PageBlue Protein Staining Solution (Thermo Fisher Scientific) according to the manufacturer instructions. Gels were imaged using a PhorImager system (Syngene) and band analysis and protein quantification were carried out using the open-source platform for biological imaging analysis Fiji (http://fiji.sc/CGI-bio/gtweb.cgi/) and the proprietary software GeneTools (Syngene).

M9: Transmission electron microscopy. The following protocol was used for the assembly of the ferritin nanocage: 1 mL of Denaturation Buffer (25 mM HEPES, pH 7.5, 350 mM NaCl, 15 mM MgCl2) was added to the cell-free reaction mixture and then heat-treated at 90°C for 10 min. Ferric sulfate heptahydrate was added drop by drop to a final concentration of 2.4 mM. The sample is then incubated overnight at 4°C. Dynamic light scattering and TEM analyses were performed to confirm the cage assembly.

The following protocol was used for TEM sample preparation: samples were adsorbed on to the carbon-coated electron microscopy grid (Formvar carbon film on 300 mesh copper grids, Electron Microscopy Science) and dried at room temperature for 5 min. For samples with negative staining, the grid was placed on a drop of uranyl acetate for 3 min and the excess stain was removed with a soft wipe. The grid was air-dried for 5 min. All grids were stored in a dry-cleaning cabinet until further use. The images were obtained in a transmission electron microscope (JEOL JEM-14000) operating at 100 kV.

M10: Statistical analysis of results. Unless otherwise specified, statistical tests were performed using a standard two-tailed t-test. The exact p value for each statistical analysis is reported directly in the figures unless p < 0.00001. The number of replicates contributing to the calculation is listed in the figure legends. All error bars and measures of central tendency are defined in the figure legends.

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

Data availability. All data used to generate Figs. 1b–d, 2b, c, 3a–e, 4a, c, and d, and Supplementary Figs. 1A–D, 2A–D, 3, 4, 5, 6, 7, and 9 in this paper are included as a Source Data File. The identified protein abundances and internal reference normalized data from the proteomics study are provided in Supplementary Data 1. This file also includes the specifications for gene ontological assignments of each protein. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE46,47 partner repository with the dataset identifier PXD018858 [https://www.ebi.ac.uk/pride/archive/projects/PXD018858]. The plasmids used to make the strains included in this study are available at [https://www.addgene.org/Cheemeng_Tan/]. Any other relevant data are available from the authors upon reasonable request.

Code availability. All custom code used to interpret and analyze the protein abundances was deposited in Github and its publicly available at [https://github.com/ccmeyer/TMT-analysis].

Received: 5 March 2020; Accepted: 27 May 2020; Published online: 19 June 2020

References
1. Brophy, J. A. N. & Voigt, C. A. Principles of genetic circuit design. Nat. Methods 11, 508–520 (2014).
2. Cardinale, S. & Arkin, A. P. Contextualizing context for synthetic biology—identifying causes of failure of synthetic biological systems. Biotechnol. J. 7, 856–866 (2012).
3. Purnick, P. E. M. & Weiss, R. The second wave of synthetic biology: from modules to systems. Nat. Rev. Mol. Cell Biol. 10, 410–422 (2009).
4. Bosdriesz, E., Molenaar, D., Teusink, B. & Bruggeman, F. J. How fast-growing bacteria robustly tune their ribosome concentration to approximate growth-rate maximization. FERS J 2029–2044 (2015).
5. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. 189, 113–130 (1986).
6. Lim, H. J. et al. Cell-free synthesis of functional phospholipase A1 from Serratia sp. Biotechnol. Biofuels 9, 1–7 (2016).
7. Thorning, L., Dondapati, S. K., Stech, M., Wisniewska, D. A. & Kubick, S. High yield production of ‘difficult-to-express’ proteins in a continuous exchange cell-free system based on CHO cell lysates. Sci. Rep. 7, 1–15 (2017).
8. Pardeé, K. et al. Paper-based synthetic gene networks. Cell 159, 940–954 (2014).
9. Takahashi, M. K. et al. A low-cost paper-based synthetic biology platform for analyzing gut microbiota and host biomarkers. Nat. Commun. 9, 3347 (2018).
10. Borkowski, O. et al. Cell-free prediction of protein expression costs for growing cells. Nat. Commun. 9, 1457 (2018).
11. Chappell, J., Jensen, K. & Freumont, P. S. Validation of an entirely in vitro approach for rapid prototyping of DNA regulatory elements for synthetic biology. Nucleic Acids Res. 41, 3471–3481 (2013).
12. Sun, Z. Z., Yeung, E., Hayes, C. A., Noireaux, V. & Murray, R. M. Linear DNA for rapid prototyping of synthetic biological circuits in an Escherichia coli based TX-TL cell-free system. ACS Synth. Biol. 3, 387–397 (2014).
13. Contreras-Llano, L. E. & Tan, C. High-throughput screening of biomolecules using cell-free gene expression systems. Synth. Biol. 3, 85012 (2018).
14. Ding, X., Contreras-Llano, L. E., Morris, E., Miao, M. & Tan, C. Minimizing context dependency of gene networks using artificial cells. Appl. Mater. Interfaces 10, 30137–30146 (2018).
15. Hong, S. H. et al. Improving cell-free protein synthesis through genome engineering of Escherichia coli lacking release factor 1. ChemBioChem 16, 844–853 (2015).
16. Schoborg, J. A., Clark, L. G., Choudhury, A., Hodgman, C. E. & Jewett, M. C. Yeast knockout library allows for efficient testing of genomic mutations for cell-free protein synthesis. Synth. Syst. Biotechnol. 1, 2–6 (2016).
17. Jiang, X. et al. Reduction of protein degradation by use of protease-deficient mutants in cell-free protein synthesis system of Escherichia coli. J. Biosci. Bioeng. 93, 151–156 (2002).
18. Goerke, A. R., Loening, A. M., Gambhir, S. S. & Swartz, J. R. Cell-free metabolic engineering promotes high-level production of bioactive Gauussia princeps luciferase. Metab. Eng. 10, 187–200 (2008).
19. Michel-Reydellet, N., Calhoun, K. & Swartz, J. Amino acid stabilization for cell-free protein synthesis by modification of the Escherichia coli genome. Metab. Eng. 6, 197–203 (2004).
20. Calhoun, K. A. & Swartz, J. R. Total amino acid stabilization during cell-free protein synthesis reactions. J. Biotechnol. 123, 193–203 (2006).
21. Tsalikova, T. et al. GroEL and GroES increase the specific enzymatic activity of newly-synthesized rhodanese if present during in vitro transcription/translation. Biochemistry 32, 3377–3380 (1993).
22. Ying, B. W., Taguchi, H., Kondo, M. & Ueda, T. Co-translational involvement of the chaperonin GroEL in the folding of newly translated polypeptides. J. Biol. Chem. 280, 12035–12040 (2005).
23. Niwa, T., Kamamori, T., Ueda, T. & Taguchi, H. Global analysis of chaperone effects using a reconstituted cell-free translation system. Proc. Natl Acad. Sci. USA 109, 8937–8942 (2012).
24. Agashe, V. R. et al. Function of trigger factor and DnaK in multidomain protein folding: Increase in yield at the expense of folding speed. Cell 117, 199–209 (2004).
25. Underwood, K. A., Swartz, J. R. & Puglisi, J. D. Quantitative polysome analysis identifies limitations in bacterial cell-free protein synthesis. Biotechnol. Bioeng. 91, 425–435 (2005).
26. Zhang, Y., Huang, Q., Deng, Z., Xu, Y. & Liu, T. Enhancing the efficiency of cell-free protein synthesis system by systematic titration of transcription and translation components. Biochem. Eng. J. 138, 47–53 (2018).
27. Desoye, B. J., Gerbasi, V. R., Thomas, P. M., Kelleher, N. L. & Jewett, M. C. A highly productive, one-pot cell-free protein synthesis platform based on genomically recoded Escherichia coli. Cell Chem. Biol. 26, 1743–1754 (2019).
28. Martin, R. W. et al. Cell-free protein synthesis from genomically recoded bacteria enables multisite incorporation of noncanonical amino acids. Nat. Commun. 9, 1203 (2018).
29. Rozovkov, A. et al. Characterization of the metabolic burden on Escherichia coli DH1 cells imposed by the presence of a plasmid containing a gene therapy sequence. Biotechnol. Bioeng. 88, 909–915 (2004).
Acknowledgements

We thank Michele Salemi and Brett Phinney from the UC Davis Proteomics core facility for their technical assistance with the mass spectrometry sample preparation and processing. We thank Rupali Reddy Pasula for technical assistance in performing and analyzing the TEM imaging of ferritin. We appreciate the discussion of the paper with the members of the Tan lab. This work is supported by a UC-MEXUS Doctoral Fellowship to L.E.C.-L., and partially supported by the National Science Foundation under Grant No. (1806366 and 1808237). The mass spectrometry experiments were funded by the UC Davis Campus Research Core Facilities (CRCF) grant 1510OD21801-01.

Author Contributions

L.E.C.-L., C.M., and C.T. wrote the paper. L.E.C.-L., C.M., and C.T. conceived the work. L.E.C.-L. and C.M. performed all experiments. C.M. analyzed the mass spectrometry data. L.E.C.-L. performed protein purification. Y.L. performed reaction buffer and protocol optimization. M.S. and S.L. performed the TEM imaging of ferritin. M.L.L. gave technical advice.

Competing interests

The University of California Davis has applied for a patent with L.E.C.-L., C.M., and C.T. listed as named inventors (USPTO Serial No: 62/868,790) covering the process for making the cell-free system described in this study. Y.L., M.S., S.L., and M.L.L. declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-16900-7.

Correspondence

correspondence

Peer review information

Nature Communications thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information

is available at http://www.nature.com/reprints

Publisher’s note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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