ACS Synthetic Biology

Article

High-throughput optimization cycle of a cell-free ribosome assembly and protein synthesis system

Filippo Caschera, Ashty S. Karim, Gianluca Gazzola, Anne D'Aquino, Norman H. Packard, and Michael C. Jewett

ACS Synth. Biol., **Just Accepted Manuscript** • DOI: 10.1021/acssynbio.8b00276 • Publication Date (Web): 24 Oct 2018

ACS Publications

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

High-throughput optimization cycle of a cell-free ribosome assembly and protein synthesis system

Filippo Caschera[#, a,b,c], Ashty S. Karim[#,a,b,c], Gianluca Gazzola[e], Anne E. d'Aquino [b,c,d], Norman H. Packard[f], Michael C. Jewett[a,b,c,e,g,h,*]

[a]Department of Chemical and Biological Engineering, [b]Chemistry of Life Processes Institute, and [c]Center for Synthetic Biology, [d]Interdisciplinary Biological Sciences Program, Northwestern University, Evanston, IL 60208, USA

[e]Rutgers Center for Operations Research, Rutgers Business School, 100 Rockafeller Road, Piscataway, NJ 08854, USA

[f]ProtoLife, Inc., 57 Post St. Suite 908, San Francisco, CA 94104, USA

[g]Robert H. Lurie Comprehensive Cancer Center and [h]Simpson Querrey Institute, Northwestern University, Chicago, IL 60611, USA

[#] These authors contributed equally

***Correspondence**: Michael C. Jewett, Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208, USA

E-mail address: m-jewett@northwestern.edu

Tel: 1 847 467 5007; **Fax:** 1 847 491 3728

Postal Address for all authors:

2145 Sheridan Road, Tech E-136

Evanston, IL 60208, USA

Abstract

Building variant ribosomes offers opportunities to reveal fundamental principles underlying ribosome biogenesis and to make ribosomes with altered properties. However, cell viability limits mutations that can be made to the ribosome. To address this limitation, the *in vitro* integrated synthesis, assembly and translation (iSAT) method for ribosome construction from the bottom up was recently developed. Unfortunately, iSAT is complex, costly, and laborious to researchers, partially due to the high cost of reaction buffer containing over 20 components. In this study, we develop iSAT in *E. coli* BL21Rosetta2 cell lysates, a commonly used bacterial strain, with a cost-effective poly-sugar and nucleotides monophosphate-based metabolic scheme. We achieved a ten-fold increase in protein yield over our base case with an evolutionary design of experiments approach, screening over 490 reaction conditions to optimize the reaction buffer. The computationally-guided, cell-free, high-throughput technology presented here augments the way we approach multi-component synthetic biology projects and efforts to repurpose ribosomes.

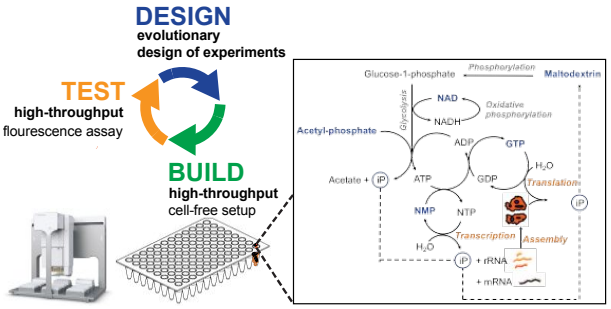
Keywords

synthetic biology, systems biology, metabolism, ribosomes, *in vitro*, iSAT, evolutionary design of experiments, machine learning

For Table of Contents Use Only

High-throughput optimization cycle of a cell-free ribosome assembly and protein synthesis system

Filippo Caschera[#, a,b,c], Ashty S. Karim[#,a,b,c], Gianluca Gazzola[e], Anne E. d'Aquino [b,c,d], Norman H. Packard[f], Michael C. Jewett[a,b,c,e,g,h,*]



The bacterial ribosome is a macromolecular machine evolutionarily optimized for the template-guided, sequence-defined polymerization of amino acids into proteins. By building ribosomes, synthetic biology efforts seek to elucidate a new understanding of the science of protein synthesis through creation – in the sense of Feynman's dictum, "*What I cannot create, I do not understand*." These efforts aim to reveal fundamental principles that underlie the operation and assembly of the ribosome complex and translation [1-3], design and build minimal cells to understand the origins of life [4-5], and enable evolution to select for ribosomes that have enhanced functions, such as altered chemical and functional properties [6-8]. However, manipulation of ribosomes in bacteria is often limited by cell viability constraints. *In vitro* assembly, or reconstruction of *E. coli* ribosomes from purified native ribosomal components into functionally active 30S and 50S ribosomal subunits, is a promising alternative to study the ribosome [9-11]. Until recently, however, *in vitro* assembly of *E. coli ribosomes* has been limited because conventional ribosome reconstitutions are non-physiological, and ribosomes reconstituted with *in vitro* transcribed ribosomal RNA (rRNA) are essentially non-functional[12-13].

To address the limitation, our lab has developed over the course of the last several years, the *in vitro* integrated synthesis, assembly and translation (iSAT) method for ribosome construction [14-17]. In a single reaction, the iSAT method constructs ribosomes in a cell-free, ribosome-free (S150) extract by transcribing DNA encoding ribosomal RNA (rRNA), and then processing and assembling transcribed rRNA and ribosomal proteins (r-proteins) into ribosomes that translate reporter proteins. Recent work on iSAT has dramatically improved the platform by optimizing extract preparation methods [16], tuning rRNA transcription [15], identifying and alleviating substrate limitations [17], and using macromolecular crowding and reducing agents [14].

The iSAT system is a unique platform, which could be potentially used for bottom-up construction of minimal cells [5, 18-19]. In this context, we recently demonstrated the ability to build functionally active ribosomes using iSAT in giant liposomes. The liposomes were prepared using

double emulsion template, and compartmentalized *in vitro* protein synthesis was analyzed using spinning disk confocal microscopy [20]. This was the first time that a cell-free transcription and translation system where the DNA molecule encoding the formation of ribosomes has been encapsulated in a model cell-like compartment, *i.e.* liposome. While this was an important step towards the construction of minimal cells, iSAT is still complex, costly and laborious to researchers, partially due to the high cost of reaction buffer containing over 20 components. To make the system more accessible, we hypothesized that the price per reaction could be decreased by using different chassis for lysate preparation with different energy regeneration metabolisms, yet it is unclear if metabolic enzymes present in the S150 extracts used for iSAT are sufficient to support ribosome synthesis, assembly and translation.

Historically, cell-free systems have been complex and expensive molecular mixtures, owing to many different chemicals and high-energy phosphate compounds that drive energy regeneration[21-24]. Glucose, PEP, pyruvate, 3-PGA, cellulose, etc. have all been evaluated for CFPS with some promising results. Recently, a high yielding, cost-effective metabolic scheme [25-26] was developed in lysates made from *E. coli* BL21Rosetta2, a common *in vivo* production strain, which could potential bring the cost of iSAT down. However, the traditional iSAT system has been developed with *E. coli* MRE600, a common strain for ribosome study [16], due to its low RNase I activity. We wondered if we could activate iSAT with a simplified and cost-effective energy metabolism in lysates from the *E. coli* BL21Rosetta2, a B strain phylogenetically divergent from MRE600[27].

We therefore set out to: (1) develop the iSAT system in *E. coli* BL21Rosetta2 cell lysates, a more commonly used bacterial strain, (2) create a simplified and cost-effective, poly-sugar and nucleotides monophosphate-based metabolic scheme that fuels iSAT, and (3) optimize the reaction buffer for cell-free protein synthesis in these lysates through high-throughput, combinatorial optimization over the 20 experimental components described in **Table 1**. While the "upstream processes" of cell-free extracts such as strain selection, cell growth, and lysis

conditions are key parameters impacting transcription and translation capabilities, we chose to limit the scope of this study to the "downstream processes", the physiochemical conditions of the *in vitro* reactions. We did this because a key challenge is that finding an effective experimental design for the optimization of a 20-dimensional experimental space is hard. Conventional approaches to the design-of-experiments problem typically aim at reducing the number of experimental parameters (i.e., components) to explore, in order to make the exhaustive search of the resulting lower-dimensional experimental space feasible. Our approach consists, instead, of a form of Evolutionary Design-of-Experiments (EDoE)[28-32], where predictive modeling and artificial intelligence guide the iterative selection of small batches or "generations" of experiments to perform, with each generation corresponding to a different small sample of the actual full-dimensional space (**Supplementary Figure S1**). This approach also allows us to make observations of the complex chemical interactions occurring in the cell-free system.

Table 1: **The 20 components included in iSAT reaction buffer tested for *E. coli* BL21Rosetta2 lysates.** Each of the 20 components are placed in one of 8 categories. Four concentrations of each component (Level 1 through 4) are represented here as varied concentration. These are used to change each component individually in the experimental setup. The pivots of all generations are also shown. The concentration of each component for a given pivot is listed in each generation's column. These values are fixed for 19 components with the remaining component being varied at the specified varied concentrations.

	Category	Component	Unit	Varied Concentration				Pivot Generation							
				Level 1	Level 2	Level 3	Level 4	1.1	1.2	2	3	4	5	6	7
1	Phosphate donor	AMP	mM	0	0.65	1.3	7.3	1.3	1.3	0	0.65	0	0	0	0
2	Phosphate donor	UMP	mM	0	0.41	0.82	4.59	0.82	0.41	4.59	0.82	0.82	0.41	0	4.59
3	Phosphate donor	CMP	mM	0	0.41	0.82	4.59	0.82	4.59	0.41	0.82	0.41	0	0	0.82
4	Phosphate donor	GTP	mM	0	0.72	1.43	8.03	1.43	0.72	8.03	8.03	8.03	8.03	8.03	8.03
5	Phosphate donor	Acetyl-phosphate	mM	0	0.16	0.32	1.79	0.32	0	0.16	0.16	0.32	0.32	0.32	0.16
6	Phosphate donor	K2HPO4/KH2PO4	mM	0	2.41	4.82	27.02	4.82	4.82	4.82	2.41	4.82	2.41	2.41	0
7	iP recycling	Maltodextrin	mM	0	16.26	29.8	44.16	2.53	42.56	29.8	29.8	44.16	44.16	29.8	16.26
8	Energy Source	3-PGA/PEP 1:1.7	mM	0	6.24	12.48	69.87	12.48	0	69.87	69.87	69.87	0	0	0
9	Cytomim	Folinic Acid	µg/mL	0	14.8	29.59	165.71	29.59	165.71	14.8	29.59	29.59	0	14.8	14.8
10	Cytomim	Putrescine	mM	0	0.92	1.85	10.34	1.85	0	0.92	0.92	1.85	1.85	0	0
11	Cytomim	Spermidine	mM	0	0.64	1.28	7.14	1.28	7.14	1.28	7.14	1.28	1.28	0	0
12	Salt	K-Glu	mM	0	14.46	200	250	8.93	250	0	200	14.46	14.46	14.46	0
13	Salt	K-OX	mM	0	1.28	2.55	14.29	2.55	1.28	2.55	1.28	0	0	0	0
14	Salt	Mg-Glu	mM	0	3.28	6.56	36.72	6.56	36.72	3.28	6.56	6.56	3.28	6.56	0
15	Translation	tRNA	µg/mL	0	82.02	164.03	918.57	164.03	0	82.02	164.03	0	164.03	0	0
16	Translation	Amino acids	mM	0	1.04	2.08	11.64	2.08	11.64	1.04	1.04	2.08	1.04	0	0
17	Cofactor	cAMP	mM	0	0.33	0.66	3.7	0.66	3.7	3.7	0.33	0.66	0.66	0	0
18	Cofactor	CoA	mM	0	0.13	0.25	1.41	0.25	1.41	0.13	0.25	0	0	0.13	0
19	Cofactor	NAD	mM	0	0.16	0.32	1.79	0.32	0	0.16	0	1.79	1.79	1.79	1.79
20	Reducing agent	DTBA	mM	0	0.96	1.93	10.80	1.93	10.8	1.93	1.93	0	0.96	10.8	1.93

In this work, we use EDoE to optimally implement a complex metabolic scheme to fuel the iSAT molecular interactions (**Figure 1**). The experimental design is based on a library of small molecules used to revitalize endogenous enzymes already present in the S150 cytoplasmic extract [16] which includes a new poly-sugar metabolic scheme. The new metabolic scheme was designed taking inspiration from an efficient and cost-effective poly-sugar based metabolism for ATP-regeneration [25-26, 33], and nucleotides triphosphate regeneration [34]. The molecular composition of this novel metabolic scheme has not been described before. We exploited the power of liquid-handling robotics to build cell-free iSAT reactions, which can then be tested for reporter protein translation and optimized via EDoE. With this powerful approach, we could overcome a barrier of complexity given by the many molecular component interactions involved in the ribosome assembly and protein synthesis *in vitro*, which simplifies reaction set-up. We achieved a ten-fold increase of protein yield of our base case with our EDoE approach. The computationally-guided, cell-free, high-throughput technology presented here alters the way we can approach complex, multi-component synthetic biology projects, providing a path forward for improving cell-free efforts in: *in vitro* ribosome engineering [35], minimal cell synthesis [18], quorum sensing [36], gene circuits optimization [37], metabolic engineering [38-39], biocatalyst discovery [40], directed evolution [41], glycosylation[42], non-standard amino acid incorporation [43-44], and human protein synthesis [45].

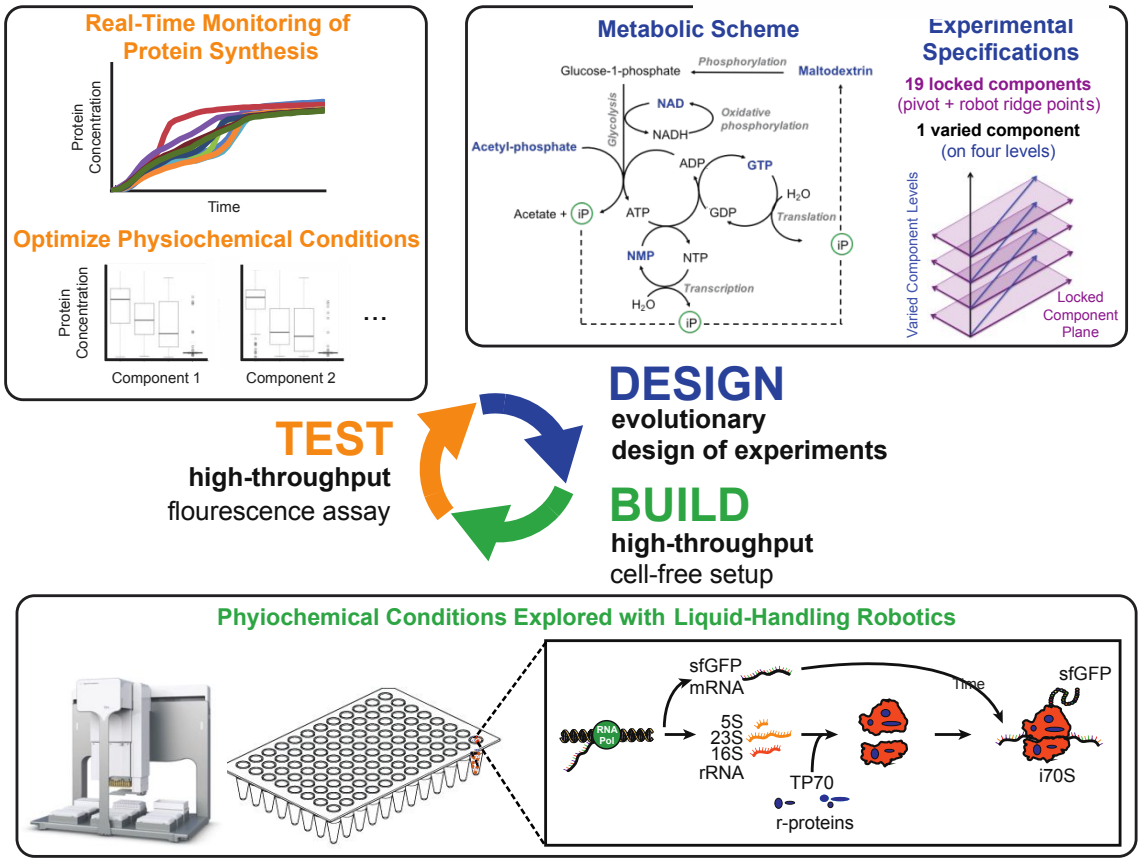


Figure 1: **Overview schematic of iterative Design-Build-Test platform for complex metabolic optimizations.**

Results

The goal for this work was to develop a robust, cost-efficient iSAT system by studying the interactions of molecular components present in during iSAT reactions. Specifically, we wanted to develop iSAT in lysates from the commonly-used *E. coli* BL21Rosetta2 strain with a poly-sugar substrate, taking advantage of nucleotide recycling. To achieve this goal, we used an evolutionary design of experiments (EDoE) approach, and liquid- handling robotics, to efficiently explore the interactions between each of the iSAT system components and carry out several rounds of optimization. This allowed us to develop and improve iSAT in BL21Rosetta2 cell lysates with a cost-efficient energy metabolism.

Design of the experimental space

We wanted to develop a poly-sugar metabolic scheme with the iSAT system in *E. coli* BL21 Rosetta2 cell lysates to make iSAT simpler and cost-effective. This metabolic scheme was chosen in order to activate complex cell-free metabolism and has been shown to work in other cell-free systems [46]. This scheme consists of the synthesis of nucleotides triphosphates from nucleotide monophosphates coupled to glycolysis activation upon hydrolysis of maltodextrin from inorganic phosphate, which is the byproduct of cell-free protein synthesis [25-26]. This design bypasses an energy re-generation system using substrate level phosphorylation, which is important to reduce the overall cost of the cell-free reaction [26], but also in the design and integration of the sub-systems of minimal cells (*i.e.,* container, information, and metabolism), an important related research area [18].

To activate such a metabolic scheme, we investigated 20 potentially beneficial small molecules. We chose to create a design space that covered these 20 components subdivided into 8 categories, each one with a specific function needed for ribosome assembly and protein synthesis (**Table 1**). The first group of components is the phosphate donor category, which

contains the molecules that upon hydrolysis release inorganic phosphate (iP) in the reaction environment. To follow, maltodextrin is considered the iP recycling element, incorporating the phosphate in glucose-1-phosphate, and activates glycolytic pathways *in vitro* for ATP synthesis and regeneration[18, 25-26]. A mixture of 3-PGA and PEP was investigated as the energy source for ATP synthesis [21, 47]. In addition, molecules of the PANOx-SP/Cytomim energy regeneration systems—cytoplasmic concentrations of spermidine, putrescine, and folinic acid—were included in the screening process as well [23-24]. While some of these components may not be required, we included them as they have been shown previously to be helpful. For instance, seminal *in vitro* translation from Ehrenberg and Kurland[48-49] and early work from our lab[23] show that putrescine is helpful though modest. Moreover, typical components such as: salts, co-factors and purified ribosomal proteins (a fixed parameter, therefore not reported in the table), necessary for enzyme functionalities and ribosome assembly [50-51], along with tRNA and amino acids important for translation and *in vitro* protein synthesis were included. Besides, the reducing agent DTBA that was recently demonstrated to be important for the iSAT system [15, 21, 23] was comprised in the mixture.

While the concentration of each component can in theory be varied continuously, the resulting experimental space would be impossible to access. We therefore decided to define the experimental space with each component varied across 4 concentration levels: low, medium-low, medium-high and high (**Table 1**). This yields an experimental space with a total of 4^{20} (~$1.1*10^{12}$) possible experiments, which is orders of magnitude above what is possible to be executed exhaustively. Thus, our experimental exploration focused on 8 small subsets of this design space, each defined by a collection of experiments located in the neighborhood of a different "pivot" experiment in the space (**Table 1**). This neighborhood is defined by all experiments obtained varying one component on the four concentration levels (**Table 1**), while keeping all of the other components constant at the same concentration level as in the pivot, as imposed by our robotic configuration. We started with two exploratory generations (generations 1.1 and 1.2) where the

pivots were manually chosen. EDoE subsequently estimated a predictive model on all experimental results and exploited this model to select the pivot for the following generation; this process was iterated five more times, for a total of eight generations (**Figure 1**). Note that we are using the term "generation" to refer to a batch of experiments whose results will be used to update the evolutionary model. This usage is consistent with evolutionary machine learning and should not be confused with other notions of biological (or chemical) generations.

Improving iSAT protein yield through iterated EDoE

Over the course of seven generations of experiments we established iSAT in *E. coli* BL21 Rosetta2 cell lysates, implemented a novel metabolic scheme, and improved protein yield 10-fold from the base case. To initially populate our EDoE dataset, medium-high and high concentration levels of components dominate the pivots of generations 1.1 and 1.2 (**Table 1**). Higher concentrations were selected in part because traditional iSAT in *E. coli* MRE600 uses similar concentrations. However, such molecular component configurations resulted in low response values (**Figure 2**; **Supplementary Figure S2**) in the BL21Rosetta2-based system.

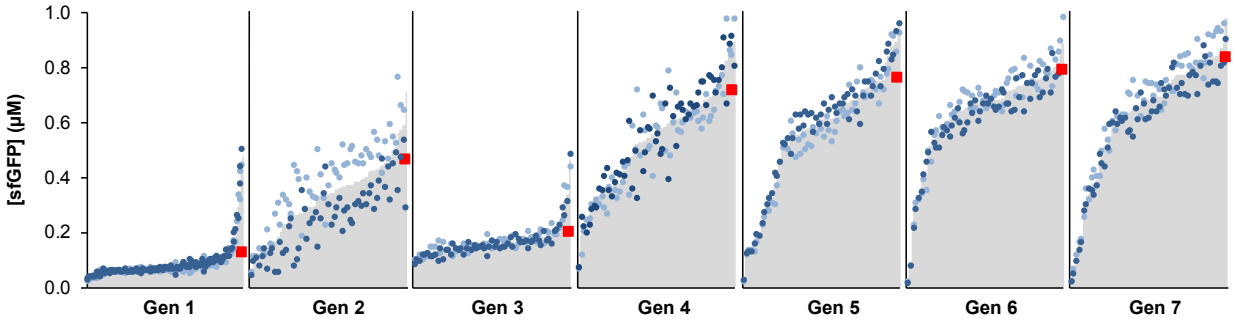


Figure 2: Response distribution of each generation of experiments. Response is defined as the fluorescence signal from sfGFP present, converted into a μM scale. The x-axis represents each unique experiment. The responses of generations 1.1 and 1.2 are displayed here as a single "Generation 1" distribution. The mean response for each experiment is plotted in grey based on two independent replicates. The response from the individual replicates are plotted as dark blue and light blue dots for each experiment. The right axis of each generation has a red square to indicate the 90% quantile level for that distribution.

With the EDoE approach, we observed an overall increase in response in the subsequent generations, with the most substantial boosts obtained in generations 4–7. Indeed, the synthesized sfGFP from our iSAT system increased from ~0.1 μM in generations 1.1-1.2 to ~1 μM in the last generation (**Figure 2**; **Supplementary Figure S2)**. The increased protein yields resulted from variation of the metabolic scheme, given by the concentration levels of the components in the pivots, and their neighborhoods, in different generations. The pivots showed a progressive convergence toward low or zero concentration levels in later generations. Overall, this indicates that several of the components used to design the initial pivots for high-throughput screening can be omitted, therefore providing a simplified mixture for energizing *in vitro* protein synthesis using *in situ* self-assembled ribosomes. However, two molecular components, GTP and NAD, converge to consistently high values in generation 4 and beyond. In particular, GTP is important for the translation process [52] and NAD is a cofactor in glycolysis and metabolism [53]. All data generated is provided in **Supplementary Table S2**.

Dependence of response on individual components

We next developed an intuition about the most crucial small molecules for iSAT in BL21Rosetta2 cell lysates by investigating the impact that different concentration levels of individual components had on the experimental response, as measured in the 480 experiments that made up all generations (**Figure 3**). An important observation is that certain molecules appear to have a neutral or negative effect on the response at any concentration, and their omission may therefore be actually beneficial for the system. This was not intuitive *a priori* based on previous literature of similar systems. In particular, it seems that nucleotide monophosphates can be omitted from the reaction mixture when the nucleotide triphosphate, GTP, is present at high quantities. Sufficient quantities of total nucleotide phosphates must be present to sustain *in vitro* transcription (**Supplementary Figure S3**) and in turn protein production. Keeping an

adequate total pool of nucleotide phosphates to generate initiate RNA polymerization and chemical energy (ATP) to sustain *in vitro* protein synthesis with the iSAT system in BL21Rosetta2 cell lysates is essential. In addition, GTP can be considered an essential high-energy molecule extremely relevant for the metabolic scheme presented herein. This molecule is important for translation and in the regeneration of nucleotides triphosphates[54].

Phosphate donor molecules, such as acetyl-phosphate also appeared to be an important component to increase response. Indeed, higher protein yields were observed with 0.32 mM acetyl-phosphate as compared to 0 mM. Acetyl-phosphate is important as high-energy phosphate donor for ATP and GTP-regeneration systems [21, 54] and has been shown to be useful for efficient NTPs and dNTPs regeneration in *E. coli* lysates [55]. Conversely, increasing the concentration of the phosphate buffer K_2HPO_4/KH_2PO_4 resulted in a gradual decrease in response. Indeed, if inorganic phosphate is accumulated at high concentration in the cell-free reaction, it inhibits the cell-free reaction, mainly by sequestering magnesium that is necessary for protein biosynthesis [56]. However, a low concentration is slightly beneficial. The inorganic phosphate would trigger the hydrolysis of maltodextrin to activate the glycolytic pathway in the crude extract [26]. Interestingly, as the system does not need expensive molecules such as 3PGA and PEP to re-generate ATP, we could conclude that it is energized mainly by the ATP produced through the glycolytic pathway using the polysaccharide (maltodextrin) as the carbon source, in addition to GTP and acetyl-phosphate high-energy phosphate donor molecules.

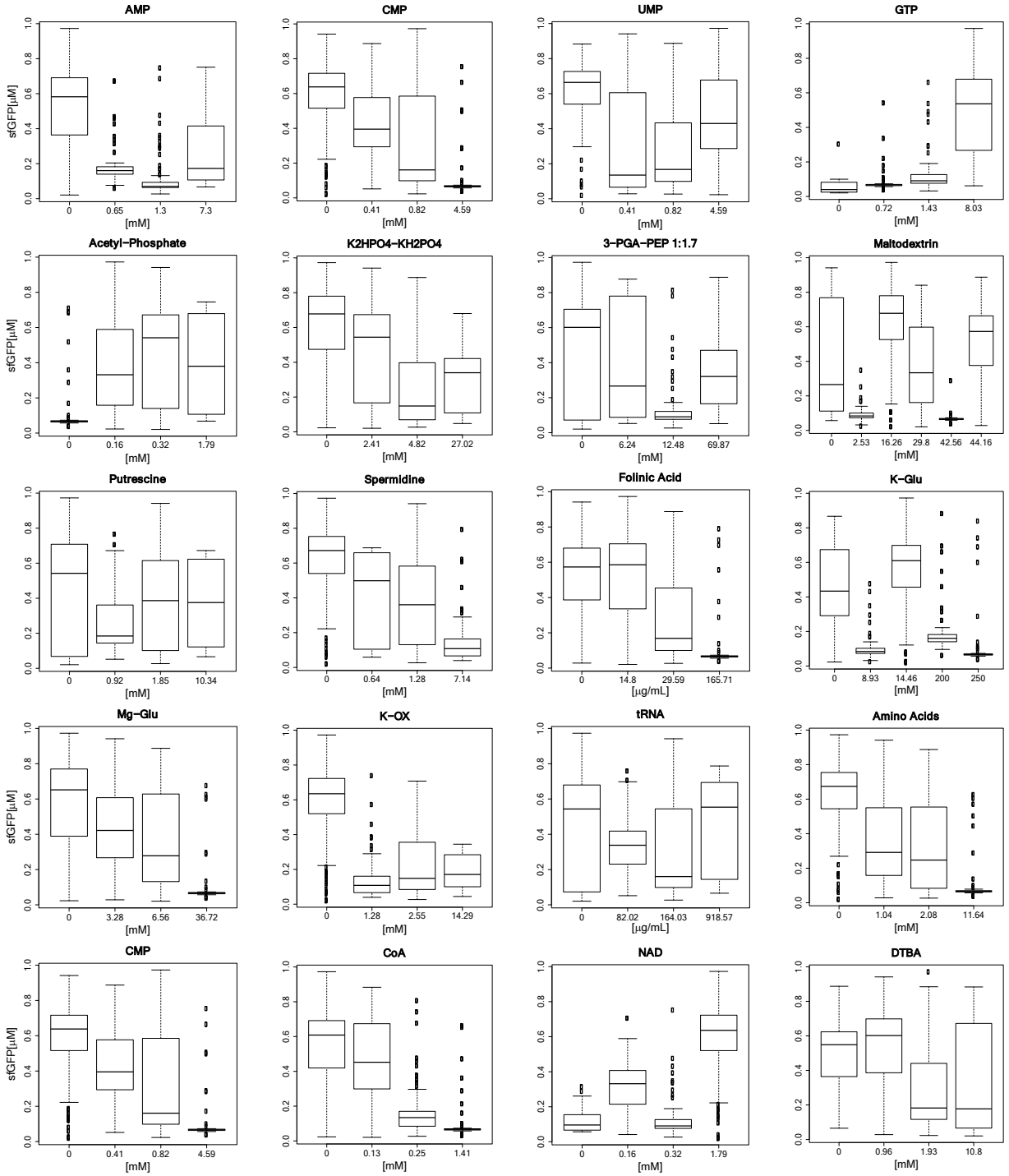


Figure 3: **Dependence of iSAT response on individual experimental components.** Each of the 20 components tested in iSAT is plotted in a separate graph. The x-axis represents concentration levels and the y-axis the corresponding conditional response distributions. Results are aggregated over the experiments performed in all generations.

Concerning polyamines, i.e. putrescine and spermidine, we observed a general and gradual decrease in the sfGFP synthesis at increasing concentrations. The same trend was observed for folinic acid, which is a molecular component used in the design of the reaction buffer of cell-free expression systems based on 3-PGA [25-26]. The response shows a somewhat non-linear dependence on K-glutamate, with higher response at 0 mM or at 14 mM. The dependence of the response on Mg-glutamate is instead more linear, with a decline of the response proportional to the increase in Mg-glutamate in the system. Magnesium salts are important for ribosome stability in *E. coli* cell-free expression system [57], as well as enzymes functionalities, and their concentration must be tuned to an optimal value to avoid detrimental effects on the system [58]. It should be mentioned that putrescine, spermidine, K-glutamate, and Mg-glutamate are components already included in the system during crude extract preparation, and therefore present in the S150 cell extract (see material and methods). This could explain higher response values when such components are not in the chemical mixture making the reaction buffer for protein synthesis. Moreover, it also appears that the omission of K-oxalate from the reaction buffer can result in higher system response. Normally, this salt is important to inhibit the reverse reaction of the phosphoenolpyruvate synthase with PEP or 3-PGA used as high-energy phosphate molecule donor [21, 59]. Therefore, this finding suggests that the S150 system can by-pass the substrate level phosphorylation for ATP regeneration.

An overarching design principle that emerged was that components involved directly in the translation of proteins need to be finely-tuned. For instance, the response tends to be quite sensitive to differences in the concentration of tRNA. This molecule is already present in the crude extract after preparation and already charged with its correspondent amino acid. Furthermore, the response decreases with higher amino acids concentrations. cAMP and CoA show a similar trend. However, at its highest concentration (1.79 mM), NAD boosted protein synthesis and therefore appears to be an essential component for the iSAT system. NAD is an important cofactor involved in glycolysis and in turn the maltodextrin-based metabolism to energize *in vitro* protein synthesis.

The reducing agent, DTBA, has been recently demonstrated important in the optimization of the iSAT system [14]. We observed that the addition at low concentration is also beneficial for the efficiency of the system presented here.

Composition of iSAT reaction conditions achieving highest protein expression and their kinetic profiles

The composition of iSAT reaction conditions achieving highest protein expression is shown in **Table 2**. These conditions show similar features. In particular, they are characterized by: AMP, K-OX, and CoA at 0 mM (with the exception of 1 experiment with CoA at 0.13 mM); GTP at 8 mM; NAD at 1.79 mM; Acetyl-phosphate at 0.32 mM (with the exception of 2 experiments at 0.16 mM); K-Glu at 14.46 mM (with the exception of one experiment at 200 mM). The analysis of these reactions highlights the importance of GTP, Acetyl-phosphate, NAD, UMP and maltodextrin as molecular framework fueling the iSAT system prepared from *E. coli* BL21Rosetta2.

Table 2: iSAT reaction conditions achieving top 9 highest protein expression mean response. The mean response values are obtained by averaging out the twenty repeats within each of the two plates and then averaging out the two within-plate means; between-plate response standard deviation, calculated over the two within-plate means.

Pivot Generation	7	5	5	4	7	6	5	4	4
Mean Response (RFU)	2594.5	2510	2373	2366	2358.5	2355	2341	2338.5	2322
% Noise	0.11	0.03	0.06	0.14	0.11	0.16	0.05	0.06	0.01
AMP	0	0	0	0	0	0	0	0	0
UMP	4.59	0.41	0.41	0.82	4.59	0	0.41	0.82	0.82
CMP	0.82	0	0	0.41	0.82	0	0	0.41	0.41
GTP	8.03	8.03	8.03	8.03	8.03	8.03	8.03	8.03	8.03
Ac.phos	0.16	0.32	0.32	0.32	0.16	0.32	0.32	0.32	0.32
K_2HPO_4/KH_2PO_4	0	2.41	2.41	4.82	0	2.41	2.41	4.82	4.82
Maltodextrin	16.26	0	16.26	44.16	16.26	16.26	44.16	44.16	44.16
3-PGA:PEP1:1.7	0	0	0	69.87	0	0	0	6.24	0
Folinic.Acid	14.8	0	0	29.59	14.8	14.8	0	29.59	29.59
Putrescine	0	1.85	1.85	1.85	0	0	1.85	1.85	1.85
Spermidine.	0	1.28	1.28	0	0	0	1.28	1.28	1.28
K.Glu	14.46	14.46	14.46	14.46	200	14.46	14.46	14.46	14.46
K.OX	0	0	0	0	0	0	0	0	0
Mg-Glu	0	3.28	3.28	6.56	0	6.56	3.28	6.56	6.56
tRNA	0	164.03	164.03	0	0	0	164.03	0	0
Amino acids	0	1.04	1.04	2.08	0	0	1.04	2.08	2.08
cAMP	0	0.66	0.66	0.66	0	0	0.66	0.66	0.66
CoA	0	0	0	0	0	0.13	0	0	0
NAD	1.79	1.79	1.79	1.79	1.79	1.79	1.79	1.79	1.79
DTBA	1.93	0.96	0.96	0	1.93	10.8	10.8	0	0

In addition, we performed a kinetic analysis to observe potential, subtle differences between our top 9 experiments (**Figure 4**). From the protein time-courses, we observed a window between 5 and 10 hours where iSAT reactions show different rates. Most likely, these rates depend on the ability to regenerate chemical energy to sustain RNA transcription, ribosome self-assembly, and translation. We suspect that the energy regeneration scheme is a major player here because in each condition with maltodextrin metabolism present (all except Top 2) we observe translation rates that plateau before increasing again during that 5- to 10-hour window. This observed plateau or lag-phase was not observed using the traditional iSAT system design[14-16]. Potentially, the maltodextrin-based metabolic scheme and slow energy release could also cause the difference in rate from traditional iSAT systems prepared with MRE600. The sharpest

protein production increases after this lag-phase appear in Top 1, Top 5, and Top 6 experiments which have no added amino acids or cAMP. Besides these observations there few obvious trends from the physiochemical environment that explain these kinetic differences. Further studies that connect RNA transcription, ribosome self-assembly, and translation could clarify how physiochemical conditions affect the kinetics iSAT.

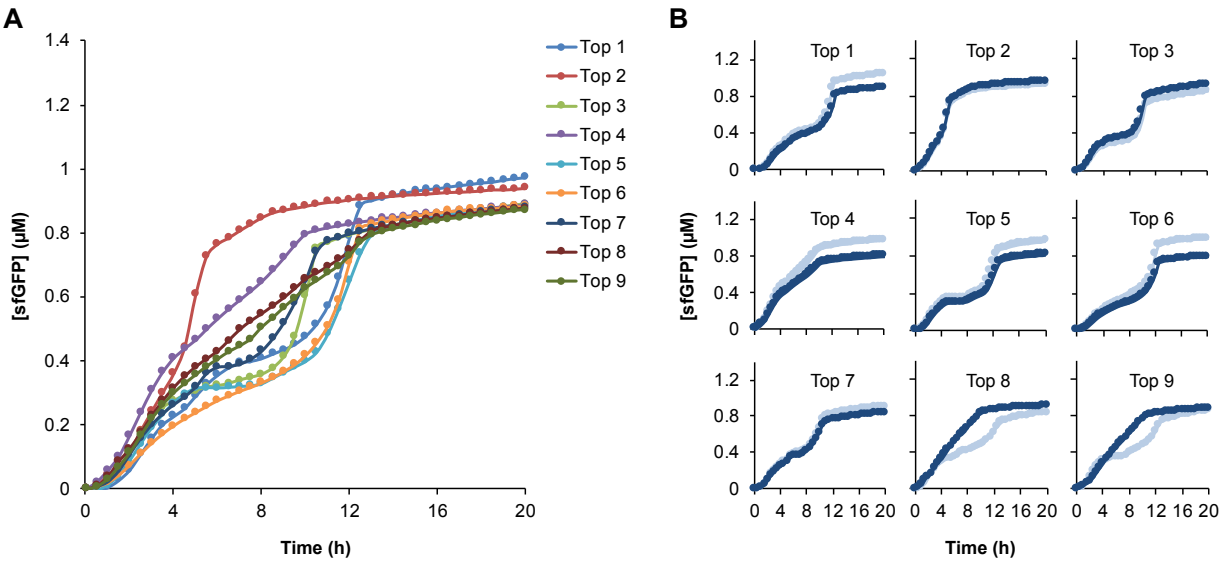


Figure 4: **Kinetics of iSAT reactions achieving top 9 highest protein expression mean response.** (A) Mean measurements of sfGFP reporter protein (μM) collected at 30-min intervals for each of the top 9 experiments in duplicate found with EDoE. (B) Each of the top 9 experiments are individually plotted to show variability between replicate reactions (two independent reactions performed).

Variation of magnesium and DNA templates concentrations

A magnesium optimization of S150 extracts for reporter protein synthesis activity in iSAT reactions is vital for *in vitro* transcription/translation[16]. Therefore, following the computationally-guided EDoE optimizations, we decided to carry out an additional magnesium optimization to improve activity. We focused on optimizing buffer conditions that resulted in the top 4 protein synthesis yields (**Figure 4**). In addition to magnesium, we explored the effects of changing DNA concentration, since DNA concentrations used in our high-throughput experimentation was 2nM and previous efforts showed that increasing DNA concentration might yield enhanced iSAT

activity[14-16]. For each of the top 4 experiments, we explored the effects of modifying the magnesium glutamate and plasmid DNA concentrations to see if iSAT activity could be improved in the context of the new metabolic scheme (**Supplementary Figure S4**). To optimize these two parameters, a lattice experimental design was carried out. Based on the need for greater rRNA transcription, pT7rrnB (plasmid DNA) concentration was varied with 2, 4, 8 and 10 nM of DNA. Magnesium glutamate concentration was altered by testing a range of concentrations: 0, 2, 3.3, 4, 5, 6, 6.6, 8 and 9 mM. Standard 15 µL batch reactions were prepared by hand and performed at 37 °C for 20 hours, with varying magnesium glutamate concentrations for each reaction. Collectively, these experiments allowed us to map component landscapes to explore global and local optima. Global optima for magnesium glutamate and DNA concentrations for each iSAT conditions were determined. The optimum concentration of DNA was as follows: for top experiment 1 the optimum was 2 nM, for top experiment 2 the optimum was 2 nM, for top experiment 3 the optimum was 8-10 nM and for top experiment 4 the optimum was 2 nM. For all experiments, the original magnesium glutamate concentration from each of the top 4 experiments was used. Taken together, our results show that the cell-free framework and its barrier-free access to reaction conditions is well-suited for rapidly acquiring physiochemical landscapes to assess and optimize pathway performance. This joins an emerging body of literature highlighting the value of cell-free systems for prototyping biological systems[60-61].

High-dimensional inter-component synergy

With all of our experimental data at hand, we set out to understand components having the largest impact of iSAT activity. This is important because it teaches us which components might synergistically work together to enable high activity. **Figure 5** tackles this analysis from a regression modeling perspective by showing the structure of a conditional inference tree, trained on all experimental data collected through generation 7 of the data. The tree, estimated via the

ctree function in the R package *partykit*, represents a partition of the experimental space into regions with homogeneous experimental response. Each path from the root node to a given leaf node represents a different sequence of recursive bisections of the experimental space. Each bisection is defined by a specific experimental component (shown in ovals), selected to maximize the statistical association between a candidate component and the experimental response. The specific concentration of this component to maximize this association is shown on the two outgoing arcs. This concentration is selected to maximize the standardized difference between the response means in the two resulting data subsamples. Each leaf node represents a different region of the experimental space (shown in boxes) and contains a box plot of the response distribution of the experiments belonging to that region. From this tree analysis, we find that the highest-response experiments are, on average, mainly contained in nodes 15 and 16. This region is defined by the presence of five components: GTP, NAD, K-Ox, 3-PGA-PEP, and Maltodextrin. Thus, these components represent “key components” with “key concentrations” in specific intervals to improve iSAT performance.

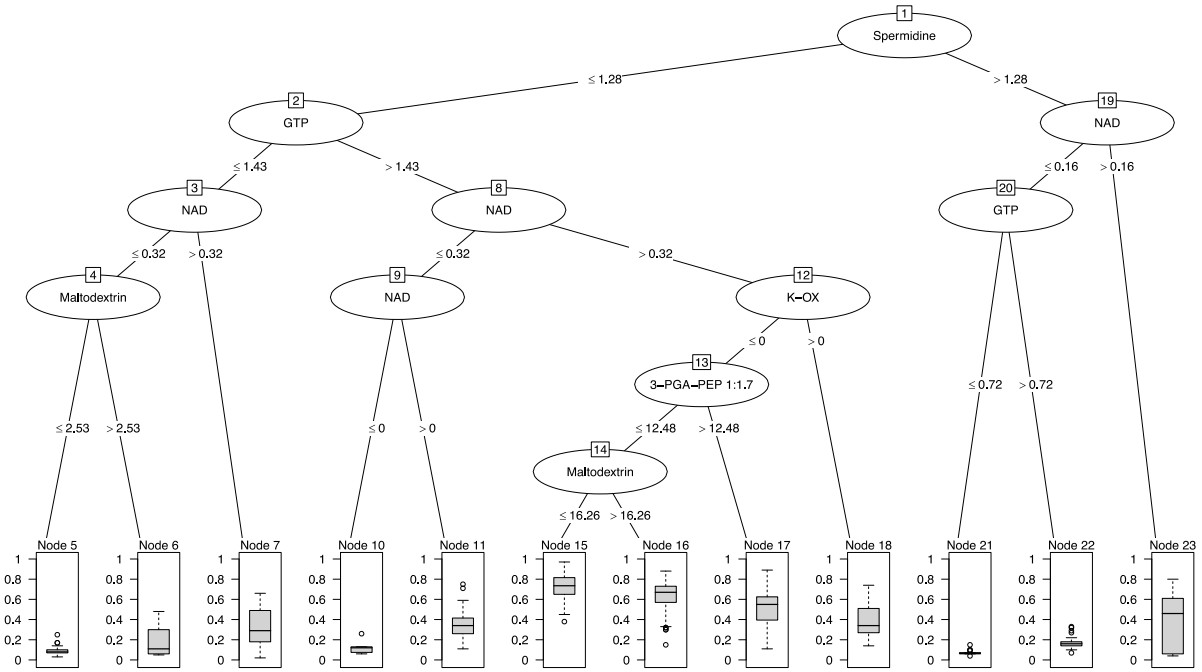


Figure 5: **Conditional inference regression tree, trained on all experimental data.** Terminal nodes show the box plots of the experimental response, conditioned on the experimental region defined by specific experimental components at specific concentrations; regions containing high-response experiments are defined by at least 4/5 experimental components.

Further insight into the impact of these key components and key concentrations on experimental response can be found in **Supplementary Figure S5**. This figure shows experimental response distributions conditioned on the number of key components at key concentrations in an experiment and illustrates how a substantial boost in experimental response is obtained only when several of these conditions are satisfied simultaneously. The average response, for example, fluctuates around 0.10 if two or less conditions are satisfied, but progressively grows to 0.30, 0.49, and 0.68 as the number of satisfied conditions increases to three, four, and five. This indicates that these key components are synergistic, that is, they yield high response only by working together. Identifying which components are key for iSAT performance allows us to define these components as essential when trying to minimize which components we include in iSAT to minimize the cost of a reaction. Indeed, with this knowledge in

hand we were able to decrease the cost of this new iSAT reaction 4-fold from the state-of-the-art reaction (**Supplementary Figure S6**).

Discussion

In this work, we have *(i)* established lysates from a common *E. coli* bacterial strain, BL21Rosetta2, as a novel platform for the iSAT system, which allows *in vitro* study of ribosome assembly and function of a new bacterial strain, *(ii)* integrated a novel simplified metabolic scheme for iSAT that exploits maltodextrin as non-phosphorylated energy source, and (iii) implemented nucleotides triphosphate regeneration for *in vitro* transcription. This is an important point as it shows a metabolic scheme with energy regeneration, transcription, translation, ribosome construction, and components necessary for self-maintenance [5, 18, 62] – a system that could be integrated as a sub-system of a minimal cell. Herein, the challenges were (i) the activation of such a complex metabolic scheme using the iSAT system, which more than using self-assembled ribosomes, is also diluted three times more than conventional cell-free protein synthesis systems. Therefore, these results could be also important to scale-up cell free ribosome synthesis for medical and industrial applications [63], such as the synthesis of peptidomimetics.

In addition, we have demonstrated that our EDoE approach can be adapted to the technical constraints of a robotic workstation, which in this work substantially limited the exploration of the experimental components, unlike in previous applications of EDoE [29-30]. Our EDoE algorithm was able to optimize the iSAT system into an *E. coli* strain sub-optimal for studying ribosome assembly, discovering complex inter-component synergies, while also decreasing the cost of the cell-free reaction 4-fold, i.e. from ~20 cents to 5 cents per reaction (**Supplementary Figure S6**). This approach also allowed us to identify top-performing conditions that we might not have tested otherwise. For example, the top-performing condition contained no added AMP in the reaction yet had high transcription and translation levels. It was surprising that

sufficient rRNA and mRNA transcription could occur. One possibility is that NMPs and NTPs still remain after dialysis potentially sequestered by enzymes in the cell lysates which could also indicate the presence of a nucleotide recycling system capable of regenerating these components. This observation warrants further investigation of the metabolic processes occurring in cell-free lysates.

Reaction optimization may be specific to which target proteins are made in the reactions. Because we used a single target protein, sfGFP, to test the yield of the iSAT protein synthesis protocol, the concentrations selected by the EDoE algorithm, albeit optimal for this particular protein, may be sub-optimal for other proteins. This is a form of “overfitting”, and can only be addressed with further experiments, using other target proteins. In addition, we believe that further optimization of the system should comprise the design of the DNA templates, either in the length and sequence of the regulatory parts and spacer sequences [30, 47, 64]. Moreover, studying the effects of strain selection, cell growth, and lysis conditions on iSAT performance could further inform development of iSAT systems in the future. Looking forward, we anticipate that new cost-effective iSAT reactions fueled by new energy regeneration schemes discovered here, could facilitate unraveling the systems biology of ribosome biogenesis, constructing minimal cells from defined components, and engineering ribosomes with new functions

Methods

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Corresponding author, Michael Jewett (m-jewett@northwestern.edu).

Bacterial strains and plasmids

E. coli BL21Rosetta2 (NEB) was used in creating S150 cell lysates for use in all iSAT experimental generations. Plasmid pJL1-sfGFP was used for reporter expression, and plasmid pT7AM552A was used for rRNA operon expression.

S150 cell extract and component preparation

E. coli BL21Rosetta2 cells were grown in 2x YTP media (16 g l^{-1} tryptone, 10 g l^{-1} yeast extract, 5 g l^{-1} NaCl, 7 g l^{-1} potassium phosphate monobasic, 3 g l^{-1} potassium phosphate dibasic). S150 extract, *E. coli* 70S ribosomes, total protein of 70S ribosomes (TP70) and T7 RNA polymerase (RNAP) were prepared as previously reported [15]. The amino acids mixtures used by the robotic workstation were prepared as previously described.[65] A single batch of extract was used for this study.

iSAT reaction setup

The typical iSAT system is composed of 22 components listed in **Supplementary Table S1** with traditional concentrations. In the new metabolic scheme explored for *E. coli* BL21Rosetta2 lysates, the iSAT reaction was composed of *E. coli* S150 crude extract (3.7 μg/mL), rRNA operon plasmid (pT7AM552A, 4.0 nM), pJL1-sfGFP (4.0 nM), T7 RNA polymerase (36 μg/mL), total protein of 70S ribosomes (TP70, 300 nM), and the 20 different components listed in **Table 1**. These 20 components, plus the homemade amino acids mixture, make up the reaction buffer necessary for fueling *in vitro* protein synthesis.

Robotic liquid-handling reaction setup and exploration of the experimental space using pivots

An Agilent BRAVO liquid-handling robot (Agilent Technologies) was used to carry out the iSAT reaction buffer setup. The liquid-handling workstation has nine plate decks and a 96-pipette tip head movable in the x-y-z directions. The workstation was programmed using VWorks™ Automation Control Software (BioNex Solutions, Inc.) to pipette different arrangements of reagents. Concentration levels for each component was made through serial dilution of stock vials of each of the 20 components using the liquid-handling system (both the 96ST and 96LT heads). Each component was diluted with 50 mM HEPES pH 7.2 buffer. After initial component dilutions are made, the reagents are pooled to together in 96-well flat-bottom plates. The exploration compatible with our robotic constraints is conducted by choosing a "pivot" experiment, and then performing alterations on the pivot. The alterations are defined by keeping 19 components fixed at the pivot concentrations (**Table 2**) while the concentration of the remaining component is varied one level at the time (the concentration levels are highlighted using a color code in **Table 1**). For example, well A1 contains all reagents except AMP, and A1 through A4 contain AMP at each of the four concentration levels (low, medium-low, medium-high, and high) listed in **Table 1**; these complete reaction buffers are then mixed with S150 extract, purified ribosomal proteins, DNA, etc. to complete the iSAT reaction mixture. The set of all single-component alterations of a pivot make up the pivot "neighborhood". The pivot neighborhood is comprised of 60 different experiments, since each component may be varied to three alternative values besides that component's pivot value. The generation of experiments determined by a particular pivot contains all the experiments in the pivot neighborhood, and in addition, 20 replicates of the pivot experiment, yielding 80 experiments fitting on a 96-well plate. Reactions are run in duplicate, and sfGFP fluorescence was measured at 37°C in sealed 500 µL tubes (Biorad) using a real time PCR machine.

GFP Quantification

The kinetic data and yields of sfGFP were measured through excitation at 485 nm while measuring emission at 528 nm with a 510 nm cutoff filter. The fluorescence response of sfGFP was converted to concentration according to a standard curve as previously described[66].

RNA Quantification

iSAT reactions were performed with radioactive ^{3}H-UTP (8 μM) supplemented in addition to the standard concentrations of NMPs and NTPs (determined by the condition tested). We used trichloroacetic acid (TCA) to precipitate radioactive iSAT samples. Radioactive counts from TCA-precipitated samples was measured by liquid scintillation to then quantify total RNA yields as previously reported (MicroBeta2; PerkinElmer)[67-68].

Modeling and design of experiments

The use of modeling in the EDoE process is illustrated in **Figure 1 and Supplementary Figure S1**. A graphical interface to modeling and EDoE tools analogous to those described in **Supplementary Figure S1** can also be found online.[32] Each generation begins with the collection of experimental results for that generation. The experiments for the following generation are designed by first building a predictive model from the data for all completed generations, including the current one. The predictive model takes an experiment as input, and outputs a prediction for the experimental result expected for that experiment, i.e., a value on the modeled experimental response surface. The predictive model is used to explore the experimental space, through a process of "virtual execution" of many randomly sampled pivots, combined with hill-climbing of the model response surface, as described below. The trade-off between random exploration and hill-climbing is varied each generation as described in the detail below, but at the end of each generation's exploratory process four good candidate pivots are automatically chosen, and the experimentalist selects between these four (on the basis of intuition combined with expert

knowledge). The pivot selected through this process is then used to determine the following generation, as described above.

The model constructed has the form of an ensemble of 25 single-hidden-layer, feed-forward neural networks, each network initialized with different random initial weights and trained via the *nnet* package in R. Each network has 20 input nodes, corresponding to the 20 dimensions (components) of the experimental space, and one output node, to predict the experimental response given values for all inputs. Prior to training, the 20 replicates of each pivot experiment are collapsed into one individual data point, whose response value is the mean of the response value across the 20 replicates. After training, a network may be used to predict experimental response for an arbitrary set of inputs, i.e. for any candidate experiment. The ensemble prediction is obtained by taking the mean of the predictions of each of the 25 networks in the ensemble.

Predictive models must be carefully controlled by a process of regularization, to avoid overfitting. We regularized our network ensemble models by exploring a range of model hyper-parameters in a bootstrapping process. The hyper-parameters explored were the number of hidden nodes for the networks (values 2, 5, 10, 20, 40), the value of the weight-decay term (values 0.01, 0.05, 0.1, 0.25, 0.5, 1), and the number of iterations of the back-propagation training algorithm (values 100, 500, 1000, 2000). For each combination of these hyper-parameters, the trained network ensemble was bootstrapped via Monte-Carlo cross-validation, on 20 independent random partitions of the data from all experiments and measured responses up to the last completed generation into training and cross-validation sets (80% training / 20% validation). Each hyper-parameter combination was assigned a score corresponding to the mean across partitions of the Pearson's correlation between predicted and observed response values in the validation set.

Once the hyper-parameter combination with the highest score has been selected through the cross-validation process, a predictive model with such hyper-parameter values is trained on the entire data from all experiments and measured responses up to the last completed generation.

Given the predictive model, any given candidate pivot may be evaluated by computing its predicted experimental response as well as that for each experiment in its neighborhood. The predicted response associated to that pivot is then taken to be the 90% quantile of the distribution of predicted responses for the set of experiments including the pivot and its neighborhood.

Given a predictive model built from the selected hyper-parameters, we can proceed to choosing good predicted experiments as candidate pivots for the next generation. These candidate pivots are chosen by combining random sampling with two techniques that manage the trade-off between random exploration and exploitation of information gathered in previous generations: (i) a steepest-ascent algorithm to hill-climb the predicted response surface to reach local maxima (ii) an algorithm to constrain experiment choice based on "experimental distance", defined as the mean pairwise Euclidean distance between each experiment in the set that includes the candidate pivot and its neighborhood and each experiment in the set of already performed experiments, calculated after normalizing the range of each component to a [0, 1] scale. The details of the experimental choice process vary from generation to generation, based on the complexity of the predicted response surface, and the evaluation of experimental results for each generation. A summary of the experimental process for different generations follows:

- Generation 1.1-1.2: chosen by the experimental team.
- Generation 2-3: randomly sample 250 candidate pivots, then select the four pivots with best predicted response among those having experimental distances falling between the median and the 3rd quartile of the distribution. The experimental team finally chooses one from these four.
- Generation 4: Run the steepest-ascent algorithm from 500 randomly sampled initial pivots; then randomly sample four candidate pivots from the local maxima with predicted response above the median of the distribution and experimental distance falling between the median and the 3rd quartile of the distribution. The experimental team finally chooses one from these four.

- Generation 5: Run the steepest-ascent algorithm from 5000 randomly sampled initial pivots; randomly sample 250 pivots from the local maxima with predicted response above the median of the distribution, with probabilities proportional to their respective predicted response; then select the four candidate pivots with best predicted response among those having experimental distance falling between the median and the 3rd quartile of the distribution. The experimental team finally chooses one from these four.
- Generation 6: Run the steepest-ascent algorithm from 20000 randomly sampled initial pivots; randomly sample 250 pivots from the local maxima with predicted response above the 3rd quartile of the distribution, with probabilities proportional to their respective predicted response; then select the four candidate pivots with best predicted response among those having experimental distance falling between the 1st and the 3rd quartile of the distribution. The experimental team finally chooses one from these four.
- Generation 7: Run the steepest-ascent algorithm from 80000 randomly sampled initial pivots; randomly select 250 pivots from the local maxima with predicted response above the 90% quantile of the distribution, with probabilities proportional to their respective predicted response; then select the four candidate pivots with best predicted response. The experimental team finally chooses one from these four.

There was one experiment in generation 2 that appeared to have extremely high response; subsequent experiments have not validated its repeatability, so that experiment has been deleted from this presentation of results. The data point corresponding to this severely noisy experiment was, however, used in the model training process, causing bias in the predictions for generation 3.

Supporting Information

The supporting files contains Supplementary Table S1, Supplementary Figure S1-S6, and

Supplementary Table S2 (.csv).

Author Contributions

Conceptualization, F.C. and M.C.J.; Reagents preparation, F.C. A.S.K.; Methodology, F.C., A.S.K., and G.G.; Software, G.G. and N.H.P.; Validation, A.S.K.; Investigation, F.C., A.S.K., and A.E.D.; Writing – Original Draft, F.C. and A.S.K.; Writing – Review & Editing, F.C., A.S.K., G.G., A.E.D., N.H.P., and M.C.J.; Funding Acquisition, M.C.J.; Resources, G.G. and N.H.P.; Supervision, N.H.P. and M.C.J.

Acknowledgements

This work was supported by the Army Research Office W911NF-16-1-0372 (to M.C.J.), National Science Foundation grant MCB-1716766 (to M.C.J.), the Human Frontiers Science Program RGP0015/2017 (to M.C.J.), the David and Lucile Packard Foundation (to M.C.J.), and the Camille Dreyfus Teacher-Scholar Program (to M.C.J.). A.S.K. and A.E.D. are supported by NSF Graduate Research Fellowships. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the DoD or the U.S. Government.

References

1. Spedding, G., *Ribosomes and protein synthesis, a practical approach*. Oxford University Press: Oxford, 1990.
2. Erlacher, M. D.; Chirkova, A.; Voegele, P.; Polacek, N., Generation of chemically engineered ribosomes for atomic mutagenesis studies on protein biosynthesis. *Nat Protoc* **2011,** *6* (5), 580-92.
3. Polacek, N., The ribosome meets synthetic biology. *Chembiochem* **2011,** *12* (14), 2122-4.
4. Forster, A. C.; Church, G. M., Towards synthesis of a minimal cell. *Mol Syst Biol* **2006,** *2*, 45.
5. Jewett, M. C.; Forster, A. C., Update on designing and building minimal cells. *Curr Opin Biotechnol* **2010,** *21* (5), 697-703.
6. Cochella, L.; Green, R., Isolation of antibiotic resistance mutations in the rRNA by using an in vitro selection system. *Proc Natl Acad Sci U S A* **2004,** *101* (11), 3786-91.
7. Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W., Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **2010,** *464* (7287), 441-4.
8. Wang, K.; Neumann, H.; Peak-Chew, S. Y.; Chin, J. W., Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat Biotechnol* **2007,** *25* (7), 770-7.
9. Traub, P. N., M., Structure and function of E. coli ribosomes. V. Reconstitution of functionally active 30S ribosomal particles from RNA and proteins. *Proc Natl Acad Sci U S A* **1968,** *59* (3), 8.
10. Maki, J. A.; Culver, G. M., Recent developments in factor-facilitated ribosome assembly. *Methods* **2005,** *36* (3), 313-20.
11. Nierhaus, K. D., F., Total reconstitution of functionally active 50S ribosomal subunits from Escherichia coli. *Proc. Nat. Acad. Sci. USA* **1974,** *71* (12), 5.
12. Green, R.; Noller, H. F., In vitro complementation analysis localizes 23S rRNA posttranscriptional modifications that are required for Escherichia coli 50S ribosomal subunit assembly and function. . *RNA* **1996,** *2* (10), 1011--1021.
13. Semrad, K.; Green, R., Osmolytes stimulate the reconstitution of functional 50S ribosomes from in vitro transcripts of Escherichia coli 23S rRNA. *RNA* **2002,** *8* (4), 401--411.
14. Fritz, B. R.; Jamil, O. K.; Jewett, M. C., Implications of macromolecular crowding and reducing conditions for in vitro ribosome construction. *Nucleic Acids Res* **2015,** *43* (9), 4774-84.
15. Fritz, B. R.; Jewett, M. C., The impact of transcriptional tuning on in vitro integrated rRNA transcription and ribosome construction. *Nucleic Acids Res* **2014,** *42* (10), 6774-85.
16. 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.
17. Liu, Y.; Fritz, B. R.; Anderson, M. J.; Schoborg, J. A.; Jewett, M. C., Characterizing and alleviating substrate limitations for improved in vitro ribosome construction. *ACS Synth Biol* **2015,** *4* (4), 454-62.
18. Caschera, F.; Noireaux, V., Integration of biological parts toward the synthesis of a minimal cell. *Curr Opin Chem Biol* **2014,** *22*, 85-91.
19. Adamala, K. P.; Martin-Alarcon, D. A.; Guthrie-Honea, K. R.; Boyden, E. S., Engineering genetic circuit interactions within and between synthetic minimal cells. *Nat Chem* **2016,** *9*, 431-439.
20. Caschera, F.; Lee, J. W.; Ho, K. K.; Liu, A. P.; Jewett, M. C., Cell-free compartmentalized protein synthesis inside double emulsion templated liposomes with in vitro synthesized and assembled ribosomes. *Chem Commun (Camb)* **2016,** *52* (31), 5467-9.

21. Calhoun, K. A.; Swartz, J. R., Energy systems for ATP regeneration in cell-free protein synthesis reactions. *Methods Mol Biol* **2007,** *375*, 3-17.
22. Caschera, F., Bacterial cell-free expression technology to in vitro systems engineering and optimization. *Synth Syst Biotechnol* **2017,** *2* (2), 97-104.
23. Jewett, M. C.; Swartz, J. R., Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnol Bioeng* **2004,** *86* (1), 19-26.
24. Jewett, M. C.; Calhoun, K. A.; Voloshin, A.; Wuu, J. J.; Swartz, J. R., An integrated cell-free metabolic platform for protein production and synthetic biology. *Mol Syst Biol* **2008,** *4*, 220.
25. Caschera, F.; Noireaux, V., Synthesis of 2.3 mg/ml of protein with an all Escherichia coli cell-free transcription-translation system. *Biochimie* **2014,** *99*, 162-8.
26. Caschera, F.; Noireaux, V., A cost-effective polyphosphate-based metabolism fuels an all E. coli cell-free expression system. *Metabolic Engineering* **2015,** *27*, 29-37.
27. Kurylo, C. M.; Alexander, N.; Dass, R. A.; Parks, M. M.; Altman, R. A.; Vincent, C. T.; Mason, C. E.; Blanchard, S. C., Genome Sequence and Analysis of Escherichia coli MRE600, a Colicinogenic, Nonmotile Strain that Lacks RNase I and the Type I Methyltransferase, EcoKI. *Genome Biology and Evolution* **2016,** *8* (3), 742-752.
28. Forlin M.; Poli I.; De March D.; Packard N.H.; Gazzola G.; Serra R., Evolutionary experiments for self-assembling amphiphilic systems. *Chemometrics and Intelligent Laboratory Systems* **2008,** *90*, 153-160.
29. Caschera, F.; Gazzola, G.; Bedau, M. A.; Bosch Moreno, C.; Buchanan, A.; Cawse, J.; Packard, N.; Hanczyc, M. M., Automated discovery of novel drug formulations using predictive iterated high throughput experimentation. *PLoS One* **2010,** *5* (1), e8546.
30. Caschera, F.; Bedau, M. A.; Buchanan, A.; Cawse, J.; de Lucrezia, D.; Gazzola, G.; Hanczyc, M. M.; Packard, N. H., Coping with complexity: machine learning optimization of cell-free protein synthesis. *Biotechnol Bioeng* **2011,** *108* (9), 2218-28.
31. Cawse, J. N.; Gazzola, G.; Packard, N., Efficient discovery and optimization of complex high-throughput experiments. *Catalysis Today* **2011,** *159* (1), 55-63.
32. ProtoLife Optimization and Discovery for Complex Systems. www.protolife.com (accessed September 1, 2018).
33. Wang, Y.; Zhang, Y. H., Cell-free protein synthesis energized by slowly-metabolized maltodextrin. *BMC Biotechnol* **2009,** *9*, 58.
34. Calhoun, K. A.; Swartz, J. R., An economical method for cell-free protein synthesis using glucose and nucleoside monophosphates. *Biotechnology Progress* **2005,** *21* (4), 1146-1153.
35. Soye, B. J. D.; Patel, J. R.; Isaacs, F. J.; Jewett, M. C., Repurposing the translation apparatus for synthetic biology. *Curr Opin Chem Biol* **2015,** *28*, 83-90.
36. Wen, K. Y.; Cameron, L.; Chappell, J.; Jensen, K.; Bell, D. J.; Kelwick, R.; Kopniczky, M.; Davies, J. C.; Filloux, A.; Freemont, P. S., A Cell-Free Biosensor for Detecting Quorum Sensing Molecules in P. aeruginosa-Infected Respiratory Samples. *ACS Synth Biol* **2017,** *6* (12), 2293-2301.
37. Garamella, J.; Marshall, R.; Rustad, M.; Noireaux, V., The All E-coli TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology. *Acs Synthetic Biology* **2016,** *5* (4), 344-355.
38. Dudley, Q. M.; Karim, A. S.; Jewett, M. C., Cell-free metabolic engineering: biomanufacturing beyond the cell. *Biotechnol J* **2015,** *10* (1), 69-82.
39. Karim, A. S.; Jewett, M. C., A cell-free framework for rapid biosynthetic pathway prototyping and enzyme discovery. *Metab Eng* **2016,** *36*, 116-126.
40. Kim, T. W.; Chokhawala, H. A.; Hess, M.; Dana, C. M.; Baer, Z.; Sczyrba, A.; Rubin, E. M.; Blanch, H. W.; Clark, D. S., High-throughput in vitro glycoside hydrolase (HIGH) screening for enzyme discovery. *Angew Chem Int Ed Engl* **2011,** *50* (47), 11215-8.

41. Mak, W. S.; Tran, S.; Marcheschi, R.; Bertolani, S.; Thompson, J.; Baker, D.; Liao, J. C.; Siegel, J. B., Integrative genomic mining for enzyme function to enable engineering of a non-natural biosynthetic pathway. *Nat Commun* **2015,** *6*, 10005.
42. Kightlinger, W.; Lin, L.; Rosztoczy, M.; Li, W.; DeLisa, M. P.; Mrksich, M.; Jewett, M. C., Design of glycosylation sites by rapid synthesis and analysis of glycosyltransferases. *Nat Chem Biol* **2018**.
43. Hong, S. H.; Kwon, Y. C.; Martin, R. W.; Des Soye, B. J.; de Paz, A. M.; Swonger, K. N.; Ntai, I.; Kelleher, N. L.; Jewett, M. C., Improving Cell-Free Protein Synthesis through Genome Engineering of Escherichia coli Lacking Release Factor 1. **2015,** *16*, 844-853.
44. Martin, R. W.; Des Soye, B. J.; Kwon, Y. C.; Kay, J.; Davis, R. G.; Thomas, P. M.; Majewska, N. I.; Chen, C. X.; Marcum, R. D.; Weiss, M. G.; Stoddart, A. E.; Amiram, M.; Ranji Charna, A. K.; Patel, J. R.; Isaacs, F. J.; Kelleher, N. L.; Hong, S. H.; Jewett, M. C., Cell-free protein synthesis from genomically recoded bacteria enables multisite incorporation of noncanonical amino acids. *Nat Commun* **2018,** *9* (1), 1203.
45. Oza, J. P.; Aerni, H. R.; Pirman, N. L.; Barber, K. W.; ter Haar, C. M.; Rogulina, S.; Amrofell, M. B.; Isaacs, F. J.; Rinehart, J.; Jewett, M. C., Robust production of recombinant phosphoproteins using cell-free protein synthesis. *Nature Communications* **2015,** *6*, 8168.
46. Swartz, J. R., Transforming Biochemical Engineering with Cell-Free Biology. *Aiche J* **2012,** *58* (1), 5-13.
47. Shin, J.; Noireaux, V., Efficient cell-free expression with the endogenous E. Coli RNA polymerase and sigma factor 70. *Journal of biological engineering* **2010,** *4*, 8.
48. Ehrenberg, M., Design and Use of A Fast and Accurate in Vitro Translation System in Ribosomes and Protein Synthesis: A Practical Approach. *Design and Use of A Fast and Accurate in Vitro Translation System in Ribosomes and Protein Synthesis: A Practical Approach* **1990**, 101-129.
49. Johansson, M.; Bouakaz, E.; Lovmar, M.; Ehrenberg, M., The kinetics of ribosomal peptidyl transfer revisited. *Mol Cell* **2008,** *30* (5), 589-98.
50. Lietzke, R.; Nierhaus, K. H., Total reconstitution of 70S ribosomes from Escherichia coli. *Methods Enzymol* **1988,** *164*, 278-83.
51. Culver, G. M.; Noller, H. F., Efficient reconstitution of functional Escherichia coli 30S ribosomal subunits from a complete set of recombinant small subunit ribosomal proteins. *RNA* **1999,** *5* (6), 832-43.
52. Matsubayashi, H.; Ueda, T., Purified cell-free systems as standard parts for synthetic biology. *Curr Opin Chem Biol* **2014,** *22*, 158-62.
53. Fothergillgilmore, L. A., The Evolution of the Glycolytic Pathway. *Trends Biochem Sci* **1986,** *11* (1), 47-&.
54. Spirin, A. S., High-throughput cell-free systems for synthesis of functionally active proteins. *Trends in biotechnology* **2004,** *22* (10), 538-45.
55. Alissandratos, A.; Caron, K.; Loan, T. D.; Hennessy, J. E.; Easton, C. J., ATP Recycling with Cell Lysate for Enzyme-Catalyzed Chemical Synthesis, Protein Expression and PCR. *ACS Chem Biol* **2016,** *11* (12), 3289-3293.
56. Kim, T. W.; Oh, I. S.; Keum, J. W.; Kwon, Y. C.; Byun, J. Y.; Lee, K. H.; Choi, C. Y.; Kim, D. M., Prolonged cell-free protein synthesis using dual energy sources: Combined use of creatine phosphate and glucose for the efficient supply of ATP and retarded accumulation of phosphate. *Biotechnol Bioeng* **2007,** *97* (6), 1510-5.
57. Failmezger, J.; Nitschel, R.; Sanchez-Kopper, A.; Kraml, M.; Siemann-Herzberg, M., Site-Specific Cleavage of Ribosomal RNA in Escherichia coli-Based Cell-Free Protein Synthesis Systems. *PLoS One* **2016,** *11* (12), e0168764.
58. Caschera, F.; Noireaux, V., Compartmentalization of an all-E. coli Cell-Free Expression System for the Construction of a Minimal Cell. *Artif Life* **2016,** *22* (2), 185-95.

59. Kuem, J. W.; Kim, T. W.; Park, C. G.; Choi, C. Y.; Kim, D. M., Oxalate enhances protein synthesis in cell-free synthesis system utilizing 3-phosphoglycerate as energy source. *Journal of Bioscience and Bioengineering* **2006,** *101* (2), 162-165.
60. Moore, S. J.; MacDonald, J. T.; Wienecke, S.; Ishwarbhai, A.; Tsipa, A.; Aw, R.; Kylilis, N.; Bell, D. J.; McClymont, D. W.; Jensen, K.; Polizzi, K. M.; Biedendieck, R.; Freemont, P. S., Rapid acquisition and model-based analysis of cell-free transcription-translation reactions from nonmodel bacteria. *Proc Natl Acad Sci U S A* **2018,** *115* (19), E4340-E4349.
61. Marshall, R.; Maxwell, C. S.; Collins, S. P.; Jacobsen, T.; Luo, M. L.; Begemann, M. B.; Gray, B. N.; January, E.; Singer, A.; He, Y.; Beisel, C. L.; Noireaux, V., Rapid and Scalable Characterization of CRISPR Technologies Using an E. coli Cell-Free Transcription-Translation System. *Mol Cell* **2018,** *69* (1), 146-157 e3.
62. Luisi, P. L.; Ferri, F.; Stano, P., Approaches to semi-synthetic minimal cells: a review. *Naturwissenschaften* **2006,** *93* (1), 1-13.
63. Voloshin, A. M.; Swartz, J. R., Efficient and scalable method for scaling up cell free protein synthesis in batch mode. *Biotechnol Bioeng* **2005,** *91* (4), 516-21.
64. Chizzolini, F.; Forlin, M.; Cecchi, D.; Mansy, S. S., Gene Position More Strongly Influences Cell-Free Protein Expression from Operons than T7 Transcriptional Promoter Strength. *Acs Synthetic Biology* **2014,** *3* (6), 363-371.
65. Caschera, F.; Noireaux, V., Preparation of amino acid mixtures for cell-free expression systems. *Biotechniques* **2015,** *58* (1), 40-3.
66. Hong, S. H.; Kwon, Y. C.; Martin, R. W.; Des Soye, B. J.; de Paz, A. M.; Swonger, K. N.; Ntai, I.; Kelleher, N. L.; Jewett, M. C., Improving cell-free protein synthesis through genome engineering of Escherichia coli lacking release factor 1. *Chembiochem* **2015,** *16* (5), 844-53.
67. Schoborg, J. A.; Hodgman, C. E.; Anderson, M. J.; Jewett, M. C., Substrate replenishment and byproduct removal improve yeast cell-free protein synthesis. *Biotechnol J* **2014,** *9* (5), 630-40.
68. Albayrak, C.; Swartz, J. R., Cell-free co-production of an orthogonal transfer RNA activates efficient site-specific non-natural amino acid incorporation. *Nucleic Acids Res* **2013,** *41* (11), 5949-63.