ACS Synthetic Biology

Open Access
This article is licensed under CC-BY 4.0

pubs.acs.org/synthbio
Research Article

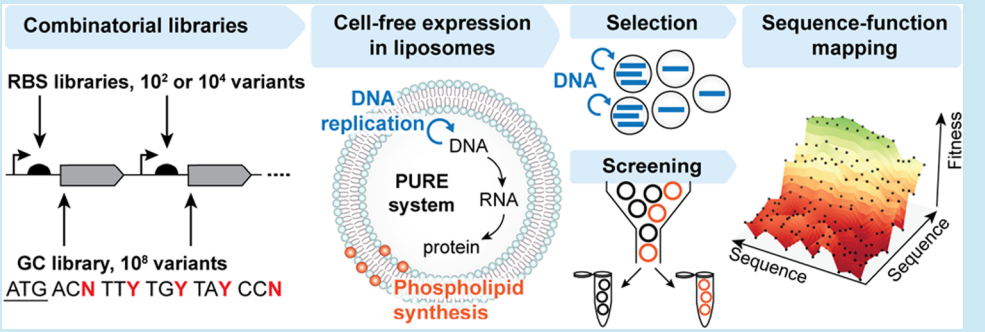
Combinatorial Optimization of Protein Systems in Synthetic Cells

Published as part of ACS Synthetic Biology special issue "Cell-Free Systems".

Marijn van den Brink, Nico J. Claassens, and Christophe Danelon*

Cite This: *ACS Synth. Biol.* 2026, 15, 3285−3299
Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: *In vitro* reconstitution of protein systems—e.g., metabolic pathways, genetic circuits, or biosensors—often requires optimization to enhance their activity. Combinatorial DNA libraries that simultaneously target multiple genes allow for a holistic optimization strategy by studying the interplay between the systems' components, which may reveal DNA variants that would be hidden when testing each element in isolation. Here, we screen large populations of synthetic vesicles that express combinatorial DNA variants of a DNA self-replicator or a phospholipid synthesis pathway. We simultaneously vary the strengths of multiple RBSs or synonymously mutate the first codons of multiple genes to explore the effects of the protein translation rates directly on the functionality of the two core synthetic cell modules. We isolated high performers through DNA self-selection or functional screening by fluorescence-activated cell sorting. Long-read sequencing of the fittest variants revealed the optimal RBS strengths and base substitutions in the first codons and indicated which genes were most impactful in regulating the functionality of the protein systems. Single-mutation data were used to predict the fitness of combinatorial variants, which was compared with the experimental fitness observed. The theoretical fitness of combinatorial variants was extremely predictive for the two-gene library of the DNA replicator but less for the larger pathway library. Altogether, our approach exemplifies how combinatorial testing can be expanded from single proteins to multiprotein systems, which can in the future be extended to the evolutionary engineering of even larger genetic and metabolic networks, and eventually an entire artificial cell.

KEYWORDS: *combinatorial DNA libraries, synthetic cell, directed evolution, cell-free gene expression, DNA self-replication, epistatic interactions*

INTRODUCTION

Cell-free gene expression offers an ideal playground to explore biological designs. It finds applications in prototyping and biomanufacturing and has emerged as a platform to build a synthetic cell.[1,2] The reconstitution of metabolic pathways, genetic circuits, or biosensors often requires optimization steps to improve functional performance. For example, the engineering of regulatory parts, such as promoters, terminators, and ribosome binding sites (RBSs), has become a routine method to control gene expression—and thereby protein concentrations—*in vivo* and *in vitro*, thus enhancing a certain objective function, for instance, maximizing the yield of a desired enzymatic product.[1,3]

Large libraries of DNA variants can be assayed by high-throughput screening of femto- to picoliter cell-free expression reactions encapsulated in lipid vesicles or droplets.[4−13] While many examples of such *in vitro* evolution screens have been reported for the optimization of single genes, multivariate optimization involving multiple proteins is more challenging due to the vast combinatorial space and unpredictable epistatic effects. Examples of multiprotein system optimization targets include (i) the production of proteins at stoichiometric amounts within a macromolecular complex, (ii) the balanced

Received: February 25, 2026
Revised: June 9, 2026
Accepted: July 1, 2026
Published: August 4, 2026



ACS Publications
© 2026 The Authors. Published by American Chemical Society

https://doi.org/10.1021/acssynbio.6c00166
ACS Synth. Biol. 2026, 15, 3285−3299

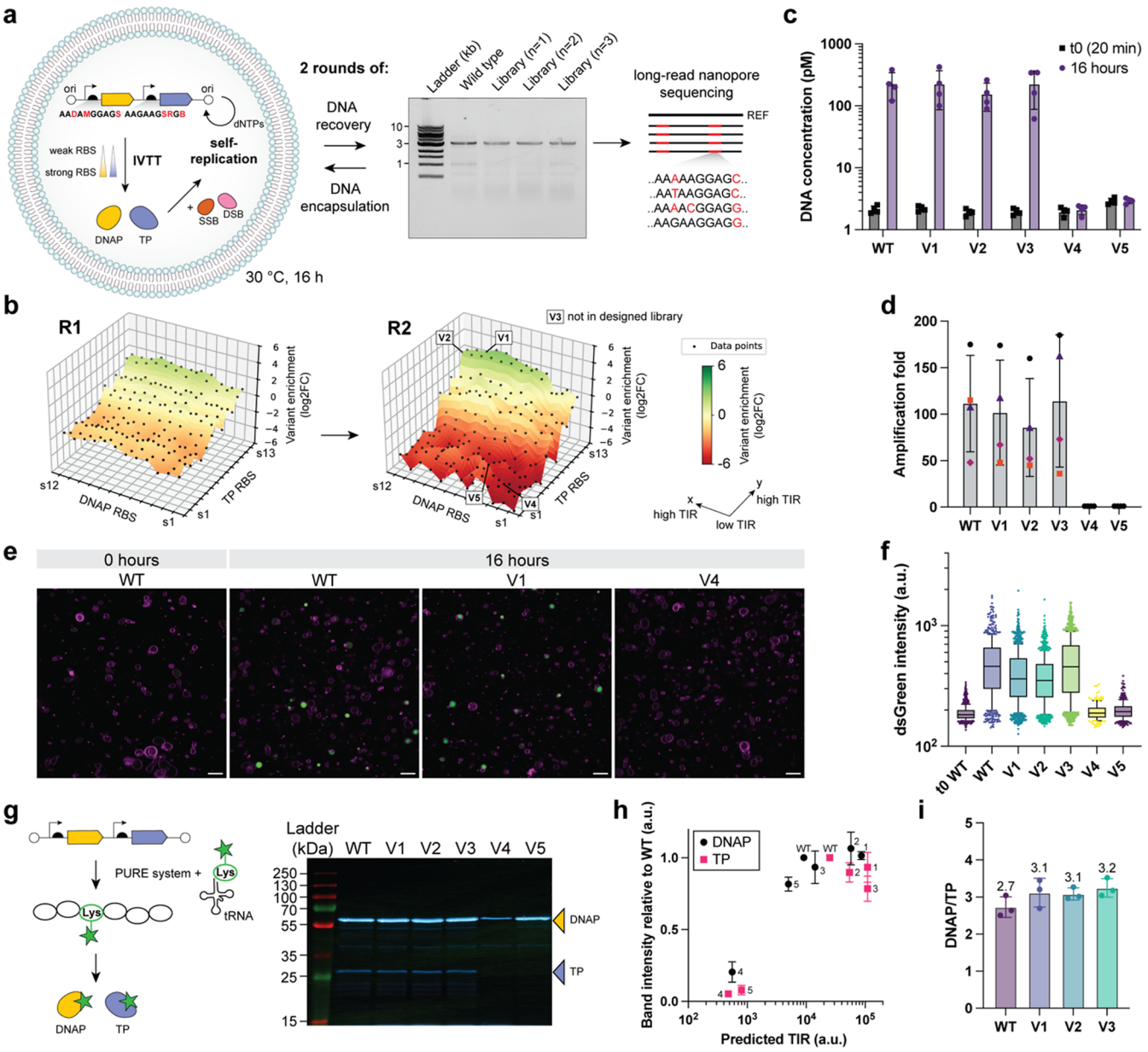


Figure 1. Selection of RBS variants through DNA self-replication. (a) Schematic illustration of *in vitro* transcription-translation and DNA self-replication of the RBS library. DNAP and TP were expressed in liposomes by PURE system. SSB and DSB were supplied as purified proteins. DNA from the liposome samples was recovered by PCR, verified by gel electrophoresis, and analyzed by nanopore sequencing. (b) 3D landscapes displaying the $\log_2$FC in read frequency of combinatorial variants with both RBSs mutated. The RBS variants on the *x*- and *y*-axes are ordered from low to high predicted TIR. The landscapes show the $\log_2$-transformed average FC of three biological replicates. The number of reads covering the 156 variants is 28,978 in R0 (preselection), 10,140 to 16,998 per replicate in R1, and 3900 to 6060 per replicate in R2. Data from the individual repeats are shown in Figure S3. (c) Quantification of DNA concentration by qPCR for clonal DNA variants V1−5 before replication (t_0, incubation 20 min) and after replication (16 h) at 30 °C (n = 4 biological replicates). (d) Amplification fold of the data in panel c, calculated as the DNA concentration at 16 h divided by the concentration at 20 min. Symbols represent the different biological replicates (n = 4). (e) Confocal fluorescence images of liposomes with different clonal DNA variants. Magenta, Cy5-conjugated lipids; green, dsGreen; Scale bars: 10 μm. (f) Quantification of the dsGreen intensity of dsGreen-positive liposomes (>129 a.u.) from confocal microscopy images. Number of dsGreen-positive liposomes: 599 (t_0 WT), 587 (WT), 1056 (V1), 1040 (V2), 1255 (V3), 170 (V4), and 332 (V5). (g) SDS-PAGE of bulk (no liposomes) PURE reaction samples from clonal DNA variants with GreenLys protein labeling reagent. The fainter bands for TP compared to DNAP could be the result of the lower number of lysine residues of TP (27) compared to DNAP (64). (h) Band intensities from the SDS-PAGE gel in panel g relative to the wild type versus the corresponding predicted TIRs (n = 3). The DNA variant is indicated for each data point. (i) Ratio of corrected band intensities for DNAP over TP. Corrected band intensities were calculated by dividing the adjusted band intensity (intensity with background subtracted) by the number of lysines per protein (64 for DNAP, 27 for TP).

production of enzymes involved in a pathway to enhance metabolic fluxes and avoid accumulation of (toxic) intermediate products, (iii) fine-tuning gene expression profiles for efficient allocation of resources, and (iv) the regulation of genetic networks to promote the integration of and coordinate different biological functions. Living organisms navigate such complexity through natural evolution across extremely long timespans. Accelerating the optimization of multigene

https://doi.org/10.1021/acssynbio.6c00166
ACS Synth. Biol. 2026, 15, 3285−3299

functions through directed evolution would require approaches that reduce the problem of combinatorial explosion, for instance, by generating 'small but smart' libraries, and implementing selection procedures that retain the best-performing designs.[14−16]

In this study, we performed *in vitro* evolution of two genetic modules, each encoding multiple proteins involved in DNA replication or phospholipid synthesis. The genomes were expressed with PURE system (Protein synthesis Using Recombinant Elements)[17] in liposome-based synthetic cells. We here define 'synthetic cell' as an umbrella term for the compartmentalized expression of genes encoding a cellular function. We investigated the effects of modified RBS sequences or the GC content of the first codons on the functionality of the two genetic systems. It is known from previous studies that translation initiation is often the rate-limiting step in bacterial translation,[3,18] and that the initiation rate also controls protein yield in PURE system.[19,20] However, it remains unclear how changing the translation rates of multiple genes impacts the overall expression pattern in PURE system, possibly distorted by epistatic interactions between genes competing for translation resources (e.g., tRNAs, tRNA synthetases, amino acids). It is also unknown whether control over protein amounts is effective in regulating the functionality of the protein systems in synthetic cells.

To address these questions, we expressed semirational, combinatorial DNA libraries in synthetic cells. The fittest variants were isolated either through selection by DNA self-replication or through a FACS-based liposome screening method with a reporter to detect the output molecule of the phospholipid synthesis pathway. Long-read sequencing analysis of the fittest variants revealed the optimal RBS sequences and bases in the first codons of the genes. The data also suggested which genes were most impactful in regulating the functionality of the modules and whether the effects of multiple mutations were predictable or unexpected. Altogether, these results lay the groundwork for future evolutionary engineering efforts by demonstrating how an increased DNA template complexity can be harnessed in synthetic cell systems.

RESULTS

Design of a Combinatorial Library of RBS Variants in a DNA Replicator

Protein-primed DNA replication was established in liposomes by the *in vitro* expression of two proteins from the *Bacillus subtilis* bacteriophage Phi29[21] (Figure 1a). The Phi29 DNA polymerase (DNAP) and terminal protein (TP) were produced from a linear template flanked with the origins of replication from the Phi29 genome. DNAP and TP form a 1:1 heterodimer and recognize the terminal origins of replication, where TP forms a covalent bond with the first deoxyadenosine monophosphate (dAMP) of the new 5′ DNA strand.[22] The polymerase then elongates the TP-primed DNA strand with strand-displacement of the original DNA. The process is further facilitated by the addition of purified double-stranded binding protein (DSB) and single-stranded binding protein (SSB) to unwind the DNA at the origins and stabilize displaced single-stranded DNA, respectively.

The effects of changing the absolute and stoichiometric amounts of DNAP, TP, and DNA on replication efficacy have been studied using purified proteins in simple buffer conditions.[23−25] However, it is unknown how the production rates of DNAP and TP impact DNA replication in this reconstituted autocatalytic network, where protein concentrations vary over time and DNA is also the substrate of RNA polymerases, which may collide with DNAP.[26] To test a range of protein production levels, we mutated the RBSs for each of the two coding sequences on the template. We targeted the RBSs, because (i) protein production level is better correlated with translation rate than transcription rate in PURE system,[19,20] and (ii) translation initiation is often a major rate-limiting step in protein translation in *Escherichia coli*,[18,27] thus possibly also in PURE system.[28] Because the strength of an RBS is influenced by the ∼35 nucleotides upstream and downstream of the start codon,[18,29−31] we designed a specific set of RBS sequences for each of the protein coding sequences using RBS Calculator.[18] The library was designed to maximize the range of predicted translation initiation rates (TIRs) per gene, while keeping the number of mutations between one and five base pair substitutions relative to each wild-type RBS (Table S1). The combinatorial library consisted of 12 RBS variants for DNAP and 13 RBS variants for TP; thus, in total, 156 combinatorial variants. We constructed the library using our recently published protocol for multiplex one-step single-stranded DNA annealing protein (SSAP)-mediated plasmid diversification (MOSAIC).[32] In short, this method relies on plasmid recombineering in an *E. coli* strain equipped with recombination machinery and requires just one coelectroporation of the target plasmid and degenerate ssDNA oligos encoding the mutations of interest. After one round of MOSAIC, 33% of the DNA library variants harbored two mutated RBSs (Figure S1). The remaining fraction of the library consisted of variants with only one modified RBS (46%) and the wild type (22%).

Self-Selection of RBS Variants through DNA Replication

To establish a genotype-phenotype link, the library of DNA replicators was expressed by PURE system in liposomes at an average DNA copy number of one per liposome (λ = 1, Methods section). After 16 h of incubation at 30 °C, quantitative PCR (qPCR) analysis confirmed 10 to 100-fold DNA amplification (Figure S2a,b). The resulting population of full-length DNA was amplified from the liposome sample by PCR and re-encapsulated in liposomes for a second round of *in vitro* transcription-translation-replication (Figures 1a and S2c). Nanopore sequencing of the initial library and the recovered DNA after rounds one and two revealed changes in combinatorial variant frequencies (Figure 1b). Certain combinations of RBS variants were enriched ($\log_2$ fold change ($\log_2$FC) > 0), while others decreased in frequency ($\log_2$FC < 0) or disappeared from the library (Figures 1b and S3). Overall, the sequence landscape showed a gradual increase of fitness, defined as $\log_2$FC, with increasing predicted TIR. The most enriched variants harbored strong RBSs for both TP and DNAP. Interestingly, the fitness was mostly dependent on the predicted TIR of TP and less of DNAP (Figure S4). This could mean that either TP is required in higher amounts and/or the TP RBS library covers a lower expression range than predicted. Specifically, for both DNAP and TP, the $\log_2$FC of the RBS variants increased with the predicted TIR up to a certain TIR, whereafter the $\log_2$FC plateaued (TP) or slightly decreased (DNAP) (Figure S4). Some RBS variants led to unpredictable results, indicated by a rugged pattern, most pronounced for TP RBSs s8-s11, possibly due to inaccurate predictions of the TIRs for some RBSs (Figures S3 and S4).

https://doi.org/10.1021/acssynbio.6c00166
ACS Synth. Biol. 2026, 15, 3285−3299

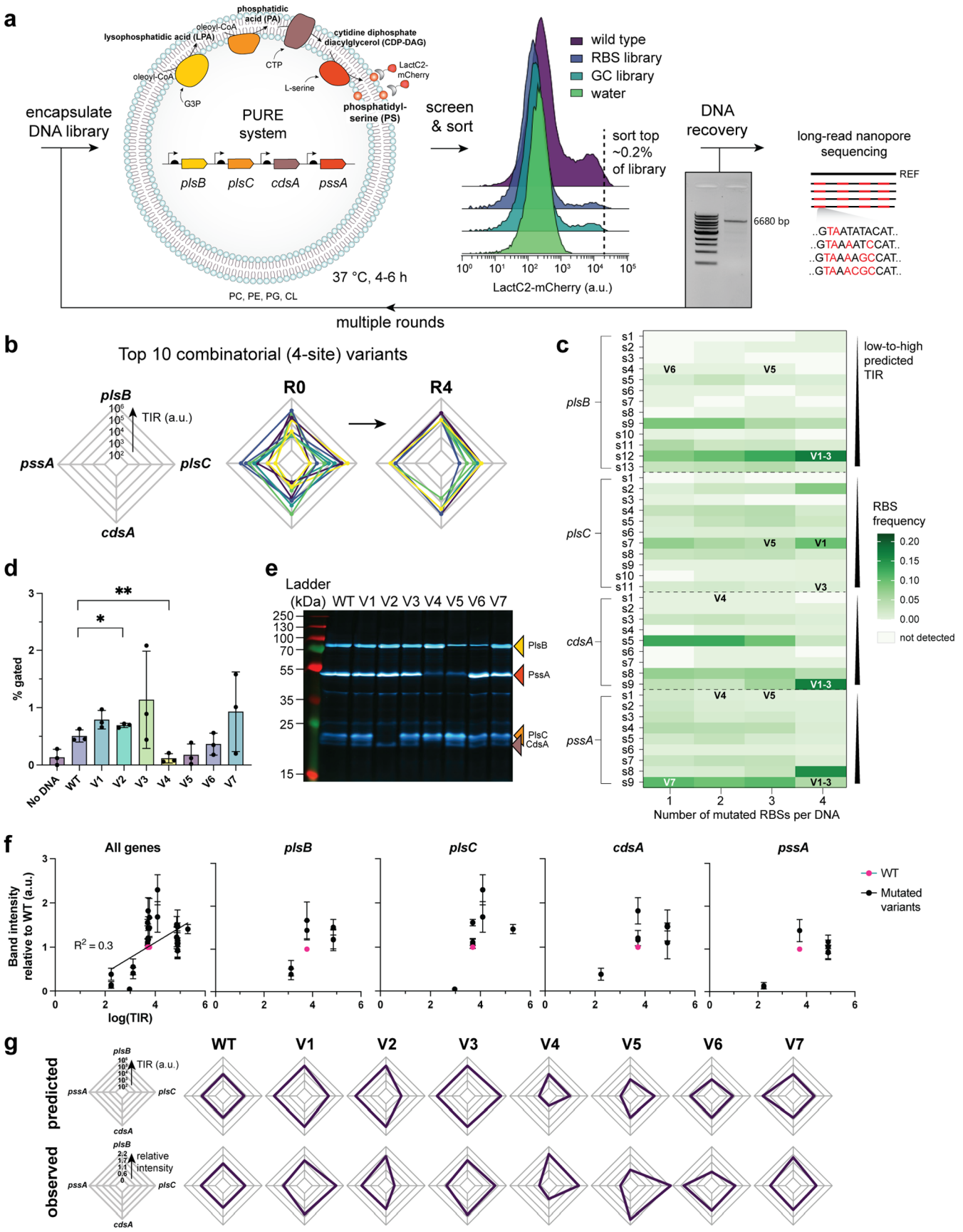


Figure 2. Activity-based selection of DNA designs for PS production from a library of four-site combinatorial variants. (a) Schematic illustration of the workflow. Liposomes express four enzymes from a DNA template with diversified RBSs. The enzymes form a pathway to produce PS, measured by the LactC2-mCherry probe. Four subsequent rounds of sorting the liposomes with the highest LactC2-mCherry signal and re-encapsulation of the recovered DNA were performed prior to long-read sequencing. (b) Radar charts displaying the top 10 combinatorial variants with the highest

https://doi.org/10.1021/acssynbio.6c00166
ACS Synth. Biol. 2026, 15, 3285−3299

Figure 2. continued

frequency in the initial library (R0) and after four rounds of sorting. The line colors represent different 4-site variants. The variants are listed in Table S6. (c) Heatmap displaying the frequency of each RBS variant. The frequency was calculated relative to the library fractions with N number of mutated RBSs per DNA. Number of reads per library fraction (each column in the heatmap): 9123 (1-site), 8510 (2-site), 3119 (3-site), and 697 (4-site). (d) Analysis of clonal DNA variants from the RBS library by flow cytometry. The percentage of gated liposomes was calculated using a similar gate as the one used for sorting the DNA library. Unpaired t-test: * $P \leq 0.05$; ** $P \leq 0.01$. $n = 3$ biological replicates. (e) SDS-PAGE with bulk PURE system reactions (no liposomes) containing the GreenLys reagent for cotranslational labeling of proteins expressed from clonal DNA variants. (f) Band intensities were measured from the SDS-PAGE gels in panel e and Figure S15a, normalized to the wild type, and plotted against the corresponding predicted TIRs ($n = 3$ biological replicates for PlsB and PssA, and $n = 2$ for PlsC and CdsA). The appended curve is the linear regression for all genes combined, $R^2 = 0.3$, $P < 0.0001$. (g) Radar charts displaying the predicted TIRs and observed relative protein expression levels (averages of replicates).

Three biological replicates resulted in similar frequency changes, indicating high reproducibility of the self-selection system behavior (Figure S3). Furthermore, while mutations could be introduced during in-liposome replication by the Phi29 DNA polymerase or during DNA recovery by PCR, no mutations outside the targeted RBS regions were detected in the DNA population with frequency ≥1%, demonstrating that the observed variations in DNA replicator fitness exclusively come from the mutations in the RBS sequences.

Reverse Engineering of Clonal Self-Replicator Variants Reveals High Performance

We then sought to validate that the observed changes in variant frequency were proportional to the ability to self-replicate (i.e., the fitness of each DNA variant). Therefore, we constructed a few clonal variants from the sequence landscape, and we assessed their self-replication ability in liposomes using qPCR (Table S2 and Figure 1c,d). Two high-frequency variants, V1 and V2, showed high replication folds. In contrast, V4 and V5, two variants in the "valley" of the sequence landscape, did not replicate at all. We tested another DNA variant, V3, which had the same TP RBS as V1 but a DNAP RBS that was not part of the theoretical library. This DNAP RBS variant was most likely generated as a byproduct during library generation with MOSAIC or linearization of the library with PCR, as it was already present before encapsulation in liposomes. V3 started at a 50-fold lower frequency than V1 in the initial library (R0), but V3 was enriched 3.2 ± 0.5 fold more than V1 in R2 (mean ± standard deviation, $n = 3$). Indeed, clonal V3 also replicated well (Figure 1c,d).

In addition to the population-level qPCR analysis, we then investigated the performance of each replicator variant at the single-liposome level using confocal fluorescence microscopy (Figure 1e). We detected DNA-replicating liposomes based on the intensity of the fluorescent dye dsGreen, which binds specifically to double-stranded DNA. We plotted the average dsGreen intensity per liposome (Figures 1f and S5) and found results similar to those of the qPCR data, corroborating that the variant frequency change ($\log_2$FC) is indeed an accurate measure for its fitness.

Lastly, we directly measured the amounts of synthesized proteins from the different RBS variants to examine their correlation with fitness. The proteins were expressed in bulk reactions in the presence of tRNA preloaded with fluorescently labeled lysine, and we quantified the fluorescence intensity of the bands after SDS-PAGE (Figures 1g,h and S6). V1−3 showed clear bands for DNAP and TP, with intensities similar to the wild type, despite having higher predicted TIR values for both DNAP (14,117−86,579 a.u.) and TP (54,294−110,313) (comparison to wild type (WT): DNAP: 9123; TP: 25,228). This observation coincides with the plateauing fitness ($\log_2$FC) at high predicted TIRs for DNAP and TP in the library data set (Figure S4) and with the similar DNA amplification folds for the high-TIR clonal variants (Figure 1d−f). The estimated ratio of molecules DNAP over TP, derived by normalizing the band intensity to the number of lysine residues per protein, increased slightly from 2.7 for the wild type to 3.1−3.2 for the library variants V1−3 (Figure 1i). In contrast, V4 and V5, which have lower TIRs for both DNAP and TP, showed reduced DNAP and TP expression levels, most likely impairing their ability to replicate. Notably, we here quantified the end-point expression level, which may differ from the protein yield and stoichiometry at earlier time points in the reaction.

Design of a Combinatorial RBS Library of a 4-Gene Phospholipid Synthesis Pathway

Next, we performed multivariate optimization of a four-gene phospholipid synthesis pathway, generating a cascade of reactions catalyzed by membrane-associated or transmembrane proteins (Figure 2a). We selected the four upstream enzymes of the *E. coli* Kennedy phospholipid pathway: PlsB, PlsC, CdsA, and PssA. These enzymes convert oleoyl-coenzyme A (oleoyl-CoA) and glycerol-3-phosphate (G3P) to phosphatidylserine (PS) in the presence of the cosubstrates cytidine triphosphate (CTP) and L-serine. The amount of final product PS is positively correlated with the fluorescence intensity of the PS-specific reporter LactC2-mCherry, enabling the quantitative assessment of PS production and membrane incorporation at the single-liposome level.[33]

While *in vivo*, the phospholipid synthesis pathway is mainly regulated post-translationally by an (unknown) allosteric regulator of PlsB activity[34] and metabolic regulation of phosphatidylglycerol (PG) and phosphatidylethanolamine (PE),[35] we reasoned that a minimal synthetic approach based on translational control could modulate phospholipid production *in vitro*. We hypothesized that the yield of PS production depends on the stoichiometry of the expression of the four enzymes, as observed in other enzymatic cascades.[14,36,37] Previously, it was found that phosphatidic acid (PA), the product of PlsC, significantly accumulates in the Kennedy pathway reconstituted in liposomes, suggesting potential for further optimization.[33] However, the yield of PS has a nontrivial dependence on the four RBS sequences, as the rate-limiting step is unknown and unexpected effects of mutating RBSs could arise from protein folding, membrane insertion/binding, allosteric regulation, inhibitory effects, or resource competition.

To tackle this multivariate problem, we generated a 4-site combinatorial RBS library by MOSAIC consisting of 13 (PlsB) × 11 (PlsC) × 9 (CdsA) × 9 (PssA) = 11,583 variants (Table S3). With a MOSAIC editing efficiency of ∼50% per target locus, we obtained a library that consisted for 11% of variants

https://doi.org/10.1021/acssynbio.6c00166
ACS Synth. Biol. 2026, 15, 3285−3299

with all RBSs mutated, and the remainder of the library consisted of library fractions where zero, one, two, or three RBSs were mutated (Figure S7a,c). Using nanopore sequencing, we detected all designed variants that had one or two mutated RBSs, and 84% and 35% of all designed variants with three and four mutated RBSs, respectively, but more of the designed variants are likely covered in the generated library as suggested by rarefaction plots (Figure S8). Hereafter, we name the fractions of the library with N mutated target loci (RBSs) "N-site fraction" (e.g., 1-site fraction).

Enrichment of Library Variants with High PS Yield through Screening by FACS

We encapsulated the linearized library in liposomes at an average DNA copy number per liposome $\lambda = 0.2$ (Methods). Functionality screening of the liposomes was performed by measuring the fluorescence intensity of the PS probe LactC2-mCherry with flow cytometry. We typically screened ∼3−6 million liposomes, of which 2−13% showed PS production (Figures S9 and S10). We calculated the screening coverage of the DNA library as shown in Table S4, taking into account that 13 ± 8% of a clonal liposome population is active (i.e., the performance of the wild type, Figure S9a) (mean ± standard deviation, $n = 5$ biological replicates). The screening coverage was approximately 3×10^3-fold for the 1-site fraction (42 variants), 3×10^2-fold for the 2-site fraction (656 variants), 3×10^1-fold for the 3-site fraction (4,518 variants), and 6-fold for the 4-site combinatorial library fraction (11,583 DNA variants). Using FACS, we sorted the liposomes with the highest LactC2-mCherry intensity, less than 1% of all liposomes, corresponding to 10,000−30,000 liposomes. We amplified the DNA from the sorted liposome sample by PCR and analyzed the DNA sequences by nanopore sequencing. We performed four subsequent rounds of sorting, DNA amplification, and re-encapsulation in liposomes.

The combinatorial (4-site) genotypes of the top ten most frequent DNA variants are presented in Figures 2b and S12. Due to the relatively low sequencing coverage of the 3-site and 4-site library fractions in the presort library (R0) (on average 4 ± 4 (3-site) and 2 ± 2 (4-site) reads per variant), we used the read frequency as a measure of fitness instead of calculating the fold change in frequency relative to R0. Comparing the initial library to that after four rounds of sorting, we see a strong shift toward high protein expression levels for PlsB, CdsA, and PssA, while a range of RBS strengths was present for PlsC. To obtain a more holistic view of the full data set, we then plotted the average frequency of each RBS for each fraction of the library (Figures 2c and S12). These results support the observation that high-fitness variants contain *plsB*, *cdsA*, and *pssA* RBSs with high predicted TIRs (*plsB*: 21,409−112,986 au; *cdsA*/*pssA*: 31,879−78,824 au), but a more even distribution of frequencies is observed for *plsC* RBSs. An exception is s5 from *cdsA*, which has a relatively high frequency considering its TIR of 4501.

The wider range of RBS strengths for PlsC among the top performers can be interpreted in the light of our previous study reporting that, among the intermediate products of the Kennedy pathway, only PA accumulated in significant amounts.[33] The conversion of PA by CdsA limits the flux, rendering PS production poorly sensitive to the PA production rate and thus also to the amount of PlsC.

Reverse Engineering of Clonal Variants of the Phospholipid Synthesis Pathway

Due to the repeated PCR steps involved, some mutations accumulated with a frequency of >0.04 at off-target locations. To validate the fitness of DNA variants as predicted by the sequencing data, we constructed three clonal variants of expected high fitness with strong RBSs for *plsB*, *cdsA*, and *pssA* and varying strengths of the *plsC* RBS (V1−3) (Figure 2g (**top row**) and Table S5) and two clonal variants of expected low fitness, one with weak *cdsA* and *pssA* RBSs (V4) and one with weak *plsB* and *pssA* RBSs and stronger *plsC* RBS (V5). All DNA variants led to PS production in liposomes (Figures S13 and S14). However, the population sizes of the top-producing liposomes differed among the variants. Specifically, we quantified the percentage of liposomes in the strict LactC2-mCherry gate, similar to the sorting gates in the library enrichment phase (i.e., the selection pressure) (Figure 2d). As expected, we found that V1−3 performed well, with 1.4- to 2.2-fold larger gated population sizes (mean values across three replicates) compared to that of the wild type (unpaired t-test for WT versus V2 was significant ($P = 0.048$), but those for WT versus V1 or V3 were not). In contrast, V4 and V5 performed poorly. We also found suboptimal performance of a variant with a predicted weak RBS for *plsB* (V6) and good performance of a variant with a predicted strong RBS for *pssA* (V7), as expected from their frequencies in the sorted library, as shown in Figure 2c. We also compared the V3, V5, and WT designs on the single-liposome level using confocal fluorescence microscopy. The data suggest improved PS synthesis of V3 over the WT and decreased activity for V5 (Figure S14). Altogether, the experimentally determined performances of the clonal variants are in agreement with our expectations on the basis of the sequenced libraries.

Next, we quantified the relative protein expression levels for V1−7 using a GreenLys SDS-PAGE assay (Figures 2e−g and S15) and interpreted the efficiency of the lipid biosynthesis pathway considering these enzyme amounts. This assay confirmed that PS is still produced with very low PlsC expression (V2). However, a reduction in the production of PlsB (V6), or CdsA and PssA (V4), or PlsB and PssA in combination with an increase in PlsC production (V5) reduced PS production (unpaired t-test for WT versus V4 was significant ($P = 0.007$), but those for WT versus V5 or V6 were not, Figure 2d). High expression of PlsC (V1) did not significantly change PS production compared with low expression (V2). Note that CdsA migrates faster than expected based on its molecular weight (31 kDa), as described earlier.[38]

We found a weak positive linear correlation between the predicted log TIR and the observed protein expression level as quantified from GreenLys SDS-PAGE ($R^2 = 0.3$) (Figure 2f). We observed a few instances suggesting that a change in the production level of a given protein could influence the production levels of other proteins whose RBSs remained the same (Figure 2f, plots for separate genes). Hence, the same TIR may give different band intensities for different genetic backgrounds. For instance, in V4, upon reducing CdsA and PssA amounts by introducing weak RBSs, the expression levels of PlsB and PlsC increased despite their wild-type RBSs being unmodified (Figure S15c). Furthermore, a *plsC* RBS with TIR 12,360 au led to a higher PlsC production in combination with the downregulation of PlsB and PssA (V5) than in combination with strong RBSs for PlsB, CdsA, and PssA (V1) (Figure S15c). In contrast, a strong reduction in PlsC

https://doi.org/10.1021/acssynbio.6c00166
ACS Synth. Biol. 2026, 15, 3285−3299

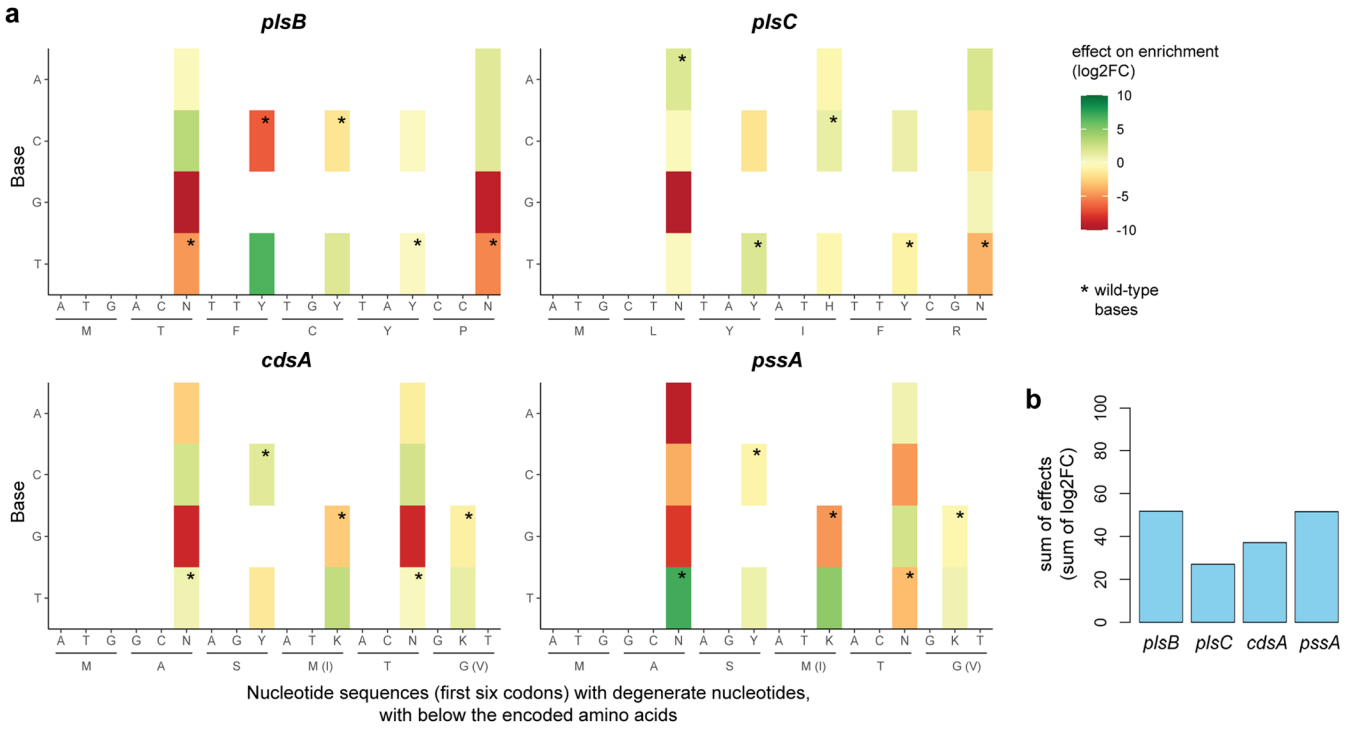


Figure 3. Base-specific effects at specific nucleotide positions in the first six codons of the phospholipid synthesis pathway genes. Data are presented for the 1-site library fraction in R4, but the results across the four library fractions are similar (plots available in the accompanying data set in the 4TU.ResearchData repository). Number of sequencing reads: 5827 (*plsB*), 10,254 (*plsC*), 14,594 (*cdsA*), and 3350 (*pssA*). (a) The effect on enrichment (log_2FC) was calculated by the fold change of the mean frequency of variants with a certain base at the specific nucleotide position over the mean frequency of variants with the other tested bases. The WT bases are indicated with an asterisk. The encoded amino acids in brackets match the non-WT bases. (b) Sum of absolute per-base effects on enrichment for each gene.

production in V2 compared to V1 and V3 did not result in a higher expression of the other three proteins. These findings suggest that mutating RBSs of multiple genes leads to unpredictable behavior of the overall expression pattern in PURE system.

Mutations within the First Six Codons Modulate Phospholipid Synthesis

We applied a second strategy to modulate the translation initiation rates of the four enzymes in the phospholipid synthesis pathway. The GC content of the sequence directly downstream of the start codon affects the translation rate *in E. coli*, most likely through the tendency of GC-rich sequences to form stronger secondary mRNA structures, which hinder translation initiation.[31] This relationship has recently been observed in PURE system as well.[19,39,40] Inspired by these studies, we varied the GC content of the first six codons between 28 and 56% for *plsB*, 22−50% for *plsC*, and 33−61% for *cdsA* and *pssA*. Only the third base of the codons was mutated such that the mutations were silent, except for the fourth and sixth codons of *cdsA* and *pssA*. A methionine was mutated to isoleucine at the fourth codon of both genes to study a potential positive or negative effect of introducing a second start codon close to the canonical one. At the sixth codon, glycine was mutated to valine to expand the range of GC content while accepting a potential effect of amino acid substitution.

The theoretical DNA library consisted of 127 (PlsB) × 191 (PlsC) × 127 (CdsA) × 127 (PssA) = 391,241,153 combinatorial (4-site) DNA variants, which is over 33,000-fold larger than the RBS library. Based on the performance of the wild type (Figure S9a), we obtained approximately 3 × 10^2-fold screening coverage for the 1-site fraction (572 variants), while only ≤1-fold coverage for the fractions with multiple modified target loci (calculation explained in Table S7). However, since many DNA variants have the same GC content, we screen the combinatorial GC variants many times (Table S7). Therefore, the screening was not an exhaustive search of all combinatorial DNA variants, but of all GC contents.

After four rounds of sorting, we found that some DNA variants were enriched and many variants were depleted from the library (Figure S16). Generally, most DNA variants were enriched in all library fractions, where each library fraction has a different number (N) of mutated loci per DNA (N = 1, 2, 3, or 4). This observation suggests that these variants were beneficial in different genetic contexts. However, there was no trend visible toward high or low GC contents for the different genes. The most enriched GC variants and their predicted TIRs are given in Table S8.

We also investigated the base-specific effects of each randomized position on PS synthesis (Figure 3a). Site-specific preferences were identified for the tested bases, although no overall preference for A or T over G or C was visible. Interestingly, the effect of the base G at the third position of the second codon was found to be very negative for all genes. This negative effect of the G at this specific position was also discovered in a recent systematic *in vivo* study involving an ultradeep characterization of a large range of randomizations at the 5′-UTR and the start of the coding sequence.[31] Thus, this shows an analogy between *in vivo* and *in vitro* conditions. A few other bases were found to have a severe effect: G in the sixth codon of *plsB*, G in the fifth codon of *cdsA*, and A in the

https://doi.org/10.1021/acssynbio.6c00166
ACS Synth. Biol. 2026, 15, 3285−3299

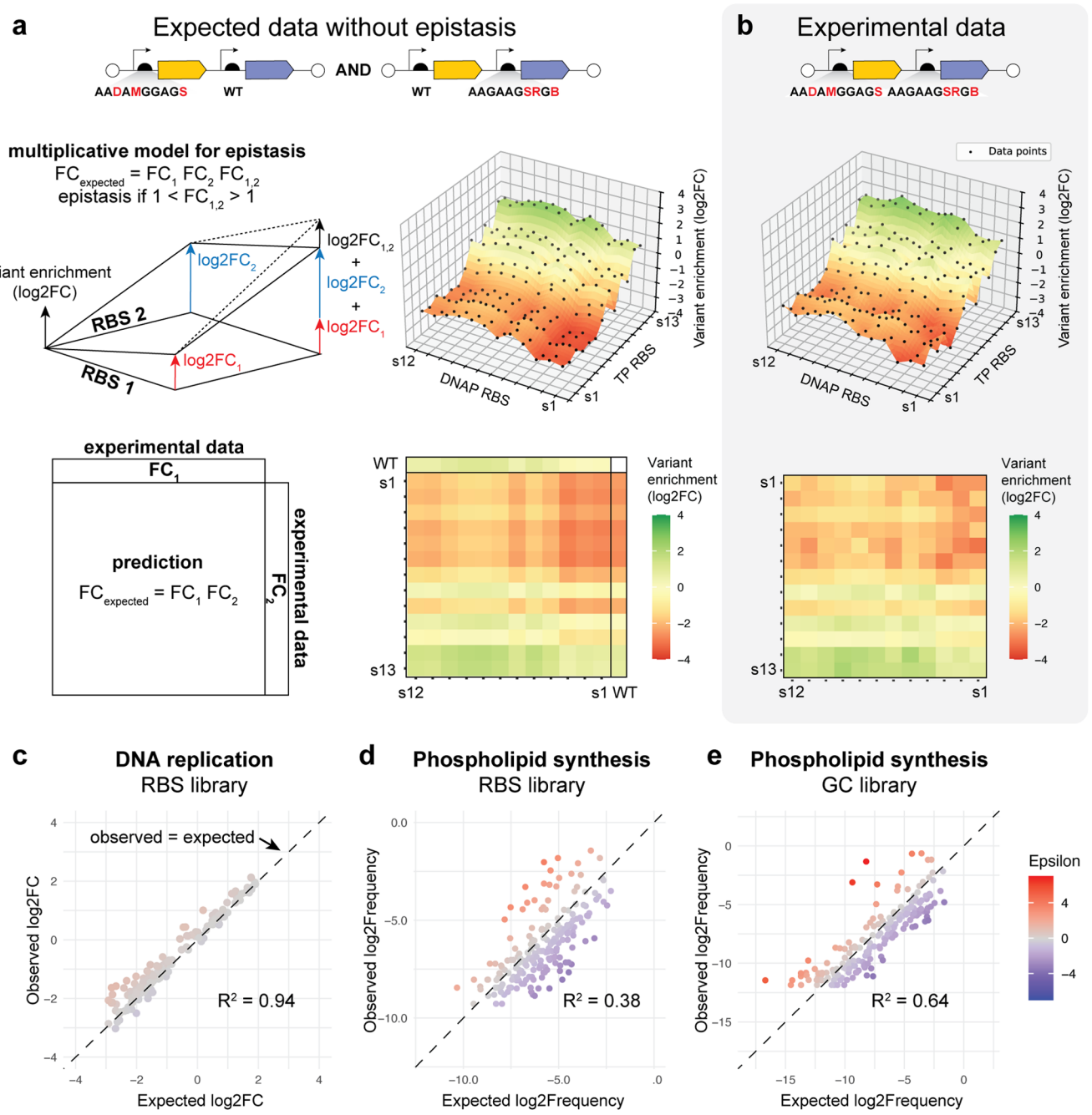


Figure 4. Predicting the fitness of combinatorial variants based on single-mutated variants. (a) Expected combinatorial variant enrichment (log_2FC) for the DNA replication RBS library. Prediction was based on a multiplicative model using the enrichment data of DNA variants with one WT and one modified RBS (1-site library fraction). The average of three biological replicates is plotted, where R0 contains 39,928 reads, and R1 contains 33,834 to 56,684 reads per replicate. (b) Experimental data of variant enrichment for the combinatorial library of RBSs of the DNA replicator (average of 3 replicates). (c) Observed versus expected log_2FC for the DNA replication RBS library (average of 3 replicates) for all 156 combinatorial variants. (a–c) Data for selection round 1 (R1) were used because, unlike R2, all DNA variants were detected by Nanopore sequencing, offering a better resolution for low-fitness variants. (d, e) Observed versus expected log_2frequency for the phospholipid synthesis RBS library (d) and GC library (e). Data are from R4 (one biological replicate). Prediction is based on a 1-site library fraction. Data points: 205 of the 656 (d) and 189 of the 121,158 (e) pairwise combinatorial variants. (c–e) Dashed line indicates $y = x$. Epsilon is the difference between the observed and expected fitness, i.e., epsilon = $log_2FC_{observed} - log_2FC_{expected}$ (c), or epsilon = $log_2frequency_{observed} - log_2frequency_{expected}$ (d, e).

second codon of *pssA*. Lastly, changing the second methionine to isoleucine at the fourth codon seems beneficial for both *cdsA* and *pssA*.

To determine the total impact of the base substitutions for each gene on PS synthesis, we summed the effects of each base. The sum was the highest for *plsB* and *pssA*, suggesting that the mutations in these genes had the largest effect on PS synthesis. In contrast, mutations in the *plsC* gene had the smallest effect on PS production of all genes (Figure 3b). This is in agreement with the RBS library screening, which showed that both low and high amounts of PlsC promoted PS synthesis (Figure 2d).

We further investigated the effect of the mutations for the most impactful genes, *plsB* and *pssA*, by constructing clonal variants with bases yielding a more positive effect for PlsB

https://doi.org/10.1021/acssynbio.6c00166
ACS Synth. Biol. 2026, 15, 3285−3299

(V8−11) or a more positive (V10) or negative (V11) effect for PssA (Table S9). SDS-PAGE analysis revealed that the PssA expression level was lower with V11, as expected (Figure S15c). However, no major increases in the PlsB or PssA quantity were observed. Nevertheless, a deletion upstream of the *plsB* gene of V11 and a deletion upstream of the *pssA* gene of V9, which were likely introduced during the cloning of these variants, may have affected the TIR of the proteins, undermining a direct effect of the targeted nucleotides for these two variants.

Predicting Fitness of Combinatorial Variants Based on Single Variants

Lastly, we examined how predictable the combinatorial variants from the three tested libraries were. Understanding whether combinations of mutations lead to expected or unexpected effects can shed light on the presence of epistatic interactions and clarify whether testing variants in a combinatorial manner is necessary compared to evaluating them individually. Epistasis occurs when the observed fitness (here, variant frequency) deviates from the expected fitness when multiple mutations are combined. For example, the effect of a combination of two RBSs may be larger or smaller than the effect of the first RBS multiplied by the effect of the second RBS (multiplicative model), such that the fitness of the double mutant cannot directly be predicted from analyzing each position in isolation. Epistatic effects may, for example, appear when multiple genes compete for translation resources. For example, changing the TIR of some genes may indirectly influence the expression level of other genes (improving or inhibiting) by modifying the utilization of tRNAs and tRNA synthetases, a phenomenon that is still poorly understood in cell-free protein synthesis systems.

We first studied the presence of epistatic interactions for the DNA replication RBS library by comparing the observed fitness of the combinatorial variants in the 2-site library fraction with the expected fitness derived from the single variants in the 1-site library fraction. We calculated the expected FC of the combinatorial variants without any epistatic effects by multiplying the FC of each individual RBS (Figure 4a). We then compared the expected and observed sequence landscapes ($\log_2$FC, Figure 4b). Both are very similar, with an R^2 of 0.94 (Figure 4c), and none of the variants exhibited significant epistatic deviations (multiple *t*-tests, Figure S17), suggesting that no epistatic interactions governed DNA replication. Alternative calculations using the 2-site library fraction only (taking the average fold change) or using frequency instead of FC led to similar conclusions (Figure S18). Therefore, the combinatorial variants of this DNA replication library were extremely predictable, and if there was any competition for translation resources, it had minimal effect on DNA self-replication.

Similarly, we predicted the fitness of the pairwise combinatorial DNA variants of the phospholipid synthesis pathway libraries. Here, we calculated the expected frequency instead of the expected $\log_2$FC because of the lower sequencing depth of the presort, 2-site library fraction (24 ± 13 reads per variant for the RBS library) compared to the DNA replication library (186 ± 100 reads per variant) (mean ± standard deviation). The correlation between the expected versus observed fitness for the two PS synthesis libraries was weaker than that for the DNA replication library (R^2 = 0.38 for the RBS library, 0.64 for the GC library) (Figure 4d,e). Alternative calculations using the 2-site library fraction only or using FC instead of frequency led to similar conclusions (Figure S19). This means that the fitness of these combinatorial variants is less predictable than for the DNA replicators, possibly due to epistatic interactions and/or the more complex sorting workflow leading to lower screening/sequencing coverage, which may increase the level of noise in the sequencing data set. Nevertheless, it is reasonable that more complex protein systems like the PS synthesis pathway, which involve more genes than the DNA replication system, are more subjected to epistasis affecting gene expression levels, as we have also observed from the GreenLys protein gels in Figure S15: RBSs lead to different protein production levels for different genetic backgrounds.

DISCUSSION

Targeted combinatorial randomization strategies have been extensively applied to optimize expression regulatory sequences for multiprotein systems *in vivo*, for instance, genetic circuits[41] and metabolic pathways,[14] but not in cell-free systems, where the screening of combinatorial libraries is only starting to be explored and has so far only targeted single proteins.[13,42,43] Here, we used combinatorial library designs targeting the RBSs or the first six codons of coding sequences to tune cell-free expression from multigene constructs. Through self-selection or functionality screening, followed by next-generation sequencing (NGS) retrieval of the high-performing variants, we identified variants regulating the functionality of its genetic module at the system's level. The proposed holistic engineering concept will aid in integrating biological functions in synthetic cells.[44,45]

Several factors may bias the abundance of variants after selection, independent of the mutational effects: the large inherent heterogeneity in functionality within clonal populations of liposomes,[46] differences between variant frequencies in the input library (Figures S3a and S12), and the FACS sorting accuracy of only 80% as determined earlier.[47,48] We assumed that each DNA variant produces an output signal distribution of identical shape (e.g., width of the distribution), leading to a single output quantity that represents its fitness (i.e., amplification fold, or LactC2-mCherry intensity). Increasing the sampling resolution of the variant distributions through the implementation of multiple sorting gates could improve the inference of the fitness value and provide insights into the variant's mechanism of action.[49] In addition, quantitative fitness information for medium- and low-fitness variants is valuable, especially when combined with active learning-assisted directed evolution. In active learning, machine learning models are used to efficiently navigate large combinatorial sequence spaces, and their predictive accuracy could be improved by training on quantitative data from a diverse sequence space.[42,50−52] Another improvement could be the implementation of unique molecular identifiers (UMIs) during the first 1−2 cycles of PCR to correct for changes in variant frequencies during DNA recovery.[53−55] Despite potential sources of 'noise', we demonstrated the functional selection of high-performing variants in one selection round of DNA replication, and within four rounds of sorting for the phospholipid synthesis pathway.

The simultaneous optimization of multiple genes led to mutations that were kilobases apart. Therefore, we used long-read nanopore sequencing for sequence-to-function mapping. Despite the relatively low single-read per-base sequencing

https://doi.org/10.1021/acssynbio.6c00166
ACS Synth. Biol. 2026, 15, 3285−3299

accuracy (99.0−99.7%),[56] quality control measures applied to the targeted mutagenesis loci allowed for the detection and counting of DNA variants with minimal mutational distance (i.e., down to one nucleotide difference), as validated earlier.[32] Thus, we linked the mutations within single full-length reads without additional DNA processing steps, such as adding barcodes to identify the DNA variants[43] or UMIs for deriving consensus sequences from multiple reads.[57]

The library populations were dominated by variants containing one or more wild-type target loci due to the choice for recombineering as the library generation method, with an editing efficiency of ∼50% per locus. To deal with the different starting frequencies among the library fractions (where each fraction contains variants with one mutated locus, or two mutated loci, etc.), we only compared variant frequencies within the same library fraction. Improving the recombineering editing efficiency or adopting PCR-based diversification methods for library generation might, in the future, lead to higher screening and sequencing coverage of the variants with all target loci mutated.

Our experiments revealed that different target loci have distinct levels of control over the system's performance. In future optimization campaigns, mutagenesis should preferentially target these most impactful regions to navigate the fitness landscape efficiently, while the expression of other genes could be fixed at a lower rate to save translation resources. We also found that coexpressing multiple proteins with RBSs having theoretical strengths (much) higher than the wild-type RBS did not lead to large increases in protein production compared to the wild type (Figures 1h and 2f). Instead, the most significant increases in the amounts of expressed proteins (maximum 1.8 to 2.3-fold relative to the wild type) were obtained in combination with downregulating the coexpression of another protein (Figure S15(V4 and V5)), which appears as a more powerful way to balance protein systems and metabolic fluxes than the sole implementation of strong RBSs. More extensive quantitative proteomics analysis would be needed to determine the precise genetic context effects on gene expression and to potentially generate certain design rules that could help optimize expression patterns of multigene constructs in PURE system. While some of these findings most likely also apply to other multiprotein systems, the discovery of specific target loci that have a higher impact on fitness remains empirical.

Lastly, pairwise epistasis calculations showed high predictability for the RBS library for self-selection of a DNA replicator and lower predictability for the phospholipid synthesis pathway libraries (Figure 4). These findings underscore the importance of combinatorial scanning over testing each mutation in isolation for different protein systems and/or selection procedures, especially for larger multigene systems. However, combinatorial testing comes at the expense of a larger library size. As mentioned above, active learning strategies can accelerate the search for good designs in a large sequence space.

Other types of libraries could be explored to improve the expression and function of protein systems in synthetic cells. Possible targets include promoters,[20,58] terminators, operator sites, gene order and orientation, promoter insulators (i.e., spacers upstream of a promoter sequence),[59] and operon (multicistronic) designs.[20,60] For example, enhanced transcription can improve protein production,[20] but excess mRNA molecules can sequester initiation factors of the translation machinery, thereby reducing protein synthesis rate.[2,61] Furthermore, protein sequence regions known to govern protein activity, such as active sites and binding sites, could be mutated. Lastly, mutating the type of amino acids in the N- and C-terminal regions, or nucleotide patches that might induce ribosome frameshifting (X/XXY/YYZ) or stalling, might increase the efficiency of protein folding and full-length synthesis.[2,61]

In the future, the system's level optimization workflow presented here could benefit from the sorting of liposomes based on microscopy images instead of FACS,[48] followed by multiomics characterization.[62−64] An alternative strategy to variant screening is to couple gene-encoded protein systems to DNA replication, such that a higher fitness of the desired system promotes DNA replication.[43,45,65] In pursuing the development of a synthetic cell, establishing a holistic approach that harnesses combinatorial optimization of multiprotein systems will surely be needed to manage the increasing number of possible genetic designs.

METHODS

Reagents and Equipment

Chemicals were purchased from Sigma-Aldrich unless stated otherwise. Milli-Q water used in buffers and solutions had a resistivity of 18.2 MΩ. DNA concentrations were measured using Qubit dsDNA HS (High Sensitivity) Assay Kit and the Qubit 4 Fluorometer (Invitrogen). The Eppendorf Eporator electroporator was used at 1.8 kV. Plasmid isolation was performed with PureYield Plasmid Miniprep System (Promega). PCR-amplified DNA was purified using the QIAquick PCR & Gel Cleanup Kit (Qiagen).

Biological Resources

Bacterial strains *E. coli* K-12 MG1655 and *E. coli* K-12 DH5α were used. pORTMAGE-Ec1 was a gift from the George Church lab (Addgene plasmid #138474; http://n2t.net/addgene:138474; RRID:Addgene_138474). Plasmid G435 was constructed by introducing a point mutation (to remove a PmeI recognition site) in G340, whose synthesis was described earlier.[66] Plasmid G555 (Addgene plasmid #216483; http://n2t.net/addgene:216483; RRID:Addgene_216483) was constructed as described earlier.[32] Plasmids and DNA oligos are listed in Tables S10 and S11, respectively. LactC2-mCherry was recombinantly expressed and purified as previously described.[47] SSB and DSB were expressed and purified as described in Soengas et al.[67] and Mencía et al.,[25] respectively, and stored at 10 mg/mL in 50 mM Tris, pH 7.5, 1 mM EDTA, 7 mM β-mercaptoethanol, 50% glycerol, and 60 mM (SSB) or 100 mM (DSB) ammonium sulfate. Purified proteins were aliquoted and stored at −80 °C.

Design and Synthesis of DNA Libraries

The RBS libraries for plasmids G435 (encoding DNAP and TP) and G555 (encoding PlsB, PlsC, CdsA, and PssA) were designed using the RBS calculator[18] in the "Optimize Expression Levels" mode (https://salislab.net/software/design_rbs_library_calculator) with parameters: host organism, *E. coli*; target minimum/maximum translation initiation rates, 1/1,000,000; genomic RBS sequence, mRNA sequence from transcription start until (and including) stop codon or terminator. The G555 GC library was designed manually by changing primarily the third bases of the codons to vary GC content of the first six codons of each gene between 22% and 61% (depending on the target locus), while maintaining the originally encoded amino acids (with two exceptions as indicated). The libraries were generated by MOSAIC, a multiplex one-step plasmid diversification protocol, which we recently developed.[32] In this method, DNA oligo libraries are coelectroporated with the target plasmid into *E. coli* cells expressing the single-stranded DNA annealing protein CspRecT from the pORTMAGE-Ec1 plasmid. During plasmid replication in the host cells, the oligos anneal to their target sites and integrate into the

https://doi.org/10.1021/acssynbio.6c00166
ACS Synth. Biol. 2026, 15, 3285−3299

plasmid sequence, creating a library of plasmids. The (degenerate) DNA oligos used in this study are listed in Table S11. For the G435 RBS library, a single electroporation yielded approximately 1600 colonies, from which the plasmids were isolated. For the G555 RBS library, four electroporation reactions were performed in parallel, yielding a total of 3700 colonies. For the G555 GC library, two parallel electroporation reactions yielded a total of 1800 colonies. The plasmids were isolated from each parallel reaction separately.

PCR was performed on the purified plasmids to linearize the DNA and simultaneously purify the target DNA from the pORTMAGE-Ec1 plasmid. The PCR reaction mix contained 1× Phusion HF buffer, 200 μM dNTPs, 500 nM of each primer (491 ChD and 492 ChD), 2 U Phusion High-Fidelity DNA polymerase, and 200 ng of plasmid in a 100 μL reaction volume. The mixture was split into 2 × 50 μL, and the thermocycling reaction was performed as follows: 98 °C for 30 s, 20−25 cycles of (98 °C for 5 s and 72 °C for 3 min (G435) or 6 min 20 s (G555)), 72 °C for 5 min. Ten units of DpnI were added to each 50 μL of PCR mix and incubated for 1 h at 37 °C to digest the plasmids. The expected PCR products orip2p3 (3214 bp, derived from G435) and oriPL (6680 bp, derived from G555) were excised from a 0.7% agarose gel and further purified using QIAquick PCR & Gel Cleanup Kit (Qiagen). The DNA concentration was measured using the Qubit fluorometer. The four oriPL RBS library PCR products were then combined at equimolar amounts to increase library coverage. Similarly, the two oriPL GC library PCR products were combined.

Synthesis of Clonal DNA Variants

Clonal plasmid variants were also generated using MOSAIC with a few modifications. Instead of one electroporation, three subsequent rounds of electroporation and plasmid isolation were performed. The *E. coli* strain DH5α harboring pORTMAGE-Ec1 was used instead of MG1655 pORTMAGE-Ec1 to prevent the emergence of smaller plasmids due to recombination during repeated transformation steps. Electroporation was performed with 10−50 ng of plasmid mixture and 2.5 μM of each DNA oligo variant added to the reaction. The plasmids were isolated from colonies grown on ampicillin-supplemented agar plates (50 μg/mL) or from liquid cell culture when the number of colonies on the plates was lower than 50. After three subsequent electroporation steps for mutagenesis, DH5α cells were chemically transformed by heat shock using 3 ng of purified plasmid mixture containing the target plasmids and pORTMAGE-Ec1 to obtain clonally pure plasmids. The transformed cells were plated on ampicillin-supplemented plates to pick single colonies for subsequent plasmid isolation and sequence validation by nanopore sequencing.

In-Liposome Gene Expression for DNA Self-Replication

For the expression of DNA libraries in liposomes, swelling solutions were prepared at a volume of 10 μL with 5 μL of PURE*frex* 2.0 (GeneFrontier) Solution I, 0.5 μL of Solution II, 1 μL of Solution III, 0.25 U/μL SUPERase·In RNase inhibitor (Invitrogen), 300 μM dNTP Mix (Thermo Fisher), 20 mM ammonium sulfate, 750 μg/mL purified SSB, 210 μg/mL purified DSB, and 50 pM linear orip2p3 DNA. During this study, a change in the commercial PURE*frex* 2.0 solution resulted in impaired DNA replication. Therefore, swelling solutions with clonal DNA variants consisted of an adjusted PURE composition, which yielded better replication compared to the standard kit: 2.5 μL of custom PURE*frex* Solution I lacking tRNAs and NTPs (PFC-Z1608-EX) (i.e., half of the recommended volume), 0.5 μL of Solution II, 1 μL of Solution III, 0.25 U/μL SUPERase·In, 300 μM dNTP Mix (Thermo Fisher), 20 mM ammonium sulfate, 750 μg/mL purified SSB, 210 μg/mL purified DSB, 50 pM linear orip2p3 DNA, and 1/3rd of the recommended concentrations of NTPs and tRNAs.

The preparation of lipid-coated beads was performed as described earlier[46] using 50 mol % DOPC, 36 mol % DOPE, 12 mol % DOPG, 2 mol % 18:1 cardiolipin, 0.01 mol % DOPE-Cy5, and 1 mass % DSPE-PEG(2000)-biotin. Five to six milligrams of lipid-coated beads were added to 10 μL of the swelling solution, and the tubes were tumbled at 4 °C for 1 h. After four freeze−thaw cycles by dipping the tubes in liquid nitrogen and thawing on ice, 3 μL of liposomes were transferred to PCR tubes containing 0.25 μL of DNase I (2 U/μL) using a cut pipet tip. The liposomes were incubated in a thermocycler at 30 °C for 16 h to enable gene expression, DNA replication, and degradation of nonencapsulated DNA by DNase I. Negative control samples were incubated at 30 °C for 20 min to activate DNase I. The liposomes were then incubated at 75 °C for 15 min to heat-inactive DNase I before amplifying the intraliposomal DNA by (quantitative) PCR.

Verification of In-Liposome DNA Replication Using Quantitative PCR

Quantitative PCR reaction mix was set up at 10 μL volume and contained 5 μL of PowerUp SYBR Green Master Mix (Applied Biosystems), 400 nM of each primer (976 ChD and 977 ChD), and 1 μL of heat-inactivated liposome sample diluted 100-fold in Milli-Q water. The thermocycling was performed on a QuantStudio 5 Real-Time PCR instrument (Thermo Fisher) as follows: 50 °C for 2 min, 94 °C for 5 min, 45 cycles of (94 °C for 15 s, 56 °C for 15 s, and 68 °C for 30 s), 68 °C for 5 min, followed by a melting curve from 65 to 95 °C. The DNA concentration was determined using a standard curve of the wild-type DNA with 10-fold dilutions ranging from 0.0001 to 100 pM and the QuantStudio Design and Analysis software v2.6.0 (Thermo Fisher).

DNA Recovery from Liposomes after DNA Replication

The DNA was recovered from the liposomes by PCR amplification. The PCR reaction mix (100 μL) contained 1× Phusion HF buffer, 200 μM dNTPs, 500 nM of each primer (491 ChD and 492 ChD), 2 U Phusion High-Fidelity DNA polymerase, 5% DMSO, and 10 μL of liposome sample diluted 100-fold in Milli-Q water. The mixture was split into 2 × 50 μL, and the thermocycling reaction was performed as follows: 98 °C for 30 s, 30 cycles of (98 °C for 5 s and 72 °C for 3 min), 72 °C for 5 min. The expected PCR products (3214 bp) were excised from a 0.7% agarose gel and further purified using QIAquick PCR & Gel Cleanup Kit (Qiagen). The DNA concentration was measured using a Qubit fluorometer.

In-Liposome Gene Expression for Phospholipid Synthesis

The lipid film swelling solutions were prepared at a volume of 20 μL with 10 μL of PURE*frex* 2.0 Solution I, 1 μL of Solution II, 2 μL of Solution III, 0.5 U/μL SUPERase·In, 1 mM CTP, 0.5 mM G3P, 0.5 mM L-serine, 5 mM β-mercaptoethanol, and 10 pM linear oriPL DNA. Lipid-coated beads were prepared and added to the swelling solution as described above, except that the mass of beads was doubled, and the tubes were tumbled at 4 °C for 1 h. After four freeze−thaw cycles, 15 μL of liposomes and 0.5 μL of DNase I (2 U/μL) were transferred to PCR tubes containing a dried film of oleoyl-CoA (Avanti Polar Lipids) to reach a final concentration of 176 μM oleoyl-CoA. The liposomes were incubated in a thermocycler at 37 °C for 6 h (R1), 5 h (R2), or 4 h (R3−R5 and clonal variants).

Liposome Screening and Sorting with FACS

For each library (RBS or GC library), two identically prepared 15-μL liposome samples were screened and sorted using the BD FACSMelody cell sorter (BD Biosciences). Fifteen microliters of liposomes were diluted in 1.3 mL of buffer containing 20 mM HEPES-KOH, pH 7.6, 180 mM potassium glutamate, 14 mM magnesium acetate, and 300 nM LactC2-mCherry. The diluted liposomes were filtered using the 35 μm nylon mesh of a cell strainer cap of 5 mL round-bottom polystyrene test tubes (Falcon). LactC2-mCherry fluorescence was measured using a 561 nm laser line and an emission filter of 610/20 nm. The photon multiplier tube voltages were set at 342 V for FSC, 405 V for SSC (with SSC threshold 359 V), and 461 V for LactC2-mCherry. First, a gate was set on FSC and SSC to select the main liposome population (~60−80% of all events) (Figure S20). For each sample, 5000−30,000 liposomes with the highest LactC2-mCherry signal were sorted in purity mode, becoming more stringent in the later sorting rounds. The sorted liposomes from two identical samples were combined. The recorded data was further processed using Cytobank (https://community.cytobank.org/). Here,

https://doi.org/10.1021/acssynbio.6c00166
ACS Synth. Biol. 2026, 15, 3285−3299

in addition to the gating based on FSC and SSC, events with LactC2-mCherry signal lower than 2 au were discarded (Figure S20).

DNA Recovery from Liposomes Sorted by FACS

To recover the DNA from the sorted liposomes, the sorted fractions were spun down at 12,000*g* for 3 min, whereafter the supernatant was removed. The remaining ∼10 μL was incubated at 75 °C for 15 min to heat-inactivate DNase I. PCR reactions were prepared in a 100-μL reaction volume by mixing 1× Xtreme buffer, 400 μM dNTPs, 300 nM of each primer (1459 ChD and 1460 ChD), 2 units of KOD Xtreme Hotstart DNA polymerase, and 10 μL of heat-inactivated liposome solution. The mixture was split into 2 × 50 μL, and the thermocycling reaction was performed as follows: 98 °C for 2 min, 35 cycles of (98 °C for 10 s and 68 °C for 6 min 40 s). The expected PCR products (6616 bp) were excised from a 0.7% agarose gel and further purified using QIAquick PCR & Gel Cleanup Kit (Qiagen). A second PCR reaction was performed with ∼60 ng of the purified DNA using primers 1301 ChD and 1518 ChD to restore the full-length DNA template (6680 bp), and the amplicon was purified as described above. The DNA concentration was measured using a Qubit fluorometer.

Nanopore Sequencing and Analysis

DNA samples were diluted to a concentration of 34 ng/μL (orip2p3) or 71 ng/μL (oriPL). Nanopore sequencing was performed by Plasmidsaurus (Oregon, US). First, analysis of the raw fastq sequencing reads was carried out using the Galaxy web platform[68] (https://usegalaxy.org). The reads were filtered on length (3020−3300 nt for orip2p3, 6400−7000 nt for oriPL) using the *Filter FASTQ reads by quality score and length* tool,[69] and mapped to the wild-type sequences using *Map with minimap2*.[70] *Minimap2* was executed using default parameters, except *retain at most INT secondary alignments*, which was set to 0. For oriPL data, the *gap open penalty deletion* and *gap open penalty insertion* were set to 16 and 48, respectively.

The mapped reads were further processed using custom R scripts (R v4.4.1, RStudio v2023.09.1 + 494). A quality control was applied to the mutated regions of interest using a per-base quality cutoff of 50; this quality control method was validated as described earlier.[32] The variant frequency was calculated by dividing the number of reads of a given variant by the total number of reads within the corresponding *N*-site library fraction (i.e., the library fraction containing all sequences with *N* mutated target loci). The variant enrichment (defined as fold change, FC) was calculated by dividing the frequency of a variant by the frequency of that variant in the initial library before sorting (R0). The variant enrichment was plotted as the $\log_2$FC ($\log_2$ fold change). The sum of frequencies per DNA variant within each *N*-site library fraction, as plotted in Figures 2c, S12, and S16, was divided by the number (*N*) of mutated target loci per DNA to normalize the total frequency within each *N*-site library fraction to 1.

Positional Base-Specific Effects

The positional base-specific effects on variant enrichment were calculated as previously described.[31] The mean frequency of all DNA variants containing a specific base at a specific position was divided by the mean frequency of all DNA variants containing another base at that position. The effect on enrichment was plotted as the $\log_2$FC.

Fitness Predictions of Combinatorial Variants without Epistasis

The expected fitness of combinatorial variants was calculated by multiplying the fitness of each individual variant. As the frequency can be seen as the expected probability to sort and sequence a DNA variant, the Bayes' law about conditional probability can be used to support the choice for the multiplicative model, as suggested by a previous *in vitro* screening study.[13] DNA variants with less than three reads in the 2-site library fraction were excluded from the analysis.

For the DNA replicator library, the expected fitness and the difference between the observed and expected fitness (epsilon) were calculated as follows

$$\mathrm{FC}_{\mathrm{expected}} = \mathrm{FC}_1 \times \mathrm{FC}_2 \tag{1}$$

where FC_i is the FC of the single variant in the 1-site library fraction.

$$\mathrm{epsilon} = \log_2\mathrm{FC}_{\mathrm{observed}} - \log_2\mathrm{FC}_{\mathrm{expected}} \tag{2}$$

where $\mathrm{FC}_{\mathrm{observed}}$ is the FC of the combinatorial variant in the 2-site library fraction.

For the PS synthesis libraries, we used frequency instead of FC due to the lower sequencing depth of the presort libraries

$$\mathrm{freq}_{\mathrm{expected}} = \mathrm{frequency}_1 \times \mathrm{frequency}_2 \tag{3}$$

where $\mathrm{frequency}_i$ is the frequency of the single variant relative to all variants with that target site mutated in the 1-site library fraction. The difference between the observed and expected fitness is thus defined as

$$\mathrm{epsilon} = \log_2\mathrm{frequency}_{\mathrm{observed}} - \log_2\mathrm{frequency}_{\mathrm{expected}} \tag{4}$$

where $\mathrm{frequency}_{\mathrm{observed}}$ is the frequency of the combinatorial variant relative to all variants with those two target sites mutated in the 2-site library fraction.

We repeated the calculation of the expected fitness using the 2-site library fraction instead of the 1-site library fraction, and we reached identical conclusions (Figures S18 and S19). The expected fitness and epsilon were calculated for the DNA replicator and PS synthesis libraries as follows

$$\mathrm{FC}_{\mathrm{expected,2\text{-}site}} = \mathrm{FC}_1 \times \mathrm{FC}_2 \tag{5}$$

where FC_i is the average FC of all variants in the 2-site library fraction with mutation *i*.

$$\mathrm{epsilon} = \log_2\mathrm{FC}_{\mathrm{observed}} - \log_2\mathrm{FC}_{\mathrm{expected,2\text{-}site}} \tag{6}$$

where $\mathrm{FC}_{\mathrm{observed}}$ is the FC of the combinatorial variant in the 2-site library fraction.

$$\mathrm{freq}_{\mathrm{expected,2\text{-}site}} = \mathrm{frequency}_1 \times \mathrm{frequency}_2 \tag{7}$$

where $\mathrm{frequency}_i$ is the sum of frequencies of all variants in the 2-site library fraction with mutation *i*.

$$\mathrm{epsilon} = \log_2\mathrm{frequency}_{\mathrm{observed}} - \log_2\mathrm{frequency}_{\mathrm{expected,2\text{-}site}} \tag{8}$$

where $\mathrm{frequency}_{\mathrm{observed}}$ is the frequency of the combinatorial variant relative to all variants in the 2-site library fraction.

Variance Calculation for Epistatic Deviations with Error Propagation

The variance in epistatic deviation ($\mathrm{FC}_{\mathrm{observed}} - \mathrm{FC}_{\mathrm{expected}}$) was calculated as follows

$$\mathrm{var}(\mathrm{FC}_{\mathrm{observed}} - \mathrm{FC}_{\mathrm{expected}}) = \mathrm{var}(\mathrm{FC}_{\mathrm{observed}}) + \mathrm{var}(\mathrm{FC}_{\mathrm{expected}}) \tag{9}$$

where $\mathrm{var}(\mathrm{FC}_{\mathrm{expected}})$ was calculated as

$$\mathrm{var}(\mathrm{FC}_{\mathrm{expected}}) = (\mathrm{var}(\mathrm{FC}_1) + \mathrm{FC}_1{}^2) \times (\mathrm{var}(\mathrm{FC}_2) + \mathrm{FC}_2{}^2) - \mathrm{FC}_1{}^2 \times \mathrm{FC}_2{}^2 \tag{10}$$

Off-Target Mutations

The presence of off-target mutations (outside RBS regions) was analyzed by comparing the mapped reads of the DNA recovered from liposomes with the initial library in Integrative Genomics Viewer (IGV) v2.16.0 software using a coverage allele-fraction threshold of 0.01 (DNA replication) or 0.04 (PS synthesis libraries).

GreenLys SDS-PAGE

Relative quantification of protein production was performed by running bulk PURE system reactions. Reactions were set up at 10 μL volume, containing 5 μL of PURE*frex* 2.0 Solution I, 0.5 μL of Solution II, 1 μL of Solution III, 0.25 U/μL SUPERase·In, 0.5 μL of 5-fold diluted FluoroTect $\mathrm{Green}_{\mathrm{Lys}}$ in vitro Translation Labeling System (Promega), and 1 nM linear DNA template. The samples were incubated at 30 °C (DNA replication proteins) or 37 °C

https://doi.org/10.1021/acssynbio.6c00166
ACS Synth. Biol. 2026, 15, 3285−3299

(phospholipid synthesis enzymes) for 2 h, unless indicated otherwise. Each sample was supplemented with 0.25 μL of RNase A Solution (4 mg/μL) (Promega) and incubated for 30 min at 37 °C. 5 μL of the sample was mixed with 1.9 μL of solution containing 1× Laemmli SDS sample buffer, nonreducing (Thermo Fisher), and 10 mM dithiothreitol. The samples were incubated for 5 min at 70 °C to denature the proteins and run on an SDS-PAGE gel consisting of a 12% polyacrylamide resolving gel and a 4% stacking gel. Fluorescently labeled proteins were visualized with a Typhoon fluorescence gel imager (Amersham Biosciences) using a 488 nm laser and a 520 nm band-pass emission filter. Band intensities (adjusted band volume with background subtracted) were quantified using Image Lab v6.1.0 (Bio-Rad Laboratories) relative to the wild-type DNA sample.

Confocal Microscopy and Image Analysis

Samples for confocal fluorescence microscopy were prepared by mixing 1−2 μL of liposomes with 8 μL of buffer containing 20 mM HEPES-KOH, pH 7.6, 180 mM potassium glutamate, 14 mM magnesium acetate, and 5× dsGreen (Lumiprobe). Using a cut pipet tip, the 10-μL sample was transferred to a BSA-functionalized custom-made glass imaging chamber. The sample was incubated at 4 °C for 1−1.5 h before imaging for liposome sinking. Image acquisition was performed with the A1R Laser scanning confocal microscope (Nikon) with an SR Apo TIRF 100× oil immersion objective (NA 1.49) and NIS-Elements v5.30.07 software (Nikon). Imaging was performed using laser line 488 nm with emission filter 525:50 nm (dsGreen), laser line 561 nm with emission filter 595/50 nm (LactC2-mCherry), and laser line 640 nm with emission filter 700/75 nm (DOPE-Cy5). Acquisition settings and adjustments of brightness and contrast were identical for multiple images when comparing fluorescence intensities. Using Fiji (ImageJ2 v2.9.0),[71] the dsGreen signal inside liposomes was quantified by detecting particles on the binarized dsGreen channel, followed by measuring the mean pixel intensity for each particle. The quantification of LactC2-mCherry intensity was performed using SMELDit (https://github.com/DanelonLab/SMELDit).

Statistics for the Number of DNA Molecules Per Liposome

Following Poissonian partitioning, the probability of encapsulating k DNA molecules per liposome is

$$P(k) = \frac{\lambda^k}{k!}e^{-\lambda} \tag{11}$$

where λ is the expected average number of DNA molecules per liposome, which is calculated as

$$\lambda = \frac{\pi N_A C d^3}{6} \tag{12}$$

where N_A is the Avogadro constant, C is the concentration of DNA, and d is the average liposome diameter.

With 10 pM DNA and a liposome diameter of 4 μm ($\lambda = 0.2$), the probability of encapsulating one DNA molecule per liposome is 0.16, and the probability of encapsulating more than one molecule is 0.02. Experimental validation that encapsulation is mostly clonal at 10 pM and 50 pM was performed previously.[47] A more accurate approximation of λ values and encapsulation probabilities can be derived by accounting for the experimentally measured distribution of the liposome diameters (https://github.com/DanelonLab/ACS), but the main conclusions remain valid.

Visualizations

Radar charts, heatmaps, and rarefaction plots were generated with R v4.4.1 in RStudio v2023.09.1+494. 3D sequence landscapes were generated with Python in Google Colaboratory. To generate the 3D landscapes, the variants with zero sequencing reads were assigned a $\log_2$ fold change identical to the lowest finite number in that data set. Other graphs were made by using Prism v10.0.3 and Illustrator v28.5.

Statistics

Bar height in bar plots is the mean value across replicates (the individual data points are often appended), and error bars indicate the standard deviation (SD). Box-whisker plots in Figures 1f, S5b, and S9b display the median (middle line), the 75/25 percentiles at the boxes, and the 90/10 percentiles at the whiskers. Statistical tests (unpaired t-test, Figures 2d and S14) were performed using GraphPad Prism v10.0.3 and were significant if the P-value (two-tailed) ≤ 0.05. When indicated, this P-value threshold was corrected for multiple comparisons using the Bonferroni correction. R^2 is the coefficient of determination. The P-value in linear regression (Figure 2f) indicates whether the slope is significantly different from zero. The number of biological replicates n (i.e., separate liposome samples) is stated in the figure captions.

■ ASSOCIATED CONTENT

Data Availability Statement

The data are available in the main manuscript and Supporting Information. Raw data, processed data, and the code for data processing underlying this article are available in the 4TU.ResearchData repository at 10.4121/1d798f6e-7071-4b7d-9d79-7425e4b9d247.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.6c00166.

Figures presenting data of nanopore sequencing, agarose gel electrophoresis, SDS-PAGE, fluorescence microscopy, FACS, and qPCR; tables listing plasmids, DNA oligos, DNA variant sequences, and variant read counts (PDF)

P values, confidence intervals, standard deviation, variance, and signal-to-noise ratios of the fold change and/or epistatic deviation for the DNA replication enrichment experiment (Data S1) (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Christophe Danelon − *Department of Bionanoscience, Kavli Institute of Nanoscience, Delft University of Technology, 2629 HZ Delft, The Netherlands; Toulouse Biotechnology Institute (TBI), Université de Toulouse, CNRS, INRAE, INSA, 31077 Toulouse, France;* orcid.org/0000-0002-0961-6640; Email: danelon@insa-toulouse.fr

Authors

Marijn van den Brink − *Department of Bionanoscience, Kavli Institute of Nanoscience, Delft University of Technology, 2629 HZ Delft, The Netherlands;* orcid.org/0009-0006-9286-0746

Nico J. Claassens − *Laboratory of Microbiology, Wageningen University, 6708 WE Wageningen, The Netherlands;* orcid.org/0000-0003-1593-0377

Complete contact information is available at: https://pubs.acs.org/10.1021/acssynbio.6c00166

Author Contributions

M.v.d.B. and C.D. conceived the project; C.D. and N.J.C. supervised the project and acquired funding; M.v.d.B. and C.D. designed the experiments; M.v.d.B. performed experiments and data analysis and created visualizations; and M.v.d.B. and C.D. wrote the paper with input from N.J.C.

https://doi.org/10.1021/acssynbio.6c00166
ACS Synth. Biol. 2026, 15, 3285−3299

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Ana Restrepo Sierra for her contribution to the conceptualization and design of the RBS libraries. We are grateful to Ilja Westerlaken for purifying the LactC2-mCherry proteins, Federico Ramirez Gomez for acquiring and analyzing microscopy images of the phospholipid-producing liposomes and for performing a GreenLys-based protein quantification experiment, and Laura Sierra Heras for the adjusted PURE system composition. We also thank Liedewij Laan and Stefan Loonen for fruitful discussions, GeneFrontier for sponsoring our research, and Miguel de Vega and Alicia del Prado from the Centro de Biología Molecular Severo Ochoa (Madrid) for providing purified SSB and DSB proteins. M.v.d.B., C.D., and N.J.C. acknowledge financial support from the "BaSyC—Building a Synthetic Cell" Gravitation grant (024.003.019) of The Netherlands Ministry of Education, Culture and Science (OCW) and The Netherlands Organization for Scientific Research (NWO). C.D. also acknowledges funding from Agence Nationale de la Recherche (ANR-22-CPJ2-0091-01). N.J.C. also acknowledges funding by the NWO-SUMMIT grant "Evolving life from non-life (EVOLF)" (SUMMIT.1.004).

■ REFERENCES

(1) Hunt, A. C.; Rasor, B. J.; Seki, K.; et al. Cell-free gene expression: methods and applications. *Chem. Rev.* **2025**, *125*, 91−149.
(2) De Capitani, J.; Mutschler, H. The long road to a synthetic self-replicating central dogma. *Biochemistry* **2023**, *62*, 1221−1232.
(3) Nieuwkoop, T.; Finger-Bou, M.; van der Oost, J.; Claassens, N. J. The ongoing quest to crack the genetic code for protein production. *Mol. Cell* **2020**, *80*, 193−209.
(4) Bernath, K.; Hai, M.; Mastrobattista, E.; et al. In vitro compartmentalization by double emulsions: sorting and gene enrichment by fluorescence activated cell sorting. *Anal. Biochem.* **2004**, *325*, 151−157.
(5) Nishikawa, T.; Sunami, T.; Matsuura, T.; Ichihashi, N.; Yomo, T. Construction of a gene screening system using giant unilamellar liposomes and a fluorescence-activated cell sorter. *Anal. Chem.* **2012**, *84*, 5017−5024.
(6) Fujii, S.; Matsuura, T.; Sunami, T.; Kazuta, Y.; Yomo, T. In vitro evolution of α-hemolysin using a liposome display. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 16796−16801.
(7) Fujii, S.; Matsuura, T.; Sunami, T.; et al. Liposome display for in vitro selection and evolution of membrane proteins. *Nat. Protoc.* **2014**, *9*, 1578−1591.
(8) Uyeda, A.; Watanabe, T.; Kato, Y.; et al. Liposome-based in vitro evolution of aminoacyl-tRNA synthetase for enhanced pyrrolysine derivative incorporation. *ChemBioChem* **2015**, *16*, 1797−1802.
(9) Uyeda, A.; Nakayama, S.; Kato, Y.; Watanabe, H.; Matsuura, T. Construction of an in vitro gene screening system of the *E. coli* EmrE transporter using liposome display. *Anal. Chem.* **2016**, *88*, 12028−12035.
(10) Murase, Y.; Nakanishi, H.; Tsuji, G.; Sunami, T.; Ichihashi, N. In vitro evolution of unmodified 16S rRNA for simple ribosome reconstitution. *ACS Synth. Biol.* **2018**, *7*, 576−583.
(11) Seo, K.; Hagino, K.; Ichihashi, N. Progresses in Cell-Free In Vitro Evolution. In *Cell-free Production: System Development*; Lu, Y.; Jewett, M. C., Eds.; Springer Nature Switzerland: Cham, 2023; pp 121−140.
(12) Gan, R.; Cabezas, M. D.; Pan, M.; et al. High-throughput regulatory part prototyping and analysis by cell-free protein synthesis and droplet microfluidics. *ACS Synth. Biol.* **2022**, *11*, 2108−2120.
(13) Scheele, R. A.; Lindenburg, L. H.; Petek, M.; et al. Droplet-based screening of phosphate transfer catalysis reveals how epistasis shapes MAP kinase interactions with substrates. *Nat. Commun.* **2022**, *13*, No. 844.
(14) Jeschek, M.; Gerngross, D.; Panke, S. Combinatorial pathway optimization for streamlined metabolic engineering. *Curr. Opin. Biotechnol.* **2017**, *47*, 142−151.
(15) Höllerer, S.; Desczyk, C.; Muro, R. F.; Jeschek, M. From sequence to function and back − High-throughput sequence-function mapping in synthetic biology. *Curr. Opin. Syst. Biol.* **2024**, *37*, No. 100499.
(16) Packer, M. S.; Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **2015**, *16*, 379−394.
(17) Shimizu, Y.; Inoue, A.; Tomari, Y.; et al. Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **2001**, *19*, 751−755.
(18) Salis, H. M.; Mirsky, E. A.; Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **2009**, *27*, 946−950.
(19) Chizzolini, F.; Forlin, M.; Martín, N. Y.; et al. Cell-free translation is more variable than transcription. *ACS Synth. Biol.* **2017**, *6*, 638−647.
(20) 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 Synth. Biol.* **2014**, *3*, 363−371.
(21) van Nies, P.; Westerlaken, I.; Blanken, D.; et al. Self-replication of DNA by its encoded proteins in liposome-based synthetic cells. *Nat. Commun.* **2018**, *9*, No. 1583.
(22) Blanco, L.; Prieto, I.; Gutiérrez, J.; et al. Effect of NH4+ ions on φ29 DNA-protein p3 replication: formation of a complex between the terminal protein and the DNA polymerase. *J. Virol.* **1987**, *61*, 3983−3991.
(23) Blanco, L.; Lázaro, J. M.; de Vega, M.; Bonnin, A.; Salas, M. Terminal protein-primed DNA amplification. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12198−12202.
(24) Esteban, J. A.; Blanco, L.; Villar, L.; Salas, M. In vitro evolution of terminal protein-containing genomes. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2921−2926.
(25) Mencía, M.; Gella, P.; Camacho, A.; de Vega, M.; Salas, M. Terminal protein-primed amplification of heterologous DNA with a minimal replication system based on phage φ29. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 18655−18660.
(26) Elías-Arnanz, M.; Salas, M. Bacteriophage φ29 DNA replication arrest caused by codirectional collisions with the transcription machinery. *EMBO J.* **1997**, *16*, 5775−5783.
(27) Mutalik, V. K.; Guimaraes, J. C.; Cambray, G.; et al. Quantitative estimation of activity and quality for collections of functional genetic elements. *Nat. Methods* **2013**, *10*, 347−353.
(28) Doerr, A.; de Reus, E.; van Nies, P.; et al. Modelling cell-free RNA and protein synthesis with minimal systems. *Phys. Biol.* **2019**, *16*, No. 025001.
(29) Kudla, G.; Murray, A. W.; Tollervey, D.; Plotkin, J. B. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* **2009**, *324*, 255−258.
(30) Nieuwkoop, T.; Terlouw, B. R.; Stevens, K. G.; et al. Revealing determinants of translation efficiency via whole-gene codon randomization and machine learning. *Nucleic Acids Res.* **2023**, *51*, 2363−2376.
(31) Höllerer, S.; Jeschek, M. Ultradeep characterisation of translational sequence determinants refutes rare-codon hypothesis and unveils quadruplet base pairing of initiator tRNA and transcript. *Nucleic Acids Res.* **2023**, *51*, 2377−2396.
(32) van den Brink, M.; Althuis, T. Y.; Danelon, C.; Claassens, N. J. MOSAIC: A highly efficient, one-step recombineering approach to plasmid editing and diversification. *ACS Synth. Biol.* **2025**, *14*, 3880−3889.
(33) Blanken, D.; Foschepoth, D.; Serrão, A. C.; Danelon, C. Genetically controlled membrane synthesis in liposomes. *Nat. Commun.* **2020**, *11*, No. 4317.

https://doi.org/10.1021/acssynbio.6c00166
ACS Synth. Biol. 2026, 15, 3285−3299

(34) Noga, M. J.; Büke, F.; van den Broek, N. J. F.; et al. Posttranslational control of PlsB is sufficient to coordinate membrane synthesis with growth in *Escherichia coli*. *mBio* **2020**, *11*, No. e02703-19, DOI: 10.1128/mBio.02703-19.
(35) Saha, S. K.; Satomi, N.; Matsuzaki, H.; Shibuya, I.; Matsumoto, K. A regulatory mechanism for the balanced synthesis of membrane phospholipid species in *Escherichia coli*. *Biosci. Biotechnol. Biochem.* **1996**, *60*, 111−116.
(36) Jeschek, M.; Gerngross, D.; Panke, S. Rationally reduced libraries for combinatorial pathway optimization minimizing experimental effort. *Nat. Commun.* **2016**, *7*, No. 11163.
(37) Zelcbuch, L.; Antonovsky, N.; Bar-Even, A.; et al. Spanning high-dimensional expression space using ribosome-binding site combinatorics. *Nucleic Acids Res.* **2013**, *41*, No. e98.
(38) Scott, A.; Noga, M. J.; de Graaf, P.; et al. Cell-free phospholipid biosynthesis by gene-encoded enzymes reconstituted in liposomes. *PLoS One* **2016**, *11*, No. e0163058.
(39) Hagino, K.; Ichihashi, N. In vitro transcription/translation-coupled DNA replication through partial regeneration of 20 aminoacyl-tRNA synthetases. *ACS Synth. Biol.* **2023**, *12*, 1252−1263.
(40) GeneFrontier Corporation. Influence of 5′UTR and N-terminal ORF region on protein synthesis efficiency in PURExfrex|Posters| Resources|PURExfrex|, 2023. https://purefrex.genefrontier.com/resources/posters/P040_CFSC2023.html.
(41) Schaerli, Y.; Isalan, M. Building synthetic gene circuits from combinatorial libraries: screening and selection strategies. *Mol. BioSyst.* **2013**, *9*, 1559−1567.
(42) Pandi, A.; Diehl, C.; Kharrazi, A. Y.; et al. A versatile active learning workflow for optimization of genetic and metabolic networks. *Nat. Commun.* **2022**, *13*, No. 3876.
(43) Di Meo, T.et al. Functional Selection in A Population of Synthetic Cells with A Minimal Metabolism. 2025 DOI: 10.1101/2025.08.25.672132.
(44) Abil, Z.; Danelon, C. Roadmap to building a cell: an evolutionary approach. *Front. Bioeng. Biotechnol.* **2020**, *8*, No. 927, DOI: 10.3389/fbioe.2020.00927.
(45) Abil, Z.; Giaveri, S.; Erb, T. J.; Rothschild, L. J. Integrating metabolism and evolution towards the realization of synthetic life. *Nat. Rev. Bioeng.* **2025**, *3*, 9−10.
(46) Blanken, D.; van Nies, P.; Danelon, C. Quantitative imaging of gene-expressing liposomes reveals rare favorable phenotypes. *Phys. Biol.* **2019**, *16*, No. 045002.
(47) Abil, Z.; Sierra, A. M. R.; Danelon, C. Clonal amplification-enhanced gene expression in synthetic vesicles. *ACS Synth. Biol.* **2023**, *12*, 1187−1203.
(48) van den Brink, M.; Stam, M.; Claassens, N. J.; Danelon, C. Image-based phenotypic sorting of synthetic cells. *Sci. Adv.* **2026**, *12*, eaed0860.
(49) Peterman, N.; Levine, E. Sort-seq under the hood: implications of design choices on large-scale characterization of sequence-function relations. *BMC Genomics* **2016**, *17*, No. 206.
(50) Wu, Z.; Kan, S. B. J.; Lewis, R. D.; Wittmann, B. J.; Arnold, F. H. Machine learning-assisted directed protein evolution with combinatorial libraries. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 8852−8858.
(51) Yang, J.; Lal, R. G.; Bowden, J. C.; et al. Active learning-assisted directed evolution. *Nat. Commun.* **2025**, *16*, No. 714.
(52) Vornholt, T.; Mutný, M.; Schmidt, G. W.; et al. Enhanced sequence-activity mapping and evolution of artificial metalloenzymes by active learning. *ACS Cent. Sci.* **2024**, *10*, 1357−1370.
(53) Fu, G. K.; Hu, J.; Wang, P.-H.; Fodor, S. P. A. Counting individual DNA molecules by the stochastic attachment of diverse labels. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 9026−9031.
(54) Kivioja, T.; Vähärautio, A.; Karlsson, K.; et al. Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **2012**, *9*, 72−74.
(55) Schmitt, M. W.; Kennedy, S. R.; Salk, J. J.; et al. Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 14508−14513.
(56) *Nanopore Sequencing Accuracy*, Oxford Nanopore Technologieshttps://nanoporetech.com/platform/accuracy.
(57) Zurek, P. J.; Knyphausen, P.; Neufeld, K.; Pushpanath, A.; Hollfelder, F. UMI-linked consensus sequencing enables phylogenetic analysis of directed evolution. *Nat. Commun.* **2020**, *11*, No. 6023.
(58) Komura, R.; Aoki, W.; Motone, K.; Satomura, A.; Ueda, M. High-throughput evaluation of T7 promoter variants using biased randomization and DNA barcoding. *PLoS One* **2018**, *13*, No. e0196905.
(59) Carr, S. B.; Beal, J.; Densmore, D. M. Reducing DNA context dependence in bacterial promoters. *PLoS One* **2017**, *12*, No. e0176013.
(60) Yoshida, A.; Kohyama, S.; Fujiwara, K.; Nishikawa, S.; Doi, N. Regulation of spatiotemporal patterning in artificial cells by a defined protein expression system. *Chem. Sci.* **2019**, *10*, 11064−11072.
(61) Li, J.; Zhang, C.; Huang, P.; et al. Dissecting limiting factors of the Protein synthesis Using Recombinant Elements (PURE) system. *Translation* **2017**, *5*, No. e1327006.
(62) Kosuri, S.; Goodman, D. B.; Cambray, G.; et al. Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 14024−14029.
(63) Su, P.-R.; You, L.; Beerens, C.; et al. Microscopy-based single-cell proteomic profiling reveals heterogeneity in DNA damage response dynamics. *Cell Rep. Methods* **2022**, *2*, No. 100237.
(64) You, L.; Su, P. R.; Betjes, M.; et al. Linking the genotypes and phenotypes of cancer cells in heterogenous populations via real-time optical tagging and image analysis. *Nat. Biomed. Eng.* **2022**, *6*, 667−675.
(65) Sierra Heras, L.; Danelon, C. Autocatalytic Selection of Gene Functions in Synthetic Cells. *Commun. Biol.* **2026**, DOI: 10.1038/s42003-026-10372-z.
(66) Abil, Z.; Sierra, A. M. R.; Stan, A. R.; et al. Darwinian evolution of self-replicating DNA in a synthetic protocell. *Nat. Commun.* **2024**, *15*, No. 9091.
(67) Soengas, M. S.; Gutiérrez, C.; Salas, M. Helix-destabilizing activity of φ29 single-stranded DNA binding protein: effect on the elongation rate during strand displacement DNA replication. *J. Mol. Biol.* **1995**, *253*, 517−529.
(68) Afgan, E.; Nekrutenko, A.; Grüning, B. A.; et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res.* **2022**, *50*, W345−W351.
(69) Blankenberg, D.; Gordon, A.; Von Kuster, G.; et al. Manipulation of FASTQ data with Galaxy. *Bioinformatics* **2010**, *26*, 1783−1785.
(70) Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **2018**, *34*, 3094−3100.
(71) Schindelin, J.; Arganda-Carreras, I.; Frise, E.; et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **2012**, *9*, 676−682.