Rational Design of a Synthetic Entner-Doudoroff Pathway for Improved and Controllable NADPH Regeneration

Chiam Yu Ng¹, Iman Farasat¹, Costas D. Maranas¹, and Howard M. Salis¹²,†

¹Department of Chemical Engineering
²Department of Biological Engineering, Pennsylvania State University, University Park, PA 16802
†Corresponding author: salis@psu.edu

Abstract

NADPH is an essential cofactor for the biosynthesis of several high-value chemicals, including isoprenoids, fatty acid-based fuels, and biopolymers. Tunable control over all potentially rate-limiting steps, including the NADPH regeneration rate, is crucial to maximizing production titers. We have rationally engineered a synthetic version of the Entner-Doudoroff pathway from Zymomonas mobilis that increased the NADPH regeneration rate in E. coli MG1655 by 25-fold. To do this, we combined systematic design rules, biophysical models, and computational optimization to design synthetic bacterial operons expressing the 5-enzyme pathway, while eliminating undesired genetic elements for maximum expression control. NADPH regeneration rates from genome-integrated pathways were estimated using a NADPH-binding fluorescent reporter and by the productivity of a NADPH-dependent terpenoid biosynthesis pathway. We designed and constructed improved pathway variants by employing the RBS Library Calculator to efficiently search the 5-dimensional enzyme expression space and by performing 40 cycles of MAGE for site-directed genome mutagenesis. 624 pathway variants were screened using a NADPH-dependent blue fluorescent protein, and 22 were further characterized to determine the relationship between enzyme expression levels and NADPH regeneration rates. The best variant exhibited 25-fold higher normalized mBFP levels when compared to wild-type strain. Combining the synthetic Entner-Doudoroff pathway with an optimized terpenoid pathway further increased the terpenoid titer by 97%.
1. Introduction

Most metabolic reactions that produce industrially important compounds depend on electron-carrying cofactors, such as NADH and NADPH. In particular, NADPH plays a vital role in the biosynthesis of drugs (Chemler et al., 2010; Gunnarsson et al., 2004; Lee et al., 2010), chiral alcohols (Bastian et al., 2011; Shen and Liao, 2013), fatty acids and biopolymers (Hong et al., 2003; Lim et al., 2002; Rathnasiri et al., 2012; Sanchez et al., 2006), while also being required for lipid biosynthesis, biomass formation, and cell replication (Alberts et al., 2002; Smolke, 2009). As a result, the regeneration rate of NADPH is often the rate-limiting step for the over-production of desired chemicals, while maintaining robust cellular growth. Therefore, increasing NADPH regeneration rates can increase both pathway productivities and product yields (Chemler et al., 2010; Fuhrer et al., 2005; Gunnarsson et al., 2004; Kabus et al., 2007; Martinez et al., 2008; Sanchez et al., 2006; Siedler et al., 2012; Walton and Stewart, 2004). Here, our objective is to develop a modular, drop-in pathway that rapidly regenerates NADPH, and provides control over redox supply levels, to increase the productivity of NADPH-dependent metabolic pathways.

In *E. coli*, the three major sources of NADPH regeneration are the pentose phosphate pathway (PPP), tricarboxylic acid (TCA) cycle, and the transhydrogenase system (Sauer et al., 2004). To increase NADPH regeneration rates, a common strategy has been to re-direct carbon flux through PPP by the deletion of *pgi* or *pfkB* (Chemler et al., 2010; Chin et al., 2009; Siedler et al., 2011); and by the over-expression of glucose-6-phosphate dehydrogenase (*zwf*) or 6-phosphogluconate dehydrogenase (*gnd*) (Chin and Cirino, 2011; Lee et al., 2007; Lim et al., 2002). Following these approaches, titers of leucocyanidin and thymidine, both limited by NADPH availability, were improved by up to 3.8-fold (Chemler et al., 2010) and 4.85-fold (Lee et al., 2010), respectively. However, the resulting release of carbon dioxide within PPP (Vital-Lopez et al., 2006) lowers product carbon yield and the growth defect caused by a *pgi* deletion limits productivity (Charusanti et al., 2010; Fong et al., 2006; Vital-Lopez et al., 2006). To overcome this challenge, it is possible to redirect carbon flux through the Entner-Doudoroff (ED) pathway, which regenerates NADPH without a concomitant carbon loss.

The Entner-Doudoroff (ED) pathway combines the enzymes glucose-6-phosphate dehydrogenase (*zwf*), 6-phosphogluconolactonase (*pgl*), 6-phosphogluconate dehydratase (*edd*), and 2-keto-3-deoxygluconate-6-phosphate (KDPG) aldolase (*eda*) to convert glucose 6-
phosphate into two units of pyruvate, while generating equimolar amounts of ATP, NADH, and NADPH (Figure 1). In contrast, the well-known Embden-Meyerhof-Parnas (EMP) glycolysis pathway performs the same conversion, but produces two moles each of ATP and NADH. There are several additional, and important, differences between these otherwise substitutable glycolytic pathways. First, the lower amount of ATP synthesis causes the ED pathway to become highly exergonic, favoring catalysis in the forward direction (Flamholz et al., 2013). As a result, the ED pathway has been shown to require 3.5-fold less enzyme to achieve the same EMP pathway flux, implying a similar reduction in the cost of assembling the catalytic machinery. Second, bacterial strains that rely on the ED pathway to perform glycolysis generally produce more NADPH than their anabolic demand (Conway, 1992; Flamholz et al., 2013; Fuhrer and Sauer, 2009). To supplement ATP synthesis, like other facultative organisms, ED-dependent bacteria carry out aerobic respiration and catabolize additional non-glycolytic substrates (Flamholz et al., 2013; Fuhrer et al., 2005). Finally, when both the EMP and ED pathways are available in the same organism, the ED pathway often fulfills an alternative role. For example, the conditionally expressed ED pathway in E. coli evolved to carry out gluconate metabolism (Chin and Cirino, 2011; Jiao et al., 2003; Liu et al., 2013; Matsushita et al., 1997; Peekhaus and Conway, 1998; Siedler et al., 2011; Zhao et al., 2004).

Engineering the natural E. coli ED pathway may not enable tunable control over its NADPH regeneration rate, due to endogenous layers of transcriptional, translational, and allosteric regulation. Instead, a promising strategy is to heterologously express a highly active version of the pathway from a different organism (Alper and Stephanopoulos, 2009; Martin et al., 2003). We therefore selected the highly active ED pathway from Zymomonas mobilis. This organism relies solely on the ED pathway for glycolysis, has a high sugar uptake rate, has a high regeneration rate of ATP and NAD(P)H, and produces large amounts of ethanol that surpasses many yeast strains (Kalnenieks et al., 2014; Sprenger, 1996). The high glycolytic flux of Z. mobilis has been attributed to the high turnover numbers, minimum allosteric control, and high expression levels of its ED enzymes (Conway et al., 1991; Kalnenieks, 2006; Sprenger, 1996). Its glucose 6-phosphate dehydrogenase (zwf) enzyme is known to regenerate both NADH and NADPH, enabling autonomous redox balancing (Fuhrer and Sauer, 2009). To the best of our knowledge, a complete Z. mobilis Entner-Doudoroff pathway has not yet been expressed in E.coli.
In this study, we designed, constructed, and systematically optimized a synthetic Entner-Doudoroff pathway as a drop-in module that significantly increases a bacterial host's NADPH regeneration rate. Using computational optimization and biophysical models, we rationally designed two operon sequences to heterologously express the four-enzyme ED pathway as well as phosphoglucose isomerase (pgi) to obtain maximum control over their expression levels (Figure 2). We constructed and assembled the resulting 8.9-kbp genetic system, and integrated it into the *E. coli* MG1655-derived genome. We then efficiently explored the 5-dimensional expression space by employing the RBS Library Calculator to design optimized genome mutations (Farasat et al., 2014) together with multiplex automated genome engineering (MAGE) mutagenesis to implement the genome mutations (Wang et al., 2009), generating libraries of $10^6$ ED pathway-genome variants. Using a NADPH-dependent fluorescent protein, we screened 624 ED pathway-genome variants for high NADPH regeneration rates, and then extensively characterized 22 re-integrated pathways by measuring *in vivo* NADPH regeneration rates and NADPH-dependent biosynthesis rates. As a result, an optimized ED pathway increased NADPH-dependent fluorescence by 25-fold and increased the production titer of an already optimized carotenoid biosynthesis pathway by 97%.
Figure 1. EMP versus ED Pathways. The major glycolytic pathways in *E. coli*, showing the Embden-Meyerhof-Parnas (EMP) and Entner-Doudoroff (ED) pathways. Native *E. coli* genes are shown in light blue color. The optimized heterologous genes in our synthetic bacterial operons are shown highlighted in yellow boxes.
Figure 2. Design and construction of ED1.0 and combinatorial ED variants. The synthetic ED-tetAR operons were designed using the Operon Calculator, with the Zm-zwf, Zm-pgi, Zm-edd, Zm-eda and Zm-pgl genes from Zymomonas mobilis ZM4 as input. Zm-zwf and Zm-pgi were grouped into the first operon and Zm-edd, Zm-eda and Zm-pgl were grouped into the second operon. Both operons were under the control of promoter P_tac. This genetic system was integrated into the chromosome of E. coli EcNR2 strain between tonB and yciL locus, resulting in strain ED1.0. 16-variant RBS library (dRBS_i, i = 1, 2, ..., 5) for each gene i was designed using RBS Library Calculator and introduced into strain ED1.0 using MAGE mutagenesis resulting in a large combinatorial population.

2. Materials and methods

Chemicals were obtained from Sigma-Aldrich Co. (St. Louis, MO) and VWR International (Radnor, PA). Enzymes were purchased from New England Biolabs Inc. (Ipswich, MA). E. coli TOP10 strain (Invitrogen), Pir116 strain (TransforMax™ EC100D™ pir-116) and E. coli K12 ER2267 (LacIq) strain (NEB) were used for plasmid construction and propagation. Plasmid pQE-mBFP (Hwang et al., 2012) was obtained from Dr. Geun-Joong Kim’s lab (Chonnam National University, South Korea). Tetracycline-resistance gene cassette tetAR was obtained from Dr John Roth’s Lab (UC Davis, CA). DNA fragment and oligonucleotide synthesis were performed by Integrated DNA Technologies Inc. (Coralville, IA) and GeneArt (Regensburg, Germany). Gene sequencing was performed by QuintaraBio (Boston, MA) and the Penn State Genomics Core. Unless stated otherwise, M9 minimal media is 1x M9 salt (Sigma-
Aldrich Co.), 2 mM magnesium sulfate, 100 µM calcium chloride, supplemented with 0.4% w/v glucose, 0.34 g/L thiamine and 0.05 g/L leucine, adjusted to a pH of 7.4. For EcNR2 or EcIF15 derived strains, biotin was added to a final concentration of 250 µg/L.

2.1. Design of synthetic operons and codon optimization

Starting from amino acid sequences, gene and operon designs were carried out using a recently developed design procedure for synthetic operons (Operon Calculator v0.50) (unpublished data). Synonymous codon selections for each coding sequence were initially determined by using weighted random choice from a custom codon usage table, followed by multi-objective optimization according to several design criteria. The custom codon usage table heavily weighted synonymous codons from the 164 most highly translated E. coli MG1655 protein coding sequences, as predicted by the RBS Calculator v2.0 (Espah Borujeni et al., 2014), while eliminating rare codons. Additional synonymous codon mutations were made according to the following criteria: (i) 5' untranslated regions and the beginnings of each protein coding sequence were co-optimized to achieve high translation rate capacities i.e. at least a 100,000 translation initiation rate on the RBS Calculator v1.1 proportional scale, (ii) the translation initiation rates of internal start codons were minimized, (iii) Shine-Dalgarno-like ribosomal pause sequences (Li et al., 2012) were removed, (iv) repetitive sequences, inverted repeat, and selected restriction enzyme recognition sequences were removed, and (v) 5' and 3' untranslated regions had minimum necessary lengths. Optimized ribosome binding sites and protein coding sequences for the five enzymes were assembled into two bacterial operons. Transcription of both operons is initiated by an IPTG-inducible P_tac promoter, and terminated by BBa_B0021 and BBa_K780000 transcriptional terminators from the Registry of Standard Biological Parts.

2.2. Strain and plasmid construction

The parent strain for initial genome integrations and ED optimization is E. coli EcNR2, derived from E. coli MG1655 with modifications bioA/bioB::λ-Red-bla ΔmutS::cat (Wang et al., 2009). The parent strains for re-integration and characterization of selected ED pathways are EcNR2 and EcIF15, which is a derivative of EcHW2f (Wang et al., 2009) with a mutated dxs ribosome binding site (Supplementary Table S1) (Farasat et al., 2014). The EcNR2 Δpgi strain was constructed by inserting two consecutive stop codons into the pgI coding sequence with co-
selection MAGE (Supplementary Materials and Methods S1.1). The synthetic operons described in section 2.1 were divided into three segments (ZmED1, ZmED2 and ZmED3) for gene synthesis. The first two segments were further divided into six gBlocks Gene Fragments each with lengths of about 500 bp. Both ZmED1 (2417 bp) and ZmED2 (2318 bp) were then assembled using Gibson Assembly (Gibson et al., 2009). The third segment (ZmED3) was synthesized by GeneArt (Life Technologies). Plasmid pCN-LED was constructed by assembling ZmED1, ZmED2 and ZmED3 into a separately constructed vector backbone (pCN-L) that contained an R6K origin, a chloramphenicol (Cm) resistance marker, and the lacI gene. Once assembled, pCN-LED contained the two ED bacterial operons flanked by 40 bp sequences and restriction sites to enable recombination into the tonB/yciL intergenic region within the E. coli MG1655 genome. Additionally, the tetAR operon was then amplified from Salmonella typhimurium LT2 TT25401 strain and inserted into plasmid pCN-LED. After sequence verifying the resulting 11.8 kb plasmid pCN-Ledt (pCN-065), the recombination cassette (ED-tetAR) was amplified by PCR and integrated into the EcNR2 genome at the tonB/yciL intergenic region using lambda red recombination approach (see section 2.4). In order to increase tetA expression required for counter-selection with nickel salts or fusaric acid during co-selection MAGE (Podolsky et al., 1996), additional MAGE genome mutagenesis was employed to increase the translation initiation rate of the integrated tetA marker from 228 to 48,372 au, employing the RBS Library Calculator to design RBS genomic mutations (Supplementary Materials and Methods S1.2.). The resulting strain is henceforth denoted as ED1.0.

2.3. Combining the RBS library Calculator and MAGE for pathway optimization

For each of the five enzyme coding sequences, we used the RBS Library Calculator in Search mode to design 16-variant degenerate RBS libraries with translation initiation rates that spanned between 4,500 to 61,000-fold range, while constricting degenerate nucleotides to within a 9 bp region (Farasat et al., 2014) (Supplementary Table S3). RBS libraries were encoded within 90 bp degenerate oligonucleotides with four phosphorothioated bases at the 5’ terminus (Supplementary Table S2). RBS libraries were incorporated into the ED1.0 chromosome by performing 40 manual rounds of MAGE using a pool size of 80 (16 x 5) oligonucleotides adjusted to a final concentration of 1 to 2 µM. Daily MAGE rounds were carried out by first growing ED1.0 or previously mutagenized strains in 5 mL of SOC broth, supplemented with
0.4% w/v glucose, for 12 to 16 hours at 30°C and 250 RPM shaking, followed by dilution to an initial OD\(_{600}\) of 0.05 in 5 mL of SOC broth. The first MAGE round begins by incubating cells at 30 °C at 250 to 300 RPM until their OD\(_{600}\) reached 0.5 to 0.7, followed by heat shock at 42 °C for 15 min to induce \(\lambda\)-prophage \((\text{bet, gam, exo})\) genes expression, pelleting, and washing three times with cold sterile water to induce electrocompetency. 50 µL of the oligonucleotide mixture was then added to electrocompetent cells and electroporated at 1,800 V. To begin the second MAGE round, cells were recovered in pre-warmed SOC until their OD\(_{600}\) reached 0.4 to 0.6, followed by repetition of heat shock, recovery, and electroporation. Three to four MAGE rounds were performed daily. The resulting pool of ED-genome variants were then characterized using several assays.

2.4. Chromosomal integration of Entner-Doudoroff pathway variants

ED-tetAR linear DNA cassettes were PCR amplified from selected variants and re-integrated into the chromosomes of strains EcNR2 or EcIF15 via homologous recombination using the following approach. Overnight cultures of isolated colonies were grown in SOC broth supplemented with 0.4% w/v glucose for 12 to 16 hours at 30 °C with aeration. Overnight cultures were then diluted to an initial OD\(_{600}\) of 0.05 to 0.1 in 5 ml of SOC broth and were grown at 30 °C to OD\(_{600}\) of 0.4 to 0.6. Induction of the \(\lambda\)-Red recombination proteins were performed by shifting the cultures to 42 °C water bath with shaking for 15 minutes. The cultures were subsequently chilled for 10 min, washed three times with cold sterile distilled water and finally resuspended in 150 µL of distilled water. For each transformation, 10 ng – 100 ng of linear DNA cassette was added to 50 µL of electrocompetent cells. After electroporation at 1800 V, electroporated cells were mixed with 1 mL of SOC and then incubated at 30 °C, 250 RPM before plating on LB agar plate containing tetracycline or appropriate antibiotic. Colonies with correct insertion of linear DNA were verified by PCR analysis.

2.5. Measurement of intracellular NADPH and NADP\(^+\) levels

Intracellular levels of NADPH and NADP\(^+\) were determined according to a method described previously (Chin and Cirino, 2011). Overnight culture inoculated with isolated colonies were grown in LB with 50 µg/mL Cm at 30 °C with 250 RPM shaking. 50 mL of M9 minimal media with 0.4% w/v glucose were then inoculated with overnight culture to OD\(_{600}\) of
0.1 and induced with with 0.5 mM IPTG. After 24 h of growth at 30 °C with 250 RPM shaking (OD$_{600}$ of 1.0 to 3.0), cultures were chilled in ice bath for 10 min, harvested by centrifugation at 4 °C, 10 min, 4750 RPM and resuspended to 0.5 mL of OD$_{600}$ of 30.

For NADPH (reduced form) analysis, cells were resuspended in 250 µL of 0.3 M NaOH, incubated at 60 °C for 7 min, then neutralized by 250 µL of of 0.3 M HCl and 50 µL of Tricine-NaOH (pH 8.0). For NADP$^+$ (oxidized form) analysis, cells were resuspended in 250 µL of 0.3 M HCl and 50 µL of Tricine-NaOH (pH 8.0), incubated at 60 °C for 7 min, then neutralized by 250 µL of of 0.3 M NaOH. Neutralized samples were then centrifuged at 4 °C, 60 min, 4750 RPM and the resulting supernatant was transferred to new microcentrifuge tube. 80 µL of reduced sample or 40 µL of oxidized sample mixed with 40 µL of 0.1 M NaCl was added to a 96-well microtiter plate for analysis. 2x stock solution of reaction mixture was prepared by mixing equal volume of 1.0 M Tricine-NaOH (pH 8.0), 4.2 mM thiazolyl blue tetrazolium bromide (MTT), 40 mM EDTA, 1.67 mM phenazine ethosulfate (PES), and 25 mM glucose-6-phosphate. Both reduced and oxidized samples were mixed with 80 µL of freshly prepared reaction mixture and then incubated at 37 °C for 15 min. 20 µL of 2.5 U/mL NADP$^+$-specific glucose-6-phosphate dehydrogenase (G6PDH) from *Saccharomyces cerevisiae* was added to initiate reaction and the time-course formation of reduced MTT was measured at 570 nm, 37 °C with Tecan microplate reader. The cofactor concentration was determined from rate of change in absorption when compared to calibration curve constructed by measuring NADP$^+$ standards on the same microplate.

### 2.6. Quantifying NADPH levels with a modified fluorescent reporter

Plasmid pQE-mBFP expresses a NADPH-dependent metagenomic blue fluorescent protein (mBFP) on a ColE1 vector (Hwang et al., 2012). The mBFP protein is a short chain dehydrogenase (SDR) that binds specifically to NADPH and emits fluorescence at 451 nm when excited at 395 nm, producing more fluorescence when supplied with more NADPH. We modified pQE-mBFP by replacing its original IPTG-inducible T5 promoter with a strong constitutive promoter and by increasing the translation initiation rate of mBFP to 356,786 au on the RBS Calculator proportional scale, resulting in pCN-mBFP. The pool of ED-genome strain variants, the strain ED1.0 and the EcNR2 strain were transformed with pCN-mBFP, followed by characterization of isogenic cultures using spectrophotometry (Infinite M1000, TECAN) to
record mBFP fluorescence. Overnight cultures in LB with 50 µg/mL of kanamycin (Kan) were used to inoculate 200 µL M9 minimal media with 10 µg/mL Kan in microtiter wells. Cultures were then incubated at 30 °C with high orbital shaking for 6 to 9 hours. Cells were then serially diluted into fresh, pre-warmed M9 minimal media with 10 µg/mL Kan and 1 mM IPTG and grown similarly for another 12 hours. During the IPTG-induced growth period, cell densities (OD_{600}) and mBFP fluorescence levels were recorded every 10 minutes. Specific mBFP fluorescence levels were determined by dividing each strain's background-corrected fluorescence levels by their corresponding OD_{600} values. Specific mBFP production rates were determined by the slope of mBFP fluorescence levels versus time over a region of linearly increasing fluorescence levels. For normalization, all specific mBFP production rates were divided by the specific mBFP production rate of the EcNR2 control strain. Selected ED-genome variants were sequenced according to their normalized specific mBFP production rates.

2.7. Measurement of NADPH-dependent carotenoid biosynthesis

The enzymes encoded by *crtEBI* from *Rhodobacter sphaeroides* catalyze the conversion of isopentenyl diphosphate to neurosporene, a brown carotenoid pigment. We previously constructed plasmids that express an optimally balanced *crtEBI* operon to produce high levels of neurosporene (Farasat et al., 2014). The operon's expression is controlled by either an IPTG-inducible P_{lacO1} promoter (pIF-001C or pIF-001K) or an arabinose-inducible P_{BAD} promoter (pIF-002). Selected ED-genome variants were transformed with pIF-001 or pIF-002, and cultured overnight in 25 mL of LB broth with 50 µg/mL Cm for EcIF15-derived strains or 50 µg/mL Kan for EcNR2-derived strains. Cultures were then diluted to a final OD_{600} of 0.1 in 25 mL 2X M9 minimal media supplemented with 0.4% w/v glucose, 10 µg/mL Cm and 0.2 mM IPTG. For strains transformed by pIF-002, 10 mM arabinose was added. Cultures were incubated at 37°C and 300 RPM shaking for 10 hours and pelleted by centrifugation. Cell pellets were washed, dried, and mixed with 1 mL acetone in pre-weighed microcentrifuge tubes, followed by incubation at 55°C and repeated vortexing for 20 minutes. The extraction solution was centrifuged again, and a 50 µL supernatant sample was transferred to 1 mL acetone (21-fold dilution) in a fresh microcentrifuge tube. The absorbance of neurosporene in acetone was recorded by a NanoDrop 2000C spectrophotometer at 470 nm, using pure acetone as a blank. Extracted cell pellets were dried open-capped in a 65°C oven for 2 days, equilibrated closed-
capped at room temperature, followed by recording of cell pellet masses by comparison of microcentrifuge weights before and after extraction.

3. Results

3.1. Rational Design and Construction of a Synthetic Entner-Doudoroff Pathway

We selected five enzymes from Z. mobilis ZM4 for heterologous expression in E. coli: glucose-6-phosphate dehydrogenase (ZMO0367/Zm-zwf), 6-phosphogluconolactonase (ZMO1478/Zm-pgl), 6-phosphogluconate dehydratase (ZMO0368/Zm-edd), 2-keto-3-deoxygluconate-6-phosphate (KDPG) aldolase (ZMO0997/Zm-eda), and phosphoglucose isomerase (ZMO1212/Zm-pgi). The first four enzymes constitute the ED pathway that converts glucose 6-phosphate to pyruvate and glyceraldehyde-3-phosphate, while the fifth reversibly interconverts fructose-6-phosphate to glucose-6-phosphate (Figure 1). We included the enzyme Zm-pgi to regulate metabolic flux at the major glycolysis branch point, glucose-6-phosphate. Throughout the paper, the operons consisting of the five Z. mobilis enzymes are designated as the synthetic ED operons.

Protein expression levels are regulated by several genetic elements, including promoters, ribosome binding sites (RBSs), and protein coding sequences. In natural genetic systems, changes in transcription, translation, and mRNA stability collectively control a protein's expression level. As a key strategy to optimizing the ED pathway, we developed an optimization procedure, called the Operon Calculator, that designs bacterial operon sequences with the overall objective of concentrating expression control to the fewest number of short genetic parts, while eliminating undesired genetic elements that confound our ability to control protein expression. The Operon Calculator minimizes the number of translated internal start codons, ribosome pause sequences, repetitive sequences, and undesired restriction sites, while selecting 5' UTR sequences, synonymous codon sequences, and terminators for high translation rate capacities and termination efficiencies (Section 2.1). As a result, two synthetic bacterial operon sequences were designed grouping together the enzymes Zm-zwf and Zm-pgi into the first operon, and Zm-pgl, Zm-eda and Zm-edd into the second operon (Figure 2, Supplementary Figure S1). We selected an IPTG-inducible P_{lac} promoter to transcribe both operons. We also designed the initial ribosome binding site sequences for all five enzyme coding sequences to have translation initiation rates of about 1,000 au on the RBS Calculator v1.1 proportional scale, which is a
moderate translation rate. Importantly, these RBS sequences were designed by the RBS Library Calculator (Farasat et al., 2014) so that a small number of nucleotide mutations could greatly vary the coding sequences' translation rates. In addition, we chose tetAR as our selection marker for genome integration.

The synthetic ED operons were first constructed and inserted into a pre-constructed vector (pCN-L) along with tetAR operon resulting in a 11.8 kb plasmid (pCN-LEDT) by combining DNA synthesis, DNA assembly, and molecular cloning. We then employed PCR amplification and homologous recombination to integrate the ED-tetAR operons (8.9 kb) into the EcNR2 chromosome (Section 2.2). We refer to this strain, harboring the first version of our synthetic Entner-Doudoroff pathway, as ED1.0 (Figure 2).

3.2. Characterization of the synthetic Entner-Doudoroff pathway in ED1.0

We first characterized the activity of the synthetic Entner-Doudoroff pathway in ED1.0, compared to its parent strain EcNR2, by measuring the NADPH/NADP⁺ intracellular redox ratio and by measuring the in vivo NADPH regeneration rate. Using a glucose 6-phosphate dehydrogenase assay on cell extract (Section 2.5), we found that the ED1.0 strain has a NADPH/NADP⁺ redox ratio that is 1.87-fold higher than its parent EcNR2 strain after both are cultured to the exponential growth phase using minimal M9 media (two-tailed, two-sample t-test, p-value = 0.037) (Figure 3A). We then selected mBFP, a NADPH-dependent fluorescent protein reporter, as a large consumption sink for NADPH that also serves as an observable readout for its in vivo regeneration rate. mBFP is a short-chain dehydrogenase that actively oxidizes NADPH and proportionally emits fluorescence (Hwang et al., 2012).

As an initial control to validate the mBFP assay, we over-expressed the transhydrogenase PntAB on an R6K origin vector within E. coli Pir116 together with an IPTG-inducible mBFP expression plasmid to measure its effect on mBFP fluorescence (Supplementary Materials and Methods S1.3.). For a comparison, we also created an R6K-origin plasmid that did not express any enzymes, and co-transformed this control plasmid with the mBFP expression plasmid in the same strain. We continuously measured mBFP fluorescence after IPTG induction during long-time cultures maintained in the exponential phase of growth using minimal M9 media. After a lag period, mBFP fluorescence increased linearly in time over a long duration, indicating that NADPH availability was a rate-limiting step to fluorescence emission (Supplementary Figure
We quantified NADPH regeneration rate by calculating the first-derivative (slope) of the time-course mBFP fluorescence per OD\textsubscript{600} over the linear regime, which we refer to as the mBFP fluorescence production rate (mBFP flu rate). Overexpression of PntAB together with induced mBFP expression, using 0.5 mM IPTG, increased the mBFP fluorescence production rate by 81%, compared to the control (93.6 ± 20.7 au/OD\textsubscript{600}.h vs 51.8 ± 28.9 au/OD\textsubscript{600}.h, two-tailed, two-sample t-test, p-value = 0.057) (Supplementary Figure S2B). Therefore, mBFP was found to be a quantitative reporter of the \textit{in vivo} NADPH regeneration rate.

We then modified the mBFP expression plasmid to ensure that NADPH availability was the rate-limiting factor that controls mBFP fluorescence across a larger dynamic range. We first replaced the IPTG-inducible promoter controlling mBFP expression with a constitutive promoter. We then replaced its ribosome binding site sequence with a rationally designed one to substantially increase its translation rate. The resulting plasmid (pCN-mBFP) constitutively expresses mBFP with a very high expression level (Section 2.6).

We then employed the pCN-mBFP plasmid to measure the \textit{in vivo} NADPH regeneration rate of the ED1.0 strain, compared to its parent EcNR2 strain. Transformed strains were grown in long-time cultures maintained in the exponential growth phase using minimal M9 media. mBFP fluorescence was monitored continuously. We found that the mBFP fluorescence production rate for the parent strain EcNR2 was relatively constant regardless of the addition of IPTG. In contrast, the mBFP fluorescence production rate for the strain ED1.0 was substantially higher and was further increased when adding IPTG to induce the ED pathway's expression. The highest mBFP fluorescence production rate was observed at an intermediate IPTG concentration (25 µM) and there was no statistically significant increase in mBFP production rate when additional IPTG was added (p-values > 0.1 for all pair-wise two-sample t-tests between 25 µM and 0.5 mM IPTG) (Figure 3B). In the absence of IPTG, the observed level of mBFP fluorescence production could be explained by transcriptional leakiness of the P\textsubscript{tac} promoter. Based on these results, the synthetic Entner-Doudoroff in strain ED1.0 is improving the NADPH regeneration rate by 4.8-fold when induced by 25 µM IPTG, compared to its parent strain. We expected that optimization of the ED pathway would be necessary to further increase its activity.
Figure 3. Characterization of ED1.0. (A) The NADPH/NADP⁺ redox ratios were measured for strains EcNR2 and ED1.0. Expression of the ED pathway was induced using 0.5 mM IPTG. ED1.0 had a significantly higher NADPH/NADP⁺ ratio than its parent strain (two-tailed, two-sample t-test, p-value = 0.037). Values and error bars represent the averages and s.d. of the ratios for four replicates. NADPH and NADP⁺ concentrations (μmol/ g DCW) for EcNR2 and ED1.0 are also reported (mean ± s.d. for n = 4). (B) The mBFP fluorescence production rates were measured for strains (blue bar) EcNR2 and (red bar) ED1.0. Increasing amounts of IPTG were added in separate experiments. The expression of the ED pathway in ED1.0 is controlled by an IPTG-inducible P_tac promoter. The pound sign (#) indicates a near-background mBFP production rate for EcNR2 strain at 0 mM IPTG. Values and error bars represent the means and s.d. of two replicates. *P < 0.1; **P < 0.05.

3.3. Efficient Search for Improved ED Pathway Variants in a 5-dimensional Expression Space

We initially designed the synthetic operons in ED1.0 to express all the ED pathway's enzymes with translation initiation rates of about 1,000 au on RBS Calculator proportional scale. The initial values were chosen to match the typical translation initiation rates controlling the expression of enzymes found in the native glycolysis, PPP, and ED pathways of E. coli MG1655. We anticipated that our initial translation rate guesses may not lead to the highest possible NADPH regeneration rate, and during the initial design of the synthetic operons, employed our RBS Library Calculator algorithm to design ribosome binding site sequences that could be easily mutagenized to provide large changes in translation initiation rate.

Specifically, the RBS Library Calculator in Genome Editing mode was applied to design 16-variant RBS libraries that systematically varied the translation rates for each enzyme coding sequence from about 10 to 900,000 au on the RBS Calculator proportional scale (Farasat et al.,
The resulting optimized RBS libraries contained a small number of nearby degenerate nucleotides that could be readily incorporated into the *E. coli* genome using a site-directed genome mutagenesis technique. A key advantage of this approach is the compactness of the resulting combinatorial RBS library and the broad coverage of the 5-dimensional expression space. Complete combinatorial incorporation of the five 16-variant RBS libraries will create a library of $16^5$ (1,048,576) variants that will uniformly sample a 5-sided hypercube with lengths that span at least a 10,000-fold change in expression (Figure 4).

**Figure 4. Uniform Sampling of the 5-dimensional Expression Space.** (A) For each of the enzyme, degenerate RBS (dRBS) for 16-variant RBS library was designed with RBS Library Calculator. By having just four degenerate nucleotides on the RBS sequence of *Zm-zwf*, one can uniformly vary the predicted translation rate across a 4500-fold range. (B) (Black dots) The predicted translation initiation rates for all possible combinations of the five optimized RBS libraries, showing two-dimensional slices of the 5-dimensional space. (Red circles) The predicted translation rates for the 22 selected ED-expressing genome variants and (blue squares) the initial ED1.0 strain showing the sampling of the space after 40 cycles of MAGE genome engineering. Translation rates are predicted using RBS Calculator v1.1.
Figure 5. Characterization of ED-expressing genome variants. (A) The normalized mBFP fluorescence production rates of 387 genome variants and 22 control strains (including 7 ED 1.0 mBFP, 10 WT mBFP and 5 no BFP control strains) were measured and ranked. Normalized mBFP fluorescence production rates were calculated by dividing all measurements by the average mBFP fluorescence production rate for the EcNR2 strain, which was 28.2 ± 9.4 au/OD_{600}/h. Boxplots represent the ranking distribution of ED1.0 mBFP, WT mBFP and the no mBFP control strains. (B) The growth rates of each variant were measured and appear in the same order as in A. Bar colors are (blue) ED-expressing genome variants, (green) parent strain EcNR2, (red) strain ED1.0, and (grey) negative controls EcNR2 and ED1.0 without mBFP overexpression.

We then applied oligo-mediated allelic replacement, also known as MAGE (Wang et al., 2009), to incorporate these ribosome binding site mutations directly into the *E. coli* genome. According to an allelic replacement efficiency calculation (Wang and Church, 2011), we estimated that 40 MAGE cycles were required to generate 14% of genomes with at least 4 out of
5 RBS mutations and 2.3% of genomes with all 5 RBSs mutated. Over a span of 12 days, 40 MAGE cycles were conducted using mutagenic oligonucleotides that correspond to the optimized RBS library sequences (Section 2.3). Bulk sequencing of genome pools after the 12th, 30th, and 40th cycles revealed that RBS sequences became increasingly mutated at specifically targeted nucleotide positions. We then characterized the pool of genome variants from the 40th cycle of MAGE.

We transformed the pool of ED-expressing genome variants with pCN-mBFP, our constitutively over-expressed mBFP plasmid, and then isolated 624 single colonies. We cultured them using minimal M9 media supplemented with 0.4% w/v glucose and 1 mM IPTG, and measured their mBFP fluorescence production rates. As controls, we also characterized the mBFP fluorescence production rates of strain ED1.0 and parent strain EcNR2 in each set of measurements. When varying their RBS sequences and enzyme expression levels, we expected that ED-expressing genome variants would have different NADPH regeneration rates and growth rates, due to changes in pathway flux and the accumulation of toxic intermediates (e.g. KDPG). 237 ED-expressing genome variants (38%) exhibited poor growth after induction of the ED pathway, indicating that some combinations of RBS sequences and enzyme expression levels resulted in imbalanced pathways. The remaining 387 ED-expressing genome variants displayed varied mBFP fluorescence production rates across a 710-fold range (between 0.045-fold to 32.12-fold), indicating that these optimized RBS sequence mutations greatly affected the ED pathway's enzyme expression levels, and correspondingly, its overall NADPH regeneration rate (Figure 5). 336 of these variants had lower mBFP-linked NADPH regeneration rates than the average of ED1.0, suggesting that our initially selected translation rates of 1,000 au were a suitable initial condition. Interestingly, there was no observed correlation between the growth rate of a genome variant and its mBFP-linked NADPH regeneration rate ($R^2 = 0.04$).

Next, we selected 22 ED-expressing genome variants, including 16 high mBFP-producing variants and 6 low mBFP-producing variants, for sequencing, re-integration into fresh E. coli EcNR2 genomes, and further characterization. We found that all 22 variants contained unique combinations of RBS sequences (Figure 6). 7 variants had four out of five modified RBS sequences. 4 variants had three modified RBSs, 6 variants had two modified RBSs, and 5 variants had a single modified RBS. None of the selected genome variants had all five of their RBS sequences modified. The RBS controlling Zm-pgl was modified in 73% of the selected
genome variants, while only 36% of genome variants had RBS modifications controlling $Zm$-zwf expression. Otherwise, 41%, 50%, and 62% of variants had RBS modifications controlling $Zm$-eda, $Zm$-edd, and $Zm$-pgi expression. In one variant (ED9), there was a spontaneous, non-designed mutation to the RBS controlling $Zm$-pgi expression. The translation initiation rates for all RBS modifications were calculated by the RBS Calculator's biophysical model (Figure 6). Notably, the ED3 genome variant contained the most highly translated RBS for $Zm$-zwf, but also a frame-shift mutation in the protein coding sequence that abrogated its expression, indicating that $Zm$-zwf over-expression was highly toxic.

We re-characterized the mBFP fluorescence production rates of the 22 selected ED-expressing genome variants and ranked them (Figure 6, Supplementary Table S4). Strain ED2 consistently ranked first among the selected variants with 25-fold higher mBFP fluorescence production rate, compared to the parent strain EcNR2. Notably, its translation rate profile is not significantly different from ED1.0, which ranked within the top three. The differences between ED2 and ED1.0 were a 3.7-fold higher translation rate for $Zm$-eda and a 10-fold lower translation rate for $Zm$-pgl, indicating that small changes in translation rate, and correspondingly enzyme expression level, can have a beneficial effect on NADPH regeneration rate. However, there remains ED-expressing genome variants in the larger pool with even higher mBFP fluorescence production rates and NADPH regeneration rates.

We examined the relationship between translation rates, enzyme expression levels, and NADPH regeneration rate for these 22 ED-expressing genome variants. Qualitatively, we note that higher enzyme expression levels did not always yield higher NADPH regeneration rates. However, we do not observe a quantitative pattern in contrast to our previous study, where we employed mass action kinetics to formulate such a quantitative relationship. Several factors could confound this analysis, including the activity of endogenous enzymes, competition for transcriptional or ribosomal resources, allosteric feedback control, insufficient sampling of the high-dimensional expression space, and variations in day-to-day measurements.
Figure 6. The Effects of Changing ED Enzyme Expression Levels on NADPH Regeneration Rates

(A) The normalized mBFP fluorescence production rates were measured and ranked for 22 selected ED-expressing genome variants that were re-integrated into the genome of parent strain EcNR2. Values and error bars represent the means and s.d. of 2-12 replicates. (B) The translation initiation rates controlling the expression of the five ED enzymes are shown, comparing the (red) 22 modified ED variants to (blue) ED1.0. Translation initiation rates were predicted using the RBS Calculator v1.1.
3.4. **Improving Terpenoid Biosynthesis using a Synthetic Entner-Doudoroff Pathway**

In bacteria, the methyl erythritol phosphate (MEP) pathway synthesizes the terpenoid precursors isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), consuming equimolar amounts of glyceraldehyde-3-phosphate (G3P) and pyruvate (Farmer and Liao, 2001). In *E. coli*, the enzymes within the MEP pathway primarily use NADPH as the source of reducing equivalents (Alper et al., 2005) ([Supplementary Table S5](#)), and the reaction catalyzed by the first enzyme in the MEP pathway, *dxs*, is a key rate-limiting step to precursor biosynthesis (Ajikumar et al., 2010a; Wang et al., 2009; Yang and Guo, 2014; Yuan et al., 2006). Interestingly, the Entner-Doudoroff pathway synthesizes both the carbon and redox precursors to terpenoid biosynthesis (Liu et al., 2014, 2013) ([Figure 7A](#)). As a consequence, we expected that a highly active ED pathway would increase the rate of terpenoid biosynthesis, particularly when the activities of the MEP and downstream terpenoid biosynthesis pathways were optimized such that NADPH availability became a greater rate-limiting factor. To test this hypothesis, we took advantage of our previous work that systematically optimized a carotenoid biosynthesis pathway from *Rhodobacter sphaeroides*, expressed within an engineered strain that had a substantially higher MEP pathway flux (Farasat et al., 2014).
Figure 7. Improving Terpenoid Biosynthesis using the Synthetic Entner-Doudoroff Pathway

(A) The metabolic reactions that convert glucose 6-phosphate into the carotenoid neurosporene via the Entner-Doudoroff, MEP, and carotenoid biosynthesis pathways. The enzymes crtEBI from *Rhodobacter sphaeroides* produce the brown pigment neurosporene. (B) Neurosporene production was measured when combining selected ED-expressing genome variants and control strains with expression of the crtEBI operon. ED pathways were integrated into either EcNR2 or EcIF15. The strain EcIF15 has an engineered RBS substantially increasing dxs expression. Values and error bars represent the means and s.d. of between 2 to 7 replicates.

First, we transformed selected ED-expressing genome variants with a crtEBI-expressing plasmid, which produces the carotenoid neurosporene under control of an IPTG-inducible P_{lacO1} promoter. Cultures were grown in 2X minimal M9 media with 0.4% w/v glucose and 0.2 mM IPTG in short (10 hour) batches, followed by hot acetone extraction, and measurement of their neurosporene content and dry cell mass. The strains ED13 and ED1.0 produced 79% and 43% higher neurosporene production titers, respectively, compared to the parent EcNR2 strain (Figure 7B). To compare the use of the ED pathway to alternative approaches to increasing NADPH regeneration, we knocked out expression of pgi in the parent strain EcNR2 and found that it increased neurosporene production by 24% using the same growth conditions and measurements. In contrast, the strains ED2, ED5, and ED11 produced about the same amount of neurosporene, compared to their parent EcNR2 strain (4% to 15% lower).
After these measurements, we questioned whether the relatively low neurosporene production rate was creating a sufficiently large burden on NADPH availability. If NADPH regeneration is to become a rate-limiting step in neurosporene production, it must be produced at a high rate with a high demand for NADPH. Next, we selected a previously engineered strain EcIF15 and transformed it with plasmid that expresses the neurosporene biosynthesis pathway now under control of an arabinose-inducible $P_{BAD}$ promoter to enable orthogonal transcriptional control over both the ED and $crtEBI$ pathways. EcIF15 was previously engineered to significantly over-express 1-deoxy-D-xylulose-5-phosphate synthase ($dxs$) and increase the biosynthesis of the precursors IPP and DMAPP (Farasat et al., 2014). As a result, with 10 mM arabinose induction, it produces 791% and 843% more neurosporene, compared to control strains $E. coli$ MG1655 and EcNR2, respectively (Figure 7B). We also characterized the $P_{BAD}$-$crtEBI$ pathway in EcIF15 strain expressing ED17’s ED-tetAR to confirm that the large titer change was not a result of swapping promoters (Supplementary Table S6).

We next introduced ED pathway variants into the EcIF15 strain and evaluated the synergistic effects of an increased NADPH regeneration rate together with an increased precursor biosynthesis rate. In contrast to the EcNR2 strain, the expression of any of the selected ED pathway variants (ED1.0, ED2, ED11, ED13, and ED17) in EcIF15 substantially increased neurosporene production. The variant ED11 yielded the highest improvement with a 1336% increase, compared to EcNR2, and a 97% increase, compared to the parent strain EcIF15 (Figure 7B). Interestingly, ED11 has three higher translation rates controlling $Zm$-$zwf$, $Zm$-$edd$ and $Zm$-$eda$ expression, compared to ED1.0 (Figure 7B). Based on these results, when the bottleneck in the MEP pathway was eliminated, expression of the synthetic ED pathway was able to produce more precursors and regenerate more NADPH, leading to large improvements in neurosporene production.
4. Discussion

In the field of metabolic engineering, increasing the availability of NADPH has been a significant challenge, driven by the need to supply greater amounts of reducing equivalents towards the over-production of a wide range of chemical products. To solve this challenge, previous efforts have deleted or over-expressed selected genes, such as oxidoreductases, transhydrogenases, and NAD kinases (Bastian et al., 2011; Chin and Cirino, 2011; W.-H. Lee et al., 2013a, 2013b; Lee et al., 2007; Lim et al., 2002; Rathnasingh et al., 2012; Sanchez et al., 2006; Wang et al., 2011). In a recent effort, the *E. coli* NAD\(^+\)-dependent glyceraldehyde-3-phosphate dehydrogenase (GAPDH) encoded by gapA gene was replaced with a *Clostridium acetobutylicum* gapC gene encoding for a NADP\(^+\)-dependent GAPDH (Martínez et al., 2008). Another recent study also showed that replacing the promoters of *E. coli* edd-eda operon and zwf gene with a constitutive promoter and a strong promoter, respectively, increased intracellular NADPH/NADP\(^+\)-ratio (Zhang et al., 2014).

In this study, we engineered a synthetic version of the Entner-Doudoroff (ED) pathway to rapidly regenerate NADPH. The pathway combines five enzymes from *Zymomonas mobilis*, expressed together within two synthetic bacterial operons that were rationally designed to achieve maximum expression control. Starting from the first version of the pathway, we then carried out systematic optimization of the enzymes' expression levels to improve the pathway's activity, first employing a NADPH-dependent fluorescent protein reporter to measure NADPH regeneration rates, followed by measuring the ED pathway's effect on an NADPH-dependent terpenoid biosynthesis pathway. By combining MAGE genome mutagenesis with our RBS Library Calculator algorithm, we introduced targeted genome modifications to greatly vary the ED pathway's individual enzyme expression levels and to efficiently search its 5-dimensional expression space. In principle, continued MAGE cycling will generate up to a million genome variants, though it was only necessary to characterize a much smaller number to identify ED pathway variants with greatly improved NADPH regeneration rates. As a result, one of our ED pathway variants exhibited a 25-fold higher NADPH regeneration rate, as measured by the fluorescent protein reporter, and another variant increased terpenoid biosynthesis by 97%. The synthetic ED pathway exists as a drop-in module; in principle, it can be transferred and expressed in many different bacterial hosts to increase their NADPH regeneration rate and enhance the production of NADPH-dependent products.
A novel and important aspect of our design approach was to commit, at an early stage, to integrating the ED pathway into the genome of our bacterial host before optimizing its expression levels. In comparison, pathway engineering efforts have traditionally relied upon multi-copy plasmids to over-express desired enzymes. Multi-copy plasmids can express more protein than expression cassettes inside genomes, but they require active selection (e.g. the addition of antibiotics) to maintain plasmid stability over long culture times, which is undesirable in industrial applications. When pathways are plasmid-encoded, any optimization of their expression levels may be problematic when the final version of the pathway must eventually be genome-integrated for industrial applications. Instead, the first version of our genome-integrated ED pathway (ED1.0) was remarkably stable over the course of 40 cycles of MAGE mutagenesis and during 2-day cultures. Taking advantage of genome engineering techniques, this design choice enabled us to rapidly insert, optimize, copy, and re-insert large genetic modules within and across genomes, while improving the modules' stabilities and maintaining the same copy number.

With our genome-centric strategy, we needed to ensure that our operons' transcription rates, translation rates, and mRNA stabilities could be sufficiently high to express high levels of enzyme even though the DNA copy number is very low (1 to 2 copies per cell). Our Operon Calculator applies several design criteria to ensure that the operons' mRNAs have fewer undesirable genetic elements and that the desired protein coding sequences have the potential to be translated at extremely high rates. Among our 22 sequenced ED pathway variants, the translation rates for Zm-pgl, Zm-edd, and Zm-zwf achieved extremely high translation initiation rates (500,000+ au on the RBS Calculator v1.1 proportional scale). If the ED enzymes had exhibited low turnover numbers, then such high translation rates would have been necessary to achieve high pathway activities.

In a previous study, applying a kinetic modeling formalism enabled us to determine a quantitative relationship between sequence, enzyme expression level, and pathway activity (Farasat et al., 2014). However, while our optimization strategy yielded highly productive pathway variants, we did not observe such a clear relationship for the synthetic ED pathway in the EcNR2 parent strain. Previous reports have shown that accumulation of KDPG is toxic (Barnell et al., 1990; Egan et al., 1992) and will therefore impact the organism's growth rate (Fuhrman et al., 1998), which could confound any expected relationship. Notably, amongst the
ED variants that both over-produced neurosporene and emitted large amounts of mBFP fluorescence, ED11 has a consistent increase in Zm-eda, Zm-edd, and Zm-zwf expression that should result in an overall increase in pathway flux, while ED2 has increased Zm-eda and decreased Zm-pgl expression, which should minimize the accumulation of KDPG. It appears that increased neurosporene production in EcIF15-derived strains also relies on a consistent flux through the entire ED pathway, including production of the MEP precursors pyruvate and G3P, and not only on the NADPH regeneration rate.

Finally, together with previous work (Ajikumar et al., 2010b; Alper and Stephanopoulos, 2007; Biggs et al., 2014; Du et al., 2012; Farasat et al., 2014; M. E. Lee et al., 2013; Nowroozi et al., 2014; Oliver et al., 2014; Pfleger et al., 2006; Smanski et al., 2014; Wu et al., 2013; Xu et al., 2013; Zelcbuch et al., 2013; Zhao et al., 2013), this study highlights the common challenges of performing expression optimization on multi-enzyme pathways in high-dimensional expression spaces. For enzymes with low turnover numbers, large changes in enzyme expression are needed to observe a significant change in pathway activity. In contrast, when enzymes have higher turnover numbers, as in the case of the ED pathway, smaller expression level changes can have a significant impact. Multiple enzymes work together to control a pathway's overall activity. As we observed in this study, synergistic changes in the individual enzyme expression levels are needed to create a more active and balanced pathway. Expression imbalances can affect growth rate, due to the accumulation of toxic intermediates.

Further, to avoid exhaustively searching high-dimensional expression spaces, new modeling formalisms will be needed to convert large sets of sequences and measurements into accurately predicted optimal expression levels. These challenges will continue as the field designs and optimizes longer pathways. New approaches will also be needed to combine pathway modules with predictably matched pathway fluxes, tailored to the organism's existing metabolic network. The success of these approaches together with advanced genome engineering techniques will accelerate the development of extremely large genetic systems, encoded within genomes, capable of radically redirecting carbon, energy, redox, uptake, and transport fluxes towards the production and secretion of desired products.
Acknowledgements

We thank Geun-Joong Kim for sharing the pQE-mBFP plasmid. We are also grateful to John Roth for gifting the strain TT25401 harboring the tetAR operon. We thank Thomas Wood for sharing strain MG1655. We thank Long Chen for sharing plasmid pIF-001K. We also thank Manish Kushwaha, Amin E. Borujeni, Tian Tian, and Ali R. Zomorrodi for helpful discussion. This research was supported by the Department of Energy (DE-SC10822882) to CDM and by the Air Force Office of Scientific Research (FA9550-14-1-0089), the Office of Naval Research (N00014-13-1-0074), DARPA (N66001-12-C-4017), an NSF Career Award (CBET-1253641), and a DARPA Young Faculty Award to H.M.S.

References


