SYNTHETIC BIOLOGY-GUIDED METABOLIC ENGINEERING
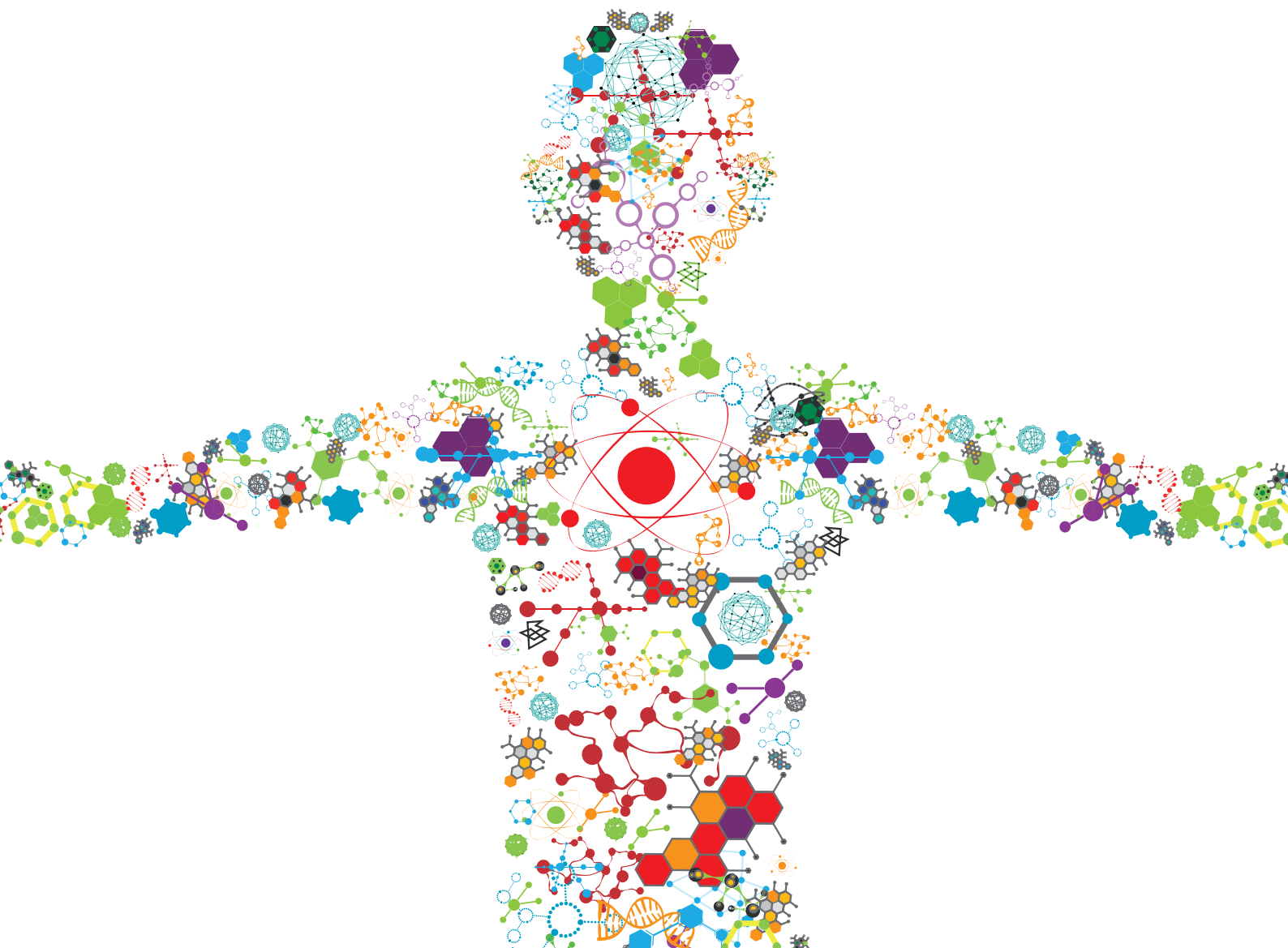
EDITED BY: Rodrigo Ledesma-Amaro, Pablo Ivan Nikel and Francesca Ceroni

frontiers Research Topics

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

SYNTHETIC BIOLOGY-GUIDED METABOLIC ENGINEERING

Topic Editors:
Rodrigo Ledesma-Amaro, Imperial College London, United Kingdom
Pablo Ivan Nikel, Novo Nordisk Foundation Center for Biosustainability
(DTU Biosustain), Denmark
Francesca Ceroni, Imperial College London, United Kingdom

Table of Contents

frontiers
in Bioengineering and Biotechnology

Check for updates

Editorial: Synthetic Biology-Guided Metabolic Engineering

Rodrigo Ledesma-Amaro [1,2*], Pablo I. Nikel [3*] and Francesca Ceroni [2,4*]

[1] Department of Bioengineering, Imperial College London, London, United Kingdom, [2] Imperial College Centre for Synthetic Biology, Imperial College London, London, United Kingdom, [3] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark, [4] Department of Chemical Engineering, Imperial College London, London, United Kingdom

Editorial on the Research Topic

Synthetic Biology-Guided Metabolic Engineering

Synthetic biology can be now considered a mature discipline. A growing number of diverse synthetic gene constructs and circuits have been designed within a wide number of organisms leading to a profound impact on different fields. This Research Topic features and reviews some of the latest progress in Synthetic Biology applications and improvements within the Metabolic Engineering portfolio, covering different aspects in the domain (e.g., bioinformatics, design of synthetic pathways, and implementation of multi-omics approaches).

Among the successes of Synthetic Biology and Metabolic Engineering, the ability to achieve smarter construct design and higher yields of valuable chemicals needs to be considered. As one example, Callari et al. engineered *Saccharomyces cerevisiae* to produce the diterpene casbene, precursor of many terpenoids of medical interest. The authors successfully achieve increased casbene titers via expression of heterologous enzymes that can boost internalization and conversion of precursors; performing dynamic control of inducible promoters also helped maximize the pathway flux toward casbene production. Another example comes from 3-hydroxypropanoic acid (3-HP), a valuable product employed in the bioproduction of several other chemicals, including bioplastics. Maury et al. achieved high product titers in batch cultivations using engineered *S. cerevisiae*. Firstly, the authors characterized a series of 34 native promoters which respond to glucose. Placing the 3-HP pathway under the control of a promoter active in absence of glucose, they then achieved decoupling of biomass and compound production, maximizing product formation. Jers et al. reviewed the state-of-the-art in 3-HP production *via* Metabolic Engineering. The review highlights how major improvements have been reached but that efforts are still needed to achieve a scalable system, mainly relying on better biochemical characterization of the synthesis pathways and bioprocessing conditions. In addition, Frost et al. present an important case of cooperation between Synthetic Biology and Metabolic Engineering approaches in the recombinant production of hemoglobin. Authors summarize the state-of-the-art and the challenges that this technology is facing while suggesting novel routes to improve the promising yeast-based production of artificial blood.

Despite the great successes, there is still a wide number of organisms that could become valuable hosts for the production of important chemicals but that still suffer from the lack of properly standardized bio-parts for their easier engineering. A valid example on this respect comes from Rajkumar et al.. The authors developed a toolkit of standardized genetic parts, namely promoters and terminators, for the genetic engineering of a non-conventional yeast with high potential, *Kluyveromyces marxianus*. Parts were isolated from the genome, "domesticated" for golden gate assembly and then cloned for characterization under different conditions. Another example is

presented by Santo-Merino et al. where the authors summarize the latest efforts in the development of novel biological parts for engineering cyanobacteria.

One of the reasons why we still miss parts and tools for a more rapid engineering of many organisms is our limited knowledge on many of them. Together with collecting new information, we need progress on how to handle and use it. García-Granados et al. suggest that in the era of the "—omics" approaches we are in, mathematical modeling and biocomputing could serve as essential tools to coordinate and predict information, providing deeper and more accessible understanding that we could use for engineering target hosts. In this respect, Goñi-Moreno and Nikel advocate the *in silico* conceptualization of metabolism and its motifs as the way forward to achieve *whole-cell biocomputations*. The authors argue that the design of merged transcriptional and metabolic circuits will not only increase the amount and type of information being processed by a given synthetic construct, but that it will also provide fundamental control mechanisms for increased reliability on synthetic implants.

Not only handling information is challenging, but also making sure that this information is accurate and the data we collect are robust. Jessop-Fabre and Sonnenschein give an overview of the efforts currently in place toward a more reproducible science. Naming novel developed bio-foundries and cloud-laboratories together with new effort for the standardization of protocols and results, the authors suggest the way Synthetic Biology should continue to invest in to achieve more reliable data acquisition. One example comes from Nadler et al.. Here, the authors develop a novel system, named CopySwitch, that allows rapid transformation and copy number modulation of any genetic construct in an easier way than previously possible in *Bacillus subtilis*.

Again, López et al. considered two different *S. cerevisiae* strains engineered for β-carotene production comparing their growth in shake-flasks and in bench-scale fed-batch fermentation (as a proxy for industrial bioprocessing conditions). Surprisingly, the two strains showed opposite behavior, highlighting the need for proper understanding of how engineered systems behave in the final, industrially relevant settings.

An important source of information and characterization is represented by transcription and translation systems (TX-TL). In their contribution, Koch et al. describe how TX-TLs offer powerful tools for the prototyping of genetic constructs and understanding of network behavior out of the cellular context. This could be of great importance in Metabolic Engineering for the modeling and prediction of the behavior of entire pathways so to optimize them prior *in vivo* engineering. However, preparation of TX-TL extracts is still challenging, and automation tools could help improve reproducibility. Combination of TX-TL technology and liposome technology has recently gained momentum, and many successful examples in this emerging field are based on genetic constructs that can be expressed in a "simplified cell" able to interact with living cells and communicate with them. On this premise, Rampioni et al. present the latest advances in synthetic cells development. The possible applications of such a technology are many, as many are the opportunities to use it for better understanding biological systems.

In all, the works collected in this Research Topic expose that we are experiencing critical times for Synthetic Biology-guided Metabolic Engineering, where deeper and better understanding of the complexity of biological systems enables more predictable designs. Walking this pathway will significantly improve the applicability of first-design principles for living organisms toward reliable cell factory design, construction and testing.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

Glucose-Dependent Promoters for Dynamic Regulation of Metabolic Pathways

Jérôme Maury[1], Soumya Kannan[2], Niels B. Jensen[1†], Fredrik K. Öberg[1†], Kanchana R. Kildegaard[1], Jochen Forster[1†], Jens Nielsen[1], Christopher T. Workman[2*] and Irina Borodina[1*]

[1] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark, [2] Department of Biotechnology and Biomedicine, Technical University of Denmark, Lyngby, Denmark

For an industrial fermentation process, it can be advantageous to decouple cell growth from product formation. This decoupling would allow for the rapid accumulation of biomass without inhibition from product formation, after which the fermentation can be switched to a mode where cells would grow minimally and primarily act as catalysts to convert substrate into desired product. The switch in fermentation mode should preferably be accomplished without the addition of expensive inducers. A common cell factory *Saccharomyces cerevisiae* is a Crabtree-positive yeast and is typically fermented at industrial scale under glucose-limited conditions to avoid the formation of ethanol. In this work, we aimed to identify and characterize promoters that depend on glucose concentration for use as dynamic control elements. Through analysis of mRNA data of *S. cerevisiae* grown in chemostats under glucose excess or limitation, we identified 34 candidate promoters that strongly responded to glucose presence or absence. These promoters were characterized in small-scale batch and fed-batch cultivations using a quickly maturing rapidly degrading green fluorescent protein yEGFP3-Cln2$_{PEST}$ as a reporter. Expressing 3-hydroxypropionic acid (3HP) pathway from a set of selected regulated promoters allowed for suppression of 3HP production during glucose-excess phase of a batch cultivation with subsequent activation in glucose-limiting conditions. Regulating the 3HP pathway by the *ICL1* promoter resulted in 70% improvement of 3HP titer in comparison to *PGK1* promoter.

Keywords: inducible promoters, dynamic regulation, yeast *Saccharomyces cerevisiae*, strain engineering, 3-hydroxypropionic acid

INTRODUCTION

Budding yeast, *Saccharomyces cerevisiae*, is widely used for the production of fuels, chemicals, food ingredients, food and beverages, and pharmaceuticals (Nielsen and Jewett, 2008; Hong and Nielsen, 2012). *S. cerevisiae* is further being exploited for production of many new products, e.g., farnesene, butanol, resveratrol or melatonin (Hong and Nielsen, 2012; Shin et al., 2012; Germann et al., 2016; Meadows et al., 2016) and many more are under development. The enabling technology in the development of yeast cell factories is metabolic engineering, which is the introduction of directed genetic modifications with the objective to improve the performance of the cell factory (Nielsen and Jewett, 2008). Development of cell factories that can meet the high requirements for yield, titer

and productivity, generally requires multiple rounds of metabolic engineering. In connection with implementation of cell factories for industrial production, it is necessary to scale-up the process, and the final process will typically involve subsequent rounds of cultivation at different conditions (**Figure 1**). A typical process could comprise a large-scale main cultivation in a so-called continuous or fed-batch mode in which at least one substrate is growth limiting, such as in the resveratrol process (Shin et al., 2012). Usually, a number of serial cultivations in batch mode precede the main cultivation step in order to propagate the inoculum. In principle, serial cultivation steps aim at building up biomass while for the final-stage cultivation it is desirable to re-direct the metabolic fluxes such that the product of interest is formed at high yield and rate (Borodina and Nielsen, 2014) (**Figure 1**). Constraints are therefore very different between the seed train, which prioritizes strain stability, efficient biomass production, process convenience and robustness, and the main cultivation step, often referred to as the "production phase," where maximal yield and product formation rate are ultimate goals. The bioprocess is therefore imposing a number of constraints on strain design that need to be taken into account early in process development.

In order to better match expression of metabolic pathways to changing conditions in the cells or in the fermentation process, a number of research groups have recently developed strategies for dynamic regulation. As opposed to static up- or down-regulations, dynamic regulation allows for re-balancing fluxes along with changing conditions. Dynamic gene expression profiles allow trade-offs between growth and production to be better managed and can help avoid build-up of undesired intermediates (Brockman and Prather, 2015b). In their review, Brockman and Prather (2015b) provide experimental demonstration of the benefits arising from dynamic control on production of a number of chemicals. Farmer and Liao improved the yield of lycopene production by 18-fold in *E. coli* by implementing a sensor responding to the build-up of acetyl-phosphate (Farmer and Liao, 2000). By modulating protein levels of glucokinase via a genetic inverter, Solomon et al. redirected glucose into gluconate production and improved titers by 30% (Solomon et al., 2012). Using a metabolic toggle switch in *E. coli* to conditionally downregulate citrate synthase expression, Soma et al. redirected acetyl-CoA into isopropanol production with an improvement on titer and yield of more than 3-fold (Soma et al., 2014). Dahl et al. used whole-genome transcript arrays to identify promoters that respond to the accumulation of a toxic intermediate of the isoprenoid biosynthetic pathway, i.e., farnesyl di-phosphate (FPP), and used the identified promoters to regulate FPP formation and improve amorphadiene production two-fold over constitutive or inducible promoters (Dahl et al., 2013). Lo et al. presented a synthetic gene circuit that decouples *E. coli* cell growth from metabolite production through a two-layered circuit based on two genetic sensor controller modules:

FIGURE 1 | Schematic view of an optimized industrial process. The top panel illustrates the progression of fermentation settings from the seed train to full scale production. The lower panel illustrates the corresponding metabolic fluxes and the switch from biomass generation to product formation.
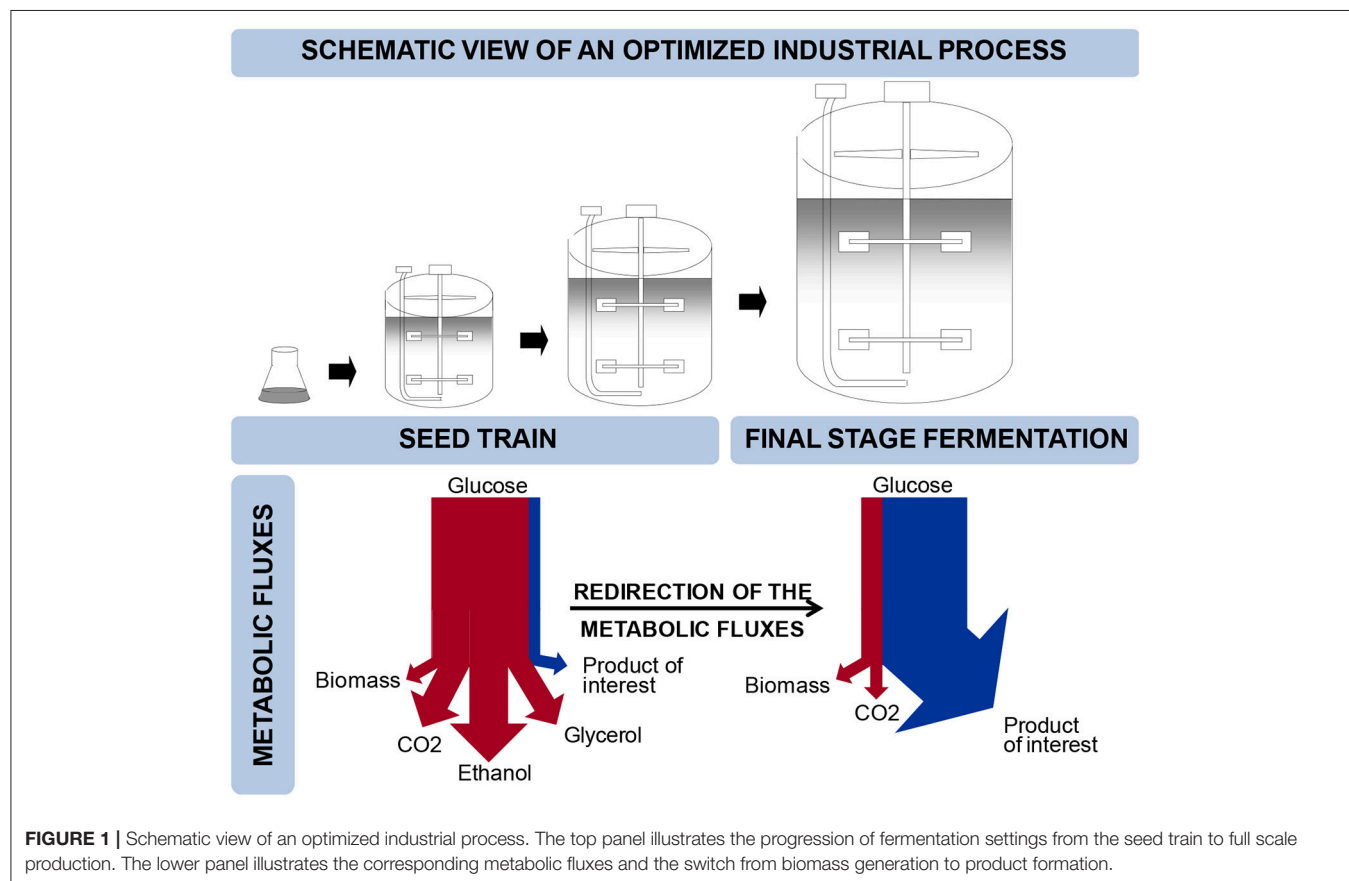
one sensing feedstock substrate, the other one sensing nutrients (Lo et al., 2016). By delaying enzyme expression until cells have depleted key nutrients and attained high cell densities, the system positively impacted conversion of hydroxycinnamic acids and oleic acid into value-added compounds (Lo et al., 2016). Conditional degradation of essential enzymes has also been exploited to stop elongation of fatty acids and improve octanoate production (Torella et al., 2013), or to increase yield and titer of myo-inositol production (Brockman and Prather, 2015a). Conditional degradation of enzymes is an efficient method for rapidly depleting an enzyme of interest even at slow growth rates where removal via dilution is slow (Brockman and Prather, 2015b).

Dynamic regulation has also been employed in a number of cases in budding yeast. On/off regulation systems like those making use of native repressible promoters like *MET3* promoter, responding to the concentration of methionine, or *HXT1* promoter reacting to glucose concentration were used to redirect the flux in the ergosterol biosynthetic pathway and improve production of isoprenoids of commercial interest (Ro et al., 2006; Asadollahi et al., 2008; Scalcinati et al., 2012). David et al. demonstrated the benefits of hierarchical dynamic control on 3-hydroxypropionic acid (3HP) production (David et al., 2016). Here, a two-layered control with a growth stage control, based on a carbon source responsive promoter, and an intracellular metabolite concentration control, based on a malonyl-CoA sensor, allowed for a sufficient build-up of biomass in the initial growth phase while gradually redirecting metabolism toward 3HP production in the production phase (David et al., 2016). David et al. report a 10-fold increase in 3HP production (David et al., 2016). More recently, dynamic control was used to establish the production of very long chain fatty acid-derived (VLCFA) compounds in *S. cerevisiae*. Production suffered from impaired biomass formation caused by deprivation of essential precursors C_{22}-CoA (Yu et al., 2017). Dynamic control, based on carbon-source regulated promoters, was used to relieve the competition between VLCFA product formation and cell-growth associated processes. This strategy divided the process in a production phase and a cell growth phase and resulted in an increase of docosanol production of almost 4-fold. After further optimization, a titer of 83.5 mg/L was achieved which represents a 80-fold improvement compared to the control strain (Yu et al., 2017).

The aim of our study is to increase the number of genetic elements that can be used by metabolic engineers for dynamic control of metabolic pathways. As industrial fermentation processes often comprise different phases (**Figure 1**), we embarked on identifying and characterizing a set of native yeast promoters that can be used to control expression of metabolic pathways in a dynamic manner, and will positively impact production of value-added chemicals in an industrially relevant context. In order for our tools to be potentially applicable to various kinds of industrial production processes, from bulk to high value-added products, it was decided to use elements already present in the cultivation medium to drive transient expression of genes. Amongst others, glucose is a nutrient often used as a carbon source. Furthermore concentrations of glucose largely differ in the seed train, usually operated as batch, and the final

stage fermentation where glucose is often fed in limiting amounts to avoid overflow metabolism. Glucose was thus identified as the element that will trigger the dynamic regulation of gene expression in our system.

Through analysis of genome-wide transcription datasets retrieved from the Gene Expression Omnibus (GEO) database (GEO database), a number of promoters were identified, cloned upstream of a quickly maturing, rapidly degrading green fluorescent protein yEGFP3-Cln2$_{PEST}$ and integrated into *S. cerevisiae* genome. Promoter activity was then characterized in microscale batch and fed-batch conditions. As a proof of the concept, the most promising promoters for transient gene expression were successfully applied to control the production of 3-hydroxypropionic acid.

MATERIALS AND METHODS

Analysis of Transcriptome Data

Transcription data was retrieved from the GEO database, using the R package GEOquery (Davis and Meltzer, 2007), for a microarray study comparing different nutrient limited chemostat cultivations of CEN.PK113-7D (accession GDS777) (Tai et al., 2005). The summarized expression values provided by the authors were log2 transformed before applying a linear model (LIMMA) to identify differentially expressed genes between the glucose-limited experiments and the other nutrient limited experiments for the aerobic series of chemostat cultivations. The differential expression profiles (log2-fold-change of glucose-limited vs. others) for *HXT1*, *TEF1*, *ADH1*, *MAL12* genes were used to find other correlated genes using the Pearson correlation coefficient and test for significance.

Cloning of the Selected Promoters Upstream of a Reporter Gene: yEGFP3-Cln2$_{PEST}$ and Integration Into *S. cerevisiae* Genome

Primers for PCR amplification and Uracil excision mediated cloning of the different promoters into the integrative vector pCFB0125 were designed and are listed as Table S1. As the length of promoter regions is not always well defined, the promoter sequence length was either defined based on previous reports or as starting approximately 1,000 bp upstream of the start codon (ATG) when no reference in the literature could be identified. In order to closely follow promoter activity, promoters were cloned upstream of a quickly maturing and rapidly degrading green fluorescent protein encoding gene: yEGFP3-Cln2$_{PEST}$ (Mateus and Avery, 2000). yEGFP3-Cln2$_{PEST}$, as reporter gene, was amplified using the following primers: uGFP_fw: ATCTGTCAU AAA ACA ATGTCTAAAGGTGAA GAATTATTC and uGFP_rv: CACGCGAUTCATATTACTTGG GTATTGCCC) with pCFB0058 (pFA6a-yEGFP3-CLN2-PEST-natMX6, EUROSCARF). Uracil excision was used as cloning method to combine the different promoter-yEGFP3-Cln2$_{PEST}$ expression cassettes together onto pCFB0125. pCFB0125 is a pESC-URA-ccdB vector devoid of its 2micron replication origin, which makes it a non replicative vector that will integrate at

the *URA3* locus upon linearization. Standard Uracil excision cloning conditions were used, as described previously (Jensen et al., 2014). After verification of successful cloning, each and every vector was linearized by digestion with restriction enzyme StuI (FastDigest®, Thermo Scientific) or EcoRV (FastDigest®, Thermo Scientific) in the case of the vector bearing promoter p*MAL32*. Digested, linearized, vectors were transformed into *S. cerevisiae* CEN.PK 113-5D. Transformants were selected on agar plates containing SC-Ura medium and were used for cultivation experiments.

Construction of 3-hydroxypropionic Acid *S. cerevisiae* Producing Strains

All strains from this study were derived from *S. cerevisiae* CEN.PK102-5B MATa ura3-52 his31 leu2-3/112 MAL2-8c SUC2 obtained from Peter Kötter (Johann Wolfgang Goethe-University Frankfurt, Germany). This strain was further modified by integration of single copy integrative vectors pCFB0380 (PTEF1-SeACS_L641P; PPGK1-ALD6; KlLEU2) and pCFB0382 (PTEF1-PDC1; SpHIS5). The latter strain is the parent strain for the 3HP producing strains of this study. Four Plasmids were constructed based on pCFB0343 (PTEF1-ACC1**; PPGK1/CaMCR; KlURA3; insertion site Easyclone X-2). More details about the Easyclone plasmid set and integration sites can be found at Jensen et al. (2014). Backbone vector fragment was prepared for USER cloning by PCR using primers vec_open_mid_ptrTEF1-PGK1 and vec_open_CaMCR and template pCFB0343. Three promoter containing fragments were prepared by PCR using *S. cerevisiae* genomic DNA as template and primers pHXT7_switch and pHXT7_rev for promoter p*HXT7*, pADH2_switch and pADH2_rev for promoter PADH2, and primers pICL1_switch and pICL1_rev for promoter p*ICL1*. Primer sequences can be found in **Table 1**.

By combining, under standard uracil exclusion cloning conditions (Jensen et al., 2014), backbone vector prepared as described above and insert PHXT7, PADH2 or PICL1, one obtains three single integrative vectors pCFB0728 (PTEF1-ACC1**; PHXT7-CaMCR; insertion site Easyclone X-2), pCFB0727 (PTEF1-ACC1**; PADH2-CaMCR; insertion site Easyclone X-2) and pCFB0729 (PTEF1-ACC1**; PICL1-CaMCR; insertion site Easyclone X-2), respectively. After transforming

strain *S. cerevisiae* CEN.PK102-5B MATa ura3-52 his31 leu2-3/112 MAL2-8c SUC2 X-3::PTEF1-SeACS_L641P; PPGK1-ALD6; KlLEU2 X-4::PTEF1-PDC1; SpHIS5 with integrative vector pCFB0343, pCFB0728, pCFB0727, or pCFB0729, 4 strains are obtained that bear pPGK1-CaMCR, pHXT7-CaMCR, pADH2-CaMCR or pICL1-CaMCR, at integration site X-2 respectively.

Cultivation

Strains bearing the different promoter-yEGFP3-Cln2$_{PEST}$ expression cassettes were cultivated in two different conditions: (1) in batch mode with mineral medium containing glucose as sole carbon source (Jensen et al., 2014); (2) in fed-batch mode with synthetic fed-batch medium for *S. cerevisiae* M-Sc.syn-1000 purchased from m2p-labs GmbH (Baesweiler, Germany). M-Sc.syn-1000 medium was supplemented with the supplied vitamins solution (final 1% v/v) and the enzyme mix (final concentration 0.5% v/v) immediately prior to use.

Mineral medium contained (L^{-1}): 7.5 g (NH4)$_2$SO$_4$, 14.4 g KH$_2$PO$_4$, 0.5 g MgSO$_4$-7H$_2$O, 22 g dextrose, 2 mL trace metals solution, and 1 mL vitamins. The pH of the medium was adjusted to 6 prior to autoclavation. Vitamin solution was added to the medium after autoclavation. Trace metals solution was added after autoclavation. Trace metals solution contained (L^{-1}): 4.5 g CaCl$_2$-2H$_2$O, 4.5 g ZnSO$_4$-7H$_2$O, 3 g FeSO$_4$-7H$_2$O, 1 g H$_3$BO$_3$, 1 g MnCl$_2$-4H$_2$O, 0.4 g Na$_2$MoO$_4$-2H$_2$O, 0.3 g CoCl$_2$-6H$_2$O, 0.1 g CuSO$_4$-5H$_2$O, 0.1 g KI, 15 g EDTA. The trace metals solution was prepared by dissolving all the components except EDTA in 900 mL ultrapure water at pH 6. The solution was then gently heated and EDTA was added. In the end, the pH was adjusted to 4, and the solution volume was adjusted to 1 L and autoclaved (121°C in 20 min). This solution was stored at + 4°C. Vitamin solution contained (L-1): 50 mg biotin, 200 mg p-aminobenzoic acid, 1 g nicotinic acid, 1 g Capantothenate, 1 g pyridoxine-HCl, 1 g thiamine-HCl, 25 g myo-inositol. Biotin was dissolved in 20 mL 0.1 M NaOH and 900 mL water is added. pH was adjusted to 6.5 with HCl and the rest of the vitamins were added. pH was re-adjusted to 6.5 just before and after adding m-inositol. The final volume was adjusted to 1 L, sterile-filtered and stored at + 4°C.

Composition of M-Sc.syn-1000, vitamin solution and enzyme mix are proprietary and were not disclosed to us.

The main difference between the mineral medium and the synthetic M-Sc.syn-1000 used here deals with the carbon source that each contains. In mineral medium used for batch cultivations, glucose is present as at a starting concentration of 20g.L^{-1} while synthetic medium M-Sc.syn-1000, used for fed-batch, contains a solubilized glucose polymer which, due to the activity of glucose releasing enzymes, allows for the slow release of glucose monomers and limits biomass production.

All cultivations were performed at 30°C with 1,000 rpm in the microbioreactor system BioLector (m2p-labs). Forty eight cultivations were run in parallel in FlowerPlates (m2p-labs) with a working volume of 1.1 mL per well. Biomass growth (light scattering units, LSU) and fluorescence (relative fluorescence units, RFU) were monitored online approximately every 12 min.

TABLE 1 | DNA sequences of PCR primers for USER cloning.

Promoter	Primer
vec_open_mid ptrTEF1-PGK1	ATGCGAUGCACACACCATAGCTTCAAAATG
vec_open_CaMCR	ATCTGTCAUAAAACAATGAGTGGTACAGGTAGATTAGC
pHXT7_switch	ATCGCAUCCGTGGAAATGAGGGGTATGC
pHXT7_rev	ATGACAGAUTTTTTGATTAAAATTAAAAAAACTTTTTGTTTTTG
pADH2_switch	ATCGCAUCGCAGGCGGGAAACCATCCA
pADH2_rev	ATGACAGAUTGTGTATTACGATATAGTTAATAGTTGATAGTTG
pICL1_switch	ATCGCAUTTCCATTCATCCGAGCGATCAC
pICL1_rev	ATGACAGAUTTTTCGTTGACTTTTTGTTATGTTATGC

Outlier Removal and Data Correction

Some cultivations were removed before further analysis due to data quality or failure of the cultivation to proceed normally.

Background Correction and Alignment of Measurements

To correct the biomass measurements for background signal, we calculated the growth rates μ via two different methods and compared them for various possible backgrounds to determine the optimal background value. The first method is derived from Poulsen et al. (2003), and is known as the log of slope (LOS) method.

Poulsen et al. (2003) note that experimental data that can be described by an exponential function sometimes contains an offset, giving a function of the form

$$b = b_0 \cdot exp(\mu t) + c \tag{1}$$

Where b is biomass and t is time, μ is growth rate, c is the background, b_0 is the initial value of of b at $t = 0$. In order to remove the background and calculate μ, take the derivative with respect to t and take the natural log to get

$$log\left(\frac{db}{dt}\right) = log(b_0\mu) + \mu t \tag{2}$$

Thus, the slope of any rectilinear section of the LOS plot will be the specific growth rate μ.

Here, we extend this past what is described by Poulsen et al. (2003) by taking the derivative of Equation (2) with respect to time and further simplifying with the chain rule to get:

$$\mu = \frac{d}{dt}\left(log\left(\frac{db}{dt}\right)\right) = \frac{d^2b}{dt} \bigg/ \frac{db}{dt} \tag{3}$$

Thus, we see that if we have a function describing the data, assuming an underlying exponential dependence on t, dividing the second derivative of this function by the first derivative will give us μ. This expression for μ is not dependent on the magnitude of b and is thus independent of background.

The second method of calculating growth rate is simply the definition of a growth rate normalized to population size:

$$\mu = \frac{db/dt}{b} \tag{4}$$

This method is commonly applied for calculating growth rate of cell populations (Ronen et al., 2002; De Jong et al., 2010; Rudge et al., 2016). As this depends directly on b, it is thus dependent on background.

It was observed that the biomass profiles of batch cultures exhibit a slight peak at the diauxic shift (Altintas et al., 2016; Figure 3B). For each biomass profile, the time of the peak was identified by determining the time at which light scattering began decreasing. If no peak was found, the peak was artificially set to 10 h for batch cultures and 33 h for fed-batch cultures. A window of 7.5 to 0.5 h before the diauxic shift peak for batch cultures and 20.5 to 1.5 h before the diauxic shift peak for

fed-batch cultures was then used to calculate growth rates via each of the two methods described above in Equations (3) and (4). These time windows were chosen as the growth curves appeared approximately exponential in these regions, which is an assumption of the LOS method. Derivatives were computed numerically from a smoothing spline fit to the data. The first and last 6 points of the calculated growth rates were trimmed to account for edge effects of the smoothing spline and the sum of squared errors (SSE) between the remaining growth rate profiles were compared. The growth rate calculation via the Equation (4) was repeated for a range of background values from 0 to 13 for batch cultures and 0 to 10 for fed-batch and compared for each value to growth rate calculated via Equation (3). The background value giving the minimum SSE for each biomass profile was chosen.

To correct for background fluorescence, we used fluorescence measurements from a control strain (p0125) cultivated under the same conditions as experimental strains. These measurements were used to define a function relating control fluorescence (autofluorescence) as a function of biomass, which we will call $f_c(b)$. For each experimental strain, each biomass measurement was input into $f_c(b)$ to obtain the autofluorescence for that amount of biomass. This autofluorescence was then subtracted from the corresponding experimental fluorescence associated with that biomass measurement. This process was carried out for all biomass/fluorescence measurements in each experimental strain, resulting in a background-corrected fluorescence profile for each replicate.

In order to compare replicates we aligned the growth curves such that landmark events occur at the same time. This is necessary as the lag phase may last different amounts of time for different cultivation experiments even when the pattern of biomass accumulation is the same after entry to exponential phase. In this case we are interested in the diauxic shift so we used this as the landmark with which to align the growth curves. As previously described we identified the diauxic shift peak and determined the difference between this landmark and a reference time, which we set to be at 15 h for batch cultures and 30 h for fed-batch cultures. This difference was subtracted from the entire set of times corresponding to both biomass and fluorescence measurements to determine the corrected times; moving forward, we only use data from corrected times greater than 0. A similar procedure for growth curve alignment is discussed by van Ditmarsch and Xavier (2011).

Promoter Activities in Small-Scale Batch Cultivations

Background-corrected biomass and fluorescence profiles were input into the PromAct model (Kannan et al., 2018) using a protein maturation rate of 0.45 h^{-1} as described for EGFP (Heim et al., 1995) and a protein degradation rate of 0.5 h^{-1} as described for yEGFP3-Cln2$_{PEST}$ (Mateus and Avery, 2000). mRNA degradation rate was set as $\frac{1}{3}$ h^{-1} based on genome-wide estimates of mRNA degradation in *S. cerevisiae* (Munchel et al., 2011; Neymotin et al., 2014).

The promoter activity profiles were cropped to 35 h for batch and 60 h for fed-batch cultures on the corrected time scale for standardization. Lag phase was identified as when the rate of change of the average biomass profile over all replicates from a given promoter first surpassed 0.5 LSU per h. The diauxic shift was identified as the time between the characteristic peak, identified as where the derivative of the average biomass profile over all replicates becomes crosses 0, to when the derivative next surpasses 0.5 LSU per h following this peak. For batch cultures, the stationary phase was also identified using the same peak detection method as with the diauxic shift identification, as a similar characteristic peak is seen. A similar method for growth phase identification was used by Altintas et al. (2016). "Early" and "Late" designations in **Figure 4** indicate the first and second half of each growth phase, respectively.

Measurement of 3-hydroxipropionic Acid Concentration

Strains expressing CaMCR (malonyl-CoA reductase from *Chloroflexus aurantiacus*) for the production of 3HP were cultivated in FlowerPlates in the BioLector, in the same conditions as described above, in batch and fedbatch modes. Samples for 3HP measurement were taken during the course of the cultivations, 3HP concentration was measured using a dedicated enzyme assay. The enzyme assay was performed as described in the supplementary materials of Borodina et al. (2015).

RESULTS

Selection of Promoters Through Analysis of Genome-Wide Transcription Datasets

In order to identify promoters with glucose-dependent expression profiles, we searched the literature for genome-wide expression datasets comparing glucose-limited and glucose excess conditions. Boer et al. (2003) characterized the specific transcriptional responses of *S. cerevisiae* cells growing at steady state in chemostats under growth limitation by carbon, nitrogen, phosphorus, or sulfur. We performed an analysis of this dataset in order to directly compare gene expression levels in the glucose limiting condition to the average of expression levels observed while not glucose limited (e.g., combining the results from the sulfur, the phosphorus and the nitrogen limiting conditions) which we describe as "glucose excess" conditions.

The 2D histogram (**Figure 2B**) displays the comparison of gene expression levels between glucose-limited and glucose excess conditions. The volcano plot (**Figure 2A**) displays the genes found significantly differently expressed between glucose-limited and glucose excess conditions. Four archetypes of gene expression were identified for their known roles and are represented in **Figure 2**: (1) *TEF1* archetype with constitutive expression in both conditions, (2) *HXT1* archetype with high expression in glucose-excess and low expression in glucose-limiting conditions, (3) *ADH2*, and (4) *MAL12* archetypes with low expression in glucose-excess and high expression in glucose limiting conditions. As some of the archetypes are

not differentially expressed in this comparison, e.g., *TEF1*, the identified genes correlated with them are most often not differentially expressed in this comparison. By contrast, genes significantly correlated to *ADH2* and *HXT1* were among the most differentially expressed in glucose-excess vs. glucose limiting conditions.

From the archetype correlation analysis, 34 genes were identified with diverse expression profiles in respect to dynamic regulation of gene expression. Promoter regions from all 34 genes were cloned (or attempted) as reported in the materials and methods section. Although not all were expected to be significantly differentially expressed, 20 genes were found significantly up-regulated in glucose limiting conditions and 5 genes were significantly down-regulated in glucose excess conditions ($p < 0.05$ after Benjamini-Hochberg correction).

Characterization of Promoter Activities in Microfermentations

Microscale Batch Cultivations

Batch cultivations were followed until stationary phase was reached. As presented in **Figure 3**, the two classical growth phases of a *S. cerevisiae* batch fermentation, where glucose is the sole carbon source, were clearly observed with a glucose growth phase followed by an ethanol growth phase separated by a diauxic shift (Monod, 1941). A few typical profiles of promoter activity are presented in **Figure 3**: constitutive expression, i.e., *TEF1*, high expression in glucose-excess conditions, i.e., *PGK1* and *TDH3*, activation at low glucose concentration, i.e., *HXK1*, *HXT7*, and *MAL12* or expression when growing on ethanol in the absence of glucose, i.e., *ADH2*, *ICL1*, and *ACS1*. None of the promoters tested showed sustained activity in stationary phase, likely due to the lack of metabolic activity in the absence of carbon sources. Promoter activities of other promoters tested can be found in Figure S1.

A heat map and cluster analysis were applied to all promoter activities in the batch condition (**Figure 4**). The cluster analysis reveals a tight correlation between promoter activities experimentally determined here and the gene expression categories defined from the analysis of genome-wide transcription data presented in **Figure 2**, which reflects that the different promoters cloned upstream of yEGFP3-Cln2$_{PEST}$ respond mostly in a similar fashion as when they are in their native context. Globally, two main branches appear, at the top of the figure promoters active in glucose excess conditions, and at the bottom can be observed promoters activating at low glucose concentration or when glucose is absent (**Figure 4**). This cluster analysis grouped profiles by archetype group in all but 6 cases where cluster grouping seemed to contradict with the gene expression data. This was observed for the promoters of *ISF1, FDH1, FOX2, RGI2, YIG1*, and *EGO4* (**Figure 4**). In the case of *EGO4*, (*RGI2 YIL057C*), and *FOX2*, this may be explained by their unclear promoter activity patterns (Figure S1). The three others, *ISF1, FDH1*, and *YIG1*, clearly behave in opposite fashion compared to what was expected from the gene expression analysis (**Figure 2**).

FIGURE 2 | Comparison of expression levels between glucose excess and glucose limiting cultivation conditions. **(A)** Volcano plot with a 2D histogram of gene counts, **(B)** plot of expression levels in glucose limiting vs. glucose excess conditions with a 2D histogram of gene counts. The four gene expression archetypes identified are represented: (1) *TEF1* archetype (red) with constitutive expression in both conditions, (2) *HXT1* archetype (green) with high expression in glucose-excess and low expression in glucose-limiting conditions, (3) *ADH2* (light blue), and (4) *MAL12* (orange) archetypes with low expression in glucose-excess and high expression in glucose limiting conditions.

Microscale Fed-Batch Cultivations

In this microscale fed-batch cultivation, realized using the glucose slow release technology Feed-In-Time (m2p-labs), two main growth phases are observed: (1) an initial phase until approximately 25 h of cultivation where residual glucose is still present at a low level (2) a glucose-limiting growth phase after 25 h of cultivation where biomass growth is limited by the slow release of glucose (**Figure 5**).

In terms of promoter activity, two major activity profiles appear: (1) promoters characterized by a higher and sustained activity in the first growth phase that remain active in the second phase, although to a lower extent (e.g., *TEF1*, *PGK1*, *TDH3*, *HXT7*, *MAL12*); and (2) promoters that strongly activate at the onset of glucose limiting phase and that quickly become at most weakly active or inactive (*ADH2*, *ACS1*, *ICL1*).

Based on the outcome of promoter activity characterization, it was decided to run a proof of concept study and test a subset of promoters for transient expression and controlled production of a commercially relevant compound, with the objective of restricting its production only in "production" phase in fed-batch and not during the seed train phases carried on as batch.

Proof of Concept Study: Dynamic Regulation of 3-hydroxypropionic Acid (3HP) Production

3HP is a platform chemical that can be converted into acrylic acid, 1,3-propanediol, malonic acid, and other valuable chemicals. In 2011, the world annual production of acrylic acid was 5000 kMT and the market size was USD 11.5 billion (Borodina et al., 2015). Acrylic acid-derived products include superabsorbent polymers used in diapers and incontinence products, plastics, coatings, adhesives, elastomers, and paints (Borodina et al., 2015). In biological systems, 3HP can be synthesized via at least four different intermediates: glycerol, lactate, malonyl-CoA or β-alanine (Kumar et al., 2013). Production of 3HP based on the β-alanine or the malonyl-CoA have been reported in *S. cerevisiae* (Chen et al., 2014; Borodina et al., 2015; Kildegaard et al., 2016). Furthermore, Kildegaard et al. reported a reduced growth rate of the strains expressing the malonyl-CoA route for 3HP production, especially in the case where pathway genes were integrated in multiple copies (Kildegaard et al., 2016).

In an attempt to control production of 3HP via the malonyl-CoA route, with the objective to drive 3HP production only in "production" phase in fed-batch and not during the seed train phases carried on as batch, three promoters were selected based on their expression profiles: p*HXT7*, p*ADH2*, and p*ICL1*. They were selected for their weak (if not absent) expression in cultivation with excess of glucose and their induction in glucose limiting conditions (**Figure 3**). Promoters of *HXT7*, *ADH2*, and *ICL1* were then cloned upstream of the gene encoding bi-functional malonyl-CoA reductase from *Chloroflexus aurantiacus* (CaMCR), which is the only heterologous enzyme necessary for producing 3HP via the malonyl-CoA route in *S. cerevisiae*. As a reference, CaMCR was expressed under the control of p*PGK1*, a promoter which is both active in glucose excess and glucose limiting conditions (**Figures 3**, **5**). When p*PGK1* is controlling the expression of CaMCR, a clear accumulation of 3HP is observed already at the early stages of growth on glucose in the batch condition (**Figure 6**). This is clearly not the case when the (glucose-) regulated promoters p*HXT7*, p*ADH2* and p*ICL1* are

FIGURE 3 | Activity of various promoters in batch cultivations. Gray shaded areas are low-growth phases. Red shading represents one standard deviation from the mean promoter activity over all replicates at each time point (see Table S2). Blue shading represents one standard deviation from the mean biomass value over all replicates at each time point.

controlling CaMCR expression, as here production of 3HP remains below detection limit of the assay (**Figure 6**). In the fed-batch cultivation condition, where all promoters are being activated, 3HP production is restored. Expressing CaMCR from the regulated *ICL1* promoter resulted in a 1.7-fold increase in 3HP production when compared to the constitutive expression from *PGK1* promoter (**Figure 6**).

It is clearly demonstrated here that production of a metabolite of interest can be turned on and off by using a set of dynamically activated promoters. Furthermore, turning off 3HP production in growth phases of batch fermentation did not affect the final production titer in the fed-batch condition, and was

even beneficial, especially when p*ICL1* controlled expression of CaMCR.

DISCUSSION

In this study, specific genome-wide expression datasets were analyzed in order to identify genes characterized by different expression patterns in glucose excess as compared to glucose limiting conditions. Four archetypes of gene expression were defined. A number of genes with expression patterns qualified as interesting were selected and their promoters cloned upstream of a quickly maturing rapidly degrading GFP. Destabilized

FIGURE 4 | Comparison of promoter activities. Clustering of dynamic promoter activity profiles summarized by the average activity in 8 non-overlapping time windows representing the growth phases in batch cultivations. Time windows were defined as follows: Early lag 0.50–1.50, Late lag 1.50–3.00, Early glucose 3.00–9.25, Late glucose 9.25–15.5, Diauxic shift 15.5–17.0, Early ethanol 17.0–20.8, Late ethanol 20.8–24.6, Stationary 24.6–35.0.

GFP variants are better suited for the study of transient gene expression as native GFP variants have particularly long half-lives. In this study, it was decided to use yEGFP3-Cln2$_{PEST}$ which is a destabilized variant of GFP that has successfully been used to monitor dynamic changes in gene expression in yeast (Mateus and Avery, 2000). After characterizing promoter activities both in batch with glucose as sole carbon source and glucose-limiting fedbatch conditions, a hand full of promoters were chosen for a proof of concept study where controlled production of the commercially relevant chemical 3HP was demonstrated.

Comparing results from the genome-wide gene expression analysis and from the characterization of promoter activity in the batch cultivation conditions, it was interesting to notice the correlation between the archetypes defined for the genome-wide gene expression analysis (**Figure 2**) and the clusters

obtained for the promoter activity characterization (**Figure 4**). This confirmed that most of the promoters cloned upstream yEGFP3-Cln2$_{PEST}$ respond according to the archetype defined for the gene they relate to. It also indicated that elements responsible for transcriptional regulation of those genes are present in the promoter sequence used in this study. This validated the promoter sequences that were chosen here. Discrepancies were observed though for six genes. For three of them, *ISF1*, *FDH1*, and *YIG1*, the promoter that was cloned triggered an expression pattern opposite to what was reported for the genome-wide gene expression analysis. This may indicate that the chosen promoter sequence is lacking regulatory elements otherwise present in the native promoter, or that DNA structure at the integration site affects activity pattern of these promoters.

Taking a look at promoter activities measured during the microscale fedbatch cultivation, it appeared that while some promoters were only activated at the onset of glucose limitation (e.g., promoters of *ACS1, ADH2, ICL1, SFC1, SPG4*), others were characterized by a much wider expression time window, with an early activation (e.g., promoters of *HXT7, MAL12, FMP3*) (**Figure 5**). This may be explained by the fact that the latter promoters are active already in the phase where residual glucose is present but at a low concentration. This early phase of the microscale fedbatch experiment is an intermediary phase caused by the technology employed here, i.e., the Feed-In-Time technology as a low amount of free glucose is initially present in the medium and acts as a carbon source until the slow enzymatic release of glucose from the polymer takes over and triggers glucose limitation. The low concentration of glucose in this early phase may still allow for activation of the above-mentioned promoters (**Figure 5**).

Surprisingly, promoters like *ADH2*, *ICL1*, or *ACS1* were characterized by a burst of expression at the onset of the glucose limiting phase. These promoter activity profiles suggest a rapid increase and decrease in their activity although they appear to maintain some activity for roughly 10 h after activation. This was not necessarily expected as the corresponding genes were clearly maintaining expression in the glucose limited continuous cultivation of the genome-wide gene expression analysis (**Figure 2**). It further underlines the necessity for testing genetic elements for controlled expression in conditions as close as possible to the intended bioprocess.

Benefits of dynamic control of metabolic pathways have been demonstrated before, as described in the introduction. Nevertheless, this study further confirms that transient activation of a metabolic pathway does not negatively impact production of a compound of interest, here 3HP, but rather benefits production performance, as shown here in the case where *ICL1* promoter restricts expression of CaMCR to the microscale fedbatch cultivation stage (**Figure 6**). Even-though other factors, e.g., timing of MCR expression, substrate availability or pathway intermediate concentrations may also contribute to the benefits observed on 3HP production, it is surprising that such a sudden and short expression timespan triggered by *ICL1* promoter can result in superior production of the compound of interest
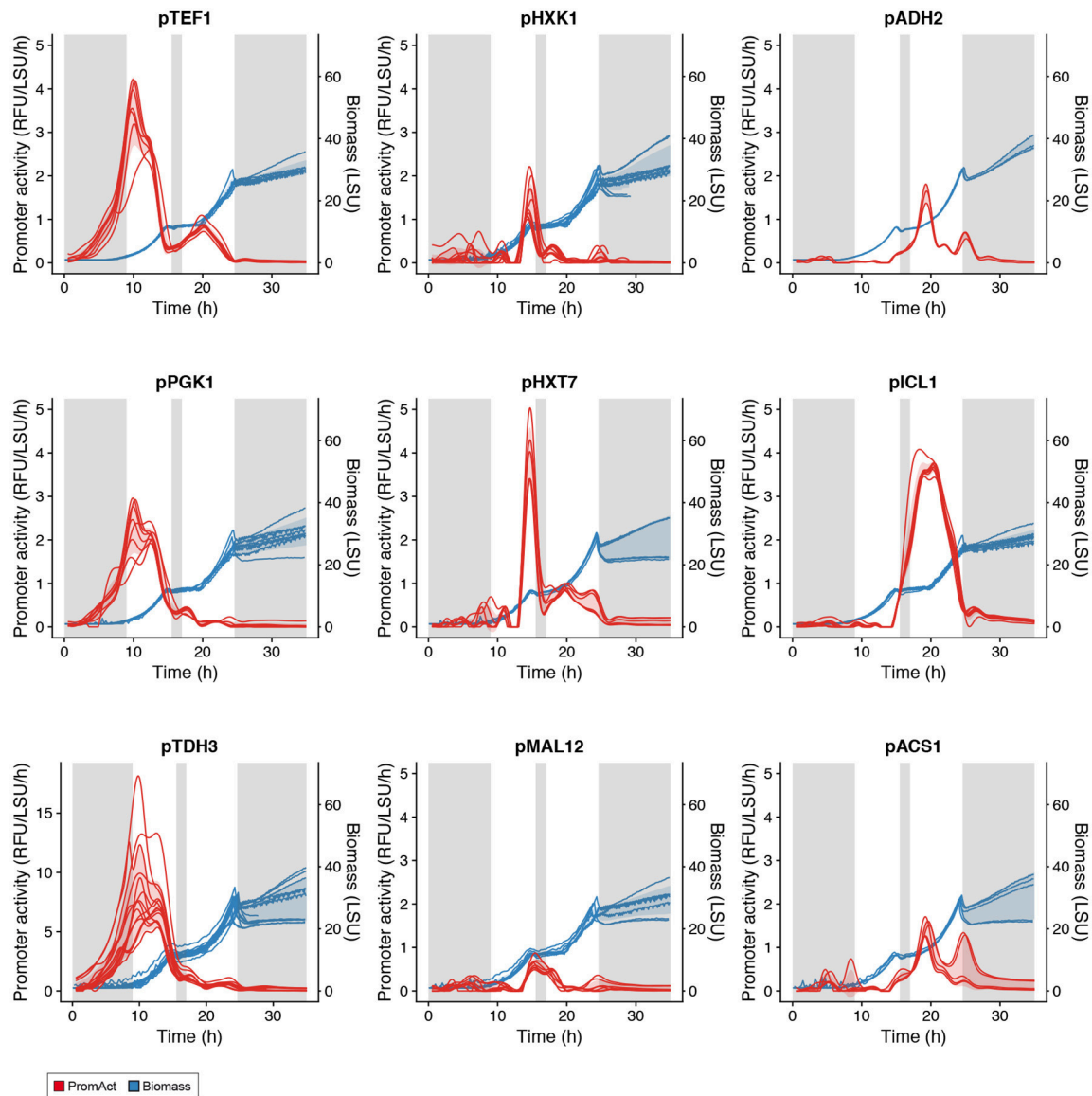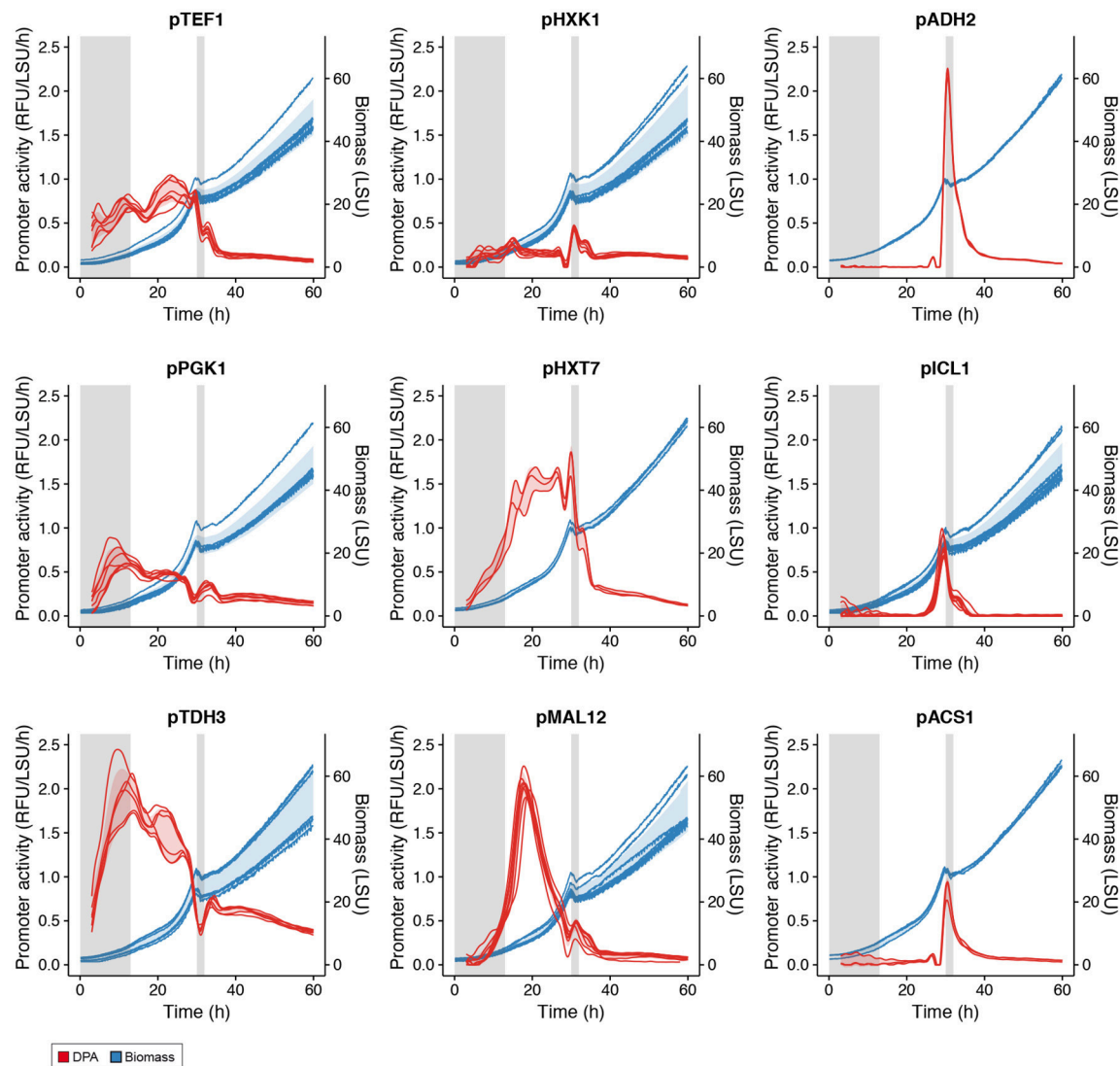
FIGURE 5 | Activity of various promoters in fed-batch cultivations. Gray shaded areas are low-growth phases. Red shading represents one standard deviation from the mean activity over all replicates (see Table S3). Blue shading represents one standard deviation from the mean biomass value over all replicates at each time point.
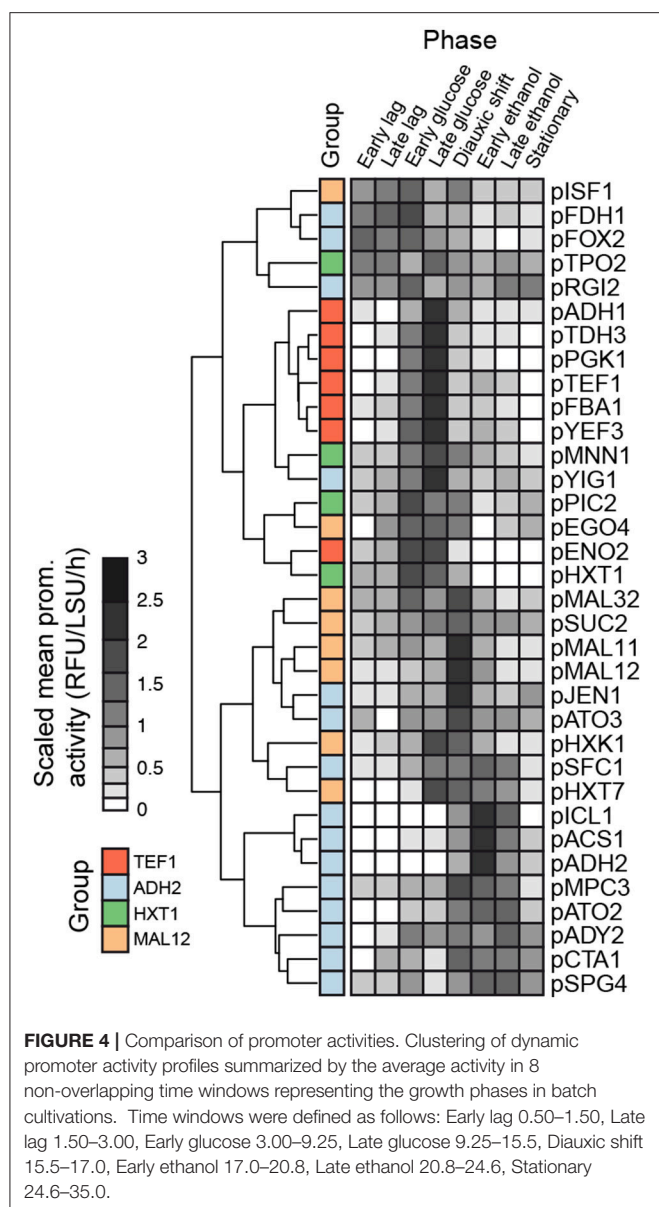
as compared to longer expression timespan as in the case of p*HXT7* or p*PGK1*. This certainly raises the question whether sudden, short but high expression bursts at the onset of changing conditions could not be sufficient to sustain efficient production of chemicals of interest. Reducing the timespan of expression of a promoter may be beneficial to the overall production process as it may restrict the loads of energy spent on transcription and translation of the specific gene only to a brief time-window. Recently Christiano et al. measured 8.8–12 h as median half lives of proteins in *S. cerevisiae* (Christiano et al., 2014). Therefore, once synthesized the functional protein would still remain in the cells even though the promoter of its gene, hence transcription and translation of that gene, would be off during most of the cultivation. Promoters like *ICL1*, *ADH2*, *ACS1*, combined with other synthetic biology tools that could expand mRNA

half-lives, e.g., sequences improving mRNA stability (Curran et al., 2013; Yamanishi et al., 2013), may certainly play an active role.

CONCLUSION

Through analysis of gene expression data, promoters showing specific responses to glucose excess or limitation were identified. DNA sequences of these promoters were cloned upstream of a quickly maturing and rapidly degrading green fluorescent protein to allow the study of promoter activities. Dynamics of promoter activities for a number of promoters in batch with glucose as sole carbon source, or glucose limiting conditions were analyzed. A subset of promoters, p*ADH2*, p*ICL1*, and p*HXT7*, were demonstrated as suitable for dynamic control

FIGURE 6 | Effect of dynamic control of gene expression on 3HP production. 3HP production in **A**: batch and **B**: fed-batch cultivations for pPGK1, pHXT7, pICL1, or pADH2 driving production of CaMCR. Error bars on 3HP levels represent standard error of the mean, $n = 3$. The gray ribbon on Biomass represents plus/minus one standard deviation. 3HP concentration is in g/L and Biomass is measured in light scattering units (LSU).

of production of a commercially relevant compound, i.e., 3HP, in *S. cerevisiae*. The set of promoters and the dynamic promoter activity profiling presented here will contribute to the molecular biology toolbox available to metabolic engineers.

AUTHOR CONTRIBUTIONS

IB and JM conceived the design of the study. JN contributed with initial ideas about the project and discussions on the possible promoter set for evaluation. JM, NJ, FÖ, and KK performed DNA work. JM generated the strains and performed small scale cultivations. JM and CW performed the transcriptome data analysis. JF participated in the design of the study and interpretation of initial results. SK and CW performed dynamic promoter activity analyses. SK, CW, and JM analyzed the data. SK, JM, IB, and CW drafted the manuscript and all authors contributed to preparing the final version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2018.00063/full#supplementary-material

REFERENCES

Altintas, A., Martini, J., Mortensen, U. H., and Workman, C. T. (2016). Quantification of oxidative stress phenotypes based on high-throughput growth profiling of protein kinase and phosphatase knockouts. *FEMS Yeast Res.* 16:fov101. doi: 10.1093/femsyr/fov101

Asadollahi, M. A., Maury, J., Møller, K., Nielsen, K. F., Schalk, M., Clark, A., et al. (2008). Production of plant sesquiterpenes in *Saccharomyces cerevisiae*: effect of ERG9 repression on sesquiterpene biosynthesis. *Biotechnol. Bioeng.* 99, 666–677. doi: 10.1002/bit.21581

Boer, V. M., De Winde, J. H., Pronk, J. T., and Piper, M. D. W. (2003). The genome-wide transcriptional responses of *Saccharomyces cerevisiae* grown on glucose in aerobic chemostat cultures limited for carbon, nitrogen, phosphorus, or sulfur. *J. Biol. Chem.* 278, 3265–3274. doi: 10.1074/jbc.M209759200

Borodina, I., Kildegaard, K. R., Jensen, N. B., Blicher, T. H., Maury, J., Sherstyk, S., et al. (2015). Establishing a synthetic pathway for high-level production of 3-hydroxypropionic acid in Saccharomyces cerevisiae via β-alanine. *Metab. Eng.* 27, 57–64. doi: 10.1016/j.ymben.2014.10.003

Borodina, I., and Nielsen, J. (2014). Advances in metabolic engineering of yeast *Saccharomyces cerevisiae* for production of chemicals. *Biotechnol. J.* 9, 609–620. doi: 10.1002/biot.201300445

Brockman, I. M., and Prather, K. L. (2015a). Dynamic knockdown of *E. coli* central metabolism for redirecting fluxes of primary metabolites. *Metab. Eng.* 28, 104–113. doi: 10.1016/j.ymben.2014.12.005

Brockman, I. M., and Prather, K. L. J. (2015b). Dynamic metabolic engineering: new strategies for developing responsive cell factories. *Biotechnol. J.* 10, 1360–1369. doi: 10.1002/biot.201400422

Chen, Y., Bao, J., Kim, I. K., Siewers, V., and Nielsen, J. (2014). Coupled incremental precursor and co-factor supply improves 3-hydroxypropionic acid production in *Saccharomyces cerevisiae*. *Metab. Eng.* 22, 104–109. doi: 10.1016/j.ymben.2014.01.005

Christiano, R., Nagaraj, N., Fröhlich, F., and Walther, T. C. (2014). Global proteome turnover analyses of the Yeasts *S. cerevisiae* and *S. pombe*. *Cell Rep.* 9, 1959–1965. doi: 10.1016/j.celrep.2014.10.065

Curran, K. A., Karim, A. S., Gupta, A., and Alper, H. S. (2013). Use of expression-enhancing terminators in *Saccharomyces cerevisiae* to increase mRNA half-life and improve gene expression control for metabolic engineering applications. *Metab. Eng.* 19, 88–97. doi: 10.1016/j.ymben.2013.07.001

Dahl, R. H., Zhang, F., Alonso-Gutierrez, J., Baidoo, E., Batth, T. S., Redding-Johanson, A. M., et al. (2013). Engineering dynamic pathway regulation using stress-response promoters. *Nat. Biotechnol.* 31, 1039–1046. doi: 10.1038/nbt.2689

David, F., Nielsen, J., and Siewers, V. (2016). Flux control at the Malonyl-CoA node through hierarchical dynamic pathway regulation in *Saccharomyces cerevisiae*. *ACS Synthet. Biol.* 5, 224–233. doi: 10.1021/acssynbio.5b00161

Davis, S., and Meltzer, P. S. (2007). GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor. *Bioinformatics* 23, 1846–1847. doi: 10.1093/bioinformatics/btm254

De Jong, H., Ranquet, C., Ropers, D., Pinel, C., and Geiselmann, J. (2010). Experimental and computational validation of models of fluorescent and luminescent reporter genes in bacteria. *BMC Syst. Biol.* 4:55. doi: 10.1186/1752-0509-4-55

Farmer, W. R., and Liao, J. C. (2000). Improving lycopene production in *Escherichia coli* by engineering metabolic control. *Nat. Biotechnol.* 18, 533–537. doi: 10.1038/75398

Germann, S. M., Baallal Jacobsen, S. A., Schneider, K., Harrison, S. J., Jensen, N. B., Chen, X., et al. (2016). Glucose-based microbial production of the hormone melatonin in yeast *Saccharomyces cerevisiae*. *Biotechnol. J.* 11, 717–724. doi: 10.1002/biot.201500143

Heim, R., Cubitt, A. B., and Tsien, R. Y. (1995). Improved green fluorescence. *Nature* 373, 663–664. doi: 10.1038/373663b0

Hong, K. K., and Nielsen, J. (2012). Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries. *Cell. Mol. Life Sci.* 69, 2671–2690. doi: 10.1007/s00018-012-0945-1

Jensen, N. B., Strucko, T., Kildegaard, K. R., David, F., Maury, J., Mortensen, U. H., et al. (2014). EasyClone: method for iterative chromosomal integration of multiple genes in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 15, 238–248. doi: 10.1111/1567-1364.12118

Kannan, S., Sams, T., Maury, J., and Workman, C. T. (2018). Reconstructing dynamic promoter activity profiles from reporter gene data. *ACS Synth. Biol.* 7, 832–841. doi: 10.1021/acssynbio.7b00223

Kildegaard, K. R., Jensen, N. B., Schneider, K., Czarnotta, E., Özdemir, E., Klein, T., et al. (2016). Engineering and systems-level analysis of *Saccharomyces cerevisiae* for production of 3-hydroxypropionic acid via malonyl-CoA reductase-dependent pathway. *Microb. Cell Fact.* 15:53. doi: 10.1186/s12934-016-0451-5

Kumar, V., Ashok, S., and Park, S. (2013). Recent advances in biological production of 3-hydroxypropionic acid. *Biotechnol. Adv.* 31, 945–961. doi: 10.1016/j.biotechadv.2013.02.008

Lo, T.-M., Chng, S. H., Teo, W. S., Cho, H.-S., and Chang, M. W. (2016). A two-layer gene circuit for decoupling cell growth from metabolite production. *Cell Syst.* 3, 133–143. doi: 10.1016/j.cels.2016.07.012

Mateus, C., and Avery, S. V. (2000). Destabilized green fluorescent protein for monitoring dynamic changes in yeast gene expression with flow cytometry. *Yeast* 16, 1313–1323.

Meadows, A. L., Hawkins, K. M., Tsegaye, Y., Antipov, E., Kim, Y., Raetz, L., et al. (2016). Rewriting yeast central carbon metabolism for industrial isoprenoid production. *Nature* 537, 694–697. doi: 10.1038/nature19769

Monod, J. (1941). *Recherches sur la Croissance des Cultures Bactériennes*. Ph.D. thesis, Université de Paris.

Munchel, S. E., Shultzaberger, R. K., Takizawa, N., and Weis, K. (2011). Dynamic profiling of mRNA turnover reveals gene-specific and system-wide regulation of mRNA decay. *Mol. Biol. Cell* 22, 2787–2795. doi: 10.1091/mbc.E11-01-0028

Neymotin, B., Athanasiadou, R., and Gresham, D. (2014). Determination of *in vivo* RNA kinetics using RATE-seq. *RNA* 20, 1645–1652. doi: 10.1261/rna.045104.114

Nielsen, J., and Jewett, M. C. (2008). Impact of systems biology on metabolic engineering of *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 8, 122–131. doi: 10.1111/j.1567-1364.2007.00302.x

Poulsen, B. R., Ruiter, G., Visser, J., and Iversen, J. J. L. (2003). Determination of first order rate constants by natural logarithm of the slope plot exemplified by analysis of *Aspergillus niger* in batch culture. 25, 565–571. *Biotechnol. Lett.* doi: 10.1023/A:1022836815439

Ro, D.-K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., et al. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 440, 940–943. doi: 10.1038/nature04640

Ronen, M., Rosenberg, R., Shraiman, B. I., and Alon, U. (2002). Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. *Proc. Natl. Acad. Sci. U.S.A.* 99, 10555–10560. doi: 10.1073/pnas.152046799

Rudge, T. J., Brown, J. R., Federici, F., Dalchau, N., Phillips, A., Ajioka, J. W., et al. (2016). Characterization of intrinsic properties of promoters. 5, 89–98. *ACS Synth. Biol.* doi: 10.1021/acssynbio.5b00116

Scalcinati, G., Knuf, C., Partow, S., Chen, Y., Maury, J., Schalk, M., et al. (2012). Dynamic control of gene expression in *Saccharomyces cerevisiae* engineered for the production of plant sesquitepene α-santalene in a fed-batch mode. *Metab. Eng.* 14, 91–103. doi: 10.1016/j.ymben.2012.01.007

Shin, S. Y., Jung, S. M., Kim, M. D., Han, N. S., and Seo, J. H. (2012). Production of resveratrol from tyrosine in metabolically engineered *Saccharomyces cerevisiae*. *Enzyme Microb. Technol.* 51, 211–216. doi: 10.1016/j.enzmictec.2012.06.005

Solomon, K. V., Sanders, T. M., and Prather, K. L. (2012). A dynamic metabolite valve for the control of central carbon metabolism. *Metab. Eng.* 14, 661–671. doi: 10.1016/j.ymben.2012.08.006

Soma, Y., Tsuruno, K., Wada, M., Yokota, A., and Hanai, T. (2014). Metabolic flux redirection from a central metabolic pathway toward a synthetic pathway using a metabolic toggle switch. *Metab. Eng.* 23, 175–184. doi: 10.1016/j.ymben.2014.02.008

Tai, S. L., Boer, V. M., Daran-Lapujade, P., Walsh, M. C., De Winde, J. H., Daran, J.-M., et al. (2005). Two-dimensional transcriptome analysis in chemostat cultures: Combinatorial effects of oxygen availability and macronutrient limitation in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 280, 437–447. doi: 10.1074/jbc.M410573200

Torella, J. P., Ford, T. J., Kim, S. N., Chen, A. M., Way, J. C., and Silver, P. A. (2013). Tailored fatty acid synthesis via dynamic control of fatty acid elongation. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11290–11295. doi: 10.1073/pnas.1307129110

van Ditmarsch, D., and Xavier, J. B. (2011). High-resolution time series of Pseudomonas aeruginosa gene expression and rhamnolipid secretion through growth curve synchronization. *BMC Microbiol.* 11:140. doi: 10.1186/1471-2180-11-140

Yamanishi, M., Ito, Y., Kintaka, R., Imamura, C., Katahira, S., Ikeuchi, A., et al. (2013). A genome-wide activity assessment of terminator regions in *Saccharomyces cerevisiae* provides a "terminatome" toolbox. *ACS Synth. Biol.* 2, 337–347. doi: 10.1021/sb300116y

Yu, T., Zhou, Y. J., Wenning, L., Liu, Q., Krivoruchko, A., Siewers, V., et al. (2017). Metabolic engineering of *Saccharomyces cerevisiae* for production of very long chain fatty acid-derived chemicals. *Nat. Commun.* 8:15587. doi: 10.1038/ncomms15587

Dynamic Control of *ERG20* and *ERG9* Expression for Improved Casbene Production in *Saccharomyces cerevisiae*

Roberta Callari, Yvan Meier, Davide Ravasio and Harald Heider*

Evolva SA, Reinach, Switzerland

Production of plant metabolites in microbial hosts represents a promising alternative to traditional chemical-based methods. Diterpenoids are compounds with interesting applications as pharmaceuticals, fragrances and biomaterials. Casbene, in particular, serves as a precursor to many complex diterpenoids found in plants from the Euphorbiaceae family that have shown potential therapeutic effects. Here, we engineered the budding yeast *Saccharomyces cerevisiae* for improved biosynthesis of the diterpene casbene. We first expressed, in yeast, a geranylgeranyl diphosphate synthase from *Phomopsis amygdali* in order to boost the geranylgeranyl diphosphate pool inside the cells. The enzyme uses isopentenyl diphosphate and dimethylallyl diphosphate to directly generate geranylgeranyl diphosphate. When co-expressing a casbene synthase from *Ricinus communis* the yeast was able to produce casbene in the order of 30 mg/L. Redirecting the flux from FPP and sterols, by means of the ergosterol sensitive promoter of *ERG1*, allowed for plasmid-based *casbene* production of 81.4 mg/L. Integration of the target genes into the yeast genome, together with the replacement of the promoter regions of *ERG20* and *ERG9* with combinations of ergosterol- and glucose-sensitive promoters, generated a titer of 108.5 mg/L of casbene. We here succeeded to engineer an improved route for geranylgeranyl diphosphate synthesis in yeast. Furthermore, we showed that the concurrent dynamic control of *ERG20* and *ERG9* expression, using ergosterol and carbon source regulation mechanisms, could substantially improve diterpene titer. Our approach will pave the way for a more sustainable production of GGPP- and casbene-derived products.

Keywords: casbene, diterpene, *ERG20*, *ERG9*, yeast, metabolic engineering, dynamic control, mevalonate pathway

INTRODUCTION

Diterpenoids represent one of the largest and most diverse classes of plant metabolites. Although in some cases they carry out important primary functions (e.g., regulation of growth and development by gibberellins), they are usually products of the secondary metabolism with specialized pathways extremely varied across the plant kingdom (Zerbe and Bohlmann, 2015). Diterpenoids are beneficial for plants because of their role in protection from abiotic stress and control of the ecological interactions with other organisms (e.g., defense against herbivores and microbial

pathogens) (Cheng et al., 2007; Tholl, 2015). Moreover, they can benefit humanity because of their industrial application as pharmaceuticals (e.g., paclitaxel, forskolin, ingenol-3-angelate, prostratin), fragrances (e.g., sclareol), and other industrial bioproducts (e.g., steviol glycosides as natural sweetners, diterpene resins for inks, and coatings) (Bohlmann and Keeling, 2008; Goyal et al., 2010; Caniard et al., 2012; Doseyici et al., 2014; Fidler and Goldberg, 2014; Howat et al., 2014; Miana et al., 2015).

Current production methods of diterpenoids rely on extraction from natural sources and chemical synthesis. Because of the low yield occurrence in the producing organisms and the structural complexity of the compounds, such methods are inefficient and environmentally costly. An attractive and environmentally friendly alternative is represented by microbial fermentation. Biosynthesis in heterologous hosts such as *Escherichia coli* or *Saccharomyces cerevisiae* can (1) reduce costs using sugar-based carbon sources, (2) increase sustainability by avoiding harvesting and extraction from natural sources, (3) increase yield and productivity using genetic manipulation of the heterologous hosts, and (4) provide enantiomerically pure products through enzymatic biocatalysis (Scalcinati et al., 2012). *S. cerevisiae* in particular is a robust host that not only offers the biosynthetic machinery needed for production of diterpenoids, but also contributes the necessary environment for expression of membrane-bound enzymes, such as cytochrome P450 hydroxylases. These p450 enzymes are frequently involved in the biosynthesis of complex plant terpenoids and are usually difficult to express in prokaryotic systems (Hamann and Møller, 2007; Kirby and Keasling, 2009).

The precursors for production of terpenes are present in the native metabolic network of *S. cerevisiae* (Chambon et al., 1991). The yeast mevalonate (MVA) pathway, through multiple rounds of condensation of isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), leads to generation of geranyl diphosphate (GPP), farnesyl diphosphate (FPP) and geranylgeranyl diphosphate (GGPP) (**Figure 1**). GPP, FPP, and GGPP represent the universal precursor units of all monoterpenes (C_{10}), sesquiterpenes (C_{15}), and diterpenes (C_{20}) respectively. However, the GGPP content in *S. cerevisiae* is rather low, due to the fact that the endogenous geranylgeranyl diphosphate synthase (GGPPS) Bts1p does not compete efficiently with enzymes directly upstream and downstream (farnesyl diphosphate synthase Erg20p and squalene synthase Erg9p) for the pools of IPP and FPP (Jiang et al., 1995). Thus, the endogenous flux of the pathway favors production of FPP, and the sterols derived from it. As a consequence, improvement of the GGPP pool is essential for efficient, high-titer diterpene production. To reach this goal different studies have focused on either the use of heterologous GGPP synthases, Erg20p mutants that are able to synthetize GGPP, or on fusions

of *BTS1* and *ERG20* (Kirby et al., 2010; Zhou et al., 2012; Ignea et al., 2015; Song et al., 2017). Such strategies have been shown to improve production of GGPP-derived compounds, but they have also demonstrated that additional improvements in substrate conversion toward diterpene production are feasible.

In this study, we developed a yeast platform for production of the diterpene hydrocarbon casbene, originally identified from castor bean (*Ricinus communis*) where it serves as a phytoalexin (Mau and West, 1994). In the last decades, the compound has gained much interest since the conversion of GGPP into casbene, catalyzed by the casbene synthase enzyme, is acknowledged as the first committed step in the biosynthesis of the diterpenoid molecular backbones jatrophane, tigliane, lathyrane, and ingenane (Kirby et al., 2010). Casbene is the precursor of many diterpenoids of medical interest, identified in a number of Euphorbiaceae (**Figure 2**). For example, Ingenol-3-angelate, found in the sap of *Euphorbia peplus,* has been approved by the FDA for the treatment of actinic keratosis (Fidler and Goldberg, 2014) whilst Prostratin, derived from *Homolanthus nutans* is a protein kinase C activator that inhibits HIV-1 infections and reduces HIV-1 latency (Miana et al., 2015). Jatrophone, isolated from extracts of *Jatropha gossypifolia* L., has shown significant anti-proliferative effects against human tumor cell lines (Kupchan et al., 1970; Theoduloz et al., 2009) and the tigliane derived diterpenoid, resiniferatoxin from *Euphorbia resinifera* is effective against a broad range of inflammatory and neuropathic pain conditions (Iadarola and Gonnella, 2013). Euphorbia factor L2 from *Euphorbia lathyris* induces apoptosis in lung cancer cell lines (Lin et al., 2017).

For efficient production of casbene in yeast, we increased the GGPP supply inside the cells by expression of a truncated, fungal GGPP synthase, from the endophytic fungus *Phomopsis amygdali*. Expression of the enzyme, producing GGPP from IPP and DMAPP directly, significantly improved GGPP biosynthesis and allowed, upon co-expression of the casbene synthase from *R. communis*, for production of casbene in the order of 30 mg/L. Similar titers have been reported previously (Kirby et al., 2010). Dynamic control of genes *ERG20* and *ERG9* by means of ergosterol- and glucose-sensitive promoters further redirected the flux to GGPP and diterpene synthesis, improving casbene production by up to 108.5 mg/L.

We envision that our approach will pave the way toward the sustainable production of various GGPP- and casbene-derived isoprenoids.

MATERIALS AND METHODS

Chemicals and Media

All chemicals were bought from Sigma-Aldrich (St. Louis Missouri, USA) unless stated otherwise. Authentic standard of cembrene was purchased from CHEMOS GmbH & Co. KG (Regenstauf, Germany). Authentic standard of casbene was received from Professor Birger Møller's laboratory at the University of Copenhagen.

LB medium for growth of *Escherichia coli* was supplied from Carl Roth GmbH + Co. KG (Karlsruhe, Germany), and was

---

Abbreviations: MVA, mevalonate; IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate; GPP, geranyl diphosphate; FPP, farnesyl diphosphate; GGPP, geranylgeranyl diphosphate; GGPPS, geranylgeranyl diphosphate synthase; *PaGGPPS*, geranylgeranyl diphosphate synthase from *Phomopsis amygdali*; *RcCBS*, casbene synthase from *Ricinus communis*; CAS, casbene; OD$_{600}$, optical density at 600 nm; PaFS, fusicoccadiene synthase from *Phomopsis amygdali*; ORF, open reading frame.
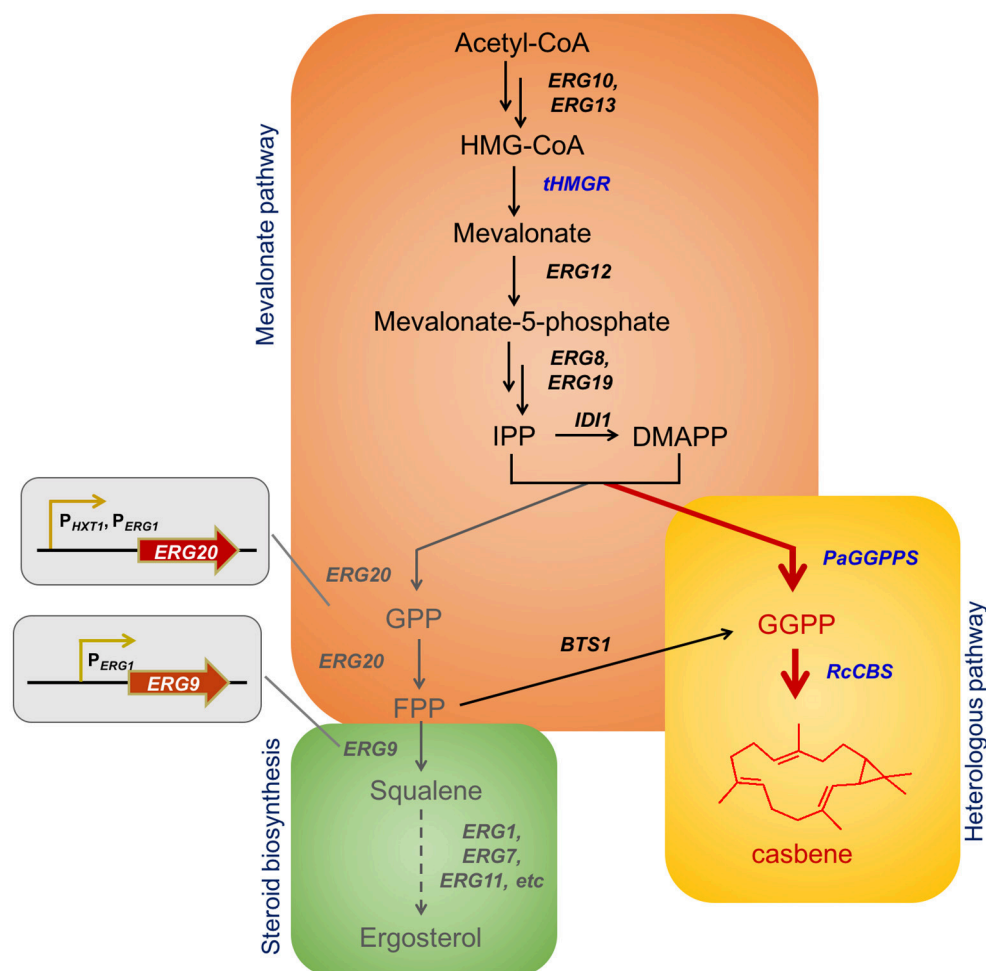
FIGURE 1 | Biosynthesis of casbene in yeast and examples of casbene derived compounds. Schematic overview of casbene biosynthesis based on engineered mevalonate pathway in *S. cerevisiae*. The new biosynthetic branch starts with *PaGGPPS*, for production of GGPP using IPP and DMAPP as sole substrates. *RcCBS* mediates casbene synthesis (in red). Endogenous genes and metabolites are shown in black. Overexpressed genes are shown in blue. Genes regulated through dynamic control (*ERG20* and *ERG9*) and genes from the steroid biosynthetic pathway are shown in gray. The truncated endogenous gene *tHMG1* and the heterologous genes *PaGGPPS* (encoding a truncated version of the fusicoccadiene synthase from *Phomopsis amygdali* serving as GGPP synthase) and *RcCBS* (encoding a truncated version of the casbene synthase from *Ricinus communis*) were overexpressed to improve casbene production. The native promoter of *ERG20* was replaced with P_{ERG1} or P_{HXT1}, respectively. The native promoter of *ERG9* was replaced with P_{ERG1}. HMG-CoA, 3-hydroxy-3-methylglutaryl coenzyme A; IPP, isopentenyl pyrophosphate; DMAPP, dimethylallyl pyrophosphate; GPP, geranyl diphosphate; FPP, farnesyl diphosphate; GGPP, geranylgeranyl diphosphate.

supplemented with 100 µg/L of ampicillin for amplification of plasmids. Yeast Extract Peptone Dextrose (YPD) medium with 20 g/L glucose was used for growth of wildtype strains prior to transformation. For preparation of pre- and main cultures we used synthetic complete (SC) drop-out medium (Formedium LTD, Hustanton, England), supplied with 6.7 g/L yeast nitrogen base, 20 g/L glucose and all amino acids necessary for the corresponding auxotrophy.

Plasmids and Strains Construction

Table 1 lists all plasmids constructed in this work. Coding sequences (**Table S1**) were synthesized by GeneArt® (Thermofisher Scientific, Zug, Switzerland) as yeast codon optimized versions. Synthetic genes were cloned via *Hind*III HF, *Sac*II restriction digestion and T4 DNA ligase (New England

Biolabs, Ipswich, Massachusetts, USA) according to standard protocols (Green and Sambrook, 2012). *E. coli* XL10 Gold (Agilent, Santa Clara, California, USA) cells were used for subcloning of genes.

S. cerevisiae strains generated throughout this study are listed in **Table 2**. All constructed strains were derived from strain NCYC 3608 (NCYC, Norwich, United Kingdom), a derivative of strain S288C modified in our labs and carrying a truncated 3-hydroxy-3-methylglutaryl coenzyme A reductase (tHMGR) in the YCT1 locus to increase the flow through the early part of the mevalonate pathway (CAS1- MATalpha ho∆0 his3∆0 leu2∆0 ura3∆0 CAT5(T91M) MIP1(-661T) SAL1 GAL2 YLL055w::loxP-pTDH3-Sc_tHMGR-tCYC1). All yeast strains were stored in 25% glycerol at −80°C. Yeast transformation was performed using the classical lithium acetate method (Gietz and Schiestl, 2007).

FIGURE 2 | Casbene is the precursor of many diterpenoids identified in plants from the Euphorbiaceae family. Euphorbia factor L2 (*Euphorbia lathyris*), resiniferatoxin (*Euphorbia resinifera*), ingenol-3-angelate (*Euphorbia peplus*), prostratin (*Homolanthus nutans*), and jatrophone (*Jatropha gossypifolia* L.).
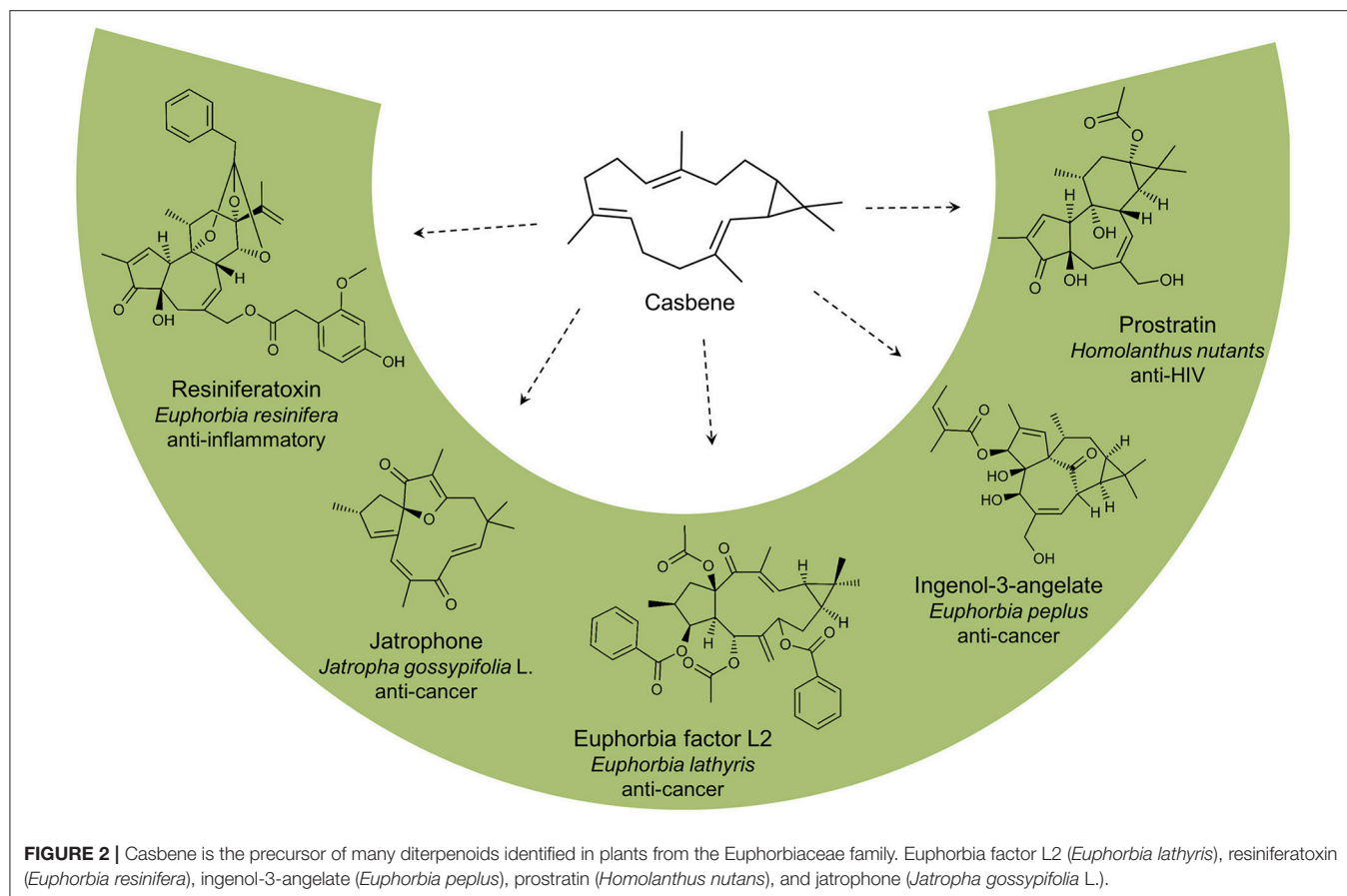
Transformants were grown on agar plates prepared with selective SC drop-out medium. For integration into the yeast genome, genes were cloned into yeast integration plasmids targeting incorporation into specific sites (**Table 1**). Transformants were verified by PCR on genomic DNA for correct insertion of heterologous genes.

The 805 bp promoter region of *ERG1* (**Table S2**) was amplified from genomic DNA of *S. cerevisiae* S288C after lysis in 30 μL 0.2% SDS at 95°C for 5 min and clarification at 14,000 g for 5 min. iProof™ High-Fidelity DNA Polymerase was used according to the manufacturer's protocol with primer pair F_ERG1p_*Spe*I/ R_ERG1p_*Sac*II. The PCR product was cut with *Spe*I/*Sac*II, and inserted into the integration vector pUG72 (Gueldener et al., 2002) close to the URA3 marker flanked by loxP sites. Marker and promoter were amplified with primers pair F_ERG9int/ R_ERG9int and transformed in yeast for homologous recombination at the ERG9 site (**Table 3**).

Plasmid pCAS4 for replacement of the promoter region of *ERG20* (**Table S2**) was kindly provided by Michael Eichenberger. The plasmid already contained up- and down-tags for homologous recombination at the ERG20 site, a URA3 marker and the promoter of *CYC1* for exchange with the promoter of *ERG20*. P_{CYC1} was released from the plasmid with *Bgl*II and *Hind*III and replaced by promoters of *ERG1* and *HXT1*. P_{ERG1} was amplified from plasmid pCAS3 with primers pair F_ERG1p_*Bgl*II and R_ERG1p_*Hind*III. P_{HXT1} was amplified from an expression vector found in our labs with primers

pair F_HXT1p_*Bgl*II and R_HXT1p_*Hind*III (**Table 3**). The PCR products were cut with *Bgl*II/*Hind*III, and inserted into the integration vector pCAS3 to replace P_{CYC1}. Strains with successful replacement of native promoters of *ERG20* and *ERG9* were verified by diagnostic PCR.

Yeast Growth

Yeast batch cultures for production of metabolites were performed with the System Duetz (EnzyScreen, Heemstede, Netherlands) in an ISF-1-W shaker (Kuhner, Birsfelden, Switzerland) at 30°C, 300 rpm, and 5 cm shaking diameter. Pre-cultures were prepared by inoculation of single colonies into 3 mL of selective liquid SC medium in triplicates. Pre-cultures were grown for 24 h at 30°C, 160 rpm. Optical density at 600 nm (OD_{600}) of a 1:40 dilution was measured in an Ultrospec 10 table top spectrophotometer (GE Healthcare, Little Chalfont, United Kingdom). Main cultures were inoculated in 2.5 mL of selective SC medium in a 24-deepwell microplate to an OD_{600} of 0.1 and incubated from 24 up to 120 h as stated.

Sample Preparation

Cells from a 2 mL volume were harvested by centrifugation at 4,000 rpm for 5 min in 2 mL screw cap tubes. Supernatants were discarded and pellets were mixed with 500 μL of methanol. Pellets were shaken for 10 min at 60°C and 1,500 rpm in a Thermo-Shaker TS-100 (Axonlab, Reichenbach an der Fils, Germany). Cell debris was removed by centrifugation at 4,000

rpm for 5 min and liquid phases were transferred to glass vials. The methanol extracts were evaporated for 24 h at room temperature under the fume hood, resolubilized in 500 μL hexane and used for analysis.

Detection of Diterpenes

Casbene and alcohols derived from FPP and GGPP were detected using gas chromatography–mass spectrometry (GC–MS). Quantification was carried out using the Agilent Technologies 7890a GC system (Agilent Technologies, Santa Clara, USA) equipped with a 5975 Mass Selective Detector (MSD). Sample volumes of 1 μL were injected in splitless mode at 250°C with the following GC program: 80°C for 2 min, raise to 220°C at 25°C/min, raise to 240°C at 3°C/min, ramp at rate 80°C/min to 300°C and held for 3 min. Helium was used as carrier gas with a constant flow rate of 1.2 mL/min. The MSD was operated at 70 eV in scan mode, with a range between 35 and 500 m/z. Diterpenes were identified by retention time and comparison with a mass spectral database (NIST version 2.0, Gaithersburg, MD, USA) and/or comparison with authentic standards. β-Caryophyllene (10 mg/L) was added in each sample as internal standard and calibration curves were generated to quantify the target compounds. In the case of geranylgeraniol an authentic standard of the compound was used, whereas casbene quantification was approximated by a standard curve generated using an authentic standard of the diterpene cembrene.

RESULTS

Establishing Casbene Production in *S. cerevisiae*

In order to assess casbene (CAS) production in yeast, the casbene synthase gene from *R. communis* (*RcCBS*) was cloned as a codon optimized version under control of the promoter of *TEF1* in a yeast expression plasmid ("pCAS1"). The N-terminal

TABLE 1 | List of plasmids constructed in this work.

Plasmid	Description	Plasmid type
pCAS1	Plasmid harboring expression cassette pTEF1-*RcCBS*-ENO2t	ARS/CEN
pCAS2	Plasmid harboring expression cassette pPGK1-*PaGGPPS*-ADH2t	ARS/CEN
pCAS3	Plasmid for ERG9 promoter exchange with insertion of the promoter region from *ERG1* gene	Integrative
pCAS4	Plasmid for ERG20 promoter exchange with insertion of the promoter region from *CYC1* gene	Integrative
pCAS5	Plasmid for ERG20 promoter exchange with insertion of the promoter region from *ERG1* gene	Integrative
pCAS6	Plasmid for ERG20 promoter exchange with insertion of the promoter region from *HXT1* gene	Integrative
pCAS7	Plasmid for pTEF1-*RcCBS*-tENO2 cassette integration into YOR092W locus	Integrative
pCAS8	Plasmid for pPGK1-*PaGGPPS*-tADH2 cassette integration into YDR122W locus	Integrative

chloroplast transit peptide, identified with the ChloroP software used at default settings (Emanuelsson et al., 1999) and needed for plastidial localization in plants, was removed (Kirby et al., 2010). Plasmid pCAS1 was introduced into yeast strain H-MEV, engineered for production of isoprenoids. H-MEV contained a chromosomally integrated copy of a truncated 3-hydroxy-3-methylglutaryl coenzyme A reductase (*tHMGR*) enabling high mevalonate pathway activity (Donald et al., 1997). The resulting strain (CAS1), was analyzed for production of casbene. Small amounts of the compound (∼ 0.8 mg/L) could be detected after 72 h of growth (data not shown).

To boost casbene production, we co-expressed a heterologous GGPP synthase from the plant-pathogenic fungus *Phomopsis amygdali*. Toyomasu et al. (2007) found that synthesis of the diterpene fusicoccadiene in *P. amygdali* is catalyzed by the unusual chimeric enzyme fusicoccadiene synthase (PaFS). This multifunctional enzyme can synthetize fusicoccadiene starting from IPP and DMAPP, through (1) a prenyltransferase domain, synthetizing GGPP from the C5 isoprene units, and (2) a

TABLE 2 | List of strains constructed in this work.

Strains	Description	MVA/sterols pathway modifications
H-MEV	S288C derivative with relatively high mevalonate pathway activity	*tHMGR*
CAS1	Strain H-MEV harboring plasmid pCAS1 and empty vector	*tHMGR*
CAS2	Strain H-MEV harboring plasmid pCAS1 and pCAS2	*tHMGR*
CAS3	H-MEV with *ERG20* under the control of P$_{ERG1}$	*tHMGR*, P$_{ERG1}$- *ERG20*
CAS4	H-MEV with *ERG9* under the control of P$_{ERG1}$	*tHMGR*, P$_{ERG1}$- *ERG9*
CAS5	H-MEV with *ERG20* and *ERG9* under the control of P$_{ERG1}$	*tHMGR*, P$_{ERG1}$- *ERG20*, P$_{ERG1}$- *ERG9*
CAS6	Strain CAS3 harboring plasmid pCAS1 and pCAS2	*tHMGR*, P$_{ERG1}$- *ERG20*
CAS7	Strain CAS4 harboring plasmid pCAS1 and pCAS2	*tHMGR*, P$_{ERG1}$- *ERG9*
CAS8	Strain CAS5 harboring plasmid pCAS1 and pCAS2	*tHMGR*, P$_{ERG1}$- *ERG20*, P$_{ERG1}$- *ERG9*
CAS9	Strain H-MEV with *ERG20* under the control of P$_{HTX1}$	*tHMGR*, P$_{HXT1}$- *ERG20*
CAS10	Strain CAS9 with *ERG9* under the control of P$_{ERG1}$	*tHMGR*, P$_{HXT1}$- *ERG20*, P$_{ERG1}$- *ERG9*
CAS11	Strain H-MEV with integrated *PaGGPPS* and *RcCBS*	*tHMGR*
CAS12	Strain CAS3 with integrated *PaGGPPS* and *RcCBS*	*tHMGR*, P$_{ERG1}$- *ERG20*
CAS13	Strain CAS9 with integrated *PaGGPPS* and *RcCBS*	*tHMGR*, P$_{HXT1}$- *ERG20*
CAS14	Strain CAS5 with integrated *PaGGPPS* and *RcCBS*	*tHMGR*, P$_{ERG1}$- *ERG20*, P$_{ERG1}$- *ERG9*
CAS15	Strain CAS10 with integrated *PaGGPPS* and *RcCBS*	*tHMGR*, P$_{HXT1}$- *ERG20*, P$_{ERG1}$- *ERG9*

terpene cyclase domain, catalyzing the cyclization of GGPP into fusicoccadiene.

We reasoned that the expression of the prenyltransferase domain from PaFS in yeast, catalyzing all steps from the C5 to the C20 precursor unit, could enhance the efficiency of this conversion, as it requires only one enzyme for the entire biosynthesis of GGPP. A construct encoding for the codon optimized prenyltransferase domain (residues 390–719 of PaFS) (Chen et al., 2016) was synthesized and cloned under control of the promoter of *PGK1* into pCAS2 yeast expression vector. pCAS2 was introduced together with pCAS1 (*RcCBS*) into H-MEV, generating strain CAS2. *PaGGPPS+RcCBS* co-expression gave rise to a peak with the same retention time and mass-spectrum of the authentic casbene standard (**Figure S1A**), while, as previously discussed, only trace amounts of the compound accumulated when expressing solely *RcCBS* (**Figure S1B**). A titer of 32 mg/L of casbene was reached in strain CAS2 (*PaGGPPS+RcCBS*) after 72 h of batch culture. The relatively high titer of casbene strongly suggested that *PaGGPPS* could considerably increase the supply of the GGPP precursor necessary for casbene production by RcCBS. This suggestion was further supported by the appearance of a pronounced peak in the chromatogram corresponding to GGOH (**Figure S1B**), the prenyl alcohol accumulating in engineered yeast cells due to endogenous phosphatase activities acting on the excess amounts of GGPP (Faulkner et al., 1999; Tokuhiro et al., 2009).

Casbene Production in Engineered Strains With Engineered Mevalonate Pathway

To further improve casbene titers, we focused on downregulation of competing metabolic pathways. PaGGPPSp uses IPP and DMAPP to catalyze the consecutive condensation steps needed for GGPP synthesis and is in competition with the native farnesyl diphosphate synthase, Erg20p for the IPP/DMAPP

pool. Moreover, FPP, produced by Erg20p, enters the ergosterol biosynthetic pathway by action of the squalene synthase Erg9p. Thus, in order to redirect the flux to diterpene production, we attempted dynamic control of either *ERG20*, or of *ERG20* in combination with *ERG9*, by replacing their native promoters with the ergosterol sensitive promoter P_{ERG1}. This approach has recently been shown to positively affect amorpha-4-11-diene production in yeast (Yuan and Ching, 2015). The promoter of *ERG1* senses intracellular ergosterol levels and adjusts transcription of downstream genes to enable production of optimal levels of this essential metabolite. We constructed three strains containing (1) *ERG20* under control of P_{ERG1}, (2) *ERG9* under control of P_{ERG1} and (3) both *ERG20* and *ERG9* under control of P_{ERG1} (**Table 2**, CAS3-5). All engineered strains also carried an integrated copy of *tHMGR* and were transformed with pCAS1 (*RcCBS*) and pCAS2 (*PaGGPPS*) to evaluate the effect of dynamic control of expression of *ERG20* and/or *ERG9* on casbene production. The resulting strains were designated as CAS6 (P_{ERG1}-*ERG20*), CAS7 (P_{ERG1}-*ERG9*) and CAS8 (P_{ERG1}-*ERG20* and P_{ERG1}-*ERG9*) (**Table 2**). CAS2, containing the native promoters of *ERG20* and *ERG9*, was used as a reference strain. Strains were analyzed for casbene production in time course experiments up to 120 h cultivation time.

Interestingly, strains with engineered *ERG20* and *ERG9* expression improved the casbene titers substantially but showed different profiles of compound accumulation (**Figure 3A**). The two strains CAS6 (P_{ERG1}-*ERG20*) and CAS8 (P_{ERG1}-*ERG20*, P_{ERG1}-*ERG9*) showed similar profiles of casbene accumulation over time. Strain CAS6 (P_{ERG1}-*ERG20*) steadily built-up casbene during 96 h of cultivation where the measured casbene titer reached 61 mg/L. This represented a significant ∼3-fold improvement in production over the reference strain CAS2, which accumulated only 21 mg/L of casbene. At 120 h the casbene titer in CAS6 dropped to ∼40 mg/L, indicating that potential degradation mechanisms or modifying enzymes start to act on the casbene molecule. Strain CAS8 (P_{ERG1}-*ERG20*, P_{ERG1}-*ERG9*) harboring dynamic control of expression of both *ERG20* and *ERG9* could reach the highest titer of 81.4 mg/L casbene after 96 h of cultivation, a ∼4-fold improvement over production in CAS2. Strain CAS7 (P_{ERG1}-*ERG9*) revealed instead the highest production after 48 h of cultivation (42 mg/L). Casbene titers steadily dropped to 18 mg/L thereafter. Finally, reference strain CAS2 (*tHMGR*) showed a gradual accumulation of casbene up 35 mg/L upon 72 h of culture, followed by a decline to 13.3 mg/L at 120 h.

Also, GGOH considerably accumulated in the engineered strains (**Figure 3B**). Strains CAS7 (P_{ERG1}-*ERG9*) and CAS2 (*tHMGR*) accumulated increasing amounts of GGOH during 72 h of cultivation, reaching up to 39.7 mg/L, and 30.3 mg/L, respectively. The titers of GGOH did not drop significantly throughout the entire cultivation period of 120 h. Strains CAS6 (P_{ERG1}-*ERG20*) and CAS8 (P_{ERG1}-*ERG20*, P_{ERG1}-*ERG9*) showed a progressive build-up of the compound reaching, upon 120 h of growth, 96.7 mg/L and 82.9 mg/L, respectively.

TABLE 3 | Oligonucleotides used for plasmid construction (restriction sites are underlined).

Name	Description
F_GGPPS_*Hind*III	cga<u>AAGCTT</u>ATGTTGTCTACTGGTTTGTCTTT GTCTCC
R_GGPPS_*Sac*II	AAGCACT<u>CCGCGG</u>TTAGTG
F_ERG1p_*Spe*I	cga<u>ACTAGT</u>GTCGAATACTACTATGACCGC
R_ERG1p_*Sac*II	acg<u>CCGCGG</u>CATGACCCTTTTCTCGATATGTT
F_ERG9int	GGTTTTGGGTTTAGTGCCTAAACGAGCA GCGAGAACACGACCACGCAGCTGAA GCTTCGTACGC
R_ERG9int	CTTCATCTCGACCGGATGCAATGCCAATT GTAATAGCTTTCCCATACTCAC TATAGGGAGACCGGC
F_ERG1p_*Bgl*II	ct<u>AGATCT</u>GTCGAATACTACTATGACCGC
R_ERG1p_*Hind*III	ctg<u>AAGCTT</u>ATGACCCTTTTCTCGATATGTT
F_HXT1p_*Bgl*II	tc<u>AGATCT</u>CAAGTGCTGATAGAAGAATACCAC
R_HXT1p_*Hind*III	gcg<u>AAGCTT</u>GATTTTACGTATATCAACTAGTTGACG
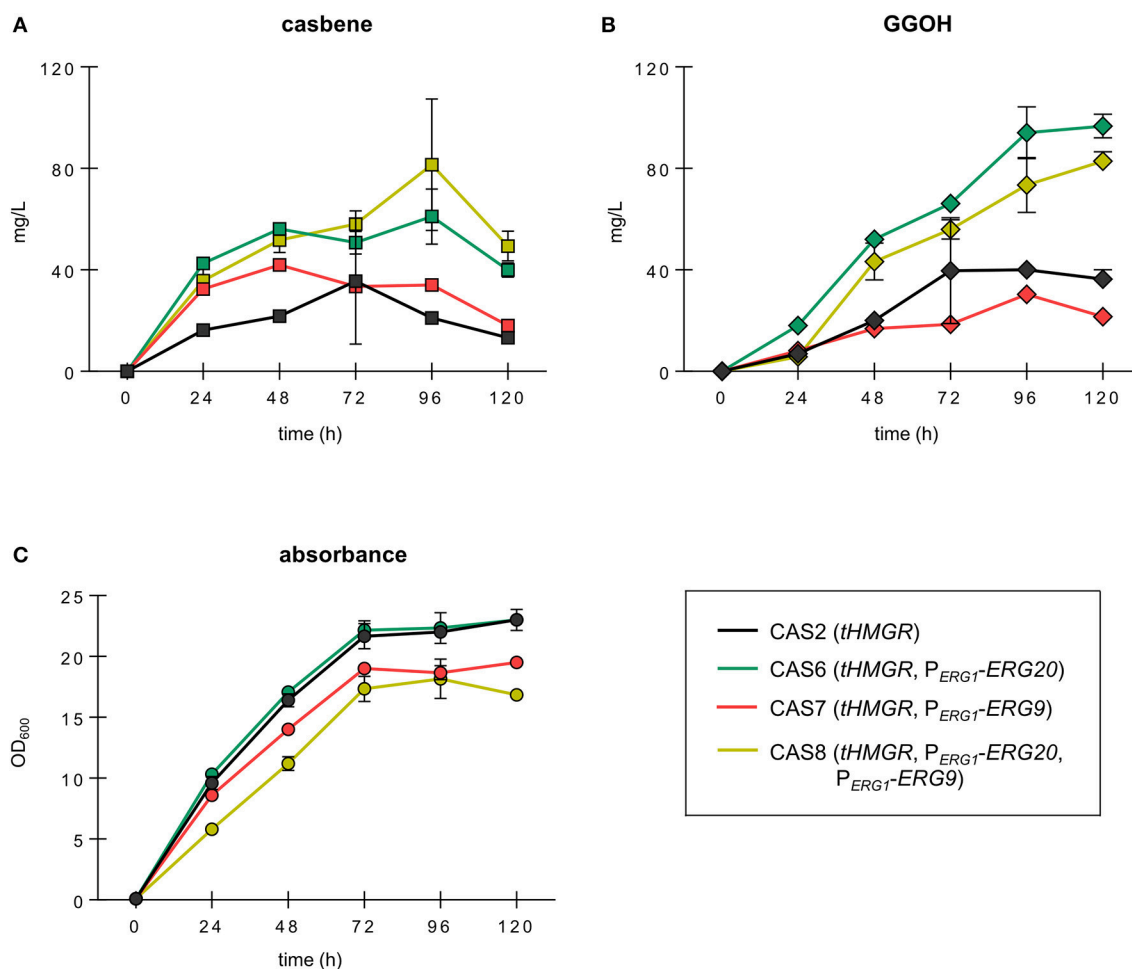
FIGURE 3 | Growth and production of casbene and GGOH in engineered strains expressing *PaGGPPS* and *RcCBS*. Time course of production of casbene **(A)** and GGOH **(B)** in strains CAS2 (*tHMGR*), CAS6 (*tHMGR*, P$_{ERG1}$-*ERG20*), CAS7 (*tHMGR*, P$_{ERG1}$-*ERG9*), and CAS8 (*tHMGR*, P$_{ERG1}$-*ERG20*, P$_{ERG1}$-*ERG9*). All strains episomally expressed *PaGGPPS* and *RcCBS*. **(C)** Growth profile of the strains determined using optical density data recorded at 600 nm (OD$_{600}$). All strains were grown in selective medium for maximally 120 h. Represented are the average and standard deviation of three independent experiments.

Strain CAS6 (P$_{ERG1}$-*ERG20*) in fact not only accumulated high amounts of both casbene and GGOH, but also showed optimal robustness, as revealed by its growth profile, which was comparable to that of reference strain CAS2 (**Figure 3C**). Strains with dynamic control of *ERG9* expression instead displayed growth retardation in comparison to reference strain CAS2. At the end of cultivation, strains CAS7 (P$_{ERG1}$-*ERG9*) and CAS8 (P$_{ERG1}$-*ERG20*, P$_{ERG1}$-*ERG9*) reached cell densities which were lower compared to strain CAS2 by 15 and 27%, respectively.

These results demonstrated the potential of dynamically controlling FPP synthesis for improved casbene production via modification of the native *ERG20* promoter.

Construction of Stable *S. cerevisiae* Strains for Casbene Production

To provide a stable chassis for casbene production, the engineered GGPP synthase *PaGGPPS* and the truncated casbene synthase *RcCBS* were stably integrated into the

yeast chromosomes. In the previous plasmid-based assays, dynamic control of solely *ERG9* expression showed the poorest performance in terms of production and was for this reason not included in the next phase of experiments. As dynamic control of *ERG20* expression was shown to be crucial for improved diterpene production, we further investigated the effect of dynamic regulation of *ERG20* by replacing its native promoter with the glucose-sensing promoter of *HXT1* (P$_{HXT1}$). It has recently been shown that using the HXT1 promoter to control *ERG9* expression could redirect the carbon flux from sterol synthesis toward sesquiterpene production (Scalcinati et al., 2012). The effect of expression of *ERG20* under control of P$_{HXT1}$ was tested either alone, or in combination with control of expression of *ERG9* by P$_{ERG1}$ (**Table 2**, strains CAS9 and CAS10).

The expression cassettes for *RcCBS* and *PaGGPPS* were integrated into five strains: CAS11 (*tHMGR*), CAS12 (*tHMGR*, P$_{ERG1}$-*ERG20*), CAS13 (*tHMGR*, P$_{HXT1}$-*ERG20*), CAS14 (*tHMGR*, P$_{ERG1}$-*ERG20*, P$_{ERG1}$-*ERG9*) and CAS15 (*tHMGR*,

P_{HXT1}-*ERG20*, P_{ERG1}-*ERG9*). All strains were analyzed for casbene and GGOH accumulation upon growth in batch culture for 96 h, as this production time yielded the highest titers of casbene in plasmid-based assays.

All engineered strains accumulated higher amounts of casbene and GGOH compared to reference strain CAS11 (24.6 mg/L of casbene and 27.6 mg/L of GGOH) (**Figure 4**). Strain CAS13 accumulated 85.5 mg/L of casbene and 50 mg/L of GGOH, whereas strain CAS12 accumulated 66.5 mg/L of casbene and 63.4 mg/L of GGOH, respectively. In strain CAS14 (*tHMGR*, P_{ERG1}-*ERG20*, P_{ERG1}-*ERG9*) accumulation of casbene did not reach the amounts observed previously (only 49.1 mg/L of casbene and 46.5 mg/L of GGOH accumulated). In strain CAS15 expression of *ERG20* controlled by P_{HXT1} combined with control of *ERG9* expression by promoter P_{ERG1} led to the highest titers of casbene (108.5 mg/L) and GGOH (79.9 mg/L) production. Interestingly, all strains, except CAS14, grew better than the reference strain CAS11 (*tHMGR* only). CAS14 (*tHMGR*, P_{ERG1}-*ERG20*, P_{ERG1}-*ERG9*) reached, in accordance with the plasmid-based expression results, lower final cell densities.

DISCUSSION

In this study we focused on the redirection of carbon flux toward casbene production by (1) introduction of a new biosynthetic branch for production of GGPP starting from IPP and DMAPP and (2) dynamic control of *ERG20* and *ERG9* expression by means of glucose- and ergosterol-sensitive promoters.

To increase the pool of GGPP in yeast, we expressed the GGPP synthase domain of the fusioccadiene synthase from *P. amygdali*. Similarly to other GGPP synthases from higher plants and bacteria (Vandermoten et al., 2009), this enzymatic domain synthesizes GGPP starting from one molecule of DMAPP and three molecules of IPP (Chen et al., 2016). In the yeast native pathway, GGPP is formed from the condensation of IPP and FPP, by action of the endogenous GGPP synthase Bts1p. FPP, in

turn, is formed by the successive condensation of DMAPP with IPP units, catalyzed by FPP synthase, Erg20p. This requirement for two enzymatic steps for the production of GGPP, represents a limiting factor, partially due to the low native activity of Bts1p (Jiang et al., 1995). Low-titer diterpene production starting from the endogenous GGPP pool has been previously reported (Zhou et al., 2012; Ignea et al., 2015) and was confirmed in this study by the very low levels of casbene produced in yeast expressing *RcCBS* (**Figure S1B**). Improved efficiency of the conversion of FPP to GGPP was shown to significantly benefit diterpene production (Zhou et al., 2012; Ignea et al., 2015). Expression of *PaGGPPS* in yeast introduced the direct conversion of IPP and DMAPP to GGPP and improved the efficiency of GGPP formation. Indeed, expression of *PaGGPS+RcCBS* lead to 32 mg/L casbene accumulation, a ~40 fold improvement compared to cells expressing solely *RcCBS* and harnessing the endogenous source of GGPP. The substantial accumulation of GGOH in those cells clearly indicated the improved efficiency of GGPP biosynthesis and suggested that the flux toward diterpene synthesis could be further optimized.

To further increase the supply of IPP and DMAPP, we restricted the flux toward FPP and sterols, by limiting the activity of Erg20p and Erg9p. These two enzymes catalyze FPP synthesis and the first step in sterol biosynthesis, respectively. Sterols are essential for yeast growth, therefore it is crucial to ensure the delivery of intermediates and end-products at levels that can guarantee proper cell function. Ergosterol, the major product of sterol biosynthesis, fulfills in fact several vital cellular functions that require balanced sterol concentrations (Parks et al., 1995). Downregulation of Ergosterol synthesis by means of weak or inducible and repressible promoters has been shown to be beneficial for isoprenoid production, while negatively affecting yeast viability (Paradise et al., 2008; Fischer et al., 2011; Scalcinati et al., 2012; Ignea et al., 2014). Therefore, we engineered a dynamic control strategy to balance metabolism between diterpene formation and cell growth. P_{ERG1} represents
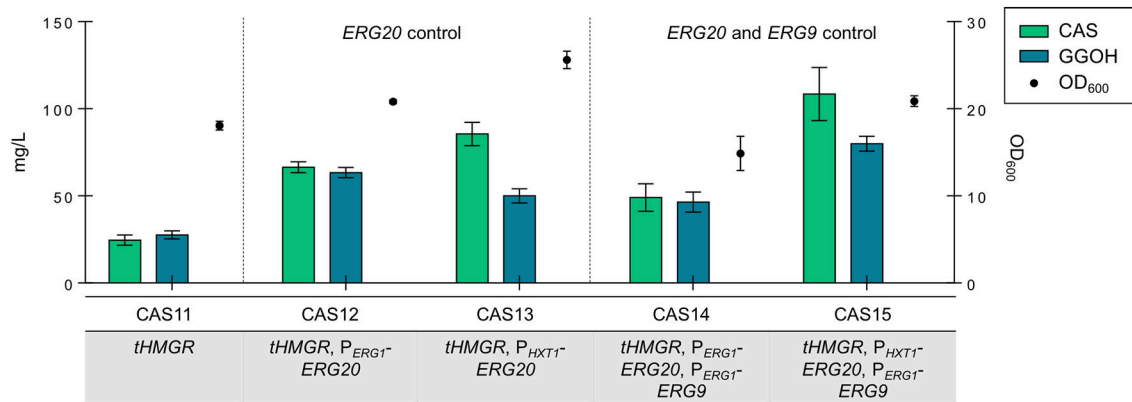
FIGURE 4 | Production of casbene and GGOH in engineered strains containing integrated *PaGGPPS* and *RcCBS*. Casbene (green bars) and GGOH (blue bars) produced by strains CAS11 (*tHMGR*), CAS12 (*tHMGR*, P_{ERG1}-*ERG20*), CAS13 (*tHMGR*, P_{HXT1}-*ERG20*), CAS14 (*tHMGR*, P_{ERG1}-*ERG20*, P_{ERG1}-*ERG9*), and CAS15 (*tHMGR*, P_{HXT1}-*ERG20*, P_{ERG1}-*ERG9*). The corresponding OD_{600} values are represented by filled circles. Engineered strains were incubated in selective SC medium for 96 h (all data: mean ± SD, $n = 3$).

an ergosterol-sensitive promoter, previously shown to efficiently restrict *ERG9* expression levels in order to boost amorphadiene production (FPP-derived) (Yuan and Ching, 2015). We applied a similar strategy to improve diterpene synthesis, extending dynamic control to *ERG20,* given the need to boost the IPP and DMAPP supply. Indeed, in our hands, the dynamic control of *ERG9* alone did not have the remarkable improvement in sesquiterpene production that was reported by Yuan and Ching (2015). The strains harboring such a modification (*tHMGR*, P$_{ERG1}$-*ERG9*) produced amounts of casbene comparable to the reference strain (*tHMGR*), at least in plasmid-based assays. Moreover, they accumulated the lowest titer of GGOH and reached lower cell densities. Enhancement of the mevalonate pathway (by *tHMGR* overexpression), together with *ERG9* down-regulation in the absence of a sesquiterpene synthase or another draining route, possibly led to accumulation of toxic concentrations of pathway intermediates, such as FPP. The toxicity of isoprenoid precursors (including FPP) has previously been reported in *E. coli* (Martin et al., 2003; Sarria et al., 2014).

Dynamic control of *ERG20* by P$_{ERG1}$ showed instead to be a significant improvement, leading to a 3-fold increase in production, without negatively affecting growth. Titers reached even 4-fold higher values than the control strain when both *ERG9* and *ERG20* expression were controlled by P$_{ERG1}$. Downregulation of *ERG20* alone, similar to the *ERG9* downregulation reported by Yuan and Ching, most likely created a metabolic balance in which gene expression was adjusted to the cellular requirement for ergosterol biosynthesis. Once the cell sensed an excess of ergosterol, *ERG20* expression was decreased, resulting in redirection of the metabolic flux from IPP/DMAPP to GGPP and non-native diterpenoids. The concerted downregulation of both *ERG20* and *ERG9*, although leading to the highest casbene accumulation, might have resulted in an excessively restricted flux toward ergosterol which would explain the growth impairment observed upon dynamic control of both *ERG9* and *ERG20*, leading to a 27% reduced cell density.

Integration of the expression constructs for *PaGGPPS* and *RcCBS* in neutral loci of the yeast genome, together with dynamic control of *ERG20* by means of the glucose-sensitive promoter of *HXT1,* further enhanced production of casbene. P$_{HXT1}$ allows for moderate expression levels when glucose is present in the medium, but leads to gene repression when glucose concentration is low or absent. This promoter was previously used for dynamic control of *ERG20* expression in strains engineered for improved production of geraniol (Zhao et al., 2017). When used for dynamic control of *ERG20*, P$_{HXT1}$ had a beneficial effect on both production and growth compared to P$_{ERG1}$. Repression of expression by P$_{HXT1}$ might be stronger than the one imposed by P$_{ERG1}$, leading to a higher supply of DMAPP and IPP for the heterologous GGPP synthase, thus explaining the higher titers. Moreover, since repression of expression mediated by P$_{HXT1}$ occurs when glucose concentration is low, its control possibly maximized the carbon flux to ergosterol and biomass, thus explaining the higher cell densities. Combining dynamic control of *ERG20* by P$_{HXT1}$ with dynamic control of *ERG9* by P$_{ERG1}$ led to the highest casbene titer of 108.5 mg/L and worked

best to guarantee a proper flux distribution between cell growth and diterpene synthesis.

Proper adjustment and usage of carbon source sensitive promoters, to improve casbene or in general diterpenoid production, opens up new perspectives for an extensive optimization of the culturing conditions during fermentation that were beyond the scope of this study. For example, substantial improvements in isoprenoid biosynthesis have been reported when ethanol is provided as the carbon source (Westfall et al., 2012; Zhao et al., 2017). In the future, it would therefore be interesting to analyze the effect of combining feeding with ethanol or with mixture of different carbon sources with *ERG20* expression under control of P$_{HXT1}$ for casbene production.

Overall these results showed remarkable improvements in diterpene production via an engineered heterologous route toward synthesis of GGPP, combined with dynamic regulation of competing pathways. However, a substantial accumulation of GGOH could still be observed in all our engineered strains, indicating that not all of the GGPP produced is channeled toward casbene biosynthesis. This may be due to the low efficiency of the casbene synthase which, similarly to other enzymes of the secondary metabolism of plants, has not evolved to synthesize high titers of products. During preparation of this manuscript, a casbene production of 160 mg/L was reported in a yeast strain expressing multiple copies of casbene synthase constructs, with improved solubility achieved by protein tagging strategies (Wong et al., 2017). Titers in our strains could most likely similarly be improved by increasing the copy number of the casbene synthase gene. Moreover, fusion between the GGPP synthase *PaGGPPS* and the casbene synthase *RcCBS* could possibly offer an enzyme with improved solubility and a more efficient substrate channeling that mimics the natural chimera fusicoccadiene synthase. Expression of such a chimera may further increase production levels.

In conclusion, although further optimization is needed, the design developed here provides a valuable strategy for sustainable production of casbene derived chemicals. Moreover, the approach can easily be adapted to synthesis of other GGPP-derived compounds.

AUTHOR CONTRIBUTIONS

RC designed experiments, performed strain constructions, extractions, analysis of samples, and drafted the manuscript. DR established the analysis methods. YM helped in construction of the integration strains. RC and HH coordinated the study. HH reviewed and edited the manuscript. All authors read and approved the final manuscript.

FUNDING

ACKNOWLEDGMENTS

The content of this research was first reported and is adapted from the PhD thesis of RC (Callari, 2018).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe. 2018.00160/full#supplementary-material

REFERENCES

Bohlmann, J., and Keeling, C. I. (2008). Terpenoid biomaterials. *Plant J.* 54, 656–669. doi: 10.1111/j.1365-313X.2008.03449.x

Callari, R. (2018). *Pathway Engineering Towards Heterologous Production of Complex Diterpenoids in Yeast*. PhD thesis. Department of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen.

Caniard, A., Zerbe, P., Legrand, S., Cohade, A., Valot, N., Magnard, J. L., et al. (2012). Discovery and functional characterization of two diterpene synthases for sclareol biosynthesis in *Salvia sclarea* (L.) and their relevance for perfume manufacture. *BMC Plant Biol.* 12:119. doi: 10.1186/1471-2229-12-119

Chambon, C., Ladeveze, V., Servouse, M., Blanchard, L., Javelot, C., Vladescu, B., et al. (1991). Sterol pathway in yeast. Identification and properties of mutant strains defective in mevalonate diphosphate decarboxylase and farnesyl diphosphate synthetase. *Lipids* 26, 633–636. doi: 10.1007/BF02536428

Chen, M., Chou, W. K., Toyomasu, T., Cane, D. E., and Christianson, D. W. (2016). Structure and Function of Fusicoccadiene Synthase, a Hexameric Bifunctional Diterpene Synthase. *ACS Chem. Biol.* 11, 889–899. doi: 10.1021/acschembio.5b00960

Cheng, A.-X., Lou, Y.-G., Mao, Y.-B., Lu, S., Wang, L.-J., and Chen, X.-Y. (2007). Plant terpenoids: biosynthesis and ecological functions. *J. Integr. Plant Biol.* 49, 179–186. doi: 10.1111/j.1744-7909.2007.00395.x

Donald, K. A., Hampton, R. Y., and Fritz, I. B. (1997). Effects of overproduction of the catalytic domain of 3-hydroxy-3-methylglutaryl coenzyme A reductase on squalene synthesis in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 63, 3341–3344.

Doseyici, S., Mehmetoglu, I., Toker, A., Yerlikaya, F. H., and Erbay, E. (2014). The effects of forskolin and rolipram on cAMP, cGMP and free fatty acid levels in diet induced obesity. *Biotech. Histochem.* 89, 388–392. doi: 10.3109/10520295.2014.883463

Emanuelsson, O., Nielsen, H., and von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* 8, 978–984. doi: 10.1110/ps.8.5.978

Faulkner, A., Chen, X., Rush, J., Horazdovsky, B., Waechter, C. J., Carman, G. M., et al. (1999). The LPP1 and DPP1 gene products account for most of the isoprenoid phosphate phosphatase activities in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 274, 14831–14837. doi: 10.1074/jbc.274.21.14831

Fidler, B., and Goldberg, T. (2014). Ingenol mebutate gel (picato): a novel agent for the treatment of actinic keratoses. *PT* 39, 40–46.

Fischer, M. J., Meyer, S., Claudel, P., Bergdoll, M., and Karst, F. (2011). Metabolic engineering of monoterpene synthesis in yeast. *Biotechnol. Bioeng.* 108, 1883–1892. doi: 10.1002/bit.23129

Gietz, R. D., and Schiestl, R. H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* 2, 31–34. doi: 10.1038/nprot.2007.13

Goyal, S. K., Samsher, and Goyal, R. K. (2010). Stevia (*Stevia rebaudiana*) a bio-sweetener: a review. *Int. J. Food Sci. Nutr.* 61, 1–10. doi: 10.3109/09637480903193049

Green, M. R., and Sambrook, J. (2012). *Molecular Cloning: A Laboratory Manual: Three-Volume Set*, 4th edn. Cold Spring Harbor Laboratory Press.

Gueldener, U., Heinisch, J., Koehler, G. J., Voss, D., and Hegemann, J. H. (2002). A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast. *Nucleic Acids Res.* 30:e23. doi: 10.1093/nar/30.6.e23

Hamann, T., and Møller, B. L. (2007). Improved cloning and expression of cytochrome P450s and cytochrome P450 reductase in yeast. *Protein Expr. Purif.* 56, 121–127. doi: 10.1016/j.pep.2007.06.007

Howat, S., Park, B., Oh, I. S., Jin, Y. W., Lee, E. K., and Loake, G. J. (2014). Paclitaxel: biosynthesis, production and future prospects. *N. Biotechnol.* 31, 242–245. doi: 10.1016/j.nbt.2014.02.010

Iadarola, M. J., and Gonnella, G. L. (2013). Resiniferatoxin for pain treatment: an interventional approach to personalized pain medicine. *Open Pain J.* 6, 95–107. doi: 10.2174/1876386301306010095

Ignea, C., Pontini, M., Maffei, M. E., Makris, A. M., and Kampranis, S. C. (2014). Engineering monoterpene production in yeast using a synthetic dominant negative geranyl diphosphate synthase. *ACS Synth. Biol.* 3, 298–306. doi: 10.1021/sb400115e

Ignea, C., Trikka, F. A., Nikolaidis, A. K., Georgantea, P., Ioannou, E., Loupassaki, S., et al. (2015). Efficient diterpene production in yeast by engineering Erg20p into a geranylgeranyl diphosphate synthase. *Metab. Eng.* 27, 65–75. doi: 10.1016/j.ymben.2014.10.008

Jiang, Y., Proteau, P., Poulter, D., and Ferro-Novick, S. (1995). BTS1 encodes a geranylgeranyl diphosphate synthase in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 270, 21793–21799. doi: 10.1074/jbc.270.37.21793

Kirby, J., and Keasling, J. D. (2009). Biosynthesis of plant isoprenoids: perspectives for microbial engineering. *Annu. Rev. Plant Biol.* 60, 335–355. doi: 10.1146/annurev.arplant.043008.091955

Kirby, J., Nishimoto, M., Park, J. G., Withers, S. T., Nowroozi, F., Behrendt, D., et al. (2010). Cloning of casbene and neocembrene synthases from Euphorbiaceae plants and expression in *Saccharomyces cerevisiae*. *Phytochemistry* 71, 1466–1473. doi: 10.1016/j.phytochem.2010.06.001

Kupchan, S. M., Sigel, C. W., Matz, M. J., Saenz Renauld, J. A., Haltiwanger, R. C., and Bryan, R. F. (1970). Jatrophone, a novel macrocyclic diterpenoid tumor inhibitor from *Jatropha gossypiifolia*. *J. Am. Chem. Soc.* 92, 4476–4477. doi: 10.1021/ja00717a066

Lin, M., Tang, S., Zhang, C., Chen, H., Huang, W., Liu, Y., et al. (2017). Euphorbia factor L2 induces apoptosis in A549 cells through the mitochondrial pathway. *Acta Pharm. Sin B* 7, 59–64. doi: 10.1016/j.apsb.2016.06.008

Martin, V. J., Pitera, D. J., Withers, S. T., Newman, J. D., and Keasling, J. D. (2003). Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* 21, 796–802. doi: 10.1038/nbt833

Mau, C. J., and West, C. A. (1994). Cloning of casbene synthase cDNA: evidence for conserved structural features among terpenoid cyclases in plants. *Proc. Natl. Acad. Sci. U.S.A.* 91, 8497–8501. doi: 10.1073/pnas.91.18.8497

Miana, G. A., Riaz, M., Shahzad-ul-Hussan, S., Paracha, R. Z., and Paracha, U. Z. (2015). Prostratin: an overview. *Mini Rev. Med. Chem.* 15, 1122–1130. doi: 10.2174/1389557515666150511154108

Paradise, E. M., Kirby, J., Chan, R., and Keasling, J. D. (2008). Redirection of flux through the FPP branch-point in *Saccharomyces cerevisiae* by down-regulating squalene synthase. *Biotechnol. Bioeng.* 100, 371–378. doi: 10.1002/bit.21766

Parks, L. W., Smith, S. J., and Crowley, J. H. (1995). Biochemical and physiological effects of sterol alterations in yeast—A review. *Lipids* 30, 227–230. doi: 10.1007/BF02537825

Sarria, S., Wong, B., García Martín, H., Keasling, J. D., and Peralta-Yahya, P. (2014). Microbial synthesis of pinene. *ACS Synth. Biol.* 3, 466–475. doi: 10.1021/sb4001382

Scalcinati, G., Knuf, C., Partow, S., Chen, Y., Maury, J., Schalk, M., et al. (2012). Dynamic control of gene expression in *Saccharomyces cerevisiae* engineered for the production of plant sesquitepene α-santalene in a fed-batch mode. *Metab. Eng.* 14, 91–103. doi: 10.1016/j.ymben.2012.01.007

Song, T. Q., Ding, M. Z., Zhai, F., Liu, D., Liu, H., Xiao, W. H., et al. (2017). Engineering *Saccharomyces cerevisiae* for geranylgeraniol overproduction by combinatorial design. *Sci. Rep.* 7:14991. doi: 10.1038/s41598-017-15005-4

Theoduloz, C., Rodríguez, J. A., Pertino, M., and Schmeda-Hirschmann, G. (2009). Antiproliferative activity of the diterpenes jatrophone and jatropholone and their derivatives. *Planta Med.* 75, 1520–1522. doi: 10.1055/s-0029-1185834

Tholl, D. (2015). Biosynthesis and biological functions of terpenoids in plants. *Adv. Biochem. Eng. Biotechnol.* 148, 63–106. doi: 10.1007/10_2014_295

Tokuhiro, K., Muramatsu, M., Ohto, C., Kawaguchi, T., Obata, S., Muramoto, N., et al. (2009). Overproduction of geranylgeraniol by metabolically engineered *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 75, 5536–5543. doi: 10.1128/AEM.00277-09

Toyomasu, T., Tsukahara, M., Kaneko, A., Niida, R., Mitsuhashi, W., Dairi, T., et al. (2007). Fusicoccins are biosynthesized by an unusual chimera diterpene synthase in fungi. *Proc. Natl. Acad. Sci. U.S.A.* 104, 3084–3088. doi: 10.1073/pnas.0608426104

Vandermoten, S., Haubruge, E., and Cusson, M. (2009). New insights into short-chain prenyltransferases: structural features, evolutionary history and potential for selective inhibition. *Cell. Mol. Life Sci.* 66, 3685–3695. doi: 10.1007/s00018-009-0100-9

Westfall, P. J., Pitera, D. J., Lenihan, J. R., Eng, D., Woolard, F. X., Regentin, R., et al. (2012). Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proc. Natl. Acad. Sci. U.S.A.* 109, E111–E118. doi: 10.1073/pnas.1110740109

Wong, J., de Rond, T., d'Espaux, L., van der Horst, C., Dev, I., Rios-Solis, L., et al. (2017). High-titer production of lathyrane diterpenoids from sugar by engineered *Saccharomyces cerevisiae*. *Metab. Eng.* 45, 142–148. doi: 10.1016/j.ymben.2017.12.007

Yuan, J., and Ching, C. B. (2015). Dynamic control of *ERG9* expression for improved amorpha-4,11-diene production in *Saccharomyces cerevisiae*. *Microb. Cell Fact.* 14:38. doi: 10.1186/s12934-015-0220-x

Zerbe, P., and Bohlmann, J. (2015). Enzymes for synthetic biology of ambroxide-related diterpenoid fragrance compounds. *Adv. Biochem. Eng. Biotechnol.* 148, 427–447. doi: 10.1007/10_2015_308

Zhao, J., Li, C., Zhang, Y., Shen, Y., Hou, J., and Bao, X. (2017). Dynamic control of *ERG20* expression combined with minimized endogenous downstream metabolism contributes to the improvement of geraniol production in *Saccharomyces cerevisiae*. *Microb. Cell Fact.* 16:17. doi: 10.1186/s12934-017-0641-9

Zhou, Y. J., Gao, W., Rong, Q., Jin, G., Chu, H., Liu, W., et al. (2012). Modular pathway engineering of diterpenoid synthases and the mevalonic acid pathway for miltiradiene production. *J. Am. Chem. Soc.* 134, 3234–3241. doi: 10.1021/ja2114486

Check for
updates

Models for Cell-Free Synthetic Biology: Make Prototyping Easier, Better, and Faster

Mathilde Koch[1], Jean-Loup Faulon[1,2,3] and Olivier Borkowski[1,2]*

[1] Micalis Institute, INRA, AgroParisTech, University of Paris-Saclay, Jouy-en-Josas, France, [2] Systems and Synthetic Biology Lab, CEA, CNRS, UMR 8030, Genomics Metabolics, University Paris-Saclay, Évry, France, [3] SYNBIOCHEM Center, School of Chemistry, Manchester Institute of Biotechnology, University of Manchester, Manchester, United Kingdom

Cell-free TX-TL is an increasingly mature and useful platform for prototyping, testing, and engineering biological parts and systems. However, to fully accomplish the promises of synthetic biology, mathematical models are required to facilitate the design and predict the behavior of biological components in cell-free extracts. We review here the latest models accounting for transcription, translation, competition, and depletion of resources as well as genome scale models for lysate-based cell-free TX-TL systems, including their current limitations. These models will have to find ways to account for batch-to-batch variability before being quantitatively predictive in cell-free lysate-based platforms.

Keywords: mathematical model, cell-free, prototyping, resource competition, transcription, translation, metabolism

INTRODUCTION

All the processes required to produce proteins in bacteria can be performed by adding DNA to a cell-free platform. After lysis of living cells, transcription, translation, degradation, and protein folding continue to operate as they do *in vivo* (Hodgman and Jewett, 2012; Sun et al., 2013; Takahashi et al., 2015a). Metabolic pathways like glycolysis or pentose phosphate pathway remain active and are used to regenerate ATP and maximize protein production over time (Kim and Swartz, 2001; Calhoun and Swartz, 2005). Protein production outside of the cell simplifies gene expression with well-defined parameters, easy to control inputs, faster time scale, and less numerous unknown interactions. As a result, many laboratories use cell-free as a prototyping platform to characterize expression of single proteins or complex metabolic pathways (Takahashi et al., 2015b; Wu et al., 2017; Borkowski et al., 2018). Mathematical models dedicated to cell-free emerged to predict protein production and understand the limits of this new platform. Cell-free properties are close to living organisms as the same processes take place in both systems, yet significant differences exist. For example, molecular crowding (Spruijt et al., 2014) and resources distribution (Sun et al., 2013) are significantly altered in cell-free and there is no resource competition with the host. Such differences oblige synthetic biologists to adapt the models already developed for living cells. This short review focuses on the recent deterministic models developed to understand lysate-based cell-free platforms and used to predict the behavior of simple or complex pathways (**Table 1**). Those models pave the way for efficient metabolic engineering in the emerging field of cell-free synthetic biology.

TABLE 1 | Deterministic models developed to understand cell-free.

Modeling strategy	Problems tackled	Level of detail	Description	Strength	Weakness	References
ODE	Protein production in cell-free	Simple*	Michaelis-Menten for the translation processes, as well as for degradation	Simple, quantitative	Parameters values are experiment dependent	Karzbrun et al., 2011
ODE	Protein production in cell-free	Simple*	Simple description of the transcription and translation processes. One parameter summarize each	Simple, qualitative	Parameters values are experiment dependent	Stögbauer et al., 2012; Siegal-Gaskins et al., 2014
ODE	Protein production in cell-free. Resource competition in cell-free	Complex*	Detailed modeling: Particular focus on the translation process and the competition for the translation machinery	Generalization power quantitative	Parameters values are experiment dependent	Underwood et al., 2005; Borkowski et al., 2018
ODE	Protein production in cell-free. Resource competition in cell-free	Simple*	Binding, unbinding and elongation bundled in one parameter. Accounts for number of RNAP, ribosomes, and promoter/RBS strengths	Easily adaptable to a new situation or phenomena	Necessitates parameter determination for new situation	Gyorgy and Murray, 2016; Halter et al., 2018; Voyvodic et al., 2018
ODE	Protein production in cell-free. Resource depletion in cell-free	Complex*	Accounts for all known processes, including NTP consumption and degradation	Models biochemical phenonema precisely so generalisable	Important parameter identification/estimation needed	Siegal-Gaskins et al., 2013; Tuza et al., 2013; Nieß et al., 2017; Matsuura et al., 2018; Moore et al., 2018
Constraint based	Protein production with cell-free. Metabolism in cell-free	Complex*	Modification of *E. Coli* metabolic model to account for cell-free constraints	Accounts for the full metabolism, can use constraints based methods such as FBA	Defining the objectives, require deeper knowledge of reactions in cell-free	Vilkhovoy et al., 2018

Simple stands for one protein produced and limited amount of parameters (<10) Complex stands for more than one protein produced or/and large amount of parameters (more than 10).

TRANSLATION AND TRANSCRIPTION PROCESSES IN CELL-FREE

Lysate-based cell-free consists of a crude cell extract supplemented with buffer, amino acids, NTP, NAD, PEG, tRNA, and metabolic intermediates (Sun et al., 2013). A major advantage of cell-free is the absence of host regulations (Hodgman and Jewett, 2012), allowing circuits to function in isolation and an easy quantitative description of gene expression. A constitutively expressed gene in cell-free exhibits specific patterns at the translation and transcription levels. Protein production can be divided in 4 phases: in phase 1, the production rate increases over time, in phase 2, the production rate is constant during a 30 min/1 h, in phase 3 the production rate decreases slowly and eventually in phase 4 the production rate is null (Siegal-Gaskins et al., 2014; **Figure 1A**). A similar 4 phases pattern is observed with the mRNA concentration (Siegal-Gaskins et al., 2014; **Figure 1B**). ODE Models describing transcription, translation, mRNA, and protein degradation processes at various scales have been successfully used to predict mRNA and protein dynamics in lysate-based systems (Stögbauer et al., 2012; Tuza et al., 2013; Siegal-Gaskins et al., 2014; Nieß et al., 2017; Borkowski et al., 2018; Matsuura et al., 2018; Moore et al., 2018). DNA concentrations are usually considered constant: degradation is neglected as plasmid DNA or protected

linear template are used, and replication is considered not to happen since no dNTPs are added to the reaction mix.

Numerous models, with varying degrees of complexity, try to reproduce those production phases observed in cell-free reactions.

A simple model based on only 4 reactions and 10 parameters is sufficient to fit the full mRNA and protein dynamics during the first hours of reaction (Karzbrun et al., 2011). The transcription process is reduced to one step in which the RNA polymerase binds to the DNA; the rate of mRNA production depends only this binding rate and the DNA length. Similarly, the translation process is described as one binding step of the ribosome on the mRNA with the rate of protein production depending only on the binding rate and the mRNA length. This model is appropriate for the first hours before the consumption of resources and/or the waste accumulation (e.g., ATP degradation, toxic metabolites…) cause the reaction to stop (Siegal-Gaskins et al., 2014). A simple way to simulate the slow decay in synthesis is the consumption of the NTP over time. The transcription reaction slows down and eventually stops (Stögbauer et al., 2012; Tuza et al., 2013). The decrease in the NTP concentration (Kim and Swartz, 2001; Jewett et al., 2008; Moore et al., 2018) is an efficient method to obtain a decreasing transcription over time and simulate protein and mRNA production in cell-free but no experimental data either confirms or denies this approach. The accumulation of inactive
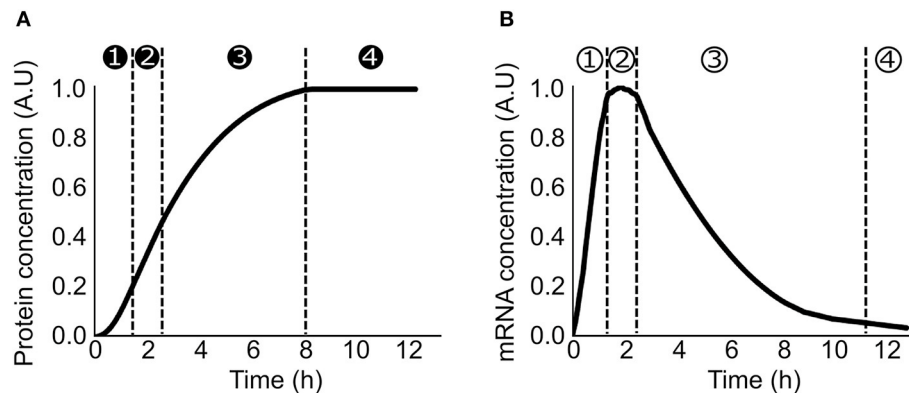
FIGURE 1 | Production of a constitutively expressed gene in cell-free **(A)** Protein concentration over time in a cell-free platform. ❶ the production rate increase over time ❷ the production rate is constant ❸ the production rate decreases ❹ the production rate is null **(B)** mRNA concentration over time in a cell-free platform ① the concentration increases over time ② the concentration is constant ③ the concentration decreases over time ④ the concentration is close to zero.

RNA polymerase/ribosome (Failmezger et al., 2016; Moore et al., 2018), accumulation of toxic metabolites (Kim and Swartz, 2000), or increase of the relative RNases concentration compare to the total amount of mRNA (Siegal-Gaskins et al., 2014) are possible other explanations of the arrest of protein production after 8 h. Voyvodic et al. (2018) added terms corresponding to decreasing resources for protein production and accumulation of toxic by-products as a reduction in production rates parametrized by a Michaelis-Menten like ratio, as an elegant way to account for the slowing production rate. As all cell-free models trying to account for the end of production after 8 h, the main issue is identifying the exact cause for decreasing production.

Models using Michaelis–Menten kinetics also succeed to capture protein production pattern in cell-free (Stögbauer et al., 2012). Those models precisely captured the observable mRNA and proteins dynamics in cell-free while remaining relatively coarse-grained. The model of Tuza et al. add extra steps in the transcription (and translation) process with a reversible binding of the RNA polymerase (ribosome) on the DNA (mRNA) followed by a reversible binding of the first NTP (amino acid) and eventually an irreversible elongation step. Moore et al. also developed a model accounting for reversible binding, unbinding and elongation steps, sharing the NTP energy source. Those more detailed descriptions of the transcription and translation processes lead to accurate predictions of the data obtained in cell-free and capture additional properties (Stögbauer et al., 2012; Tuza et al., 2013; Borkowski et al., 2018; Moore et al., 2018). For example, the non-additive cost of protein production when several genes are expressed requires higher level of complexity to be predicted (Borkowski et al., 2018).

While all models presented in this section described transcription and translation processes, the main challenge they faced is proper parameter identification, as biochemical parameters can vary widely from batch to batch and from *in vivo* to cell-free systems. Currently, models often used components concentration measured *in vivo* and estimated their concentration based on the dilution factor of the *E. coli*

cytoplasm after the lysate extraction protocol, which is not entirely satisfactory.

RESOURCE COMPETITION IN CELL-FREE

Resource competition is an important phenomenon that impacts circuit behavior in cell-free systems and should be accounted for in modeling approaches.

As a fixed amount of resources is present in the cell-free extract, competition has been measured between synthetic circuits (Siegal-Gaskins et al., 2014; Borkowski et al., 2018; Moore et al., 2018). Some of the previously described models take into account the limitation of each resource and include a fixed amount of transcription and translation machineries to predict the impact of resource competition in cell-free (**Figure 2A**; Siegal-Gaskins et al., 2013; Borkowski et al., 2018; Moore et al., 2018).

A maximal protein production is measured after a few hours before resources depletion and degradation (**Figure 1**). This upper limit on production rate is the result of one or several limited resources (RNA polymerase, NTP, ribosome, elongation factors, amino acids, chaperone, tRNA synthetase, or tRNA). DNA, NTP, amino acids, and T7 RNA polymerase are directly added to the mix so their impacts on the protein production can be easily measured. Increasing DNA concentration leads to an increase of protein production until a saturation point is reached (Siegal-Gaskins et al., 2014; Borkowski et al., 2018; Voyvodic et al., 2018), and toxicity can be observed with high DNA concentration (Borkowski et al., 2018). T7 polymerase (Siegal-Gaskins et al., 2014), amino acids, tRNA and nucleotides (Shin and Noireaux, 2010) are present in excess in the cell-free mix causing no noticeable competition for these resources. Eventually, High NTP concentration negatively affects the translation process (Nagaraj et al., 2017). Natural transcription and translation machineries are less controlled as they are added via the crude extract. Indirect measurements using competition for resources between two plasmids are

FIGURE 2 | Resource competition in cell-free **(A)** Competition for resources between two genes expressed in cell-free. The transcription and translation machineries amounts are fixed **(B)** Production of energy and amino acids via the core metabolic network describing glycolysis, pentose phosphate pathway, the TCA cycle and the Entner-Doudoroff pathway.
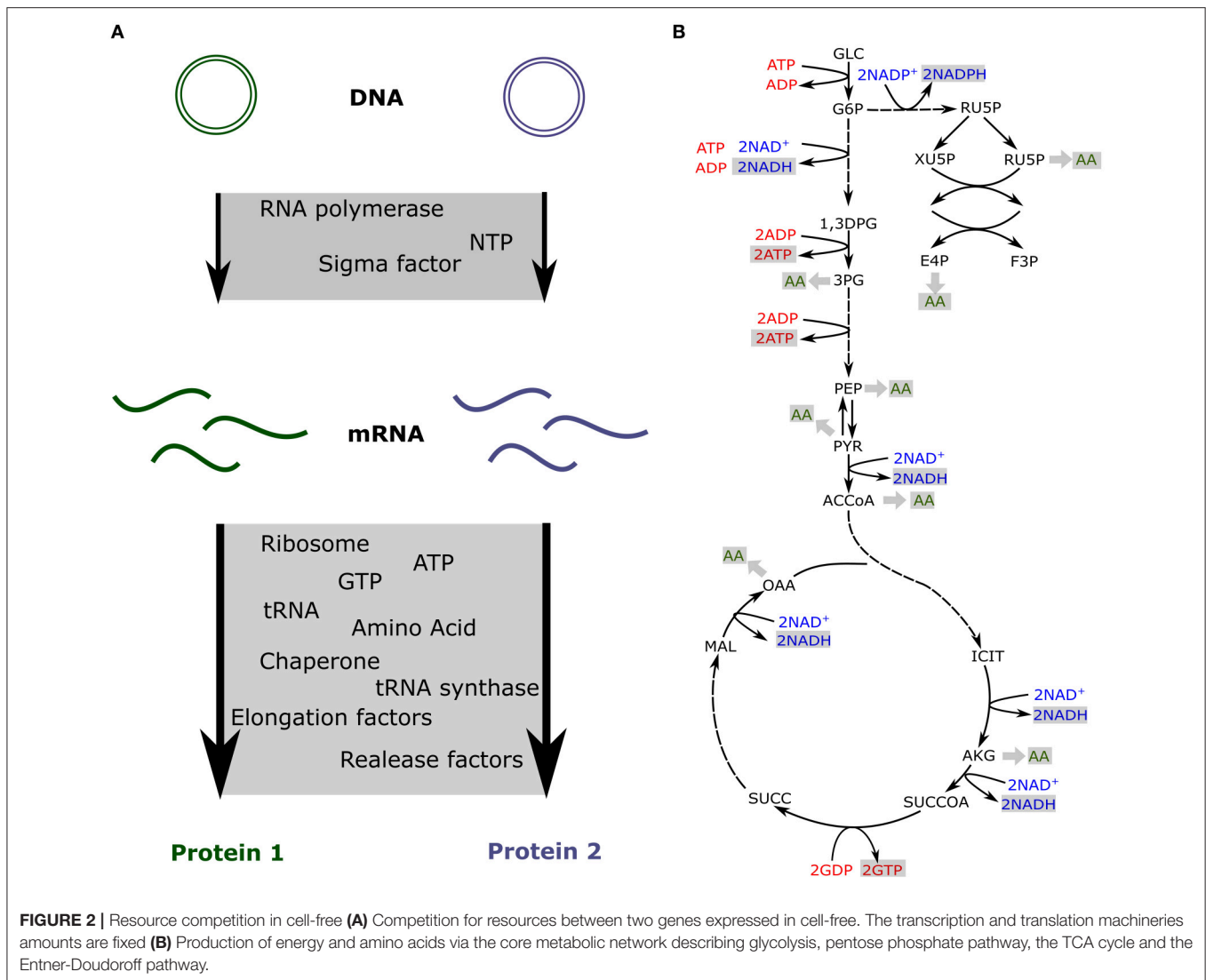
used to deduce competition for transcription and translation machineries (Underwood et al., 2005; Siegal-Gaskins et al., 2014; Gyorgy and Murray, 2016; Borkowski et al., 2018; Moore et al., 2018). The main source of competition can be at the transcription and/or translation level depending of the extract and the level of protein produced (Underwood et al., 2005; Li et al., 2014; Siegal-Gaskins et al., 2014; Gyorgy and Murray, 2016; Borkowski et al., 2018; Moore et al., 2018; Voyvodic et al., 2018). Parametrization of the models using the appropriate RNA polymerase and ribosomes concentrations and binding/unbinding rates allows an accurate description of the resource competition and to fit properly the production of several proteins expressed concurrently in cell-free (Siegal-Gaskins et al., 2014; Borkowski et al., 2018; Moore et al., 2018). Accounting for RNAP and ribosomes sharing between parts can also be leveraged to minimize the number of experiments required to fit parameters and obtain a predictive model (Halter et al., 2018). While not the main focus of this review, parameter estimation or identification is a major hurdle of

detailed models, and techniques from systems biology (e.g., Lillacci and Khammash, 2010) can be used to tackle this issue.

The models presented in this sections, while being able to account for transcription, translation and resource competition from a lack of generalizability due both to variability in experimental conditions as batches can differ greatly, and to scarcity of biochemical work measuring those parameters in cell-free setting as has been estimated from *in vivo* measurements.

METABOLISM IN CELL-FREE

The models presented in this section are constraint based, so as to take the whole metabolism into account and not the circuit in isolation as done in the previous sections.

In cell-free platforms, translation, and transcription are not the only active processes. Glycolysis, pentose phosphate pathway, TCA cycle, Entner-Doudoroff pathway, and amino

acid biosynthesis are still producing ATP, reducing equivalents, and amino acids (Kim and Swartz, 2001; Jewett and Swartz, 2004a,b). The previously described models account for the resources competition for a fixed amount of transcription and translation resources and usually do not include any metabolite production or consumption. Such an approach is quite limiting for metabolic pathway prototyping as those circuits also compete for metabolites (Wu et al., 2017; Borkowski et al., 2018). Constraint-based models have been used to simulate metabolites production and consumption when various proteins are produced at different levels (Vilkhovoy et al., 2018; **Figure 2B**). This model coupled transcription and translation processes with the availability of metabolic resources. Flux balance analysis was adapted to cell-free conditions with the objective function being the maximization of the protein translation rate. Growth associated reactions were removed and cell-free specific deletions were added from *E. coli* metabolic model, leading to 264 reactions and 146 species (Vilkhovoy et al., 2018). The stoichiometric network was adjusted to cell-free and fluxes were constrained by experimental measurements of glucose, nucleotides, amino and organic acid consumption and production rates. The transcription and translation were bound by Michaelis-Menten formula with a maximum transcription and translation rate depending on RNA polymerase concentration, RNA polymerase elongation rate, gene length, promoter strength, and the ribosome concentration, a polysome amplification constant, the ribosome elongation rate, the protein length, and the RBS strength, respectively. The energy efficiency was calculated using the ATP cost by transcription and translation processes. Transcription and translation rates are subject to resource constraints encoded by the metabolic network (**Figure 2B**). This model efficiently predicts proteins production and simulates optimal flux distribution in cell-free metabolic network. It makes predictions possible for metabolic engineering in cell-free as metabolites produced or consumed by a pathway will be accounted for via its energy efficiency.

Constraint based modeling for cell-free systems is an interesting field that would need further developments from the research community, both to include cell-free specific constraints and reactions, as well as to account for dynamic behavior such as metabolite exhaustion in cell-free systems.

CONCLUSION

Cell-free appeared as the ideal platform for circuits prototyping. It accelerates characterization and avoids the impact of the host on the circuit behavior. Models can be easily parametrized and predictions are easier and more accurate than *in vivo*, for qualitative behavior. Parametrization for quantitative behavior can be tackled using techniques from systems biology. Simple models succeed to accurately predict the simultaneous production levels of multiple proteins and the competition for the limited amount of resources in cell-free. A certain level of complexity is necessary to capture competition for metabolites but produces a powerful tool for metabolic engineering. The main limit for lysate-based cell-free in metabolic engineering and modeling remains extract preparation: extract efficiency can differ strongly depending on the experimentalist leading to variability of protein production and necessity of robust controls for each new batch, as well as uncertain parameters that vary with each batch for the modeler. Preliminary control of the extract quality and tuning of model parameters on each batch is required to obtain accurate predictions and precludes generalization. A way forward to both increase reproducibility and predictive modeling in cell-free systems would be a higher degree of automation in the extract production providing robust lysate preparation at affordable price.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

REFERENCES

Borkowski, O., Bricio, C., Murgiano, M., Rothschild-Mancinelli, B., Stan, G. B., and Ellis, T. (2018). Cell-free prediction of protein expression costs for growing cells. *Nat. Commun.* 9:1457. doi: 10.1038/s41467-018-03970-x

Calhoun, K. A., and Swartz, J. R. (2005). Energizing cell-free protein synthesis with glucose metabolism. *Biotechnol. Bioeng.* 90, 606–613 doi: 10.1002/bit.20449

Failmezger, J., Nitschel, R., Sánchez-Kopper, A., Kraml, M., and Siemann-Herzberg, M. (2016). Site-specific cleavage of ribosomal RNA in *Escherichia coli*-based cell-free protein synthesis systems. *PLoS ONE* 11:e0168764. doi: 10.1371/journal.pone.0168764

Gyorgy, A., and Murray, R. M. (2016). "Quantifying resource competition and its effects in the TX-TL system," in *2016 IEEE 55th Conference on Decision and Control, CDC 2016* (Las Vegas, NV). doi: 10.1109/CDC.2016.7798775

Halter, W., Allgower, F., Murray, R. M., and Gyorgy, A. (2018). "Optimal experiment design and leveraging competition for shared resources in cell-free extracts," in *2018 Conference on Decision and Control* (Miami Beach, FL).

Hodgman, C. E., and Jewett, M. C. (2012). Cell-free synthetic biology: thinking outside the cell. *Metab. Eng.* 14, 261–269 doi: 10.1016/j.ymben.2011.09.002

Jewett, M. C., Calhoun, K. A., Voloshin, A., Wuu, J. J., and Swartz, J. R. (2008). An integrated cell-free metabolic platform for protein production and synthetic biology. *Mol. Syst. Biol.* 4:220. doi: 10.1038/msb2008.57

Jewett, M. C., and Swartz, J. R. (2004a). Mimicking the *Escherichia coli* cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnol. Bioeng.* 86, 19–26 doi: 10.1002/bit.20026

Jewett, M. C., and Swartz, J. R. (2004b). Substrate replenishment extends protein synthesis with an *in vitro* translation system designed to mimic the cytoplasm. *Biotechnol. Bioeng.* 87, 465–472 doi: 10.1002/bit.20139

Karzbrun, E., Shin, J., Bar-Ziv, R. H., and Noireaux, V. (2011). Coarse-grained dynamics of protein synthesis in a cell-free system. *Phys. Rev. Lett.* 106:048104. doi: 10.1103/PhysRevLett.106.048104

Kim, D. M., and Swartz, J. R. (2000). Prolonging cell-free protein synthesis by selective reagent additions. *Biotechnol. Prog.* 16, 385–390. doi: 10.1021/bp000031y

Kim, D. M., and Swartz, J. R. (2001). Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis. *Biotechnol. Bioeng.* 74, 309–316. doi: 10.1002/bit.1121

Li, J., Gu, L., Aach, J., and Church, G. M. (2014). Improved cell-free RNA and protein synthesis system. *PLoS ONE* 9:e106232. doi: 10.1371/journal.pone.0106232

Lillacci, G., and Khammash, M. (2010). Parameter estimation and model selection in computational biology. *PLoS Comput. Biol.* 6:e1000696. doi: 10.1371/journal.pcbi.1000696

Matsuura, T., Hosoda, K., and Shimizu, Y. (2018). Robustness of a reconstituted *Escherichia coli* protein translation system analyzed by computational modeling. *ACS Synth. Biol.* 7, 1964–1972 doi: 10.1021/acssynbio.8b 00228

Moore, S. J., MacDonald, J. T., Wienecke, S., Ishwarbhai, A., Tsipa, A., Aw, R., et al. (2018). Rapid acquisition and model-based analysis of cell-free transcription–translation reactions from nonmodel bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 115, E4340–E4349. doi: 10.1073/pnas.1715806115

Nagaraj, V. J., Greene, J. M., Sengupta, A. M., and Sontag, E. D. (2017). Translation inhibition and resource balance in the TX-TL cell-free gene expression system. *Synthet. Biol.* 2:ysx005. doi: 10.1093/synbio/ysx005

Nieß, A., Failmezger, J., Kuschel, M., Siemann-Herzberg, M., and Takors, R. (2017). Experimentally validated model enables debottlenecking of *in vitro* protein synthesis and identifies a control shift under *in vivo* conditions. *ACS Synth. Biol.* 6, 1913–1921 doi: 10.1021/acssynbio.7b00117

Shin, J., and Noireaux, V. (2010). Efficient cell-free expression with the endogenous *E. coli* RNA polymerase and sigma factor 70. *J. Biol. Eng.* 4:8. doi: 10.1186/1754-1611-4-8

Siegal-Gaskins, D., Noireaux, V., and Murray, R. M. (2013). "Biomolecular resource utilization in elementary cell-free gene circuits," in *American Control Conference* (Washington, DC).

Siegal-Gaskins, D., Tuza, Z. A., Kim, J., Noireaux, V., and Murray, R. M. (2014). Resource usage and gene circuit performance characterization in a cell-free "breadboard." *ACS Synth. Biol.* 3, 416–425. doi: 10.1021/sb40 0203p

Spruijt, E., Sokolova, E., and Huck, W. T. (2014). Complexity of molecular crowding in cell-free enzymatic reaction networks. *Nat. Nanotechnol.* 9, 406–407. doi: 10.1038/nnano.2014.110

Stögbauer, T., Windhager, L., Zimmer, R., and Rädler, J. O. (2012). Experiment and mathematical modeling of gene expression dynamics in a cell-free system. *Integr. Biol.* 4, 494–501. doi: 10.1039/c2ib00102k

Sun, Z. Z., Hayes, C. A., Shin, J., Caschera, F., Murray, R. M., and Noireaux, V. (2013). Protocols for Implementing an *Escherichia coli* based TX-TL cell-free expression system for synthetic biology. *J. Vis. Exp.* 79:e50762. doi: 10.3791/50762

Takahashi, M. K., Chappell, J., Hayes, C. A., Sun, Z. Z., Kim, J., Singhal, V., et al. (2015a). Rapidly characterizing the fast dynamics of RNA genetic circuitry with cell-free transcription–translation (TX-TL) systems. *ACS Synth. Biol.* 4, 503–515. doi: 10.1021/sb400206c

Takahashi, M. K., Hayes, C. A., Chappell, J., Sun, Z. Z., Murray, R. M., Noireaux, V., et al. (2015b). Characterizing and prototyping genetic networks with cell-free transcription-translation reactions. *Methods* 86, 60–72. doi: 10.1016/j.ymeth.2015.05.020

Tuza, Z. A., Singhal, V., Kim, J., and Murray, R. M. (2013). An in silico modeling toolbox for rapid prototyping of circuits in a biomolecular "breadboard" system. in *Proceedings of the IEEE Conference on Decision and Control* (Florence).

Underwood, K. A., Swartz, J. R., and Puglisi, J. D. (2005). Quantitative polysome analysis identifies limitations in bacterial cell-free protein synthesis. *Biotechnol. Bioeng.* 91, 425–435. doi: 10.1002/bit.20529

Vilkhovoy, M., Horvath, N., Shih, C.-H., Wayman, J., Swartz, J., and Varner, J. D. (2018). Sequence specific modeling of *E. coli* cell-free protein synthesis. *ACS Synth Biol.* 7, 1844–1857 doi: 10.1021/acssynbio.7b00465

Voyvodic, P., Amir, P., Mathilde, K., Jean-Loup Faulon, J., and Bonnet, J. (2018). Plug-and-play metabolic transducers expand the chemical detection space of cell-free biosensors. *bioRxiv [Preprint]*. doi: 10.1101/397315

Wu, Y. Y., Sato, H., Huang, H., Culler, S. J., Khandurina, J., Harish, et al. (2017). System-level studies of a cell-free transcription-translation platform for metabolic engineering. *biorxiv [Preprint]*. doi: 10.1101/172007

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Check for updates

How Synthetic Biology and Metabolic Engineering Can Boost the Generation of Artificial Blood Using Microbial Production Hosts

August T. Frost[1], Irene H. Jacobsen[1], Andreas Worberg[2] and José L. Martínez[1]*

[1] Section of Synthetic Biology, Department of Biotechnology and Biomedicine, Technical University of Denmark (DTU), Lyngby, Denmark, [2] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark (DTU), Lyngby, Denmark

Hemoglobin is an essential protein to the human body as it transports oxygen to organs and tissues through the bloodstream (Looker et al., 1992). In recent years, there has been an increasing concern regarding the global supply of this vital protein, as blood availability cannot currently meet the high demands in many developing countries. There are, in addition, several risks associated with conventional blood transfusions such as the presence of blood-borne viruses like HIV and Hepatitis. These risks along with some limitations are presented in **Figure 1** (Kim and Greenburg, 2013; Martínez et al., 2015). As an alternative, producing hemoglobin recombinantly will eliminate the obstacles, since hemoglobin-based oxygen carriers are pathogen-free, have a longer shelf-life, are universally compatible and the supply can be adjusted to meet the demands (Chakane, 2017). A stable, safe, and most importantly affordable production, will lead to high availability of blood to the world population, and hence reduce global inequality, which is a focus point of the World Health Organization for the millennium (WHO, 2018). Synthetic biology and metabolic engineering have created a unique opportunity to construct promising candidates for hemoglobin production (Liu et al., 2014; Martínez et al., 2016). This review sets out to describe the recent advances in recombinant hemoglobin production, the societal and the economic impact along with the challenges that researchers will face in the coming years, such as low productivity, degradation, and difficulties in scale-up. The challenges are diverse and complex but with the powerful tools provided by synthetic biology and metabolic engineering, they are no longer insurmountable. An efficient production of cell-free recombinant hemoglobin poses tremendous challenges while having even greater potential, therefore some possible future directions are suggested in this review.

Keywords: synthetic biology, metabolic engineering, yeast, protein production, blood substitutes, HBOC, recombinant hemoglobin

FIGURE 1 | The limitations to conventional blood transfusions. The limitations are divided into three distinct groups; (a) limitations in storage and shelf life, (b) limitations due to donor dependency, blood-borne diseases, and incompatibility of blood types, and (c) limitations due to religious beliefs.
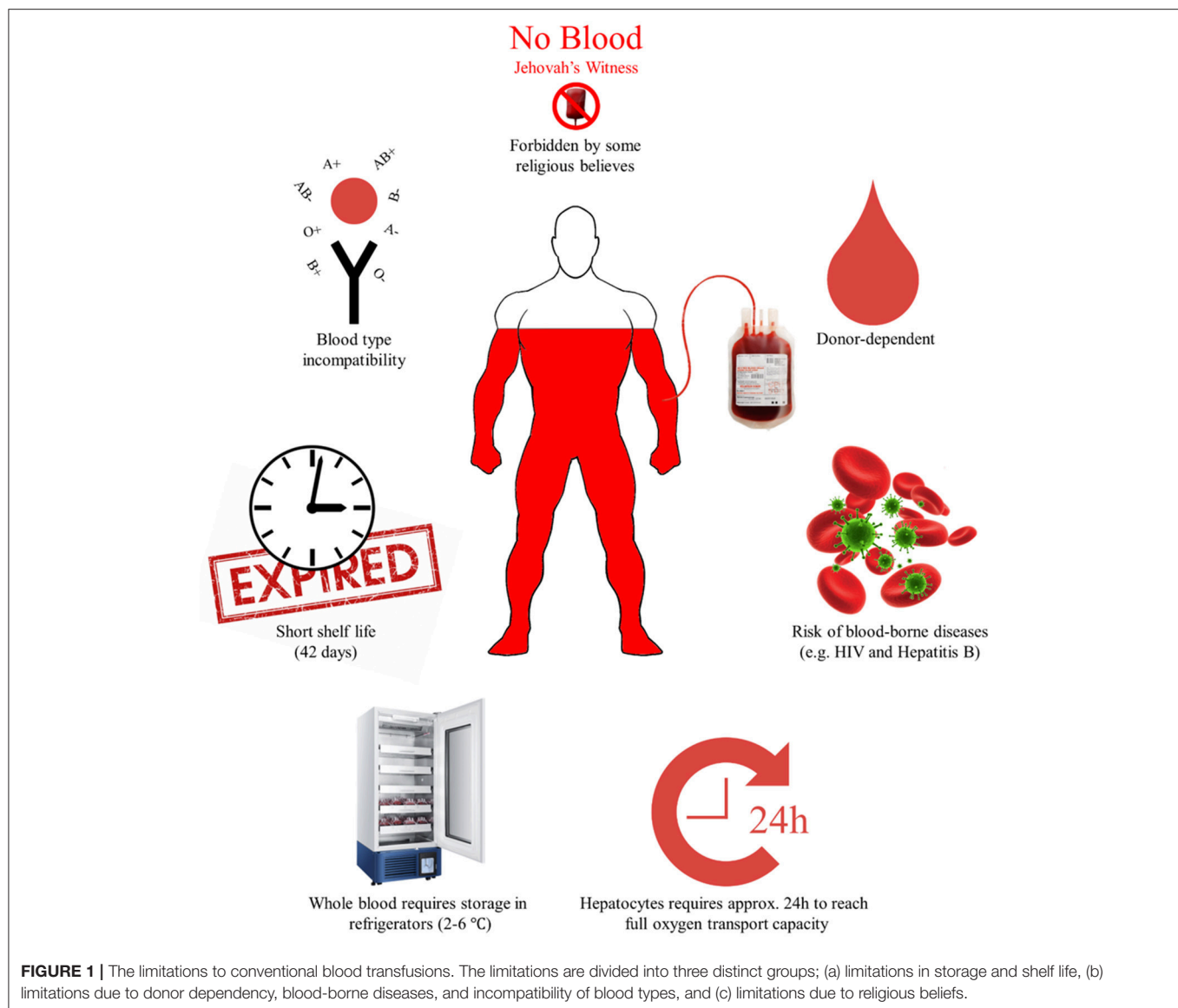
HISTORICAL PERSPECTIVE—SOCIETAL AND MEDICAL NEEDS

Since the first recorded blood transfusion in the ancient Inca civilization (fifteenth century), there have been countless attempts to create alternative blood substitutes to further improve the chance of surviving anemia (Sarkar, 2008). In the dawn of this development, elementary liquids such as sheep blood, urine, and beer were tested for the possibility to be a substitute for human blood. In the late nineteenth century, even cow milk was injected into patients with the belief that it would help regenerate white blood cells (Chen et al., 2009).

The earliest report of a hemoglobin-based blood substitute originates from the beginning of the 1930s, however it was not until 1957 that the first study on an artificial red blood cell composed by microencapsulated hemoglobin was performed (Li et al., 2005). After a stagnant period, the advancement was significantly stimulated by the onset of AIDS, hepatitis, and HIV infections in the 1980s, which imposed a radical risk to conventional blood transfusions (Kim and Greenburg, 2013; Martínez et al., 2015). Consequently, the following period was a golden age for the development of artificial substitutes that would be able to mimic the oxygen-carrying property of erythrocytes (Varnado et al., 2013; Martínez et al., 2015).

Through history, the pursuit of artificial blood substitutes has been heavily funded by the military to eliminate the supply, storage, and portability problems associated with human blood. However, the societal and medical need to find an alternative to conventional blood transfusion has escalated in recent years, especially in developing countries (Kim and Greenburg, 2013; Moradi et al., 2016). This is due to factors such as population

growth, natural disasters, decreasing donor number, population aging, terror attacks, and the risk of blood-borne pathogens threatening the supply-demand balance of human blood (Looker et al., 1992; Varnado et al., 2013; Alayash, 2014; Moradi et al., 2016; Chakane, 2017).

The development of blood substitutes furthermore has the potential to get rid of logistical barriers for pre-hospital use in acute emergency situations and remote civilian locations, by enabling long-term storage, eliminating blood type matching, supplying adequate quantities, be "pathogen-free," and providing immediate availability in catastrophic scenarios (Chakane, 2017).

RECENT ADVANCES

Recent breakthroughs in the fields of synthetic biology and metabolic engineering have substantially boosted the development of the so-called oxygen carriers, which can be divided into two main categories: perfluorocarbon-based substitutes (PFC) and hemoglobin-based oxygen carriers (HBOCs) (Mozafari et al., 2015). While the research in the PFC as blood substitutes has been partially discontinued, due to their limited oxygen transfer capacity to tissues and toxicity problems, the capacity of HBOCs to mimic red blood cells in terms of oxygen transport while being less toxic, has put the focus on them. Hemoglobin supply is, therefore, the first key issue for the successful development of the latter. The primary sources of hemoglobin for the current development of HBOCs are expired human blood, mammalian hemoglobins formed as a by-product in the meat industry (e.g., bovine hemoglobin), and recombinant hemoglobin produced in transgenic organisms (Motwani et al., 1996; Sanders et al., 1996; Varnado et al., 2013; Alayash, 2014; Moradi et al., 2016). The two first options are not the optimal choices, as their availability is too limited to allow a substantial development in HBOCs research. It is paramount for the case of the recombinant hemoglobin production, that new production hosts are designed with an increased and more efficient production capacity. Here, synthetic biology and metabolic engineering are key role players, both in terms of designing better cell factories and achieving affordable production costs.

Escherichia coli was the first choice as production workhorse. On a first attempt, a single β-globin chain with a cleavable linker was expressed, and subsequently refolded *in vitro* with a native α-globin chain and exogenous heme (Nagai and Thøgersen, 1987). In a follow-up experimental setup, α-globin and β-globin chains were co-expressed *in vivo* with endogenous heme incorporated (Shen et al., 1993). Although recombinant hemoglobin was successfully produced in bacteria, later advances focused on eukaryal hosts, as a result of the discovery that the vital functions of the hemoglobin produced by bacterial hosts were altered, most likely due to the methionine termini at the end of the globin chains. The alterations were identified to be caused by reduced Bohr effect and 2,3-BPG effects of the produced recombinant hemoglobin compared to normal human hemoglobin (Hoffman et al., 1990).

Thus, the production shifted to the model yeast *Saccharomyces cerevisiae*. Since then, synthetic biology and metabolic engineering have been applied to modify the heme biosynthetic pathway by overexpressing the genes coding for enzymes responsible for the rate-limiting steps (e.g., *HEM3*). Furthermore, the optimal globin expression balance between α and β chains was studied. In combination, the recombinant hemoglobin production was improved by an impressive 87% (Liu et al., 2014) compared to the previous attempts.

Important key issues to take into consideration, in order to get further in the advancement of HBOCs, apart from the sustained hemoglobin supply, are the safety concerns and cost-effectiveness. Non-clinical and clinical studies of HBOCs have raised safety concerns regarding hemoglobin extravasation across the blood vessel wall, scavenging of endothelial nitric oxide, oversupply of oxygen, and oxidative side reactions (Alayash, 2014). As a result, the regulatory agencies in the United States and the European Union have not yet approved any HBOCs (Chen et al., 2009; Varnado et al., 2013; Mozafari et al., 2015; Meng et al., 2018).

ECONOMICAL PERSPECTIVE AND SOCIETAL IMPACT

A study performed in the United States have estimated that it is currently possible to produce recombinant human hemoglobin at a cost of approximately $11/g. However, if the operating expenses and equipment investments are included, the cost increases to a dazzling amount of ≥$200/g. In comparison, human hemoglobin can be derived from alternative sources for roughly $2/g to $4/g (excl. post-related costs) based on fixed reimbursement prices of whole blood packs (Varnado et al., 2013).

There is, consequently, a demanding need to significantly improve the cost-effectiveness by either reducing the production cost or increasing the expression yield of recombinant hemoglobin by more than 3-fold to be cost-effective against alternatives. Synthetic biology can contribute with further optimizations to provide affordable recombinant hemoglobin at a comparable price to alternatives in the future. WHO has created a set of goals to improve global health for the world population by 2020—these are called *"HEALTH 2020"* (WHO, 2018). Recombinant hemoglobin is extremely relevant to one of the goals, which is to reduce inequality in the access to the medicine of the twenty first century and save lives.

THE CHALLENGES OF RECOMBINANT HEMOGLOBIN PRODUCTION

The evolution of synthetic biology has given researchers the possibility to synthesize new genetic elements instead of transferring them from a donor organism (Stephanopoulos, 2012). Nowadays entire genomes can be synthesized and inserted into a host organism (Kuo et al., 2018). Biotechnological research has also benefited from synthetic biology, as it synergizes with metabolic engineering. Synthetic biology provides genetic switches, vectors, characterized enzymes, and minimal hosts, all

of which minimizes the cost and the time needed for metabolic engineering (Keasling, 2012).

To successfully engineer novel cell factories a number of frameworks can be applied, one of the most popular is the *"Design-Build-Test-Learn"* cycle—**Figure 2**. The cycle is an iterative process divided into four phases, which aims at satisfying a set of predetermined specifications. The first phase is the *"Design"*-phase where the organism is designed *in silico*. That design is then constructed in the *"Build"*-phase. The third phase is the *"Test"*-phase, where the organism or construct is tested in terms of productivity, *-omics*, and growth characteristics. The final part is the *"Learn"*-phase, where the results from the *"Test"*-phase are evaluated and compared to the predetermined specifications set in the *"Design"*-phase. Often multiple iterations of the cycle are needed to achieve the optimal organism and satisfy the specifications (Ando and Martin, 2018; Jensen and Keasling, 2018). Synthetic biology and metabolic engineering contribute mainly to the first two phases (Nielsen and Keasling, 2016).

Higher Productivity

As previously described, the production of recombinant hemoglobin from a heterologous host is not yet economically feasible. Without accounting for operation cost and the expenses of establishing the production plant, the estimated costs should still be reduced by at least three-fold before being able to compete on the market. One of the challenges is low productivity. In 2016, Martìnez et al. developed a recombinant *S. cerevisiae*-based cell factory. After a few iterations of the DBTL cycle, the constructed stain was able to synthesize functional human hemoglobin with α-globin and β-globin genes expressed in the optimal ratios, being the maximum measured concentration of active human hemoglobin up to 7% of the total cell protein content (Martínez et al., 2016). In perspective, this means that this strain would be able to produce 200 kg of hemoglobin in a 100 m^3 tank, assuming the highest theoretical yield. Considering that one liter of human blood contains 150 grams of hemoglobin, the equivalent to 1.300 liters of blood would be produced in one batch. These are, however, ideal conditions in which upscaling limitations and recovery of the product are not considered. To reach these theoretical numbers at production-scale, a higher productivity needs to be achieved in laboratory-scale. Enhancing productivity in a heterologous organism is a task in the *"Design"*-phase of the DBTL-cycle. Classically a lot of work would have to be conducted, and a lot of manpower required, in order to plan and test hypotheses which might lead to higher productivity. In recent years, both the number and the variety of accurate open-end genome-scale models have increased (Nielsen and Keasling, 2016). Applying these models to the rational design can save time in the optimization of the productivity. It is now possible to find a robust, open-end genome-scale model of *S. cerevisiae,* that can be customized to have the genetic modifications implemented in previous research. Different strategies can be executed, and a simulation will reveal the success of the strategy in terms of enhanced productivity. Bottlenecks and feedback inhibitions can be identified and hence eliminated (Nielsen and Keasling, 2016).

Ultimately, only successful *in-silico* strategies will be carried out experimentally.

Intracellular Degradation and Negative Feedback Circuits

In the pursuit of higher productivities, most of the times undesired trade-offs are found as a consequence. Normally, degradation of precursors due to inherent feedback regulatory reactions is the main issue, and recombinant hemoglobin production is, unfortunately, not an exception to this matter. Heme is essential to the yeast cell for signaling and co-factor purposes, but it is also cytotoxic in higher concentrations as it can originate reactive oxygen species (ROS). To prevent accumulation and thereby toxicity, heme biosynthesis is tightly controlled and coordinated with heme degradation in well-understood processes (Martínez et al., 2016; Hanna et al., 2018). When heme is bound in hemoglobin, ROS are no longer formed, and heme will not be degraded (Liu et al., 2014).

A rapidly forming hemoglobin would therefore remove un-bound heme from the mitochondrion and hence stop degradation. Thus, the key is to have an abundance of α and β chains ready for assembly. In the *"Test"*-phase, proteomics combined with the advances in mass spectrometry, can assist the quantification of the chains and thereby identify possible limitations in the following *"Learn"*-phase (Trent, 2012). Other ways to reduce degradation involve gene knock-outs. For instance, genes involved in (i) iron-depletion initiated degradation of heme to recover the iron (Philpott and Protchenko, 2008) (ii) vacuolar miss-sorting and vacuolar degradation of mature hemoglobin (Ammerer et al., 1986; Marcusson et al., 1994; Hecht et al., 2014). Gene edition is tedious and very limiting most of the times, however the introduction and refining of the CRISPR-Cas9 system can drastically shorten the duration of the *"Build"*-phases where multiple knock-outs or insertions are needed. The method allows for gene disruption in up to five different loci in *S. cerevisiae* with a high success rate. The method is, moreover, marker-free (Jakočiunas et al., 2015).

Oxygen Limitation and Scale-Up

Hemoglobin synthesis in yeast is induced by high levels of extracellular oxygen (Martínez et al., 2016). In production-scale vessels adequate oxygen supply and transfer pose an issue due to the low solubility of oxygen in submerged cultures. To ensure a linear scale-up of oxygen availability other parameters need to be modified, e.g., agitation which can lead to shear stress (Reuss, 2008; Garcia-Ochoa and Gomez, 2009). Switching the point of view from process engineering to genetic and metabolic engineering could solve the problem. In 2015, Martìnez et al. deleted a heme activator protein (*HAP1*) in *S. cerevisiae* and found that heme production was increased due to the removal of a feedback inhibition. Another measured effect of the *HAP1* mutant is the reduction of respiration (Martínez et al., 2015). In the ideal scenario a large-scale production could be established based on a non-respiring strain where the presence of oxygen would initiate heme synthesis. The feasibility of the strategy can be predicted based on model-simulations rather than real-life cultivations in production scale.

FIGURE 2 | The Design-Build-Test-Learn cycle (DBTL-cycle). The "Design"-phase includes setting goals, choosing a host organism, using models and simulations, and planning the design of the organism. In the "Build"-phase the organism is constructed based on the planned design. The "Test"-phase covers the in-depth testing of the organism. The final phase is the "Learn"-phase where the results are analyzed and compared to the predetermined specifications.

Ethanol Production

The Crabtree-positive nature of *S. cerevisiae* (meaning its ability to produce ethanol from glucose in aerobic conditions) is one of the challenges in recombinant hemoglobin production for several reasons: (i) lower yields as carbon is diverted into ethanol production (Dai et al., 2018) and (ii) yields of hemoglobin are lower on ethanol than on glucose (Liu et al., 2014). The code to engineering a Crabtree-negative *S. cerevisiae* strain has finally been cracked by a Dai et al. who used state-of-the-art synthetic biology and molecular biology tools. The group reverse engineered a strain which had evolved under Adapted Laboratory Evolution. The created strain is the first Crab-negative *S. cerevisiae* reported to be able to grow at high glucose concentrations (Dai et al., 2018).

This particular strain in combination with the strategies stated above could create the future *S. cerevisiae*-based hemoglobin producer. Some challenges could arise as the Crabtree negative

S. cerevisiae needs mitochondrial respiration to reoxidize co-factors, whereas it is beneficial to knock-out *HAP1* to avoid respiration in recombinant hemoglobin production (Martínez et al., 2015; Dai et al., 2018). Simulations with an open-end genome-scale model will disclose whether further design is needed on co-factor reoxidizing or compartmentalization.

Recovery and Purification

Recovery and purification of hemoglobin from *S. cerevisiae* differs from the challenges described above, as it requires a new iteration of the DBTL-cycle. All the recombinant hosts designed to date, store the produced hemoglobin intracellularly. To recover the intracellular hemoglobin numerous unit operations will be required (Bulmer et al., 1996; Stocker-Majd et al., 2008). Here we will not discuss on how to solve the task by process engineering, but rather the solutions that can be created by using synthetic biology and metabolic engineering.

The rational approach sounds simple: to secrete the hemoglobin and recover it from the fermentation broth. All though *S. cerevisiae* is a desired platform for heterologous protein production, the yeast only secretes a few of its native proteins. To successfully secrete hemoglobin the protein needs the correct signal peptide, and genes involved in protein translocation should be upregulated (Tang et al., 2015; Bao et al., 2017). The combined powers of synthetic biology and metabolic engineering can decrease the amount of time and experiments needed. Combinations of genes and signal peptides can be modeled using genome-scale models to find the optima. The material for the modifications can be synthetically generated or taken directly from a BioBrick library (Constante et al., 2011) and lastly, CRIPSR-Cas9 can serve as an efficient and marker-free method for gene insertions or deletions (Jakočiunas et al., 2015). However, it is a complex task to make multimeric proteins secreted. Synthetic biology can contribute through the design of novel protein structures that would facilitate secretion, and hence the downstream processing, while keeping the functional properties intact. As an example, there are currently several strategies ongoing in order to design fusion hemoglobin molecules as a single peptide, but keeping a domain-like structure that resembles the multimeric structure.

Novel Production Hosts

Experimenting with novel organisms is no longer just a wish from the researchers: new technologies have made the wish come true. With BioBricks and minimal hosts, organisms can be constructed and customized like mechanical objects—such as airplanes (Constante et al., 2011; Keasling, 2012). BioBricks are standardized which means that they are modular and can be assembled in any way desired, and can be ordered from the Standard Registry of Biological Parts which is a growing collection (Constante et al., 2011). It is not farfetched to envision that a paradigm shift will occur in the near future of biotechnology. Researchers will be able to design the organism that they want instead of working around native shortcomings.

All of these technologies can widen the array of possible cell factories. For the production of recombinant hemoglobin this means that the production platform will no longer be restricted to *S. cerevisiae*. The performance of more organisms can be tested out both theoretically (using models) and experientially (using well-defined modular components) to find the optimal one. With the recent advances in high through-put micro-fermentation systems all-new stains can be designed, constructed and tested rapidly (Kensy et al., 2009).

CONCLUDING REMARKS

The implementation and continuous development of synthetic biology and metabolic engineering tools have enabled researchers to move the boundaries of what cell factories can do. The tools cannot only be used for creating new "super-drugs" but also to make well-known treatments more available.

Producing hemoglobin recombinantly in yeast has proven feasible. The challenges which still remain are considered solvable with the combined forces of knowledge, synthetic biology and metabolic engineering. This review has described some of the possible strategies to further improve yeast-based cell factories.

Currently, commercial production of recombinant hemoglobin is too expensive, to match the price of blood-derived hemoglobin. The cost should be further decreased by three-fold, but once strains have been optimized and large production plants have been established, hemoglobin can be produced at large scale. Yeasts are easy to cultivate and require cheap substrates in order to produce hemoglobin. It is also important to note that recombinant hemoglobin is not associated with the same risks as blood- derived hemoglobin in terms of pathogens and unstable supply.

All in all, the production of recombinant hemoglobin can ensure access to safe and affordable blood substitutes for the entire world population. Ultimately, the production of recombinant hemoglobin via synthetic biology will decrease inequality and help reach the goals set by WHO.

AUTHOR CONTRIBUTIONS

AF, IJ, and JM wrote the manuscript. AW and JM edited and supervised the final version.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Alayash, A. I. (2014). Blood substitutes: why haven't we been more successful? *Trends Biotechnol.* 32, 177–185. doi: 10.1016/j.tibtech.2014.02.006

Ammerer, G., Hunter, C. P., Rothman, J. H., Saari, G. C., Valls, L. A., and Stevens, T. H. (1986). PEP4 gene of *Saccharomyces cerevisiae* encodes proteinase A, a vacuolar enzyme required for processing of vacuolar precursors. *Mol. Cell. Biol.* 6, 2490–2499. doi: 10.1128/MCB.6.7.2490

Ando, D., and Martin, G. M. (2018). Two-scale 13C metabolic flux analysis for metabolic engineering. 1671, 333–352. doi: 10.1016/B978-0-12-374512-5.50021-8

Bao, J., Huang, M., Petranovic, D., and Nielsen, J. (2017). Moderate expression of SEC16 increases protein secretion by *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 83:16. doi: 10.1128/AEM.03400-16

Bulmer, M., Clarkson, A. I., Titchener-Hooker, N. J., and Dunnill, P. (1996). Computer-based simulation of the recovery of intracellular enzymes and its pilot-scale verification. *Bioprocess Eng.* 15, 331–337. doi: 10.1007/BF02426444

Chakane, S. (2017). *Towards New Generation of Hemoglobin-Based Blood Substitutes*. Lund: Lund University.

Chen, J.-Y., Scerbo, M., and Kramer, G. (2009). A review of blood substitutes: examining the history, clinical trial results, and ethics of hemoglobin-based oxygen carriers. *Clinics* 64, 803–814. doi: 10.1590/S1807-59322009000800016

Constante, M., Grünberg, R., and Isalan, M. (2011). A Biobrick library for cloning custom eukaryotic plasmids. *PLoS ONE* 6, 1–10. doi: 10.1371/journal.pone.0023685

Dai, Z., Huang, M., Chen, Y., Siewers, V., and Nielsen, J. (2018). Global rewiring of cellular metabolism renders *Saccharomyces cerevisiae* Crabtree negative. *Nat. Commun.* 9, 1–8. doi: 10.1038/s41467-018-05409-9

Garcia-Ochoa, F., and Gomez, E. (2009). Bioreactor scale-up and oxygen transfer rate in microbial processes: an overview. *Biotechnol. Adv.* 27, 153–176. doi: 10.1016/j.biotechadv.2008.10.006

Hanna, D. A., Hu, R., Kim, H., Martinez-Guzman, O., Torres, M. P., and Redd, A. R. (2018). Heme bioavailability and signaling in response to stress in yeast cells. *J. Biol. Chem.* 293, 12378–12393. doi: 10.1074/jbc.RA118.002125

Hecht, K. A., O'Donnell, A. F., and Brodsky, J. L. (2014). The proteolytic landscape of the yeast vacuole. *Cell. Logist.* 4:e28023. doi: 10.4161/cl.28023

Hoffman, S. J., Looker, D. L., Roehrich, J. M., Cozart, P. E., Durfee, S. L., Tedesco, J. L., et al. (1990). Expression of fully functional tetrameric human hemoglobin in Escherichia coli. *Proc. Natl. Acad. Sci.* 87, 8521–8525. doi: 10.1073/pnas.87.21.8521

Jakočinas, T., Bonde, I., Herrgård, M., Harrison, S. J., Kristensen, M., Pedersen, L. E., et al. (2015). Multiplex metabolic pathway engineering using CRISPR/Cas9 in *Saccharomyces cerevisiae*. *Metab. Eng.* 28, 213–222. doi: 10.1016/j.ymben.2015.01.008

Jensen, M. K., and Keasling, J. (2018). *Synthetic Metabolic Pathways*. New York, NY: Humana Press.

Keasling, J. D. (2012). Synthetic biology and the development of tools for metabolic engineering. *Metab. Eng.* 14, 189–195. doi: 10.1016/j.ymben.2012.01.004

Kensy, F., Zang, E., Faulhammer, C., Tan, R. K., and Büchs, J. (2009). Validation of a high-throughput fermentation system based on online monitoring of biomass and fluorescence in continuously shaken microtiter plates. *Microb. Cell Fact.* 8, 1–17. doi: 10.1186/1475-2859-8-31

Kim, H. W., and Greenburg, G. (2013). *Hemoglobin- Based Oxygen and Oxygen Cell Substitutes Carriers as Red Therapeutics*. Berlin: Springer-Verlag Berlin Heidelberg.

Kuo, J., Stirling, F., Lau, Y. H., Shulgina, Y., Way, J. C., and Silver, P. A. (2018). Synthetic genome recoding: new genetic codes for new features. *Curr. Genet.* 64, 327–333. doi: 10.1007/s00294-017-0754-z

Li, S., Nickels, J., and Palmer, A. F. (2005). Liposome-encapsulated actin-hemoglobin (LEAcHb) artificial blood substitutes. *Biomaterials* 26, 3759–3769. doi: 10.1016/j.biomaterials.2004.09.015

Liu, L., Martínez, J. L., Liu, Z., Petranovic, D., and Nielsen, J. (2014). Balanced globin protein expression and heme biosynthesis improve production of human hemoglobin in *Saccharomyces cerevisiae*. *Metab. Eng.* 21, 9–16. doi: 10.1016/j.ymben.2013.10.010

Looker, D., Abbott-Brown, D., Cozart, P., Durfee, S., Hoffman, S., Mathews, A. J., et al. (1992). A human recombinant haemoglobin designed for use as a blood substitute. *Nature* 356, 258–260. doi: 10.1038/356258a0

Marcusson, E. G., Horazdovsky, B. F., Cereghino, J. L., Gharakhanian, E., and Emr, S. D. (1994). The sorting receptor for yeast vacuolar carboxypeptidase Y is encoded by the VPS10 gene. *Cell* 77, 579–586. doi: 10.1016/0092-8674(94)90219-4

Martínez, J. L., Liu, L., Petranovic, D., and Nielsen, J. (2015). Engineering the oxygen sensing regulation results in an enhanced recombinant human hemoglobin production by *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* 112, 181–188. doi: 10.1002/bit.25347

Martínez, J. L., Petranovic, D., and Nielsen, J. (2016). Heme metabolism in stress regulation and protein production: from Cinderella to a key player. *Bioengineered* 7, 112–115. doi: 10.1080/21655979.2015.1126016

Meng, F., Kassa, T., Jana, S., Wood, F., Zhang, X., Jia, Y., et al. (2018). Comprehensive biochemical and biophysical characterization of hemoglobin-based oxygen carrier therapeutics: all HBOCs are not created equally. *Bioconjug. Chem.* 29, 1560–1575. doi: 10.1021/acs.bioconjchem.8b00093

Moradi, S., Jahanian-Najafabadi, A., and Roudkenar, M. H. (2016). Artificial blood substitutes: First steps on the long route to clinical utility. *Clin. Med. Insights Blood Disord.* 9, 33–41. doi: 10.4137/CMBD.S38461

Motwani, N., Talarico, T., Jain, S., Bajwa, W., Blackburn, R., Nwosu, V., et al. (1996). Production, purification, and characterization of recombinant human hemoglobin rainier expressed in Saccharomyces cerevisiae. *Protein Expr. Purif.* 8, 447–455. doi: 10.1006/prep.1996.0123

Mozafari, M., Ramedani, A., and Yazdanpanah, A. (2015). Artificial blood- a game changer for future medicine: where are we today? *J. Blood Disord. Transfus.* 6, 8–10. doi: 10.4172/2155-9864.1000312

Nagai, K., and Thøgersen, H. C. (1987). "Synthesis and sequence-specific proteolysis of hybrid proteins produced in Escherichia coli". *Methods Enzymol.* 153, 461–481. doi: 10.1016/0076-6879(87)53072-5

Nielsen, J., and Keasling, J. D. (2016). Engineering cellular metabolism. *Cell* 164, 1185–1197. doi: 10.1016/j.cell.2016.02.004

Philpott, C. C., and Protchenko, O. (2008). Response to iron deprivation in *Saccharomyces cerevisiae*. *Eukaryot. Cell* 7, 20–27. doi: 10.1128/EC.00354-07

Reuss, M. (2008). "Oxygen transfer and mixing: scale-up implications". *Biotechnology* 3, 185–217. doi: 10.1002/9783527620845.ch10

Sanders, K. E., Ackers, G., and Sligar, S. (1996). Engineering and design of blood substitutes. *Curr. Opin. Struct. Biol.* 6, 534–540

Sarkar, S. (2008). Artificial Blood. *Indian J. Crit. Care Med.* 12, 140–144. doi: 10.4103/0972-5229.43685

Shen, T. J., Ho, N. T., Simplaceanu, V., Zou, M., Green, B. N., Tam, M. F., et al. (1993). Production of unmodified human adult hemoglobin in *Escherichia coli*. *Proc. Natl. Acad. Sci.* 90, 8108–8112. doi: 10.1073/pnas.90.17.8108

Stephanopoulos, G. (2012). Synthetic biology and metabolic engineering. *ACS Synth. Biol.* 1, 514–525. doi: 10.1021/sb300094q

Stocker-Majd, G., Hilbrig, F., and Freitag, R. (2008). Extraction of haemoglobin from human blood by affinity precipitation using a haptoglobin-based stimuli-responsive affinity macroligand. *J. Chromatogr. A* 1194, 57–65. doi: 10.1016/j.chroma.2008.03.091

Tang, H., Bao, X., Shen, Y., Song, M., Wang, S., Wang, C., et al. (2015). Engineering protein folding and translocation improves heterologous protein secretion in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* 112, 1872–1882. doi: 10.1002/bit.25596

Trent, R. J. (2012). "*Omics*," in *Molecular Medicine: Genomics To Personalized Healthcare.* London; Waltham; San Diego, CA: Academic Press/Elsevier Inc.

Varnado, C. L., Mollan, T. L., Birukou, I., Smith, B. J. Z., Henderson, D. P., and Olson, J. S. (2013). Development of recombinant hemoglobin-based oxygen carriers. *Antioxid. Redox Signal.* 18, 2314–2328. doi: 10.1089/ars.2012.4917

WHO (2018). *What are the Key Components of Health 2020?* Available online at: http://www.euro.who.int/en/health-topics/health-policy/health-2020-the-european-policy-for-health-and-well-being/about-health-2020/what-are-the-key-components-of-health-2020 (Accessed September 17, 2018).

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

CopySwitch—*in vivo* Optimization of Gene Copy Numbers for Heterologous Gene Expression in *Bacillus subtilis*

Florian Nadler, Felix Bracharz and Johannes Kabisch*

Computer-Aided Synthetic Biology, Institute of Biology, Technische Universität Darmstadt, Darmstadt, Germany

The Gram-positive bacterium *Bacillus subtilis* has long been used as a host for production and secretion of industrially relevant enzymes like amylases and proteases. It is imperative for optimal efficiency, to balance protein yield and correct folding. While there are numerous ways of doing so on protein or mRNA level, our approach aims for the underlying number of coding sequences. Gene copy numbers are an important tuning valve for the optimization of heterologous gene expression. While some genes are best expressed from many gene copies, for other genes, medium or even single copy numbers are the only way to avoid formation of inclusion bodies, toxic gene dosage effects or achieve desired levels for metabolic engineering. In order to provide a simple and robust method to address above-mentioned issues in the Gram-positive bacterium *Bacillus subtilis*, we have developed an automatable system for the tuning of heterologous gene expression based on the host's intrinsic natural competence and homologous recombination capabilities. Strains are transformed with a linearized, low copy number plasmid containing an antibiotic resistance marker and homology regions up- and downstream of the gene of interest. Said gene is copied onto the vector, rendering it circular and replicative and thus selectable. We could show an up to 3.6-fold higher *gfp* (green fluorescent protein) expression and up to 1.3-fold higher *mPLC* (mature phospholipase C) expression after successful transformation. Furthermore, the plasmid-borne *gfp* expression seems to be more stable, since over the whole cultivation period the share of fluorescent cells compared to all measured cells is consistently higher. A major benefit of this method is the ability to work with very large regions of interest, since all relevant steps are carried out *in vivo* and are thus far less prone to mechanical DNA damage.

Keywords: *Bacillus subtilis*, copy number change, gene expression tuning, heterologous expression, natural competence

INTRODUCTION

Heterologous gene and pathway expression remains a task mostly approached by trial and error, making use of a broad palette of available optimization strategies (Stevens, 2000). The goal of most approaches is to find the optimal level of gene expression for the maximum amount of active protein without the formation of inclusion bodies. There are special cases, where inclusion body formation

is wanted for easier protein purification, but these are only applicable in combination with specific refolding strategies. In prokaryotic gene expression, optimization implies finding a suited balance between transcription and translation speed on the one hand and correct protein folding on the other hand. This can be achieved by reducing the temperature during protein production with the drawback of increased process times (Sørensen and Mortensen, 2005).

Advanced methods use libraries of constructs, transcribed and translated with different strengths by varying the promoter (Kraft et al., 2007), the ribosome binding site (Kohl et al., 2018) or by fusion with peptides aiding in correct folding (Butt et al., 2005; Kraft et al., 2007). Another approach is to vary the copy number of the genetic constructs. As shown before, the number in which a gene is present in a cell is correlated with the expression level of said gene (Lee et al., 2015). In the most prominent host *Escherichia coli*, plasmids with different copy numbers are available such as the high copy number pUC vectors (Yanisch-Perron et al., 1985) and the low copy number pBR322 based vectors (Bolivar et al., 1977) and assisted integration into the chromosome is possible through recombineering (Sharan et al., 2009). Plasmid replication is a major burden on the cell's metabolism, thus always requiring an active selection mechanism to ensure plasmid stability during production. The higher the copy number of the plasmids, the higher this burden and the less stable the plasmid DNA is propagated. For the yeast *S. cerevisiae* it was demonstrated, that a set of tunable copy number plasmids (dependent on antibiotic concentration) could balance a pathway for n-butanol production in a way that yielded a 100-fold increase in production of the desired chemical (Lian et al., 2016). In another publication, different copy number plasmids were introduced into *Bacillus methanolicus* via electroporation to show the effect of gene dosage on the expression of *gfpuv*, an amylase from *Streptomyces griseus* and the lysine decarboxylase *cadA* from *E. coli* for overproduction of cadaverine. A positive correlation between estimated plasmid copy numbers and observed expression levels could be determined (Irla et al., 2016). Since plasmid copy numbers can be affected by media composition as well as growth phase, efforts have been undertaken to uncouple expression levels from gene dosage by introducing so-called incoherent feedforward loops in the form of transcription-activator-like effector (TALE)-regulated promoters (Segall-Shapiro et al., 2018).

Contrary to the Gram-negative organism *E. coli*, which under laboratory conditions is only able to take up plasmid DNA after various pretreatment steps, *Bacillus subtilis* becomes naturally competent. Transformation is as easy as cultivating cells in a competence-inducing medium in the presence of the DNA to be integrated for a few hours (Kumpfmüller et al., 2013). To further accelerate this process, *B. subtilis* strains with inducible competence have been developed (Rahmer et al., 2015). In contrast to this easy method for transforming linear DNA for genomic integration, the transformation of circular plasmid DNA is more difficult in *B. subtilis*. One method based on naturally competent cells has the drawback of needing multimeric plasmid forms (Canosi et al., 1978). These can be generated either by PCR and ligation or by transformation of a monomeric form

into specialized *E. coli* strains (e.g., JM105 Voss et al., 2003). Most other methods of transforming *B. subtilis* with replicative plasmids require the use of special devices (e.g., electroporation), pretreatment of cells with polyethylene glycol (PEG), glycine or the preparation of protoplasts, which is tedious and requires recovery for 2 to 3 days (Vojcic et al., 2012).

The CopySwitch system presented in this work allows the rapid and easy *in vivo* increase of copy numbers for applications in *B. subtilis* without the drawbacks of classic plasmid transformation techniques. The underlying process is depicted in **Figure 1**. A construct for heterologous expression is integrated into the chromosome via natural competence. After verification of the correct insertion, a linearized plasmid, containing an origin of replication (*ori*) for *B. subtilis*, an antibiotic resistance gene and homologous regions up- and downstream of the integration site, is added to competent cells. Upon intracellular uptake, recombination occurs and all genetic elements between the two recombination sites are copied onto the plasmid, rendering it circular and therefore replicative again (Tomita et al., 2004b). Since most of the steps happen inside the cell, plasmid size becomes less limiting than with *in vitro* cloning methods. If a certain chromosomal locus is chosen for integration of genes to be expressed heterologously, the same CopySwitch backbone can be reused.

We chose two different reporter systems for demonstrating the application of CopySwitch, one being the intracellular production of GFP and the other being the secretion of a recombinant mature phospholipase C (mPLC) from *Bacillus cereus* SBUG 516. Both were placed under the acetoin-inducible promoter P_{acoA} from *B. subtilis*. The mPLC is a monomeric, phosphatidylcholine-preferring PLC (E.C.3.1.4.3, UniProt-ID P09598), containing three catalytically relevant zinc ions. It is used in the industrial degumming of oils (Jiang et al., 2015; Elena et al., 2017). Its sequence comprises of the amino acids 39–283 (mature form) and is preceded by the first 33 amino acids of the *B. subtilis amyE* gene, which function as a signal peptide for secretion (Durban et al., 2007). Green fluorescent protein production can be assessed via flow cytometry fluorescence measurements, whereas mPLC activity is evaluated by the increased absorbance at 410 nm due to accumulation of *p*-nitrophenol from the cleavage of the chromogenic substrate *para*-nitrophenylphosphorylcholine (*p*-NPPC) (Kurioka and Matsuda, 1976; Durban et al., 2007).

With the here described method we aim to address some of the shortcomings and current bottlenecks of traditional replicative plasmid transformation protocols.

MATERIALS AND METHODS

Strains and Plasmids

For propagation and maintenance of plasmids *E. coli* DH10B cells were used. For all experiments on *gfp* expression, a derivative of *Bacillus subtilis* 6051-HGW was used as a starting point, combining six different modifications previously described (Kabisch et al., 2013; Zobel et al., 2015) (see **Table 2**). For all experiments on *mPLC* (matured phospholipase C) expression, a derivative of *Bacillus subtilis* PY79 was used, combining seven

FIGURE 1 | Flow-chart of the CopySwitch method; after linearization of a replicative plasmid (via restriction or PCR), it is taken up by *Bacillus subtilis* via natural transformation. Regions homologous to the chromosome are recombined and the DNA sequence between the regions is copied onto the plasmid. Hatched blocks: homology regions; dark gray blocks: sequence of interest; white block: low copy number *ori*; Adapted from Tsuge and Itaya (Tsuge and Itaya, 2001).

knock-outs of extracellular proteases. This so-called KO7 strain is available from the Bacillus Genetic Stock Center (Columbus, OH, USA) under the BGSCID 1A1133. In our studies we used a derivative of pBS72 (Titok et al., 2003) for a low copy number *ori*. Instead of the full length *B. subtilis* origin region as it is used for example in the commercially available pHT01 system (MoBiTec, Göttingen, Germany), we shortened the region by approximately 600 base pairs (bps) by deleting two putative open reading frames ('orf-4 and orf-3) of unknown function (Titok et al., 2006) in order to reduce replicative burden.

For details on the plasmid sequences including oligonucleotides see **Supplementary Material** section.

Media and Cultivation

All standard cultivations were performed in LB containing 10 g/L NaCl (LB10; Carl Roth, Karlsruhe, Germany), except for *B. subtilis* strains carrying a resistance gene for the salt-sensitive antibiotic zeocin (InvivoGen, San Diego, CA, USA). In these cases, LB containing only 5 g/L NaCl (LB5; Carl Roth, Karlsruhe, Germany) was used. Transformation medium for *Bacillus subtilis* was prepared according to Kumpfmüller and coworkers (Kumpfmüller et al., 2013). Antibiotics were used in the following concentrations: 100 µg/ml ampicillin, 50 µg/ml (*E. coli*) or 12.5–50.0 µg/ml (*B. subtilis*) kanamycin, 100 µg/ml spectinomycin, and 40 µg/ml zeocin.

All incubations for the flow cytometry experiment were performed at 37°C and 200 - 250 rpm in a 25 mm amplitude rotary shaker (Infors GmbH, Einsbach, Germany). Thirty milliliters of preculture (15 ml 2x LB5, 1% (v/v) glucose from a 40% (w/v) stock solution, 40 µg/ml zeocin and with or without 25 µg/ml kanamycin, filled to 30 ml with sterile, desalted H_2O) were inoculated from glycerol stocks of BsFLN045 and BsFLN051, respectively. After incubation overnight, 20 ml of main culture (10 ml 2x LB5, 0.5% (v/v) glucose from a 40% (w/v) stock solution, 0.5% (v/v) acetoin from a 10% (w/v) stock solution, 40 µg/ml zeocin, and with or without 25 µg/ml kanamycin, filled to 20 ml with sterile, desalted H_2O) were inoculated with preculture to an optical density of 0.1 at 600 nm (OD_{600}) in quadruplicates for each strain. Media were prepared

as master mixes and distributed evenly into 50 ml Erlenmeyer flasks for cultivation.

Cultivations for the phospholipase C experiments were performed at 30°C and 150 rpm in a 50 mm amplitude rotary shaker (Eppendorf New Brunswick, Enfield, CT, USA). Twenty milliliters of preculture (SB medium: 32.0 g/L Tryptone, 20.0 g/L yeast extract, 5.0 g/L NaCl, 5.0 ml 1N NaOH, 42.3 mM Na_2HPO_4, 22.0 mM KH_2PO_4, 18.7 mM NH_4Cl, and 0.02 mM $CaCl_2$, supplemented with 1% (v/v) glucose from a 40% (w/v) stock solution, 40 µg/ml zeocin, and with or without 25 µg/ml kanamycin) were inoculated from glycerol stocks of BsMKA005, BsFLN064 and BsFLN066, respectively. To ensure sufficient growth overnight, precultures were incubated at 37°C. Fifty milliliters of main culture (SB containing 0.5% (v/v) of acetoin from a 10% stock solution in SB and respective antibiotic) were inoculated to reach an OD_{600} of 0.2 in triplicates for each strain. Cultivation was carried out in 500 ml Erlenmeyer flasks.

Enzymes

All restriction enzymes and Q5 proofreading polymerase (for amplicons >3 kb) were purchased from New England Biolabs (NEB, Ipswich, MA, USA). Colony polymerase chain reactions (cPCRs) were performed with standard Taq polymerase, whereas error-free amplification of small (<3 kb) fragments was performed with OptiTaq, which is a mixture of Taq and Pfu polymerase (Roboklon, Berlin, Germany). Alternatively to Q5, proofreading Polymerase X was also used for amplicons >3 kb (Roboklon, Berlin, Germany).

Molecular Biology Methods

All PCRs and restriction digestions were carried out as suggested by the supplier of the used enzymes. Sequencing reactions and oligonucleotides were purchased from Eurofins genomics (Ebersberg, Germany). If replicative plasmids were used as PCR templates and no antibiotic switch was possible after cloning, a thorough DpnI digestion (1–2 h at 37°C) was performed to reduce background through template plasmid carry-over, followed by heat inactivation (20 min at 65°C). PCR fragments were column-purified before being used for

TABLE 1 | Optimization of transformation conditions. Numbers given correspond to colony forming units.

Regeneration time	1 h			OD$_{600}$	2 h			OD$_{600}$	4 h			OD$_{600}$
Kanamycin [μg/ml]	12.5	25	50		12.5	25	50		12.5	25	50	
BsFLN049 (500 bp)	0	0	0	3.2	0	0	0	4.6	0	0	0	6.2
BsFLN050 (1,000 bp)	0	0	0	3.2	0	0	0	4.6	0	0	0	6.1
BsFLN051 (2,000 bp)	1	0	0	3.2	9	0	0	4.8	35	25	0	6.4
Neg. control (elution buffer)	0	0	0	3.0	0	0	0	5.2	1	0	0	6.0

TABLE 2 | Strains and plasmids used in this work.

Strain	Relevant genotype	Reference
EcDH10B	F$^-$ endA1 recA1 galE15 galK16 nupG rpsL ΔlacX74 Φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) λ$^-$	Invitrogen, Darmstadt, Germany
BsKO7	BsPY79 + ΔnprE ΔaprE Δepr Δmpr ΔnprB Δvpr Δbpr	Dr. Zeigler, Bacillus Genetic Stock Center (#1A1133)
BsFLN024	Δ(ydzT-ydjC)::lox72, ΔsacA::(Zeo^R, P$_{xylA}$-cre, xylR, P$_{spac}$-comS, lacI), ΔspoIIGA::lox72, ΔlytC::lox72, Δpks(A-R)::lox72, ΔamyE::(sigL-acoR)	This work
BsFLN045	BsFLN024 + Δpks(A-R)::(lox72, P$_{acoA}$-GFPmut2, lox72)	This work
BsFLN049	BsFLN045 + replicative plasmid p14071	This work
BsFLN050	BsFLN045 + replicative plasmid p14072	This work
BsFLN051	BsFLN045 + replicative plasmid p14073	This work
BsFLN064	BsMKA005 + Δpks(A-R)::(P$_{acoA}$-SP$_{amyE}$-mPLC, lox72)	This work
BsFLN066	BsFLN064 + replicative plasmid p14096	This work
BsMKA005	BsKO7 + ΔsacA::(Zeo^R, P$_{xylA}$-cre, xylR, P$_{spac}$-comS, lacI), ΔspoIIGA-sigE::lox72	PhD thesis Marianna Karava
Plasmids	For all listed plasmids see Supplemental Material for sequences	
p14067 (pJET-GFPmut2)	Chromosomal integration of P$_{acoA}$-GFPmut2 into BsFLN024	This work
p14059 (plcBReT-sh- pks500)	E. coli shuttle plasmid, contains low copy number ori for B. subtilis and two homology regions left and right of chromosomal locus to be copied (500 bp each) and kanamycin resistance	This work
p14060 (plcBReT-sh- pks1k)	E. coli shuttle plasmid, contains low copy number ori for B. subtilis and two homology regions left and right of chromosomal locus to be copied (1 kb each) and kanamycin resistance	This work
p14035 (plcBReT-sh)	E. coli shuttle plasmid, contains low copy number ori for B. subtilis and two homology regions left and right of chromosomal locus to be copied (2 kb each) and kanamycin resistance	This work
p14071 (plcBReT-sh- pks500-GFPmut2)	p14059 + chromosomal locus P$_{acoA}$-GFPmut2 copied between homology regions	This work
p14072 (plcBReT-sh- pks1k-GFPmut2)	p14060 + chromosomal locus P$_{acoA}$-GFPmut2 copied between homology regions	This work
p14073 (plcBReT-sh- GFPmut2)	p14035 + chromosomal locus P$_{acoA}$-GFPmut2 copied between homology regions	This work
p14094 (pMSE3-acoA- amyE-mPLC)	Template plasmid for amplification of P$_{acoA}$-SP$_{amyE}$-mPLC cassette	Dr. Dominique Boettcher, University of Greifswald
p14096 (plc-pks2k-mPLC)	p14035 + chromosomal locus P$_{acoA}$-SP$_{amyE}$-mPLC copied between homology regions	This work

cloning (NucleoSpin® Gel and PCR Clean-up Kit, Macherey-Nagel, Düren, Germany or Monarch® PCR & DNA Cleanup Kit, NEB, Ipswich, MA, USA). Plasmids were also column-purified with two different kits (High Pure Plasmid Isolation Kit, F. Hoffmann-La Roche AG, Basel, Switzerland or Monarch® Plasmid Miniprep Kit, NEB, Ipswich, MA, USA). For plasmid preparation from *B. subtilis*, 5 to 10 mg of lyophilized lysozyme (Sigma-Aldrich, St. Louis, MO, USA) were added to the resuspension buffer and incubated at 37°C for 10 min. All subsequent steps were done as described in the respective protocol of the plasmid preparation kit used. DNA assemblies were performed using SLiCE as described by Messerschmidt and coworkers (Messerschmidt et al., 2016). In contrast to the described method, DNA was transformed by electroporation (2.5 kV, 2 mm gap). For an assembly containing more than three DNA parts, amplifying the SLiCE assembled DNA via PCR enabled us to reduce the fragment number. Correct assembly of fragments and integration into a *B. subtilis* genome was detected by colony PCR (cPCR) and confirmed by sequencing of either plasmids or proofreading-amplified PCR products of regions of interest.

Transformation of *Bacillus subtilis*

For transformation, 0.05% (w/v) of casamino acids (CAA; Sigma-Aldrich, St. Louis, MO, USA) and either 1% (v/v) glycerol stock or 3–5% (v/v) of an overnight culture of the desired strain (grown in the same medium) were added to 1 ml of transformation medium (TM) containing a suitable antibiotic. Cells were cultivated for 5–6 h, shaking at 37°C until the broth was visibly turbid. At this point, plasmid DNA (undigested for chromosomal integration and *XhoI*-linearized for CopySwitch) was added at a final concentration of >200 fmol/ml and incubated for 30 min while shaking. For each milliliter of transformation sample, 200 µl of expression mix (2.5% (v/v) of yeast extract and CAA each) were added and, depending on the task, shaken for another 1 (chromosomal integration) to 4 (CopySwitch) h at 37°C. Afterwards, different amounts (200 to 1,000 µl) were spread onto antibiotic-containing LB-agar-plates and incubated overnight to ensure growth of distinct single colonies. To recycle the antibiotic marker in the chromosomal integration approaches, single colonies were subjected to xylose-induced expression of *cre* recombinase by growth in 300 µl of LB5-Zeo (40 µg/ml) containing 1% (v/v) of xylose for >6 h at 37°C and 1,400 rpm (Thermoshaker, Eppendorf, Hamburg, Germany). Constant aeration during this time was achieved by melting a hole into the lid of the 2 ml Eppendorf tubes used. This procedure results in recombination of *lox71* and *lox66*, thereby looping out the marker gene cassette inbetween, generating a *lox72* scar. After a dilution streaking on zeocin-containing LB5-agar-plates, single colonies were further verified via cPCR, replica-plating to ensure loss of the antibiotic marker and finally sequencing. To verify a CopySwitch, we isolated plasmid from *B. subtilis* and transformed the obtained DNA into *E. coli* to obtain sufficient amounts of plasmid for sequencing.

Enzymatic Assay For mPLC Activity
Sampling

Samples for the enzymatic assay of mPLC activity were collected after 6, 12, 24, and 48 h of cultivation. Two milliliters of culture broth were centrifuged (17.000 rcf, 4°C, 5 min) and 1.8 ml were transferred into new vessels. Sterile-filtering was omitted to ensure full enzymatic activity.

Assay Principle and Plate Reader Measurements

The underlying principle of the activity assay lies in the ability of mPLC to turn the non-natural substrate *p*-NPPC (CAS number 21064-69-7, Cayman Chemicals, Ann Arbor, MI, USA) into phosphorylcholine and *p*-nitrophenol. Absorption of *p*-nitrophenol is measured at 410 nm (Kurioka and Matsuda, 1976; Durban et al., 2007). Activity measurements were carried out with 70 µl of diluted supernatant (one part supernatant + three parts 100 mM borax-HCl buffer, pH 7.5) and 30 µl of substrate solution (100 mM *p*-NPPC dissolved in 100 mM borax-HCl buffer, pH 7.5) in 96-well microwell plates (96-well, PS, F-bottom, #655101, Greiner Bio-One GmbH, Frickenhausen, Germany). Measurements were done at 30°C for 2 h every minute and for another 5 h every 5 min afterwards. Since substrate depletion caused the reaction to plateau after several hours, only values from the first 4 h were taken for all calculations.

Data Analysis

All data conversions, analyses and plotting were performed with R Version 3.5.1 (Ihaka and Gentleman, 1996) using Wickhams tidyverse Version 1.2.1. (Wickham, 2017).

Assaying GFP Fluorescence
Sampling

Samples for flow cytometry analysis were taken after 3, 5, 7, 9, 11, 20, and 24 h of cultivation, to ensure capturing the transition between non-expressing and gene-expressing state. The sample volume was chosen to yield an OD_{600} between 1.7 and 2.6 after pelleting the cells, discarding the supernatant and resuspending in 1.5 ml 1x ClearSort Sheath Fluid (Sony Biotechnology, Weybridge, UK). Before resuspension, the cell pellets were stored at −20°C until all samples were collected and prepared for measurement. Reconstituted samples were diluted again (one part sample plus four parts 1x ClearSort Sheath Fluid), to be in a suitable range of events per second for flow cytometry measurement.

Flow Cytometry

Fluorescence in cells was detected in a Sony LE-SH800SZBCPL (Sony Biotechnology, Weybridge, UK) using a 488 nm argon laser. Photomultipliers for backscatter and FL-1 (525/50 nm) were set to 25.5 and 38.0% with a FSC-threshold of 0.20% and a window extension of 50. The FSC-diode was set on an amplification setting of 6/12 with events per second being 50,000.

Flow Cytometry Data Analysis

Areas of scattering and fluorescence signals (FSC-A, SSC-A, FL1-A) were brought to a near-normal distribution by applying an inverse hyperbolic sine (asinh) to all values. Cell agglomerates and fragments were excluded by fitting a bivariate normal distribution in the channels FSC-A and FSC-W. For the threshold between fluorescent and non-fluorescent cells, the minimum of the clearly bimodal distribution of all measured FL-1 values was chosen. For every sample, medians of the fluorescent population as well as the ratio of counts between fluorescent events and total events was calculated. The resulting values were analyzed by showing means and standard deviations of biological quadruplicates. All analyses, calculations and plotting were performed with R (Ihaka and Gentleman, 1996). Specifically, for parsing of fcs files and gating, bioconductor's (Huber et al., 2015) flowcore (Hahne et al., 2009) was used. Packages from the metapackage tidyverse, especially dplyr (Hadley et al., 2017) and ggplot2 (Wickham, 2011) were employed for summarization and presentation.

Plate Reader Measurements

Directly after all flow cytometry measurements had been done, 200 µl of each diluted sample were transferred from the flow cytometry tubes into 96-well microwell plates (Nunc™ MicroWell™ #152037, ThermoFisher Scientific, Waltham, MA, USA). Optical density was measured at 600 nm in a PHERAstar FSX (BMG Labtech, Ortenberg, Germany). Fluorescence was induced with an excitation wavelength of 485 nm and measured at an emission wavelength of 520 nm. This procedure guarantees

comparability between flow cytometry and plate reader data. All analyses, calculations and plotting were performed with R (Ihaka and Gentleman, 1996).

RESULTS

Optimizing Parameters For Successful Transformation and Gene Transfer Onto Replicative Plasmid

Transformation via natural competence requires internalization of single-stranded DNA, formation of a heteroduplex between regions of homology and expression of an antibiotic resistance marker. Using our method for replicative plasmid transformation furthermore requires time for expression of additional, plasmid-based replication machinery (e.g., RepA protein) and replication of the newly circularized plasmid. To optimize the yield of plasmid-bearing colonies, three different parameters were tested. For standard chromosomal integration, homology regions of 500 bp and a regeneration time after transformation of 1 h were employed. However, this procedure did not yield any colonies for CopySwitch, regardless of the antibiotic concentration (see **Table 1**). To increase the probability of correctly assembled replicative plasmids, longer regeneration times (1, 2, and 4 h) and longer regions of homology (500, 1,000, and 2,000 bp) were tested. The volume plated corresponds to 48 fmol of linearized plasmid in the transformation mix. All of the different conditions were tested in BsFLN045 with either plasmids containing different homology regions or elution buffer (negative control). From the colony count of these transformations, the conclusions can be made that (i) 4 h regeneration time resulted in satisfactory CopySwitch efficiency and (ii) 2,000 bp are a sufficient length for homology regions. Further, a kanamycin concentration of 25 µg/ml was superior to both 12.5 µg/ml, which resulted in false-positive clones after >17 h incubation period (data not shown) and 50 µg/ml (too high, no colonies formed at all). Higher colony forming units (cfus) were the result of prolonged incubation. In case of 2,000 bp homology and 12.5 µg/ml kanamycin a 4 h incubation resulted only in a doubling of OD_{600} but an increase of cfus by a factor of 35. Sequencing of three different, re-transformed plasmids showed no mutations of the copied locus.

Growth Curve Comparison Between One and Multiple Gene Copy Bearing Strains

Growth curves of a strain containing a chromosomal copy of P_{acoA}-*gfp* (BsFLN045) and a strain containing a chromosomal and a plasmid copy (BsFLN051) have been recorded at 37°C. During lag- and log-phase there is no visible difference between both curves. Contrary to this observation, in transient phase the plasmid-bearing strain is growing slower and does not reach an OD_{600} as high as BsFLN045 (see **Figure 2A**).

All strains of the mPLC experiment were also investigated for differences in growth behavior. At 30°C, no significant differences could be found (see **Figure 3A**). The reason for choosing 30°C instead of 37°C for
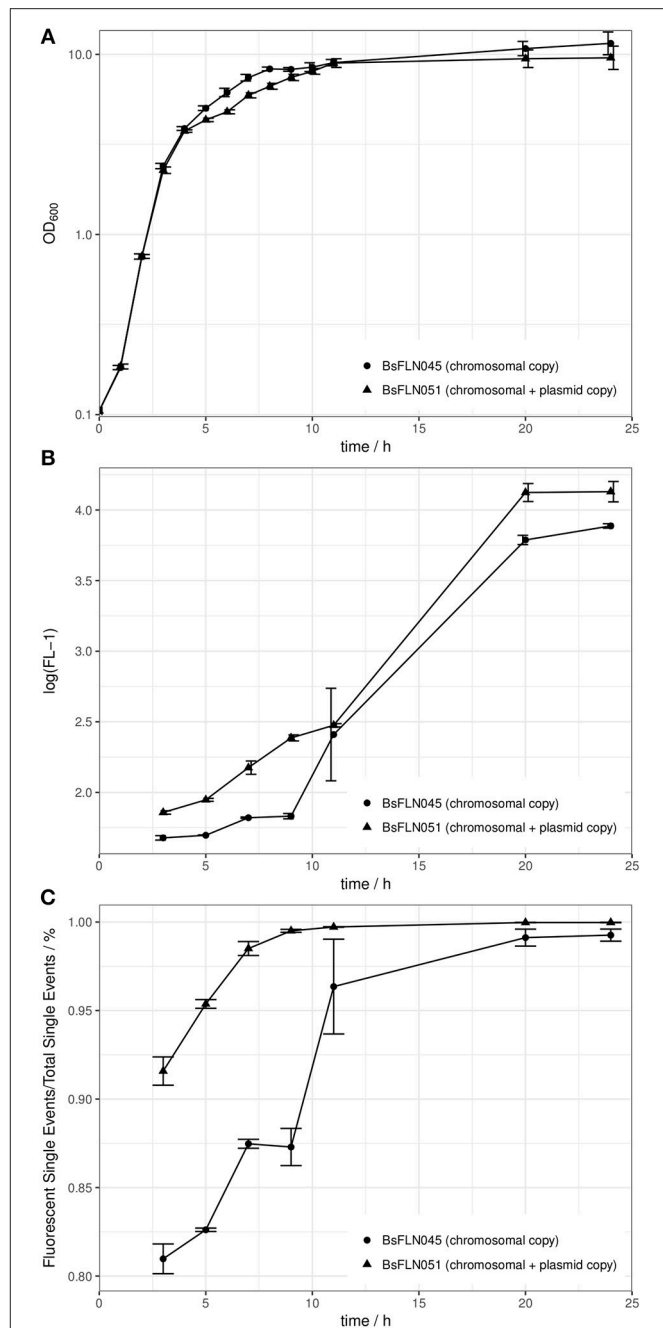
FIGURE 2 | **(A)** Averaged growth curves of strains BsFLN045 (chromosomal copy) and BsFLN051 (chromosomal + plasmid copy); **(B)** Logarithmic plot of increase in FL-1 fluorescence and assumed intracellular GFP content over cultivation time. For all points except at $t = 11$ h, the increased fluorescence of BsFLN051 is statistically significant (Student's *t*-test, alpha = 0.05). **(C)** Increase in the share of *gfp*-expressing cells over cultivation time. For all points except at $t = 11$ h the increased fluorescence of BsFLN051 is statistically significant (Student's *t*-test, alpha = 0.05). Plotted values are averages of $n = 4$ biological replicates with error bars showing standard deviation.

cultivation, lies in higher protein stability and solubility during heterologous expression at lower temperature (Schein and Noteborn, 1988).

FIGURE 3 | **(A)** Averaged growth curves of strains BsMKA005 (negative control), BsFLN064 (chromosomal copy), and BsFLN066 (chromosomal + plasmid copy); **(B)** OD_{600}-normalized slope values of phospholipase C assay as a measure of enzymatic activity. Since amount of substrate and volume of enzyme solution (= diluted crude supernatant) were constant and differences in OD_{600} at sampling time are normalized, the slope is a direct indicator for the amount of active enzyme produced by the according strain. The data used for calculation of the slope consists of 4 h of kinetic measurement. Plotted values are averages of n = 3 biological replicates with error bars showing standard deviation.

Determining *gfp* Expression Levels With Flow Cytometer and Microwell Plate Reader

To test whether our system has an influence on *gfp* gene expression, two different approaches were made. Flow cytometry was used to determine if more copies of *gfp* genes equal higher amounts of GFP protein per cell and additionally what the share of expressing cells was, compared to the total cell population.

In general, observed fluorescence in the FL-1 channel appeared to positively correlate with cultivation time. However, for BsFLN051 with additional, plasmid-based *gfp* expression, values were consistently higher than for BsFLN045 (see **Figure 2B**).

Similarly, the share of fluorescent cells is significantly higher for the plasmid carrying strain BsFLN051 for all time points except $t = 11$ h (see **Figure 2C**). At the same time, the number of non-fluorescent cells appears to be significantly lower if the plasmid is present, especially in early growth phases.

Our second approach involved the widely used standard procedure of measuring fluorescence of the whole sample in a microwell plate reader (see **Figure S1**). Since the OD_{600} was already in a suitable range (0.3–0.5), photometry samples were taken directly from the flow cytometry tubes of section "Sampling."

Application of CopySwitch to Change *mPLC* Expression Levels

After showing the principle of CopySwitch with the easy-to-monitor GFP, we applied the system to the expression of a secreted and industrially relevant enzyme. The chosen mPLC from *B. cereus* SBUG 516 cleaves the substrate *p*-NPPC and thereby releases p-nitrophenol, which can be measured as the increase of absorption at 410 nm (Durban et al., 2007). In total, four different time points were assayed (t6, t12, t24, and t48; number corresponds to hours after inoculation). Since every parameter was kept the same for every sample, the amount of enzyme in the sample directly correlates to the slope of the absorption change over time. OD_{410} values of each biological replicate were divided by their corresponding OD_{600} values before calculations to compensate for slightly different growth between replicates and strains. To make sure the increase of absorption is not due to background activity from the cultivation supernatant, each sample was additionally measured without substrate. Strain BsMKA005 without the *mPLC* gene was included as a negative control. Buffer containing substrate was measured as a autohydrolysis control. After 4 h of measurement at 30°C, no significant increase of absorption could be detected for (a) the negative control strain with and without substrate, (b) buffer with and without substrate and (c) mPLC-containing strains without substrate. All these controls strongly suggest mPLC as the sole source of *p*-NPPC-cleaving activity. After 6 h of cultivation, almost no enzymatic activity could be detected in the diluted supernatant used for the assay, indicating expression has not started yet. After 12 and 24 h of cultivation, the strain containing plasmid and chromosomal copies (BsFLN066) yields an about 1.3-fold increase of the slope (= activity) over the strain containing only a chromosomal *mPLC* copy (BsFLN064). After 48 h the activity difference drops to about 1.1-fold, still favoring BsFLN066 (see **Figure 3B**).

DISCUSSION

For CopySwitch being applicable as a tool for the *B. subtilis* community, we first had to find the right transformation parameters. Our results suggest these to be approximately 2,000 bp for each homology region, a regeneration time of 4 h and a kanamycin concentration of 25 μg/ml. The overall increased values, compared to standard chromosomal integration (500 bp, 1 h, 12.5 μg/ml) may stem from the plasmid nature of

the CopySwitch product, which requires separate replication machinery and replication of a circularized plasmid. The *ori* used in our system has a reported copy number of six units/chromosome (Titok et al., 2003). Since the used antibiotic resistance marker against kanamycin is also increased in gene copy number, this may explain the possibility of using higher antibiotic concentrations. In addition, we tried using the backbone of the high copy number plasmid pMSE3 (Silbersack et al., 2006; data not shown). These attempts did not show robust and reproducible results, likely due to the necessity of having four kb of adjacent chromosomal DNA for efficient homologous recombination on the plasmid, resulting in the increase of copy numbers of ORFs in these regions. For the *pksX* locus used in this work, this implies a copy number increase of *ymzD*, *ymcC* (both putative integral inner membrane proteins), *pksS* (cytochrome P450 of bacillaene metabolism) and *ymzB* (hypothetical protein) to several hundreds (Silbersack et al., 2006). The *pksX* locus had been chosen because it contains a secondary metabolite cluster for the polyketide bacillaene (Butcher et al., 2007), which has proven to be non-essential under laboratory conditions in our previous experiments (data not shown). Deletion of this cluster furthermore reduces the genome by 76.5 kb, which means less metabolic cost for replication and unnecessary protein production. It has already been shown elsewhere, that efficiency of recombinational transfer may be region-dependent (Tomita et al., 2004a). The integration of completely synthetic and unique regions into the chromosome could be a way to address this issue. To find out more about the metabolic stress applied by additional gene copy numbers, growth curves were recorded. There were no visible differences neither in lag-, nor in log-phase, indicating that the additional burden of plasmid replication in BsFLN051 is not interfering with growth (see **Figure 2A**). However, in the transition stage from log- to stationary phase, BsFLN051 seems to have a slight disadvantage and also does not reach the final optical density of BsFLN045. This is likely due to the nature of the acetoin-inducible promoter, which is controlled by catabolite repression, meaning expression from it only starts once residual glucose in the medium is exhausted. As shown elsewhere, the growth is impaired from the onset of protein expression. This effect is getting more pronounced as the number of promoter-gene constructs rises (Silbersack et al., 2006). With our second test system (mPLC), those effects were not observed, probably due to a lower cultivation temperature of 30°C and the omission of glucose.

Since the system is intended for fast optimization of heterologous gene expression, we performed test expressions of *GFPmut2* with (BsFLN051) and without (BsFLN045) CopySwitch. Flow cytometry indicates not only higher fluorescence of the plasmid-bearing strain on a single cell level, but also an increased share of fluorescent cells. Apparently, additional *gfp* transcription from CopySwitch plasmid p14073 appears to increase the total amount of GFP in cells. The reduction of the non-fluorescent cell population in BsFLN051 might be due to stronger deregulation of the utilized *acoA* promoter. The increased copy number of CcpA (Carbon catabolite control protein (A) binding sites in BsFLN051 compared to the single chromosomal promoter of BsFLN045

leads to a titration of CcpA repressor proteins. Since the medium contains glucose to ensure promoter repression and therefore undisturbed growth before the onset of protein production, *ccpA* plays an important role in our system (Ali et al., 2001). Additionally to flow cytometry, a photometric measurement of the whole sample was conducted in microwell plates directly after cytometry. As expected, the fluorescence intensity (normalized by OD_{600}) is following the same pattern, rising over time and being higher in the plasmid-bearing BsFLN051 (see **Figure S1**).

To strengthen our hypothesis, another experiment was performed with strains containing either no *mPLC* gene (BsMKA005), one *mPLC* gene integrated into the chromosome (BsFLN064) or *mPLC* both in the chromosome and on a low copy number plasmid (BsFLN066). The slope of the enzymatic activity (= absorption increase at 410 nm) was taken as read-out for comparison of the strains, after having been normalized against their respective OD_{600} values at the time of sampling. BsMKA005 showing no activity neither with nor without chromogenic substrate, suggests that mPLC is the enzyme responsible for absorption change in the supernatants of the remaining two strains. Since the negative control of buffer plus substrate also did not show any increase of absorption over time, autohydrolysis seems to play an insignificant role. After a cultivation time of 6 h, corresponding to the end of exponential phase, the first samples were taken. A very shallow slope indicates almost no activity for all three strains, which is in accordance to previous publications, showing maximum induction of acetoin catabolism after exponential phase (Ali et al., 2001). As expected, after 12 and 24 h, both BsFLN064 and BsFLN066 had increased levels of activity compared to BsMKA005, with BsFLN066 outcompeting BsFLN064 by 1.3-fold each time. Since activity values are OD_{600}-normalized and every parameter despite copy number has been kept constant, it is likely that CopySwitch is leading to the higher amount of enzyme in the supernatant, therefore being responsible for higher activity. This effect is still visible after 48 h of cultivation, although dropping down to 1.1-fold in favor of BsFLN066.

CONCLUSIONS

In comparison to other means of plasmid transformation in *B. subtilis*, CopySwitch has the great advantage of using naturally occurring competence. Other methods require either special *E. coli* strains for plasmid propagation, special devices for electroporation or tedious preparation of very fragile protoplasts. Additionally, all of the above-mentioned methods are either not well-suited for high-throughput applications due to lack of robustness or require expensive, specialized equipment. Since a lot of genetic engineering in *B. subtilis* is done chromosomally, this procedure is readily applicable for many strains. The CopySwitch method is as easy as growing *B. subtilis* in a competence-inducing medium, adding a linearized plasmid followed by addition of expression mix and plating. All these steps can be automated and parallelized with standard liquid-handling platforms.

In summary, our method comprises a very easy and straightforward tool for fast tuning of gene expression, which is scalable and easily automatizable. Future applications beyond the scope of this work could be an easy tuning of metabolic pathways by performing CopySwitch with complete pathways or subsets of these.

AVAILABILITY OF DATA AND MATERIALS

The flow cytometry datasets analyzed during the current study are available from the corresponding author on reasonable request.

AUTHOR CONTRIBUTIONS

FN conducted the experiment. FN and FB analyzed the data. FN, FB and JK prepared the manuscript. All authors read and approved the final manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2018.00207/full#supplementary-material

REFERENCES

Ali, N. O., Bignon, J., Rapoport, G., and Debarbouille, M. (2001). Regulation of the acetoin catabolic pathway is controlled by sigma L in *Bacillus subtilis*. *J. Bacteriol.* 183, 2497–2504. doi: 10.1128/JB.183.8.2497-2504.2001

Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., et al. (1977). Construction and characterization of new cloning vehicle. *II. A* multipurpose cloning system. *Gene* 2, 95–113. doi: 10.1016/0378-1119(77)90000-2

Butcher, R. A., Schroeder, F. C., Fischbach, M. A., Straight, P. D., Kolter, R., Walsh, C. T., et al. (2007). The identification of bacillaene, the product of the PksX megacomplex in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U. S. A.* 104, 1506–1509. doi: 10.1073/pnas.0610503104

Butt, T. R., Edavettal, S. C., Hall, J. P., and Mattern, M. R. (2005). SUMO fusion technology for difficult-to-express proteins. *Protein Expr. Purif.* 43, 1–9. doi: 10.1016/j.pep.2005.03.016

Canosi, U., Morelli, G., and Trautner, T. A. (1978). The relationship between molecular structure and transformation efficiency of some *S. aureus* plasmids isolated from *B. subtilis*. *Mol. Gen. Genet.* 166, 259–267.

Durban, M. A., Silbersack, J., Schweder, T., Schauer, F., and Bornscheuer, U. T. (2007). High level expression of a recombinant phospholipase C from *Bacillus cereus* in *Bacillus subtilis*. *Appl. Microbiol. Biotechnol.* 74, 634–639. doi: 10.1007/s00253-006-0712-z

Elena, C., Cerminati, S., Ravasi, P., Rasia, R., Peiru, S., Menzella, H. G., et al. (2017). *B. cereus* phospholipase C engineering for efficient degumming of vegetable oil. *Process Biochem.* 54, 67–72. doi: 10.1016/j.procbio.2017.01.011

Hadley, W., Romain, F., Lionel, H., and Kirill M. (2017). *dplyr: A Grammar of Data Manipulation*. Available at: https://CRAN.R-project.org/package=dplyr

Hahne, F., LeMeur, N., Brinkman, R. R., Ellis, B., Haaland, P., Sarkar, D., et al. (2009). flowCore: a Bioconductor package for high throughput flow cytometry. *BMC Bioinformatics* 10:106. doi: 10.1186/1471-2105-10-106

Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* 12, 115–121. doi: 10.1038/nmeth.3252

Ihaka, R., and Gentleman, R. (1996). R: A language for data analysis and graphics. *J. Comput. Graph. Stat.* 5:299.

Irla, M., Heggeset, T. M. B., Nærdal, I., Paul, L., Haugen, T., Le, S. B., et al. (2016). Genome-based genetic tool development for *Bacillus methanolicus*: theta- and rolling circle-replicating plasmids for inducible gene expression and application to methanol-based cadaverine production. *Front. Microbiol.* 7:1481. doi: 10.3389/fmicb.2016.01481

Jiang, X., Chang, M., Jin, Q., and Wang, X. (2015). Optimization of the degumming process for camellia oil by the use of phospholipase C in pilot-scale system. *J. Food Sci. Technol.* 52, 3634–3644. doi: 10.1007/s13197-014-1418-3

Kabisch, J., Thürmer, A., Hübel, T., Popper, L., Daniel, R., and Schweder, T. (2013). Characterization and optimization of *Bacillus subtilis* ATCC 6051 as an expression host. *J. Biotechnol.* 163, 97–104. doi: 10.1016/j.jbiotec.2012.06.034

Kohl, A., Srinivasamurthy, V., Böttcher, D., Kabisch, J., and Bornscheuer, U. T. (2018). Co-expression of an alcohol dehydrogenase and a cyclohexanone monooxygenase for cascade reactions facilitates the regeneration of the NADPH cofactor. *Enzyme Microb. Technol.* 108, 53–58. doi: 10.1016/j.enzmictec.2017.09.003

Kraft, M., Knüpfer, U., Wenderoth, R., Pietschmann, P., Hock, B., and Horn, U. (2007). An online monitoring system based on a synthetic sigma32-dependent tandem promoter for visualization of insoluble proteins in the cytoplasm of *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 75, 397–406. doi: 10.1007/s00253-006-0815-6

Kumpfmüller, J., Kabisch, J., and Schweder, T. (2013). An optimized technique for rapid genome modifications of Bacillus subtilis. *J. Microbiol. Methods* 95, 350–352. doi: 10.1016/j.mimet.2013.10.003

Kurioka, S., and Matsuda, M. (1976). Phospholipase C assay using *p*-nitrophenylphosphoryl-choline together with sorbitol and its application to studying the metal and detergent requirement of the enzyme. *Anal. Biochem.* 75, 281–289. doi: 10.1016/0003-2697(76)90078-6

Lee, M. E., DeLoache, W. C., Cervantes, B., and Dueber, J. E. (2015). A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth. Biol.* 4, 975–986. doi: 10.1021/sb500366v

Lian, J., Jin, R., and Zhao, H. (2016). Construction of plasmids with tunable copy numbers in *Saccharomyces cerevisiae* and their applications in pathway optimization and multiplex genome integration. *Biotechnol. Bioeng.* 113, 2462–2473. doi: 10.1002/bit.26004

Messerschmidt, K., Hochrein, L., Dehm, D., Schulz, K., and Mueller-Roeber, B. (2016). Characterizing seamless ligation cloning extract for synthetic biological applications. *Anal. Biochem.* 509, 24–32. doi: 10.1016/j.ab.2016.05.029

Rahmer, R., Morabbi Heravi, K., and Altenbuchner, J. (2015). Construction of a super-competent *Bacillus subtilis* 168 using the P_{mtlA}-*comKS* inducible cassette. *Front. Microbiol.* 6:1431. doi: 10.3389/fmicb.2015.01431

Schein, C. H., and Noteborn, M. H. M. (1988). Formation of soluble recombinant proteins in *Escherichia coli* is favored by lower growth temperature. *Nat. Biotechnol.* 6, 291–294. doi: 10.1038/nbt0388-291

Segall-Shapiro, T. H., Sontag, E. D., and Voigt, C. A. (2018). Engineered promoters enable constant gene expression at any copy number in bacteria. *Nat. Biotechnol.* 36, 352–358. doi: 10.1038/nbt.4111

Sharan, S. K., Thomason, L. C., Kuznetsov, S. G., and Court, D. L. (2009). Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protoc.* 4, 206–223. doi: 10.1038/nprot.2008.227

Silbersack, J., Jürgen, B., Hecker, M., Schneidinger, B., Schmuck, R., and Schweder, T. (2006). An acetoin-regulated expression system of *Bacillus subtilis*. *Appl. Microbiol. Biotechnol.* 73, 895–903. doi: 10.1007/s00253-006-0549-5

Sørensen, H. P., and Mortensen, K. K. (2005). Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb. Cell Fact.* 4:1. doi: 10.1186/1475-2859-4-1

Stevens, R. C. (2000). Design of high-throughput methods of protein production for structural biology. *Structure* 8, R177–R185. doi: 10.1016/S0969-2126(00)00193-3

Titok, M., Suski, C., Dalmais, B., Ehrlich, S. D., and Jannière, L. (2006). The replicative polymerases PolC and DnaE are required for theta replication of the *Bacillus subtilis* plasmid pBS72. *Microbiology* 152, 1471–1478. doi: 10.1099/mic.0.28693-0

Titok, M. A., Chapuis, J., Selezneva, Y. V., Lagodich, A. V., Prokulevich, V. A., Ehrlich, S. D., et al. (2003). *Bacillus subtilis* soil isolates: plasmid replicon analysis and construction of a new theta-replicating vector. *Plasmid* 49, 53–62. doi: 10.1016/S0147-619X(02)00109-9

Tomita, S., Tsuge, K., Kikuchi, Y., and Itaya, M. (2004a). Region dependent efficiency for recombinational transfer of the *Bacillus subtilis* 168 genome. *Biosci. Biotechnol. Biochem.* 68, 1382–1384. doi: 10.1271/bbb.68.1382

Tomita, S., Tsuge, K., Kikuchi, Y., and Itaya, M. (2004b). Targeted isolation of a designated region of the *Bacillus subtilis* genome by recombinational transfer. *Appl. Environ. Microbiol.* 70, 2508–2513. doi: 10.1128/AEM.70.4.2508-2513.2004

Tsuge, K., and Itaya, M. (2001). Recombinational transfer of 100-kilobase genomic DNA to plasmid in *Bacillus subtilis* 168. *J. Bacteriol.* 183, 5453–5458. doi: 10.1128/JB.183.18.5453-5458.2001

Vojcic, L., Despotovic, D., Martinez, R., Maurer, K.-H., and Schwaneberg, U. (2012). An efficient transformation method for *Bacillus subtilis* DB104. *Appl. Microbiol. Biotechnol.* 94, 487–493. doi: 10.1007/s00253-012-3987-2

Voss, C., Schmidt, T., Schleef, M., Friehs, K., and Flaschel, E. (2003). Production of supercoiled multimeric plasmid DNA for biopharmaceutical application. *J. Biotechnol.* 105, 205–213. doi: 10.1016/j.jbiotec.2003.07.004

Wickham, H. (2011). ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.* 3, 180–185. doi: 10.1002/wics.147

Wickham, H. (2017). *Tidyverse: Easily Install and Load the "Tidyverse."* Available online at: https://CRAN.R-project.org/package=tidyverse

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors. *Gene* 33, 103–119. doi: 10.1016/0378-1119(85)90120-9

Zobel, S., Kumpfmüller, J., Süssmuth, R. D., and Schweder, T. (2015). *Bacillus subtilis* as heterologous host for the secretory production of the non-ribosomal cyclodepsipeptide enniatin. *Appl. Microbiol. Biotechnol.* 99, 681–691. doi: 10.1007/s00253-014-6199-0

Gene-Expressing Liposomes as Synthetic Cells for Molecular Communication Studies

Giordano Rampioni[1], Francesca D'Angelo[1], Livia Leoni[1] and Pasquale Stano[2]*

[1] Department of Science, University Roma Tre, Rome, Italy, [2] Department of Biological and Environmental Sciences and Technologies (DiSTeBA), University of Salento, Lecce, Italy

The bottom-up branch of synthetic biology includes—among others—innovative studies that combine cell-free protein synthesis with liposome technology to generate cell-like systems of minimal complexity, often referred to as synthetic cells. The functions of this type of synthetic cell derive from gene expression, hence they can be programmed in a modular, progressive and customizable manner by means of *ad hoc* designed genetic circuits. This experimental scenario is rapidly expanding and synthetic cell research already counts numerous successes. Here, we present a review focused on the exchange of chemical signals between liposome-based synthetic cells (operating by gene expression) and biological cells, as well as between two populations of synthetic cells. The review includes a short presentation of the "molecular communication technologies," briefly discussing their promises and challenges.

Keywords: synthetic cells, bottom-up synthetic biology, molecular communications, quorum sensing, lipid vesicles (liposomes), cell-free protein synthesis

MOLECULAR COMMUNICATIONS AND SYNTHETIC CELLS (SCs)

Natural organisms coordinate their activities through communication. Isolated cells, tissue cells, as well as higher organisms, share their environment with other living forms. Tactile, physical, and especially chemical signals define in unique and complex manner the sensory world of living beings. Communications in the chemical domain are ubiquitous intercellular processes, and play important roles in all organisms.

Inspired by the already mentioned capabilities of natural organisms, a new branch of biomimetic technology has been proposed which focuses on *molecular communications* (Nakano et al., 2011, 2013). Network engineers have envisioned the exploitation of chemical exchanges as the basis for developing new types of Information and Communication Technologies (the so-called bio-chem-ICTs, **Figure 1A**). This is an exciting new arena for engineers and biologists that aims at the construction of well-characterized biological parts, devices, and systems that will process chemical information in a controlled and programmable manner, as it happens with classical electric signals. The challenge, here, relies on the ability of managing communication and information processing through chemical signals with the same mastery as nature has done for billions of years. Such a broad and innovative territory of research offers several opportunities for various approaches to synthetic biology, which needs adequate theoretical frameworks, numerical modeling strategies, and experimental methodologies. More generally, bio-chem-ICTs refers to radically new forms of computation, communication, and information processing approaches—at the nano- and micro-scale levels—based on chemical and biochemical systems (Amos et al., 2011).
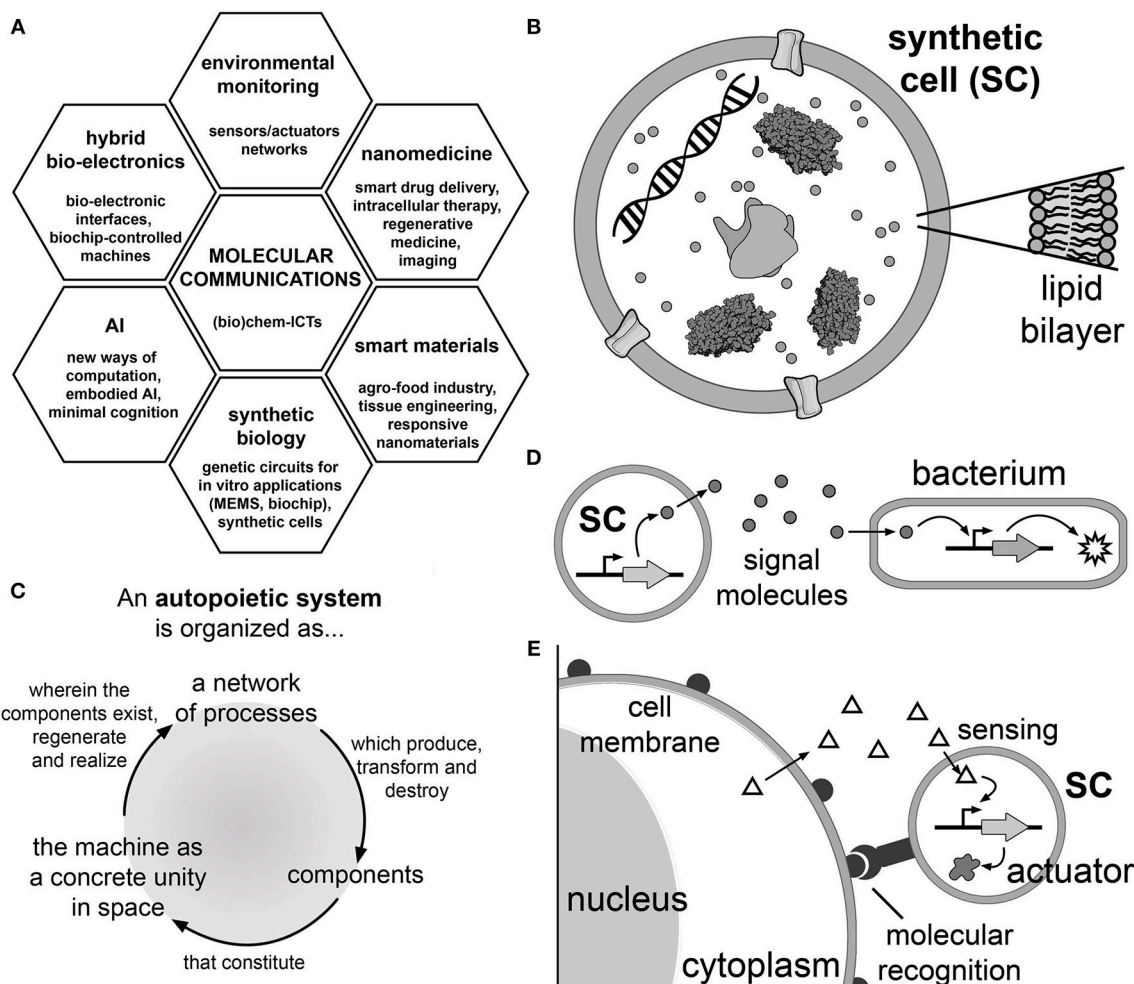
FIGURE 1 | Molecular communications based on synthetic cell (SC) technology. **(A)** Application areas of molecular communication research. Molecular communication is a (bio)chem-information and communication technology that can be applied to nanomedicine (smart drug delivery systems), smart responsive materials, synthetic biology (construction of biochips), artificial intelligence (AI), hybrid bio-electronic systems and for sensors in environmental monitoring (Nakano et al., 2013). **(B)** Synthetic cells are cell-like systems, generally built by encapsulating a number of (bio)molecular components into artificial micro-compartments. One of the possible designs focuses on liposome-based SCs operating by gene expression (Luisi, 2002; Luisi et al., 2006). With this aim, TX-TL kits produce the protein(s) of interest starting from the corresponding DNA sequence. The SC membrane can be functionalized with membrane proteins as pores (Noireaux and Libchaber, 2004) and receptors (Hamada et al., 2014); cytoskeletal proteins have been implemented as well (Maeda et al., 2012). **(C)** The principles of autopoiesis (self-production) (Varela et al., 1974), which guides the long-term goal of constructing SCs that produce all their components. Autopoiesis provides insights into the spatial and dynamical organization that a chemical system should be endowed with in order to display self-maintenance, organizational closure, homeostasis and reproduction achieved by the internal processes of manufacturing and assembling its components. **(D)** Schematic representation of a SC which produces and releases a signal molecule into the environment. The signal is perceived by a natural cell (e.g., a bacterium) that consequently activates a response (for example, a reporter protein, an enzyme operating as an actuator to perform a certain operation, including a reply signaling) (Nakano et al., 2011; Stano et al., 2012). **Table 1** reports several cases of unidirectional or bidirectional molecular communications between SCs, or between SCs and natural cells. **(E)** The vision of using SCs as smart drug delivery systems or for enzyme replacement therapy (Leduc et al., 2007). SCs, intended as a biotechnological evolution of current liposomes for drug delivery, reach and bind to the target cells by a molecular recognition mechanism and activate their internal circuits responding to chemical stimuli and consequently act, in a programmable manner, for a certain task (e.g., producing a therapeutic or diagnostic agent Ding et al., 2018; Krinsky et al., 2018, or a secondary easy-to-detect signal, etc.). The chemical stimulus can be an endogenous chemical that derives from the target cell itself (as shown in the cartoon) or from other tissues (not shown), as well as purposely-added exogenous chemicals (not shown).

Owing to our direct involvement in the field (Stano et al., 2012; Rampioni et al., 2014, 2018), and considering recent exciting reports, in this review we present and discuss the intersection between the bio-chem-ICT idea of exchanging chemical signals in a programmable way, and the bottom-up synthetic biology approach focused on the construction of cell-like systems based

on gene expression inside liposomes (Luisi, 2002; Noireaux and Libchaber, 2004; Luisi et al., 2006; Ichihashi et al., 2010; Stano et al., 2011; Nourian and Danelon, 2013; Spencer et al., 2013). For simplicity, we will shortly refer to these systems simply as "synthetic cells" (SCs, **Figure 1B**), keeping in mind that these are rather simple mimics of biological cells.

In this mini-review, the principles on which liposome-based SCs operate will be summarized, together with an explanation of the reason why they could contribute significantly to molecular communication technologies on account of their inherent possibilities in terms of design, modeling, control, programmability, and modularity. Next, recent experimental reports focused on chemical communication between SCs and natural cells (or with other SCs) will be reviewed (see also Lentini et al., 2016), while the opportunities and challenges facing this novel research arena will be discussed in the final section.

Before advancing in the discussion, two notes of warning are intended for readers unfamiliar with this research field. Firstly, the term "synthetic cell" is also used in synthetic biology to indicate *living* cells generated either by engineering biological cells (e.g., metabolic engineering, genetic optimization, or reprogramming), as well as by the transplantation of an entire synthetic genome in a living cell deprived of its own genome. Second, bottom-up synthetic biology approaches aiming at constructing cell-like systems are not restricted to liposome-based SCs. No less interesting are systems based on other types of compartments (Walde et al., 1994; Martino et al., 2012; Huang et al., 2014; Karzbrun et al., 2014; Dora Tang et al., 2015; Rideau et al., 2018), nor those based on new artificial molecules (Kurihara et al., 2011; Marguet et al., 2013; Taylor et al., 2015). Interested readers can refer to recent reviews for a broader discussion (Buddingh and van Hest, 2017; Salehi-Reyhani et al., 2017; Göpfrich et al., 2018; Schwille et al., 2018). The current review will focus only on SCs based on gene expression inside liposomes.

BASIC PRINCIPLES ON LIPOSOME-BASED SCs OPERATING VIA GENE EXPRESSION

SCs based on gene expression inside liposomes find their origin in early studies on cell models aiming at achieving minimal life-like behaviors (Morowitz et al., 1988; Luisi and Varela, 1989; Schmidli et al., 1991; Oberholzer et al., 1995a,b, 1999; Szostak et al., 2001; Luisi, 2002; Pohorille and Deamer, 2002; Mansy and Szostak, 2009). Born within the origins-of-life community, this research was intended as a means of investigating the emergence of life on Earth, more precisely by demonstrating the emergence of life as a system-level phenomenon due to a particular type of organization (the autopoietic one). Hence, the *autopoietic* (self-production) (Varela et al., 1974; Luisi and Varela, 1989; Luisi, 2003) (**Figure 1C**) and the *chemoton* theories (chemical automaton) (Gánti, 1975) are two valuable theoretical frameworks for the construction of SCs which display features of biological organisms. Starting in the first years of 2000, SCs and similar constructs became highly relevant also in the context of synthetic biology, either as tools for generating basic knowledge, or as systems designed for applied research, i.e., biotechnology and nanomedicine.

The SCs discussed in this review are liposomes, with a size ranging typically from 0.1 to 10–100 μm: they contain DNA and a cell-free gene expression system. They are made by assembling liposomes in an aqueous phase which contains all the molecules needed to be encapsulated for accomplishing protein

synthesis from a DNA template (e.g., enzymes, ribosomes, tRNAs, nucleotides, amino acids etc.). The protein synthesis machinery can derive from a cell extract or from a reconstituted system [such as the PURE system (Shimizu et al., 2001)]. Accordingly, it can be noted that SC technology is based on liposome technology (including microfluidics) and cell-free systems (including biochemical reconstitution approaches). As a result of the reactions occurring in their aqueous lumen and/or on their boundary surface, SCs can display behavior(s) typical of living cells. For example, SCs produce proteins from a corresponding gene; in turn, the synthesized protein can be an enzyme that converts substrates into products, or it can be a pore-forming protein, creating pores on the liposome membrane, or it can be a receptor that binds a signal molecule, etc. More in general, SCs can be functionalized with any chemical network of biological relevance that is functional *in vitro*.

Several reactions different from gene expression have been successfully performed inside liposomes, confirming the potentiality of SCs in terms of scope, programmability, and functionality. Some examples are: PCR and RT-PCR (Oberholzer et al., 1995a; Shohda et al., 2011; Lee et al., 2014; Tsugane and Suzuki, 2018), DNA replication (Sakatani et al., 2018; van Nies et al., 2018), and several enzymatic reactions. Moreover cytoskeletal elements have been reconstituted inside SCs (Cabré et al., 2013; Furusato et al., 2018; Litschel et al., 2018). *Ad hoc* designed gene circuits lead to SCs that can perform useful operations in a programmable way, including communication, as discussed below. SCs with the capacity of self-producing *all* their own constitutive components, and which possibly grow-and-divide as living cells do, are still missing, although interesting reports that show progress in this directions have been published (Kurihara et al., 2011).

This mini-review focuses on SCs capable of communicating with biological cells and with each other. However, other interesting research directions are under current development, including the construction of SCs with nested design (Deng et al., 2017; York-Duran et al., 2017; Hindley et al., 2018), the production of ATP inside SCs (Feng et al., 2016; Altamura et al., 2017; Lee et al., 2018), the attempts of self-producing SC parts (Schmidli et al., 1991; Kuruma et al., 2009; Scott et al., 2016; Li et al., 2017; Exterkate et al., 2018), and the shift from isolated SCs to "SC communities", including tissue-like structures (Carrara et al., 2012; Hadorn et al., 2013; Booth et al., 2016).

SCs THAT EXCHANGE CHEMICAL SIGNALS: A BOTTOM-UP SYNTHETIC BIOLOGY PLATFORM FOR MOLECULAR COMMUNICATIONS

SCs based on gene expression inside liposomes can be useful tools for developing molecular communication technologies (Stano et al., 2012). Current SC technology allows building simple systems capable of exchanging chemical signals, and therefore performing elementary signal processing. The idea is to design SCs capable of communicating with each other or with biological cells in a programmable manner (**Figure 1D**). This

innovative perspective has multifold theoretical and practical consequences. From the theoretical viewpoint, SCs that can regulate their internal mechanisms in response to external perturbations (the chemical signaling) are *de facto* experimental tools for investigating minimal cognitive systems (Damiano and Stano, 2018a,b). Considering the proposed extension of the Turing imitation game to the SC realm (Cronin et al., 2006), molecular communication can contribute to the determination of life-likeness criteria as referred to SCs, as recently investigated by the Sheref Mansy group (Lentini et al., 2017). In a more practical perspective, an expansion of actual drug delivery strategies can be proposed. Inspired by the scenario depicted by Leduc and collaborators (**Figure 1E**) (Leduc et al., 2007), SCs could activate internal mechanisms upon perception of chemical signals, thus acting as "intelligent" drug carriers. As an example, SCs could be targeted to specific cells (e.g., tumoural cells) by exploiting antigen-antibody recognition. Once localized, their internal genetic circuit could be activated by chemical stimuli produced by the target cell itself or by other endogenous or exogenous chemical signals. These "smart" SCs could produce and release therapeutics (or drugs) *in situ*. Note that a recent study has reported SCs (injected into the tumor) that constitutively produce a toxin against breast cancer cells (Krinsky et al., 2018). The therapeutic (or diagnostic) use of SCs is, today, still a hypothetic scenario. Nevertheless, continuous improvements in SC design and construction is expected to favor a more rapid prototyping, thus accelerating the path toward applicative purposes.

Sensors, Actuators, Controllers, and Molecular Diffusion

Like hardware robots or conventional communication devices, SCs are embodied systems composed of molecular elements that perform specific operations. Hardware components, such as sensors, controllers, and actuators (Mataric, 2007; Wang et al., 2013) have their molecular counterparts in SCs.

In the context of SCs operating by gene expression, sensors can be protein receptors or RNA aptamers that bind to a signal molecule and consequently change their conformation. This event directly or indirectly affects the "controller system," which is based on the regulation of gene expression by protein receptors or RNA aptamers (riboswitches) at the transcriptional or translational level, respectively. These mechanisms are well-understood (Alberts et al., 2014). Depending on its design, the regulatory circuit can involve a single gene or multiple genes. As a result of this sensing-and-regulation system, the synthesis of an actuator (a protein) is promoted or inhibited. In turn, the actuator operates on some further step (e.g., producing a signal molecule, catalyzing a useful reaction, creating a pore on the SC membrane, acting as a controller/regulator of another circuit, etc.). Key examples of this general mechanism will be commented on in section A Survey of Published Reports and listed in **Table 1**.

To provide SCs with communication capability, water-soluble proteins (sensors, regulators, signal-producing elements, or components of the gene expression machinery) should be either encapsulated, or synthesized in the SC lumen. This has become a standard practice, somehow, at least for some prokaryotic proteins (Stano et al., 2011). It is not trivial, instead, dealing with membrane-associated and integral membrane sensors/receptors, even if reports have shown that this is a feasible goal in SCs technology (strategies as membrane protein reconstitution Yanagisawa et al., 2011; Altamura et al., 2017; Jørgensen et al., 2017 or synthesis-from-within Kuruma et al., 2009; Hamada et al., 2014; Soga et al., 2014 have been employed). Genetic circuits of distinctive complexity have already been proven to be functional, also inside liposomes (Noireaux et al., 2003; Shin and Noireaux, 2012; Siegal-Gaskins et al., 2014).

In addition to molecular elements, in order to establish an intercellular communication channel, diffusion of the signal molecule in the outer aqueous environment should be taken into account. The signal molecule cannot be directed toward the communication partner, but it spreads in all direction, guided by the concentration gradient. Although the average behavior of many signal molecules can be foreseen, individual molecules follow an erratic path. In addition to free diffusion, for closely packed SCs, communication through gap junctions (reconstituted in liposomes) has been proposed (Ramundo-Orlando et al., 2005; Moritani et al., 2010).

A Survey of Published Reports

The pioneer experimental report on a simple cell-like system sending a signal molecule to biological cells was published by the Ben Davis group (Gardner et al., 2009). The authors encapsulated the precursors of the formose reaction inside liposomes, and observed that one class of products of the intra-vesicular reaction escaped the liposomes through a channel formed by α-haemolysin and spontaneously reacted with the borate ions present in the external medium to generate furanosyl-boronates structurally similar to the quorum sensing (QS) signal molecule AI-2, that naturally triggers bioluminescence in *Vibrio harveyi*. Remarkably, the "synthetic" signal released by the liposome was able to induce natural behavior (i.e., light emission) in this bacterium.

Despite its great interest as proof of the concept study, the SCs used by Ben Devis and co-workers were not based on gene expression, therefore they lacked those aspects of programmability and control that are peculiar to synthetic biology. Being a novel research area, literature on the liposome-based SCs which operate by gene expression to interface with natural cells (or with other SCs) is, to the best of our knowledge, limited to the six studies that are summarized in **Table 1** together with the already cited study by Gardner et al. (2009). Additional cases involving non-liposome compartments are also available (Gupta et al., 2013; Schwarz-Schilling et al., 2016; Sun et al., 2016; Niederholtmeyer et al., 2018), but these will not be discussed in this mini-review.

In 2014, Sheref Mansy and collaborators designed SCs acting as "translators" for the bacterium *Escherichia coli*, using theophylline as trigger and isopropyl β-D-1-thiogalactopyranoside (IPTG) as signal molecule (Lentini et al., 2014). These SCs are liposomes containing IPTG, the PURE system as the transcription-translation (TX-TL) machinery, and a DNA template coding for a riboswitch that, after binding

TABLE 1 | Comparative analysis of six experimental reports based on liposome SCs that send/receive chemical signals (2009–2018).

#	1	2	3	4	5	6	7
Reference No.	Gardner et al., 2009	Lentini et al., 2014	Adamala et al., 2017	Lentini et al., 2017	Tang et al., 2018	Rampioni et al., 2018	Ding et al., 2018
Year	2009	2014	2017	2017	2017	2018	2018
Directionality	SCs→ bacteria	SCs→ bacteria	SCs→ SCs	bacteria→ SCs	SCs→ SCs	SCs→ bacteria	bacteria→ SCs
Sender	SCs: DPhPC vesicles encapsulating the reagents of the formose reaction	SCs: POPC/chol vesicles encapsulating the PURE system	SCs: POPC/chol vesicles encapsulating E. coli extracts	Bacteria: V. fischeri	SCs: POPC/chol vesicles encapsulating the PURE system	SCs: POPC vesicles encapsulating the PURE system	Bacteria: E. coli
Sender details (specifications, preparation…)	Thin film hydration, sonication, freezing-thawing, extrusion (100nm)	Hydration of freeze-dried homogenized liposomes	Hydration of freeze-dried extruded (1 µm) liposomes	Hydration of freeze-dried liposomes	Hydration of freeze-dried extruded (1 µm) liposomes	Droplet transfer method (2–5 µm)	Transformed with a plasmid encoding for EsaI under T7 promoter
Production of signal molecule?	Yes	No	No	Yes	No	Yes	Yes
Actuator(s) for the production of signals or for its emission	Formose reaction	α-Haemolysin (pore formation) produced by gene expression as a response to theophylline	α-Haemolysin (pore formation), produced by gene expression as a response to arabinose or theophylline	LuxI synthase produced by the cellular own gene; E. coli converts C2-CoA into longer fatty acid chains, LuxI, produced by gene expression synthesizes the signal molecule	α-Haemolysin (pore formation), produced by gene expression as a response to 3OC6-HSL	RhlI, produced by gene expression, synthesizes the signal molecule	EsaI, produced by gene expression
Precursors	Formaldehyde, borate	None	None	Cellular metabolite precursors; SAM, C2-CoA	None	SAM, C4-CoA	Cellular metabolite precursors
Signal molecule	AI-2-like borate-esters of formose products	IPTG	IPTG (or doxycycline)	3OC6-HSL; HSLs molecules	Glucose	C4-HSL	HSL molecules
Exporting system	α-Haemolysin	α-Haemolysin	α-Haemolysin	None (diffusion through the bacterial membrane); None (diffusion through the vesicle membrane)	α-Haemolysin	None (diffusion through the vesicle membrane)	None (diffusion through the bacterial membrane)
Medium	Liquid LB medium	Liquid M9 minimal medium	Liquid HEPES buffer including KCl and MgCl2	Liquid LBS medium	Liquid HEPES buffer	Gel LB-Agar	Liquid ACB buffer
Receiver	Bacteria: V. harveyi	Bacteria: E. coli	SCs: POPC/chol vesicles encapsulating E. coli (or HeLa cell) extracts	SCs: POPC/chol vesicles encapsulating E. coli extracts	SCs: Proteinosomes made of PNIPAAm (GOx-conjugated)	Bacteria: P. aeruginosa	SCs: POPC/chol vesicles encapsulating E. coli extracts
Receiver details (specifications, preparation…)	V. harveyi MM32 mutant strain unable to detect AI-1 and to produce AI-2, but capable of detecting AI-2-type molecules.	E. coli BL21(DE3) pLysS carrying a plasmid encoding GFP under the control of a T7 promoter and a lac operator sequence.	Hydration of freeze-dried homogenized liposomes, and extrusion (1 µm)	Hydration of freeze-dried liposomes; V. harveyi strain in which light emission is increased in response to the signal molecule produced by SCs	HRP encapsulated in the proteinosomes prepared by water droplet/oil interfacial assembly, membrane cross-linking, and phase transfer	A P. aeruginosa mutant stain unable to synthesize endogenous C4-HSL (rhlI mutant) carrying the PrhIA::lux transcriptional fusione activated by the RhlR/C4-HSL complex	Droplet transfer method (10–50 µm)
Sensor	LuxP/LuxQ sensor/transducer protein	lac repressor	lac repressor (or Tet protein)	LuxR	Glucose oxidase	RhlR	EsaR

(Continued)

TABLE 1 | Continued

#	1	2	3	4	5	6	7
Response	Bioluminescence	Fluorescence (GFP)	Bioluminescence	Bioluminescence (luxA and luxB). Synthesis of a response signal: LasI-mediated AHLs synthesis from precursors (C2-HSL, SAM)	Fluorescence (resorufin, from Amplex Red via HRP catalysis)	Bioluminescence, fluorescence (mCherry)	Synthesis and export of an antimicrobial peptide (Bac2A)
Further comments	First example of SCs sending a message to natural cells	SCs act as "translator" for natural cells	Molecular communication between two liposome-based SCs populations	Bidirectional communication between SCs and natural cells (V. fischeri). The article includes the study of many other systems.	Molecular communication between two different types of SCs populations (liposomes and proteinosomes)	Unidirectional communication between SCs (sender) and P. aeruginosa (receiver)	SCs receiver respond by producing a toxic peptide that kills the sender bacteria.

Note that the first study (column 1) is not based on gene expression. Additional examples, whereby other-than-liposome compartments are used, can be found in Schwarz-Schilling et al. (2016), Sun et al. (2016), Niederholtmeyer et al. (2018), The most representative results are shown; more details can be found in the original publications.

3OC6-HSL, N-(3-oxohexanoyl)-l-homoserine lactone; ACB, buffer containing the same components of SC interiors; AI-2, autoinducer-2; C2-CoA, acetyl-coenzyme A; C4-CoA, butyryl-coenzymeA; C4-HSL, N-butanoyl-l-homoserine lactone; C8-HSL, N-octanoyl-l-homoserine lactone; chol, cholesterol; DPhPC, 1,2-di-O-phytanoyl-sn-glycero-3-O-phosphocholine; EsaR and EsaI, signal receptor and synthase from Erwinia stewartii; GOx, glucose oxidase; HRP, horseradish peroxidase; LBS, lactobacillus selective; PNIPAAm, poly(N-isopropylacrylamide) cross-linked with glucose oxidase; POPC, 1-palmitoyl-2-oleoyl-sn-glycero-3-phsphatidylcholine; SAM, S-adenosylmethionine; Tet, ten-eleven translocation protein.

to the free-diffusible molecule theophylline, activated the expression of the pore forming protein α-haemolysin. The authors demonstrated that only in the presence of theophylline, did IPTG escape the liposomes through α-haemolysin, and activate the expression of the green fluorescent protein (GFP) gene in receiver *E. coli* cells. In this way, SCs acted as chemical translators allowing *E. coli* to sense theophylline (the latter molecule cannot be normally sensed by *E. coli*).

Adamala et al. (2017) built SCs containing engineered genetic circuits and regulatory cascades. These SCs can be controlled/triggered by external signals, and can be fused together in order to bring together products of incompatible reactions. In particular, the group lead by Edward Boyden showed that by using cell lysates with transcriptional-translational activity, DNA vectors encoding genes for IPTG (or doxycycline) detection and permeable chemical inducers, as arabinose or theophylline, the arabinose (or theophylline) activates the α-haemolysin production in the first SC population, so that pre-encapsulated impermeable IPTG (or doxycycline) could be released, and thus activate a response in a second SC population.

The group of Sheref Mansy recently reported two-way chemical communication between SCs and bacteria (Lentini et al., 2017). They exploited cell extracts to generate SCs able to synthesize molecules perceived by *Vibrio fischeri*, *V. harveyi*, *E. coli*, and *Pseudomonas aeruginosa*. In particular, the expression of LuxI-like synthases inside liposomes, in the presence of acetyl coenzyme A and *S*-adenosylmethionine (SAM), resulted in the production of molecules able to activate *E. coli* and *V. fischeri*-based biosensor strains for acyl-homoserine lactone (AHLs) detection. Cell extracts operated both for TX-TL reactions and for the synthesis of some AHL precursors. Moreover, it was shown that SCs containing *ad hoc* designed genetic circuits could express QS signal molecule receptors able to trigger the expression of reporter and QS signal synthase genes (e.g., *gfp* and *luxI*), upon perception of QS signal molecules produced by bacteria. The extent to which SCs could "imitate" natural cells in term of their response to the investigated QS signal molecule was estimated by a sort of cellular Turing test (Cronin et al., 2006).

The signaling between liposome-based SCs and proteinosomes (cell-like particles made of proteins) mediated by glucose, has been recently reported by a joint work of the groups of Sheref Mansy and Stephen Mann (Tang et al., 2018). In this study, the unidirectional signaling pathway was based on: (i) liposome transmitters, containing the PURE system, a DNA plasmid carrying a chemically inducible repression switch (EsaR), a gene coding for α-haemolysin, and glucose; (ii) proteinosome receivers, consisting of a cross-linked enzymatically active glucose oxidase (Gox)-poly(*N*-isopropylacrylamide) (PNIPAAm) membrane and encapsulated horseradish peroxidase (HRP). The addition of the permeable AHL molecule *N*-(3-oxohexanoyl)-L-homoserine lactone (3OC6-HSL) triggered intravesicular α-hemolysin expression and consequent membrane pore formation in liposome-based SCs, which allowed the release of glucose contained in the aqueous lumen. Glucose oxidation on the proteinosome membrane produced hydrogen peroxide, which in

turn converted a molecule into a fluorescent output by reacting with the HRP encapsulated in proteinosome. This study provides an example of molecular communication between two different types of artificial cell-like systems.

A recent report comes from our laboratory, and it deals with unidirectional SC–*P. aeruginosa* communication, based on the QS AHL signal molecule C4-HSL (Rampioni et al., 2018). In particular, SCs were prepared by encapsulating the PURE system inside GVs prepared by the droplet transfer method (Pautot et al., 2003; Fujii et al., 2014), together with butyryl coenzyme A and SAM as precursors, and a plasmid encoding for RhlI, the synthase for C4-HSL production. SCs produced C4-HSL (a natural QS signal molecule), which was perceived by *P. aeruginosa* both in liquid medium and in gel. In particular, *P. aeruginosa* modified its gene expression pattern in response to the C4-HSL-produced by SCs, demonstrating that reprogramming of gene expression in the bacterial cell is similar when interacting with other bacteria or with SCs. The entire TX-TL mechanism was assessed by *rhlI* mRNA and RhlI protein quantification, as well as by chemical identification of the C4-HSL signal produced by the SCs. The experimental results interestingly match with previously published numerical modeling (Rampioni et al., 2014), confirming the predictive power of *in silico* simulations in SCs research.

Finally, the Tan group reported an interesting study where SCs and bacteria engaged unidirectional communication in various ways (SCs to SCs, bacteria to SCs, and SCs to bacteria) (Ding et al., 2018). In this case, the QS signal molecule was produced *via* the EsaI synthase, and was perceived by the cognate EsaR receptor. Gene expression in SCs was triggered when binding of the QS signal molecule to EsaR led to derepression of an EsaR-controlled promoter region. Quite interestingly, the authors designed SCs that produce an antimicrobial peptide (Bac2A) in response to QS signal molecules sent by bacteria—a proof of principle of the use of signal processing and actuation dynamics for the generation of SCs interfacing with natural cells. Moreover, SCs embedded in biofilms were also reported.

DIRECTIONS AND CHALLENGES FOR FUTURE WORK

The works compared in **Table 1** represent proof-of-concept pioneer works that will likely stimulate further research to expand SC capabilities related to molecular communications. In this context, several challenges and open questions can be envisaged. Some refers to mechanistic, biochemical and biological aspects, others to the capability of engineering molecular communications.

With respect to the mechanisms of molecular communication, "sender" and "receiver" SCs mainly relied on transmembrane diffusion of signal molecules. This simple approach has been effective because some QS signal molecules, such as short-tail AHLs, can cross the lipid bilayer (Pearson et al., 1999). The generation of α-haemolysin pores is a drastic (yet effective) solution that has been used to bypass the low permeability of SC membranes when non free-diffusible signal molecules have

been used (e.g., IPTG or glucose), but this causes the release of all the low-MW compounds contained inside SCs (the cut-off molecular weight value for the α-haemolysin pore is 3 kDa; Song et al., 1996). An alternative could be the use of DNA nanopores, whose properties are tunable by design (Krishnan et al., 2016). The future employment of more sophisticated import/export mechanisms based on membrane proteins will allow expanding the chemical repertoire of signal molecules secreted or perceived by SCs (e.g., peptides), thus increasing the communication capability and specificity. In this respect, ongoing progress on the functionalization of SC envelopes with integral membrane proteins is promising (see section Basic Principles on Liposome-Based SCs Operating via Gene Expression).

Looking at the biological partners of SCs for molecular communications, early studies focused on bacteria, since they are prone to genetic engineering and their intercellular communication systems have been thoroughly studied at the molecular level, especially in the case of QS systems. From a practical viewpoint, SC/bacteria communication is a technological platform for the long-term goal of interfering with bacterial populations and for therapeutic strategies that could be devised against infections. Indeed, the ability of SCs to drive gene expression in response to external cues envisages the generation of injectable SCs endowed with the ability to produce or release an antimicrobial compound only in response to a signal molecule produced by a bacterial pathogen. The study reported in **Table 1** by the Tan group (Ding et al., 2018) has provided a proof-of-principle that SCs can be generated which are able to kill bacteria by a mechanism triggered by the bacteria themselves.

Proving that SCs can communicate with eukaryotic cells is one of the next milestones, especially when nanomedicine applications are devised. This complex task could require the generation of SCs with internal operations that rely on eukaryotic signal synthesis or more complex signal reception machineries. The relevance of these approaches is that SCs could be employed as intelligent drug-delivery systems that perform a therapeutic action by extracting information from their microenvironment. As mentioned, the generation of SCs constitutively producing a tumor-killing protein (the *Pseudomonas* exotoxin A) has been recently described (Krinsky et al., 2018). Another task would involve enzyme replacement therapy (Itel et al., 2017). For example, SCs that consume excess phenylalanine could play a therapeutic role in phenylketonuria (Leduc et al., 2007). Notably, Thomas M. S. Chang proposed in a pre-liposome age the therapeutic use of enzyme-containing semi-permeable collodion capsules circulating in the bloodstream (Chang, 1964, 1972). The generation of SC interfacing *via* molecular communication with neural cells can also be imagined. The resulting hybrid bio/synthetic cell networks could also be exploited for innovative investigations of neural functions (Pinato et al., 2011).

Considering the engineering plan of networking SCs (or SCs and biological cells), the rigorous design of molecular communication channels requires a proper modeling of the physical and information levels. At the physical level stochastic diffusion plays a central role. This peculiar aspect is the ultimate limit of molecular communication (when compared to traditional electro-magnetic systems) because it is essentially a

random process. Intercellular molecular communications rely on diffusion of chemical signals under a concentration gradient. They are, therefore, slow stochastic processes; their success depends on a number of factors, like the sender/receiver ratio, their spatial arrangement, the viscosity of the medium, and the temperature. Numerical models can be useful to understand the limiting factors and the constraints operating at this (inescapable) physical level (Nakano et al., 2011, 2013). The stochastic dimension of molecular communications affects its reliability. Facing with it represents an engineering challenge. The second aspect refers to the amount of information transmitted in the molecular communication "channel," and this is a theoretical issue. To apply classical information and communication theory to such a novel scenario, "information" should be defined with respect to the type of signal molecules, number of sent/received molecules, and time-dependent concentration profile (switch-like, pulse-like, etc.). Control theory for bottom-up synthetic biology should be delineated (Del Vecchio et al., 2016). Its peculiarity stems from molecular discreteness, random timing of sending/receiving, nature of "noise," etc.

In conclusion, SCs could significantly contribute to the origin of a very novel research field based on communication with biological cells. Thanks to their modular constructive principle, their biocompatibility and programmability, SCs of the type discussed in this review have the unique ability to act as passive carriers of hydrophilic and hydrophobic drugs, and to actively drive gene expression in response to chemical stimuli from other cells and from the environment.

At present, main challenges in this field rely on our capacity of (i) designing and build multi-functional SCs based on a proper genetic circuit and auxiliary molecular parts/devices, (ii) building homogeneous populations of SCs that are stable in biological fluids, (iii) and being able to control SC behavior even in a complex and fluctuating environment, such as a human host. All these challenges will probably be solved in the near future thanks to constant improvements in SC technology (in a broad sense, i.e., not necessarily restricted to liposomes). Along this path, there will be room for developing various systems in which *in vitro* usage will generate opportunities for understanding principles of biological systems and constructing short-term devices (e.g., biosensors).

AUTHOR CONTRIBUTIONS

PS conceived the research, all authors wrote the paper.

FUNDING

REFERENCES

Adamala, K. P., Martin-Alarcon, D. A., Guthrie-Honea, K. R., and Boyden, E. S. (2017). Engineering genetic circuit interactions within and between synthetic minimal cells. *Nat. Chem.* 9, 431–439. doi: 10.1038/nchem.2644

Alberts, B., Johnson, A. D., Lewis, J., Morgan, D., Raff, M., Roberts, K., et al. (2014). *Molecular Biology of the Cell. 6th Edn*. New York, NY: W. W. Norton & Company.

Altamura, E., Milano, F., Tangorra, R. R., Trotta, M., Omar, O. H., Stano, P., et al. (2017). Highly oriented photosynthetic reaction centers generate a proton gradient in synthetic protocells. *Proc. Natl. Acad. Sci. U.S.A.* 114, 3837–3842. doi: 10.1073/pnas.1617593114

Amos, M., Dittrich, P., McCaskill, J., and Rasmussen, S. (2011). Biological and chemical information technologies. *Procedia Computer Sci.* 7, 56–60. doi: 10.1016/j.procs.2011.12.019

Booth, M. J., Schild, V. R., Graham, A. D., Olof, S. N., and Bayley, H. (2016). Light-activated communication in synthetic tissues. *Sci. Adv.* 2:e1600056. doi: 10.1126/sciadv.1600056

Buddingh, B. C., and van Hest, J. C. M. (2017). Artificial cells: synthetic compartments with life-like functionality and adaptivity. *Acc. Chem. Res.* 50, 769–777. doi: 10.1021/acs.accounts.6b00512

Cabré, E. J., Sánchez-Gorostiaga, A., Carrara, P., Ropero, N., Casanova, M., Palacios, P., et al. (2013). Bacterial division proteins FtsZ and ZipA induce vesicle shrinkage and cell membrane invagination. *J. Biol. Chem.* 288, 26625–26634. doi: 10.1074/jbc.M113.491688

Carrara, P., Stano, P., and Luisi, P. L. (2012). Giant vesicles "colonies": a model for primitive cell communities. *ChemBioChem* 13, 1497–1502. doi: 10.1002/cbic.201200133

Chang, T. M. S. (1964). Semipermeable microcapsules. *Science* 146, 524–525. doi: 10.1126/science.146.3643.524

Chang, T. M. S. (1972). *Artificial Cells*. Springfield, IL: Charles C. Thomas.

Cronin, L., Krasnogor, N., Davis, B. G., Alexander, C., Robertson, N., Steinke, J. H. G., et al. (2006). The imitation game—a computational chemical approach to recognizing life. *Nat. Biotechnol.* 24, 1203–1206. doi: 10.1038/nbt1006-1203

Damiano, L., and Stano, P. (2018a). Synthetic biology and artificial intelligence: grounding a cross-disciplinary approach to the synthetic exploration of (embodied) cognition. *Complex Syst.* 27, 199–228. doi: 10.25088/complexsystems.27.3.199

Damiano, L., and Stano, P. (2018b). Understanding embodied cognition by building models of minimal life. Preparatory steps and a preliminary autopoietic framework. *Commun. Computer Info. Sci.* 830, 73–87. doi: 10.1007/978-3-319-78658-2_6

Del Vecchio, D., Dy, A. J., and Qian, Y. (2016). Control theory meets synthetic biology. *J. R. Soc. Interface* 13:20160380. doi: 10.1098/rsif.2016.0380

Deng, N.-N., Yelleswarapu, M., Zheng, L., and Huck, W. T. S. (2017). Microfluidic assembly of monodisperse vesosomes as artificial cell models. *J. Am. Chem. Soc.* 139, 587–590. doi: 10.1021/jacs.6b10977

Ding, Y., Contreras-Llano, L. E., Morris, E., Mao, M., and Tan, C. (2018). Minimizing context dependency of gene networks using artificial cells. *ACS Appl. Mater. Interfaces* 10, 30137–30146. doi: 10.1021/acsami.8b10029

Dora Tang, T.-Y., van Swaay, D., deMello, A., Ross Anderson, J. L., and Mann, S. (2015). *In vitro* gene expression within membrane-free coacervate protocells. *Chem. Commun.* 51, 11429–11432. doi: 10.1039/c5cc04220h

Exterkate, M., Caforio, A., Stuart, M. C. A., and Driessen, A. J. M. (2018). Growing membranes *in vitro* by continuous phospholipid biosynthesis from free fatty acids. *ACS Synth. Biol.* 7, 153–165. doi: 10.1021/acssynbio.7b00265

Feng, X., Jia, Y., Cai, P., Fei, J., and Li, J. (2016). Coassembly of photosystem II and ATPase as artificial chloroplast for light-driven ATP synthesis. *ACS Nano* 10, 556–561. doi: 10.1021/acsnano.5b05579

Fujii, S., Matsuura, T., Sunami, T., Nishikawa, T., Kazuta, Y., and Yomo, T. (2014). Liposome display for *in vitro* selection and evolution of membrane proteins. *Nat. Protoc.* 9, 1578–1591. doi: 10.1038/nprot.2014.107

Furusato, T., Horie, F., Matsubayashi, H. T., Amikura, K., Kuruma, Y., and Ueda, T. (2018). De novo synthesis of basal bacterial cell division proteins FtsZ, FtsA, and ZipA inside giant vesicles. *ACS Synth. Biol.* 7, 953–961. doi: 10.1021/acssynbio.7b00350

Gánti, T. (1975). Organization of chemical reactions into dividing and metabolizing units: the chemotons. *Biosystems* 7, 15–21. doi: 10.1016/0303-2647(75)90038-6

Gardner, P. M., Winzer, K., and Davis, B. G. (2009). Sugar synthesis in a protocellular model leads to a cell signalling response in bacteria. *Nat. Chem.* 1, 377–383. doi: 10.1038/nchem.296

Göpfrich, K., Platzman, I., and Spatz, J. P. (2018). Mastering complexity: towards bottom-up construction of multifunctional eukaryotic synthetic cells. *Trends Biotechnol.* 36, 938–951. doi: 10.1016/j.tibtech.2018.03.008

Gupta, A., Terrell, J. L., Fernandes, R., Dowling, M. B., Payne, G. F., Raghavan, S. R., et al. (2013). Encapsulated fusion protein confers "sense and respond" activity to chitosan-alginate capsules to manipulate bacterial quorum sensing. *Biotechnol. Bioeng.* 110, 552–562. doi: 10.1002/bit.24711

Hadorn, M., Boenzli, E., Sørensen, K. T., De Lucrezia, D., Hanczyc, M. M., and Yomo, T. (2013). Defined DNA-mediated assemblies of gene-expressing giant unilamellar vesicles. *Langmuir* 29, 15309–15319. doi: 10.1021/la402621r

Hamada, S., Tabuchi, M., Toyota, T., Sakurai, T., Hosoi, T., Nomoto, T., et al. (2014). Giant vesicles functionally expressing membrane receptors for an insect pheromone. *Chem. Commun.* 50, 2958–2961. doi: 10.1039/c3cc48216b

Hindley, J. W., Elani, Y., McGilvery, C. M., Ali, S., Bevan, C. L., Law, R. V., et al. (2018). Light-triggered enzymatic reactions in nested vesicle reactors. *Nat. Commun.* 9:1093. doi: 10.1038/s41467-018-03491-7

Huang, X., Li, M., and Mann, S. (2014). Membrane-mediated cascade reactions by enzyme-polymer proteinosomes. *Chem. Commun.* 50, 6278–6280. doi: 10.1039/c4cc02256d

Ichihashi, N., Matsuura, T., Kita, H., Sunami, T., Suzuki, H., and Yomo, T. (2010). Constructing partial models of cells. *Cold Spring Harb. Perspect. Biol.* 2:a004945. doi: 10.1101/cshperspect.a004945

Itel, F., Schattling, P. S., Zhang, Y., and Städler, B. (2017). Enzymes as key features in therapeutic cell mimicry. *Adv. Drug Deliv. Rev.* 118, 94–108. doi: 10.1016/j.addr.2017.09.006

Jørgensen, I. L., Kemmer, G. C., and Pomorski, T. G. (2017). Membrane protein reconstitution into giant unilamellar vesicles: a review on current techniques. *Eur. Biophys. J.* 46, 103–119. doi: 10.1007/s00249-016-1155-9

Karzbrun, E., Tayar, A. M., Noireaux, V., and Bar-Ziv, R. H. (2014). Programmable on-chip DNA compartments as artificial cells. *Science* 345, 829–832. doi: 10.1126/science.1255550

Krinsky, N., Kaduri, M., Zinger, A., Shainsky-Roitman, J., Goldfeder, M., Benhar, I., et al. (2018). Synthetic cells synthesize therapeutic proteins inside tumors. *Adv. Healthc. Mater.* 7:e1701163. doi: 10.1002/adhm.201701163

Krishnan, S., Ziegler, D., Arnaut, V., Martin, T. G., Kapsner, K., Henneberg, K., et al. (2016). Molecular transport through large-diameter DNA nanopores. *Nat. Commun.* 7:12787. doi: 10.1038/ncomms12787

Kurihara, K., Tamura, M., Shohda, K.-I., Toyota, T., Suzuki, K., and Sugawara, T. (2011). Self-reproduction of supramolecular giant vesicles combined with the amplification of encapsulated DNA. *Nat. Chem.* 3, 775–781. doi: 10.1038/nchem.1127

Kuruma, Y., Stano, P., Ueda, T., and Luisi, P. L. (2009). A synthetic biology approach to the construction of membrane proteins in semi-synthetic minimal cells. *Biochim. Biophys. Acta* 1788, 567–574. doi: 10.1016/j.bbamem.2008.10.017

Leduc, P. R., Wong, M. S., Ferreira, P. M., Groff, R. E., Haslinger, K., Koonce, M. P., et al. (2007). Towards an *in vivo* biologically inspired nanofactory. *Nat. Nanotechnol.* 2, 3–7. doi: 10.1038/nnano.2006.180

Lee, K. Y., Park, S.-J., Lee, K. A., Kim, S.-H., Kim, H., Meroz, Y., et al. (2018). Photosynthetic artificial organelles sustain and control ATP-dependent reactions in a protocellular system. *Nat. Biotechnol.* 36, 530–535. doi: 10.1038/nbt.4140

Lee, S., Koo, H., Na, J. H., Lee, K. E., Jeong, S. Y., Choi, K., et al. (2014). DNA amplification in neutral liposomes for safe and efficient gene delivery. *ACS Nano* 8, 4257–4267. doi: 10.1021/nn501106a

Lentini, R., Santero, S. P., Chizzolini, F., Cecchi, D., Fontana, J., Marchioretto, M., et al. (2014). Integrating artificial with natural cells to translate

chemical messages that direct *E. coli* behaviour. *Nat. Commun.* 5:4012. doi: 10.1038/ncomms5012

Lentini, R., Yeh Martín, N., Forlin, M., Belmonte, L., Fontana, J., Cornella, M., et al. (2017). Two-way chemical communication between artificial and natural cells. *ACS Central Sci.* 3, 117–123. doi: 10.1021/acscentsci.6b00330

Lentini, R., Yeh Martín, N., and Mansy, S. S. (2016). Communicating artificial cells. *Curr. Opin. Chem. Biol.* 34, 53–61. doi: 10.1016/j.cbpa.2016.06.013

Li, J., Haas, W., Jackson, K., Kuru, E., Jewett, M. C., Fan, Z. H., et al. (2017). Cogenerating synthetic parts toward a self-replicating system. *ACS Synth. Biol.* 6, 1327–1336. doi: 10.1021/acssynbio.6b00342

Litschel, T., Ramm, B., Maas, R., Heymann, M., and Schwille, P. (2018). Beating vesicles: encapsulated protein oscillations cause dynamic membrane deformations. *Angew. Chem. Int. Ed.* 57, 16286–16290. doi: 10.1002/anie.201808750

Luisi, P. L. (2002). Toward the engineering of minimal living cells. *Anat. Rec.* 268, 208–214. doi: 10.1002/ar.10155

Luisi, P. L. (2003). Autopoiesis: a review and a reappraisal. *Naturwissenschaften* 90, 49–59. doi: 10.1007/s00114-002-0389-9

Luisi, P. L., Ferri, F., and Stano, P. (2006). Approaches to semi-synthetic minimal cells: a review. *Naturwissenschaften* 93, 1–13. doi: 10.1007/s00114-005-0056-z

Luisi, P. L., and Varela, F. J. (1989). Self-replicating micelles—A chemical version of a minimal autopoietic system. *Origins Life Evol. Biosphere* 19, 633–643. doi: 10.1007/BF01808123

Maeda, Y. T., Nakadai, T., Shin, J., Uryu, K., Noireaux, V., and Libchaber, A. (2012). Assembly of MreB filaments on liposome membranes: a synthetic biology approach. *ACS Synth. Biol.* 1, 53–59. doi: 10.1021/sb200003v

Mansy, S. S., and Szostak, J. W. (2009). Reconstructing the emergence of cellular life through the synthesis of model protocells. *Cold Spring Harb. Symp. Quan. Biol.* 74, 47–54. doi: 10.1101/sqb.2009.74.014

Marguet, M., Bonduelle, C., and Lecommandoux, S. (2013). Multicompartmentalized polymeric systems: towards biomimetic cellular structure and function. *Chem. Soc. Rev.* 42, 512–529. doi: 10.1039/c2cs35312a

Martino, C., Kim, S.-H., Horsfall, L., Abbaspourrad, A., Rosser, S. J., Cooper, J., et al. (2012). Protein expression, aggregation, and triggered release from polymersomes as artificial cell-like structures. *Angew. Chem. Int. Edit.* 51, 6416–6420. doi: 10.1002/anie.201201443

Mataric, M. J. (2007). *The Robotics Primer*. Cambridge, MA: MIT Press. Available online at: https://mitpress.mit.edu/books/robotics-primer

Moritani, Y., Nomura, S. M., Morita, I., and Akiyoshi, K. (2010). Direct integration of cell-free-synthesized connexin-43 into liposomes and hemichannel formation. *FEBS J.* 277, 3343–3352. doi: 10.1111/j.1742-4658.2010.07736.x

Morowitz, H. J., Heinz, B., and Deamer, D. W. (1988). The chemical logic of a minimum protocell. *Orig. Life Evol. Biosph.* 18, 281–287.

Nakano, T., Eckford, A. W., and Haraguchi, T. (2013). *Molecular Communications*. Cambridge: Cambridge University Press.

Nakano, T., Moore, M., Enomoto, A., and Suda, T. (2011). "Molecular communication technology as a biological ICT," in *Biological Functions for Information and Communication Technologies Studies in Computational Intelligence*, ed H. Sawai (Berlin; Heidelberg: Springer), 49–86.

Niederholtmeyer, H., Chaggan, C., and Devaraj, N. K. (2018). Communication and quorum sensing in non-living mimics of eukaryotic cells. *Nat. Commun.* 9, 5027. doi: 10.1038/s41467-018-07473-7

Noireaux, V., Bar-Ziv, R., and Libchaber, A. (2003). Principles of cell-free genetic circuit assembly. *Proc. Natl. Acad. Sci. U.S.A.* 100, 12672–12677. doi: 10.1073/pnas.2135496100

Noireaux, V., and Libchaber, A. (2004). A vesicle bioreactor as a step toward an artificial cell assembly. *Proc. Natl. Acad. Sci. U.S.A.* 101, 17669–17674. doi: 10.1073/pnas.0408236101

Nourian, Z., and Danelon, C. (2013). Linking genotype and phenotype in protein synthesizing liposomes with external supply of resources. *ACS Synth. Biol.* 2, 186–193. doi: 10.1021/sb300125z

Oberholzer, T., Albrizio, M., and Luisi, P. (1995a). Polymerase chain-reaction in liposomes. *Chem. Biol.* 2, 677–682. doi: 10.1016/1074-5521(95)90031-4

Oberholzer, T., Nierhaus, K. H., and Luisi, P. L. (1999). Protein expression in liposomes. *Biochem. Biophys. Res. Commun.* 261, 238–241. doi: 10.1006/bbrc.1999.0404

Oberholzer, T., Wick, R., Luisi, P. L., and Biebricher, C. K. (1995b). Enzymatic RNA replication in self-reproducing vesicles: an approach to a minimal cell. *Biochem. Biophys. Res. Commun.* 207, 250–257. doi: 10.1006/bbrc.1995.1180

Pautot, S., Frisken, B. J., and Weitz, D. A. (2003). Production of unilamellar vesicles using an inverted emulsion. *Langmuir* 19, 2870–2879. doi: 10.1021/la026100v

Pearson, J. P., Van Delden, C., and Iglewski, B. H. (1999). Active efflux and diffusion are involved in transport of *Pseudomonas aeruginosa* cell-to-cell signals. *J. Bacteriol.* 181, 1203–1210.

Pinato, G., Raffaelli, T., D'Este, E., Tavano, F., and Cojoc, D. (2011). Optical delivery of liposome encapsulated chemical stimuli to neuronal cells. *J. Biomed. Opt.* 16:095001. doi: 10.1117/1.3616133

Pohorille, A., and Deamer, D. (2002). Artificial cells: prospects for biotechnology. *Trends Biotechnol.* 20, 123–128. doi: 10.1016/S0167-7799(02)01909-1

Rampioni, G., D'Angelo, F., Messina, M., Zennaro, A., Kuruma, Y., Tofani, D., et al. (2018). Synthetic cells produce a quorum sensing chemical signal perceived by Pseudomonas aeruginosa. *Chem. Commun.* 54, 2090–2093. doi: 10.1039/C7CC09678J

Rampioni, G., Mavelli, F., Damiano, L., D'Angelo, F., Messina, M., Leoni, L., et al. (2014). A synthetic biology approach to bio-chem-ICT: first moves towards chemical communication between synthetic and natural cells. *Nat Comput.* 13, 333–349. doi: 10.1007/s11047-014-9425-x

Ramundo-Orlando, A., Serafino, A., Schiavo, R., Liberti, M., and d'Inzeo, G. (2005). Permeability changes of connexin32 hemi channels reconstituted in liposomes induced by extremely low frequency, low amplitude magnetic fields. *Biochim. Biophys. Acta* 1668, 33–40. doi: 10.1016/j.bbamem.2004.11.003

Rideau, E., Dimova, R., Schwille, P., Wurm, F. R., and Landfester, K. (2018). Liposomes and polymersomes: a comparative review towards cell mimicking. *Chem. Soc. Rev.* 47, 8572–8610. doi: 10.1039/c8cs00162f

Sakatani, Y., Yomo, T., and Ichihashi, N. (2018). Self-replication of circular DNA by a self-encoded DNA polymerase through rolling-circle replication and recombination. *Sci. Rep.* 8:13089. doi: 10.1038/s41598-018-31585-1

Salehi-Reyhani, A., Ces, O., and Elani, Y. (2017). Artificial cell mimics as simplified models for the study of cell biology. *Exp. Biol. Med.* 242, 1309–1317. doi: 10.1177/1535370217711441

Schmidli, P. K., Schurtenberger, P., and Luisi, P. L. (1991). Liposome-mediated enzymatic synthesis of phosphatidylcholine as an approach to self-replicating liposomes. *J. Am. Chem. Soc.* 113, 8127–8130. doi: 10.1021/ja00021a043

Schwarz-Schilling, M., Aufinger, L., Mückl, A., and Simmel, F. C. (2016). Chemical communication between bacteria and cell-free gene expression systems within linear chains of emulsion droplets. *Integr. Biol.* 8, 564–570. doi: 10.1039/C5IB00301F

Schwille, P., Spatz, J., Landfester, K., Bodenschatz, E., Herminghaus, S., Sourjik, V., et al. (2018). MaxSynBio: avenues towards creating cells from the bottom up. *Angew. Chem. Int. Ed.* 57, 3382–13392. doi: 10.1002/anie.201802288

Scott, A., Noga, M. J., de Graaf, P., Westerlaken, I., Yildirim, E., and Danelon, C. (2016). Cell-Free Phospholipid Biosynthesis by Gene-Encoded Enzymes Reconstituted in Liposomes. *PLoS ONE* 11:e0163058. doi: 10.1371/journal.pone.0163058

Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., et al. (2001). Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* 19, 751–755. doi: 10.1038/90802

Shin, J., and Noireaux, V. (2012). An *E. coli* cell-free expression toolbox: application to synthetic gene circuits and artificial cells. *ACS Synth. Biol.* 1, 29–41. doi: 10.1021/sb200016s

Shohda, K., Tamura, M., Kageyama, Y., Suzuki, K., Suyama, A., and Sugawara, T. (2011). Compartment size dependence of performance of polymerase chain reaction inside giant vesicles. *Soft Matter* 7, 3750–3753. doi: 10.1039/C0SM01463J

Siegal-Gaskins, D., Tuza, Z. A., Kim, J., Noireaux, V., and Murray, R. M. (2014). Gene circuit performance characterization and resource usage in a cell-free "breadboard." *ACS Synth. Biol.* 3, 416–425. doi: 10.1021/sb400203p

Soga, H., Fujii, S., Yomo, T., Kato, Y., Watanabe, H., and Matsuura, T. (2014). *In vitro* membrane protein synthesis inside cell-sized vesicles reveals the dependence of membrane protein integration on vesicle volume. *ACS Synth. Biol.* 3, 372–379. doi: 10.1021/sb400094c

Song, L., Hobaugh, M. R., Shustak, C., Cheley, S., Bayley, H., and Gouaux, J. E. (1996). Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* 274, 1859–1866.

Spencer, A. C., Torre, P., and Mansy, S. S. (2013). The encapsulation of cell-free transcription and translation machinery in vesicles for the construction of cellular mimics. *J. Vis. Exp.* 80:e51304. doi: 10.3791/51304

Stano, P., Carrara, P., Kuruma, Y., de Souza, T. P., and Luisi, P. L. (2011). Compartmentalized reactions as a case of soft-matter biotechnology: synthesis of proteins and nucleic acids inside lipid vesicles. *J. Mater. Chem.* 21, 18887–18902. doi: 10.1039/c1jm12298c

Stano, P., Rampioni, G., Carrara, P., Damiano, L., Leoni, L., and Luisi, P. L. (2012). Semi-synthetic minimal cells as a tool for biochemical ICT. *BioSystems* 109, 24–34. doi: 10.1016/j.biosystems.2012.01.002

Sun, S., Li, M., Dong, F., Wang, S., Tian, L., and Mann, S. (2016). Chemical signaling and functional activation in colloidosome-based protocells. *Small* 12, 1920–1927. doi: 10.1002/smll.201600243

Szostak, J. W., Bartel, D. P., and Luisi, P. L. (2001). Synthesizing life. *Nature* 409, 387–390. doi: 10.1038/35053176

Tang, T.-Y. D., Cecchi, D., Fracasso, G., Accardi, D., Coutable-Pennarun, A., Mansy, S. S., et al. (2018). Gene-mediated chemical communication in synthetic protocell communities. *ACS Synth. Biol.* 7, 339–346. doi: 10.1021/acssynbio.7b00306

Taylor, A. I., Pinheiro, V. B., Smola, M. J., Morgunov, A. S., Peak-Chew, S., Cozens, C., et al. (2015). Catalysts from synthetic genetic polymers. *Nature* 518, 427–430. doi: 10.1038/nature13982

Tsugane, M., and Suzuki, H. (2018). Reverse transcription polymerase chain reaction in giant unilamellar vesicles. *Sci. Rep.* 8:9214. doi: 10.1038/s41598-018-27547-2

van Nies, P., Westerlaken, I., Blanken, D., Salas, M., Mencía, M., and Danelon, C. (2018). Self-replication of DNA by its encoded proteins in liposome-based synthetic cells. *Nat. Commun.* 9:1583. doi: 10.1038/s41467-018-03926-1

Varela, F. G., Maturana, H. R., and Uribe, R. (1974). Autopoiesis: the organization of living systems, its characterization and a model. *Biosystems* 5, 187–196. doi: 10.1016/0303-2647(74)90031-8

Walde, P., Goto, A., Monnard, P., Wessicken, M., and Luisi, P. (1994). Oparins reactions revisited—enzymatic-synthesis of poly(adenylic acid). *J. Am. Chem. Soc.* 116, 7541–7547. doi: 10.1021/ja00096a010

Wang, Y.-H., Wei, K. Y., and Smolke, C. D. (2013). Synthetic biology: advancing the design of diverse genetic systems. *Annu. Rev. Chem. Biomol. Eng.* 4, 69–102. doi: 10.1146/annurev-chembioeng-061312-103351

Yanagisawa, M., Iwamoto, M., Kato, A., Yoshikawa, K., and Oiki, S. (2011). Oriented reconstitution of a membrane protein in a giant unilamellar vesicle: experimental verification with the potassium channel KcsA. *J. Am. Chem. Soc.* 133, 11774–11779. doi: 10.1021/ja2040859

York-Duran, M. J., Godoy-Gallardo, M., Labay, C., Urquhart, A. J., Andresen, T. L., and Hosta-Rigau, L. (2017). Recent advances in compartmentalized synthetic architectures as drug carriers, cell mimics and artificial organelles. *Coll. Surf. B Biointerf.* 152, 199–213. doi: 10.1016/j.colsurfb.2017.01.014

Improving Reproducibility in Synthetic Biology

Mathew M Jessop-Fabre and Nikolaus Sonnenschein*

The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark

Synthetic biology holds great promise to deliver transformative technologies to the world in the coming years. However, several challenges still remain to be addressed before it can deliver on its promises. One of the most important issues to address is the lack of reproducibility within research of the life sciences. This problem is beginning to be recognised by the community and solutions are being developed to tackle the problem. The recent emergence of automated facilities that are open for use by researchers (such as biofoundries and cloud labs) may be one of the ways that synthetic biologists can improve the quality and reproducibility of their work. In this perspective article, we outline these and some of the other technologies that are currently being developed which we believe may help to transform how synthetic biologists approach their research activities.

Keywords: synthetic biology, automation, reproducibility, cloud lab, biofoundry

INTRODUCTION

Science is reliant on the development of reproducible data. Despite the great importance of reproducibility, there have so far been few studies into the reproducibility of life science publications. One exception is within cancer biology, where it was reported that only 11% of landmark cancer studies could be reproduced (Begley and Ellis, 2012). In a survey of research scientists, 77% of biologists stated that they had tried and failed to reproduce someone else's result in the lab (Baker, 2016). And when asked "what percentage of published results in your field are reproducible?", biologists estimated that only 59% of results are reproducible.

Different measures have been taken to maintain high levels of rigor within the sciences. The process of peer-review is one of these measures and has so far served as a successful strategy. However, with the increasing volume, complexity, and detail of data that is being published, peer-review is arguably unable to ensure that published research is reproducible and other measures are needed prior to this step.

Synthetic biology aims to introduce engineering principles into the life sciences in the hope that it can greatly improve the reliability of the "Design-Build-Test-Learn cycle" (**Figure 1**). Any engineering discipline must be predictable and reproducible. In this perspective article, we discuss some of the ways the synthetic biology community is working towards achieving these goals.

IMPROVEMENT OF EXISTING LABORATORY PRACTICES

One way to increase reproducibility is to increase the quality of protocol reporting. Experimental protocols are often specific to a laboratory or even to a single researcher, and are frequently taught to new lab members through a combination of practical guidance and a written protocol. This form of knowledge transfer is valuable in passing on tacit information (knowledge that is difficult to pass on in written form), but can lead to variation in the way experiments are performed. The passing on of

written protocols without practical guidance can lead to even more problems, as written protocols are open to interpretation as they often contain ambiguities or rely on tacit knowledge (Miles and Lee, 2018). Online protocol editors, such as protocols.io, aim to improve the quality of protocols by creating a platform for scientists to easily share and edit protocol documents (Teytelman et al., 2016). Researchers are encouraged to upload detailed step-by-step instructions that can be followed, verified, and improved upon by researchers from other labs. In addition, videos can also be uploaded to capture some of the tacit information inherent in performing experiments. These protocols can then be linked to the methods sections of papers, and perhaps can help to improve the standardisation of common techniques and thereby improve experimental reproducibility.

Protocol editors combined with protocol management systems can provide a framework for labs to organise their research activities. The Aquarium software developed at the Klavins Lab is an example of such a system[1]. In this system a "designer" can accurately specify a protocol, which is then sent to the lab as a "job" to be performed by students, technicians, or robots. The system can be integrated with a laboratory information management system (LIMS) such as Benchling, enabling the use of materials to be tracked and giving trace-back of the items used for each experiment.

AUTOMATION

For many years, laboratory automation has promised to revolutionise biology. However, the incorporation of traditional lab automation technologies (such as liquid handling robots) by academic labs has been slight. However, many of the techniques used within synthetic biology are performed on a routine basis, and as such may be desirable targets for automation. The successful implementation of automation technology can help to improve both the throughput and reproducibility of experiments, although the high cost and lack of flexibility of traditional lab automation has so far hindered their wide-spread adoption in academia. Recently though, there have been advances in automation technology that, when combined with new protocol sharing methods and protocol management systems, may allow for researchers to gain the benefits of automation.

The pharmaceutical industry has found particular favour with automation, as assay protocols are relatively simple to automate (Bogue, 2012). The high cash flow within large companies gives them the ability to invest in high quality and highly specialized equipment. This approach is unlikely to function well in an academic environment as the high costs and low flexibility of these operations are often prohibitive. Flexible and low cost solutions have generally not been offered by established vendors of liquid handling robots. Liquid handlers from vendors such as Tecan and Hamilton are state of the art, but expensive, large, and technically complex. In addition, operation of these machines usually requires skilled technicians, and can still take several months before they are operational. Interestingly, a recent Kickstarter campaign from Opentrons has aimed to address both

the demand for a low cost and flexible liquid handler. The resulting OT-One robot has proved popular in both academia and industry, enabling Opentrons to establish itself in the field of lab automation. The more recent and advanced OT-2 liquid handler can be purchased from 4,000 USD, making it affordable to a wide range of academic labs. They have addressed the issue of flexibility through the adoption of open-source hardware and software. The relative ease of set-up and use of the Opentrons robots (programmable through a Python API, or soon through a graphical user interface) enables it to be used in a flexible manner by non-specialized workers. This means that many labs now have the opportunity to automate parts of their workflows and share their efforts throughout the community. The simple sharing of protocols will hopefully mean that other labs can easily test and verify the work of others.

Another issue that has been a problem for increased implementation of automation systems until recently, is that it has been difficult to link up devices from different manufacturers. Synthace is a start-up that aims to eliminate this issue by developing software that can control multiple robots from different vendors. An entire lab may be able to be controlled through a unified piece of software that allows for much simpler coordinated control of complex automation equipment.

Contrary to liquid handling robots, microfluidic technologies allow for the control of liquids on a microscopic scale and can also be used to automate and scale down many common laboratory procedures. Commercial microfluidic devices are becoming increasingly commonplace in the lab, with devices popular in the handling of DNA and RNA, such as fragment analysers used in next-generation sequencing workflows. The field of microfluidics is large and well-established, with over 3,000 articles published in 2017 alone (PubMed search for keyword: "microfluidic"). Microfluidics may provide a cheap and powerful alternative to traditional laboratory automation. Devices have been built that are capable of performing strain transformation and culturing, while other developments in the field have focused on DNA assembly (Ben Yehezkel et al., 2016; Gach et al., 2016; Linshiz et al., 2016). Even though microfluidic devices have been shown capable of automating protocols within synthetic biology, their adoption has been slow, and traditional liquid handlers remain the dominant method of automation, as the majority of laboratory automation has been designed to handle the array formats (such as 96 well plates) that are standard in the lab. Marrying microfluidic technologies to array-based laboratory automation can prove difficult, and is perhaps one reason for the limited integration of microfluidics into automated workflows.

Tedious and repetitive tasks are highly error-prone for humans to conduct, and automation may be seen as a way to reduce such error (Yeow et al., 2014). However, careful consideration for the requirements of any automation system must be carried out before the purchase of equipment. Counterintuitively, robots may be slower and less accurate than humans under certain conditions. Zielinski et al. (2014) reported that their newly developed pipetting protocol had up to a 3x larger coefficient of variation and took twice as long to perform when carried out by a robotic liquid handler rather than by

---

[1] https://www.aquarium.bio
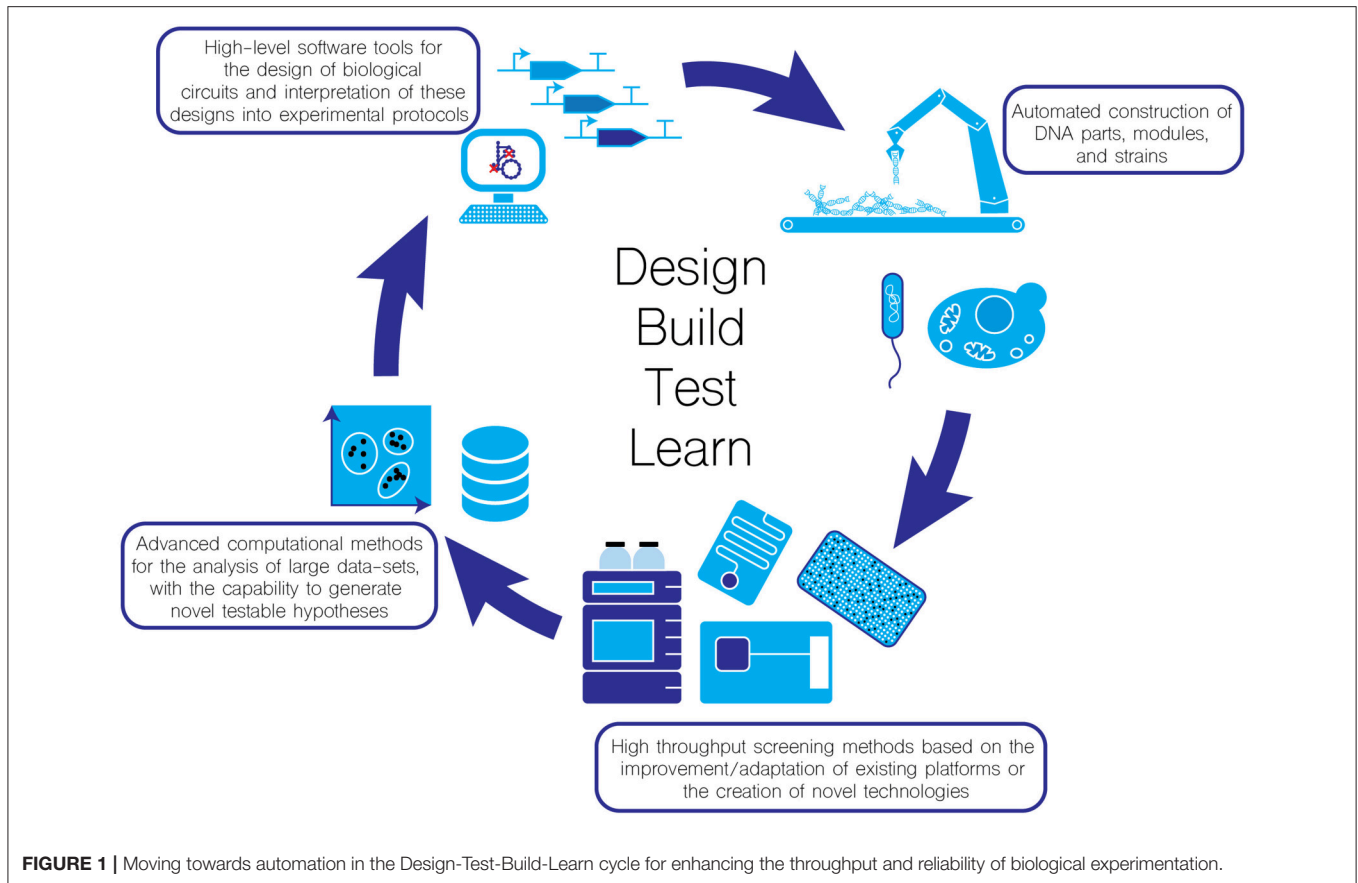
FIGURE 1 | Moving towards automation in the Design-Test-Build-Learn cycle for enhancing the throughput and reliability of biological experimentation.

a human. Traditional liquid handlers are also limited in ways such as requiring "dead volumes" of reagents inside wells, which can increase the cost of experiments. To maintain high levels of accuracy, all automation equipment must undergo strict and frequent quality assessments (Chai et al., 2013). There is an additional investment of time required for the set-up and optimisation of new protocols, which when combined with the aforementioned issues can dissuade academic labs from purchasing such systems. However, as we discuss below, there have recently been developments that aim to give researchers easier access to automation.

BIOFOUNDRIES

The issues mentioned in the previous section can prevent academic labs from investing in extensive automation. However, in the past few years there has been a movement to create fully or semi-automated labs termed "biofoundries." Biofoundries typically consist of a core laboratory that has been extensively automated to carry out a range of functions. These integrated systems are unique in their flexibility compared to standard automated pipelines. A user can submit a job (e.g., create a combinatorial library) to the biofoundry, where it will be scheduled before being carried out by the automated lab. Once a job has been successfully run, the user is sent (if applicable) the

biological output of their job along with any experimental or log data that the user requires.

Much of the work conducted within synthetic biology revolves around DNA manipulation, and as such so does the work of many of the biofoundries. One example is the Edinburgh Genome Foundry; whose focus is on manufacturing long lengths of genetic material *via* several DNA assembly methods. Their approach differs from that of a standard laboratory by only requiring a minimum of human intervention. A range of standard equipment, such as thermocyclers and incubators, are linked together by a network of robotic arms that can carry plates and samples between instruments. They are able to construct and transform complex sections of DNA into host organisms and then conduct a range of experiments on the engineered strains. The minimisation of human intervention, and shared use of reagent resources helps to bring down costs, hopefully to a point where it becomes economically viable for many researchers to outsource parts of their construction workflows to facilities like this.

The UK and US are currently leaders in the development of biofoundries, with several found throughout both countries (Chao et al., 2017). Many of these are open to accepting jobs from outside their host university institutes with the goal to serve the wider scientific community (**Table 1**). A widespread use of these facilities for design construction within synthetic biology may help improve standardisation within the field, if the

TABLE 1 | Overview of Biofoundries based at public institutions around the world.

Name of biofoundry	Host organisation	Focus	Website
Agile BioFoundry	US Department of Energy, University of California, USA	Full integration of the Design-Build-Test-Learn cycle, with advanced scale-up capabilities	https://agilebiofoundry.org
BIOFAB	University of Washington, USA	Use of the Aquarium software to control the functioning of a traditional lab. Services focus on plasmid construction and yeast strain construction.	http://www.uwbiofab.org
Concordia Genome Foundry	Concordia University, Canada	Cloning and NGS sample preparation	https://www.concordia.ca/research/casb/genome-foundry.html
DAMP lab	Boston University, USA	DNA construction, NGS preparations, protein expression and purification	https://www.damplab.org
Earlham DNA Foundry	Earlham Institute, UK	DNA assembly and insertion into a range of hosts, including plant cells	http://www.earlham.ac.uk/earlham-dna-foundry
Edinburgh Genome Foundry	Edinburgh University, UK	Assembly of large DNA fragments (>5,000 nt)	https://www.genomefoundry.org/home
GeneMill	University of Liverpool, UK	DNA construction, screening, and metabolomics workflows	https://genemill.liv.ac.uk
iBioFAB	University of Illinois, USA	Full automation of the Design-Build-Test-Learn cycle	https://www.igb.illinois.edu/research-areas/biosystems-design/research
London DNA Foundry	Imperial College, UK	Design and construction of complex DNA assemblies	http://www.londondnafoundry.co.uk
NUS Synthetic Biology Foundry	National University of Singapore, Singapore	DNA assembly, and analytical capabilities for downstream performance measurements	http://syncti.org/synthetic-biology-foundry/
SYNBIOCHEM	University of Manchester, UK	Automating the Design-Build-Test-Learn cycle for production of fine chemicals	http://synbiochem.co.uk/synbiochem-pipeline/

This is not a comprehensive list, but contains those centres and facilities that have accessible webpages.

community of biofoundries agree upon the design and adoption of operational standards.

CLOUD LABORATORIES

A similar concept to biofoundries is the "cloud laboratory." Like a biofoundry, a cloud lab is usually comprised of a central facility that is heavily automated. However, the cloud lab gives a higher level of freedom to the customer by offering a real lab and its capabilities to biologists sat behind a computer in a remote location (Check Hayden, 2014). Two Californian companies have spearheaded this movement; Transcriptic, and the Emerald Cloud Lab. Both companies offer their extensive lab automation as a service, although each in a different way. Transcriptic has constructed a series of sterile work-cells. Within each work-cell is a set of basic lab equipment such as liquid handlers, thermocyclers, fridges, incubator, and a plate reader as well as a robotic arm to move materials between machines. Customers can control the actions within a work-cell *via* an application programming interface (API). A user can create an experimental protocol through the high-level Python API, which then converts the user's protocol into the JSON-formatted Autoprotocol language developed by Transcriptic to control their automation equipment (Miles and Lee, 2018). Users receive detailed results and diagnostics on their experiments, with troubleshooting on failed experiments made easier by the available metrics for all instruments on the executed run. The Emerald Cloud Lab has taken a different approach, where they have not broken down operations into discrete and identical work-cells, but have instead focused on offering a wider and more flexible range of services to the customer. The Emerald Cloud Lab has a stronger focus on the analytical side of biology through access to high-pressure liquid chromatography and mass spectrometry, among others. Cloud labs offer the chance for experimentalists to keep full control over their process, while also providing detailed and reproducible protocols. The hope is that academics will include these protocols within their publications, and share them online where others can use or modify them. Cloud labs may also enable early-stage biotech start-ups to test their early ideas and designs without the need for investing in setting up their own laboratory beforehand or entering into an incubator too early. However, there are still limitations with these technologies. Perhaps the most important is that it may be impossible to have any one facility that can cater to the needs of all biological researchers, since scientists make use of a wide array of instrumentation in their routine experiments. A recent review into cloud labs found that within biomedical research, 89% of published papers were found to contain one or more methods that could be carried out at a cloud lab (Groth and Cox, 2017). However, only 3% of papers had all of their methods supported by a cloud lab, suggesting that most complete workflows are as yet unable to be ported to a cloud lab. While it may not be necessary for a lab to implement complete workflows at a cloud lab, it may also prove disadvantageous to only perform a small part of their protocols (for example if it requires the frequent shipping of materials). There is therefore an inevitable judgement call that must be made to determine when it becomes beneficial to outsource a protocol.

THE ROBOT SCIENTIST

The robust development of lab automation, whether in-house or through a biofoundry or cloud lab, may lead to biologists automating parts of their hypothesis generation and experimental design cycle too. While both biofoundries and cloud labs implement instructions from a human operator, the robot scientist aims to go one step further by removing the scientist from the hypothesis generation and experimental design steps. An early pioneering effort in this field was the creation of a "Robot Scientist" named Adam (King et al., 2004, 2009). The physical design of Adam resembles that of a biofoundry, with several instruments handling the execution of a set of experimental techniques. Materials are moved between these instruments by three robotic arms, and the entire set-up is enclosed in a sterile plastic enclosure. The main difference to a biofoundry is that Adam has generated its own scientific hypotheses, which it then tests experimentally, resulting in the generation of new scientific knowledge in yeast genetics. Adam is not alone. The more recent Eve Robot Scientist has also generated novel scientific insights, this time revealing new drug candidates for the treatment of *Plasmodium vivax* (Williams et al., 2015). This marriage of automation and basic hypothesis generation and testing may prove to be valuable within synthetic biology, as a researcher can be left to generate higher level hypotheses, which are then interpreted and executed by an automated facility.

STANDARDISATION

One of the key enablers of engineering disciplines is the adoption of a comprehensive set of standards (O'Connell, 1993). Such levels of standardisation are much more difficult to achieve in biological rather than electrical systems as there is a higher degree of uncertainty in the systems being engineered (Endy, 2005; Arkin, 2008; Canton et al., 2008). As an example, if one takes a regulatory element from one organism and places it in another organism then it is likely that the functionality of that element will differ. Even between different strains of the same species there is a large variation in how genetic elements are influenced by the host (Balagaddé et al., 2005; Cardinale et al., 2013). Synthetic biology requires standards that can cope with the issue of context dependency. Attempts to introduce standards, such as the biobrick standard, have been so far unable to fully deal with this issue (Decoene et al., 2018). However, promising work is being done to improve the orthogonality of biological parts (Stanton et al., 2013; Qian et al., 2017). Advances are also being made in to how biological parts are characterised and how orthogonality can be measured (Lucks et al., 2008; Mutalik et al., 2013; Ceroni et al., 2015). The creation of such standards may then help biofoundries and cloud labs to run standard building and characterisation protocols, that when combined with a developed set of operational standards can help to further increase the reliability and reproducibility of synthetic biology workflows. The standardisation of data practices is also just as vital, as the greatest of experimental results are worthless without the ability to learn from them, but such a discussion is out of the scope of this article [for a deeper discussion on data standardisation we refer the reader to Decoene et al. (2018)].

CONCLUSIONS

One important question to address in synthetic biology is, how do we increase the predictability of our designed circuits and strains? Answering this question will have wide-reaching consequences for the field, but will require a shift in how synthetic biology is carried out in academia. Although synthetic biology has already broken the mould in many respects for how science is conducted at universities, deeper changes are needed in the funding and publishing infrastructures to guide the development of standardised practices.

In the meantime, recent developments in laboratory automation may help to improve the quality of experimental protocols, as machines must be programmed with precise and unambiguous instructions and protocols can be widely distributed and validated between labs around the world. The use of third-party facilities such as cloud labs or biofoundries can reduce some of the experimental variation between researchers while also reducing some of the burden. As we now order primers (and genes) instead of making them ourselves, perhaps we will soon too routinely order engineered strains. Robust lab automation holds the potential to bring computer-aided manufacturing (CAM) to synthetic biology by coupling computer-aided design (CAD) software for biological systems to the construction of cell factories, biological computers, and novel enzymes (Raman et al., 2009; Vasilev et al., 2011; Nielsen et al., 2016; Roehner et al., 2016; Cardoso et al., 2018).

There is an ever-growing demand for (high quality) data for machine learning applications, which requires a systematic approach to data generation, management, and sharing. Automation is capable of providing this approach as it can provide detailed logs of experimental runs and of the data acquisition from instruments. The data produced by academic labs needs to be made available to the greater scientific community in standardised formats, so that the greater community can learn from each other. Data collection and management is an area that requires a lot of attention. Although we have not had the space to offer a detailed discussion here, the synthetic biology community would greatly benefit from more focus on the standardisation of data practices. Traditional funding agencies are perhaps less likely to place emphasis on infrastructure, and so scientists are less incentivised to standardise their data practices.

AUTHOR CONTRIBUTIONS

MJ-F and NS conceived the study, and contributed jointly to the research and writing of this article.

ACKNOWLEDGMENTS

REFERENCES

Arkin, A. (2008). Setting the standard in synthetic biology. *Nat. Biotechnol.* 26:771. doi: 10.1038/nbt0708-771

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature* 533, 452–454. doi: 10.1038/533452a

Balagaddé, F. K., You, L., Hansen, C. L., Arnold, F. H., and Quake, S. R. (2005). Long-term monitoring of bacteria undergoing programmed population control in a microchemostat. *Science* 309, 137–140. doi: 10.1126/science.1109173

Begley, C. G., and Ellis, L. M. (2012). Raise standards for preclinical cancer research. *Nature* 483:531. doi: 10.1038/483531a

Ben Yehezkel, T., Rival A., Raz, O., Cohen, R., Marx, Z., Camara, M., et al. (2016). Synthesis and cell-free cloning of DNA libraries using programmable microfluidics. *Nucleic Acids Res.* 44:e35. doi: 10.1093/nar/gkv1087

Bogue, R. (2012). Robots in the laboratory: a review of applications. *Ind. Robot An Int. J.* 39, 113–119. doi: 10.1108/01439911211203382

Canton, B., Labno, A., and Endy, D. (2008). Refinement and standardization of synthetic biological parts and devices. *Nat. Biotechnol.* 26:787. doi: 10.1038/nbt1413

Cardinale, S., Joachimiak, M. P., and Arkin, A. P. (2013). Effects of genetic variation on the *E. coli* host-circuit interface. *Cell Rep.* 4, 231–237. doi: 10.1016/j.celrep.2013.06.023

Cardoso, J. G. R., Jensen, K., Lieven, C., Lærke Hansen, A. S., Galkina, S., Beber, M., et al. (2018). Cameo: a python library for computer aided metabolic engineering and optimization of cell factories. *ACS Synth. Biol.* 7, 1163–1166. doi: 10.1021/acssynbio.7b00423

Ceroni, F., Algar, R., Stan, G. B., and Ellis, T. (2015). Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nat. Methods* 12:415. doi: 10.1038/nmeth.3339

Chai, S. C., Goktug, A. N., Cui, J., Low, J., and Chen, T. (2013). "Practical considerations of liquid handling devices in drug discovery," in *Drug Discovery*, ed H. A. El-Shemy (IntechOpen), 177–200. doi: 10.5772/52546

Chao, R., Mishra, S., Si, T., and Zhao, H. (2017). Engineering biological systems using automated biofoundries. *Metab. Eng.* 42, 98–108. doi: 10.1016/j.ymben.2017.06.003

Check Hayden, E. (2014). The automated lab. *Nature* 516, 131–132. doi: 10.1038/516131a

Decoene, T., De Paepe, B., Maertens, J., Coussement, P., Peters, G., De Maeseneire, S. L., et al. (2018). Standardization in synthetic biology: an engineering discipline coming of age. *Crit. Rev. Biotechnol.* 38, 647–656. doi: 10.1080/07388551.2017.1380600

Endy, D. (2005). Foundations for engineering biology. *Nature* 438, 449–453. doi: 10.1038/nature04342

Gach, P. C., Shih, S. C., Sustarich, J., Keasling, J. D., Hillson, N. J., Adams, P. D., et al. (2016). A droplet microfluidic platform for automating genetic engineering. *ACS Synth. Biol.* 5, 426–433. doi: 10.1021/acssynbio.6b00011

Groth, P., and Cox, J. (2017). Indicators for the use of robotic labs in basic biomedical research: a literature analysis. *Peer J.* 5:e3997. doi: 10.7717/peerj.3997

King, R. D., Rowland, J., Oliver, S. G., Young, M., Aubrey, W., Byrne, E., et al. (2009). The automation of science. *Science* 324, 85–89. doi: 10.1126/science.1165620

King, R. D., Whelan, K. E., Jones, F. M., Reiser, P. G., Bryant, C. H., Muggleton, S. H., et al. (2004). Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature* 427, 247–252. doi: 10.1038/nature02236

Linshiz, G., Jensen, E., Stawski, N., Bi, C., Elsbree, N., Jiao, H., et al. (2016). End-to-end automated microfluidic platform for synthetic biology: from design to functional analysis. *J. Biol. Eng.* 10:3. doi: 10.1186/s13036-016-0024-5

Lucks, J. B., Qi, L., Whitaker, W. R., and Arkin, A. P. (2008). Toward scalable parts families for predictable design of biological circuits. *Curr. Opin. Microbiol.* 11, 567–573. doi: 10.1016/j.mib.2008.10.002

Miles, B., and Lee, P. L. (2018). Achieving reproducibility and closed-loop automation in biological experimentation with an IoT-enabled Lab of the future. *SLAS Technol. Transl. Life Sci. Innov.* 23, 432–439. doi: 10.1177/2472630318784506

Mutalik, V. K., Guimaraes, J. C., Cambray, G., Mai, Q. A., Christoffersen, M. J., Martin, L., et al. (2013). Quantitative estimation of activity and quality for collections of functional genetic elements. *Nat. Methods* 10:347. doi: 10.1038/nmeth.2403

Nielsen, A. A., Der, B. S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E. A., et al. (2016). Genetic circuit design automation. *Science* 352:aac7341. doi: 10.1126/science.aac7341

O'Connell, J. (1993). Metrology: the creation of universality by the circulation of particulars. *Soc. Stud. Sci.* 23, 129–173. doi: 10.1177/030631293023001005

Qian, Y., Huang, H. H., Jiménez, J. I., and Del Vecchio, D. (2017). Resource competition shapes the response of genetic circuits. *ACS Synth. Biol.* 6, 1263–1272. doi: 10.1021/acssynbio.6b00361

Raman, S., Vernon, R., Thompson, J., Tyka, M., Sadreyev, R., Pei, J., et al. (2009). Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins* 77(Suppl. 9), 89–99. doi: 10.1002/prot.22540

Roehner, N., Young, E. M., Voigt, C. A., Gordon, D. B., and Densmore, D. (2016). Double dutch: a tool for designing combinatorial libraries of biological systems. *ACS Synth. Biol.* 5, 507–517. doi: 10.1021/acssynbio.5b00232

Stanton, B. C., Nielsen, A. A., Tamsir, A., Clancy, K., Peterson, T., and Voigt, C. A. (2013). Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nat. Chem. Biol.* 10:99. doi: 10.1038/nchembio.1411

Teytelman, L., Stoliartchouk, A., Kindler, L., and Hurwitz, B. L. (2016). Protocols.io: virtual communities for protocol development and discussion. *PLoS Biol.* 14, e1002538. doi: 10.1371/journal.pbio.1002538

Vasilev, V., Liu, C., Haddock, T., Bhatia, S., Adler, A., Yaman, F., et al. (2011). "A software stack for specification and robotic execution of protocols for synthetic biological engineering", in *3rd International Workshop on Bio-Design Automation* (San Diego, CA).

Williams, K., Bilsland, E., Sparkes, A., Aubrey, W., Young, M., Soldatova, L. N., et al. (2015). Cheaper faster drug development validated by the repositioning of drugs against neglected tropical diseases. *J. R. Soc. Interface* 12. doi: 10.1098/rsif.2014.1289

Yeow, J., Ng, P. K., Sin Tan, K., Chin, T., and Yin Lim, W. (2014). Effects of stress, repetition, fatigue and work environment on human error in manufacturing industries. *J. Appl. Sci.* 14, 3464–3471. doi: 10.3923/jas.2014.3464.3471

Zielinski, D., Gordon, A., Zaks, B. L., and Erlich, Y. (2014). iPipet: sample handling using a tablet. *Nat. Methods* 11:784. doi: 10.1038/nmeth.3028

frontiers
in Bioengineering and Biotechnology

Check for updates

New Applications of Synthetic Biology Tools for Cyanobacterial Metabolic Engineering

María Santos-Merino[1†], Amit K. Singh[1†] and Daniel C. Ducat[1,2]*

[1] MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing, MI, United States, [2] Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI, United States

Cyanobacteria are promising microorganisms for sustainable biotechnologies, yet unlocking their potential requires radical re-engineering and application of cutting-edge synthetic biology techniques. In recent years, the available devices and strategies for modifying cyanobacteria have been increasing, including advances in the design of genetic promoters, ribosome binding sites, riboswitches, reporter proteins, modular vector systems, and markerless selection systems. Because of these new toolkits, cyanobacteria have been successfully engineered to express heterologous pathways for the production of a wide variety of valuable compounds. Cyanobacterial strains with the potential to be used in real-world applications will require the refinement of genetic circuits used to express the heterologous pathways and development of accurate models that predict how these pathways can be best integrated into the larger cellular metabolic network. Herein, we review advances that have been made to translate synthetic biology tools into cyanobacterial model organisms and summarize experimental and *in silico* strategies that have been employed to increase their bioproduction potential. Despite the advances in synthetic biology and metabolic engineering during the last years, it is clear that still further improvements are required if cyanobacteria are to be competitive with heterotrophic microorganisms for the bioproduction of added-value compounds.

Keywords: cyanobacteria, metabolic engineering, synthetic biology, genome scale models, photosynthesis

INTRODUCTION

Over 1 billion years ago, cyanobacteria began to drive the rise of oxygen in Earth's atmosphere, setting the stage for the evolution of new forms of complex, multicellular organisms (Shih, 2015). More recently, these photosynthetic prokaryotes have come under increasing scrutiny for their potential to sustain the lifestyle of complex eukaryotic lifeforms, this time as potential bioproduction hosts. Cyanobacteria can efficiently harvest CO_2 as a carbon source, powering their metabolic processes by absorption of sunlight, the most abundant form of renewable energy. As fast-growing, and relatively simple bacteria, cyanobacteria hold the promise of being an ideal chassis for ambitious metabolic engineering projects. In recent years, cyanobacterial potential is being unlocked through the development of an increasing number of molecular tools (Berla et al., 2013; Carroll et al., 2018; Sun et al., 2018). These advances have been coupled with increasing capacities to manipulate endogenous genetic sequences and transfer exogenous DNA into multiple cyanobacterial strains (Berla et al., 2013; Cassier-Chauvat et al., 2016). Moreover, due to advances in sequencing technology, the genomes of >270 cyanobacterial species have been sequenced (Fujisawa et al., 2017),

greatly facilitating the application of systems-level techniques, such as transcriptomics and proteomics. These technological improvements in the manipulation of cyanobacteria are set against the backdrop of increasingly refined tools that have been developed in the fields of metabolic engineering and synthetic biology, potentially setting the stage for cyanobacteria to significantly contribute as crop species in the Twenty-first century.

Metabolic engineering can be defined as the practice of optimizing genetic and regulatory processes in cells to increase production of certain metabolic substances (Kumar and Prasad, 2011). This methodology has been used in cyanobacteria as well as other bacteria, to expand the list of products that they are able to make and increase production efficiency (Angermayr et al., 2015; Lai and Lan, 2015; Carroll et al., 2018). Traditional approaches to improve production of a specific compound were based on random mutagenesis and selection or the targeted introduction of individual genes and mutations, which requires considerable time to design and implement (Kumar and Prasad, 2011). Recently, systems metabolic engineering has emerged as new methodology for solving these issues (Nogales et al., 2013). It is based on the use of mathematic models to simulate and predict behaviors that emerge in complex systems, and has been used extensively for the improvement of microbial production (Lee et al., 2011).

In parallel, synthetic biology principles promote a bottom up approach to design biological systems, recombining defined parts or modules to restructure existing systems or build new pathways *de novo* (Sengupta et al., 2018). It is based on the construction of intricate biological systems using standardized, well-characterized, and interchangeable biological parts, or "modules" (Cheng and Lu, 2012). Collectively, these modules form "toolboxes" of components that can be used to modify an organism, including a catalog of characterized biological parts [e.g., promoters, ribosome binding sites (RBSs), riboswitches, terminator libraries], standardized methods for the assembly and manipulation of genetic components, and predictive models designed to facilitate pathway optimization (Gordon et al., 2016). Ideally, truly modular biological parts would be characterized in a manner that would allow researchers to accurately predict how they will function in the context of an interconnected system (Pasotti et al., 2012). In practice, it remains difficult to characterize parts in a manner that is completely transferrable to other organisms (Sengupta et al., 2018), because they are typically characterized by a specific research group in particular environment (Decoene et al., 2018). Another challenge to the transfer of parts is that biological components are usually non-orthogonal and may interact with genes, proteins, and metabolites of the chassis organism (Fu, 2013), or have their function influenced by host pathway components (Wang et al., 2013). Still, largely by using specific parts and strategies developed in model heterotrophic organisms, useful synthetic biology components have been recently adapted and debugged for use in cyanobacteria.

This review focuses on recent engineering tools and strategies developed for metabolic engineering of cyanobacteria as hosts for generating added-value compounds while using solar energy and CO_2 as inputs. We discuss the current challenges and opportunities for the application of synthetic biology principles in cyanobacteria. Finally, we highlight the potential for genome-scale models as tools to assist cyanobacterial engineering.

CYANOBACTERIA AS HOST FOR BIOMOLECULE PRODUCTION

Cyanobacteria stand out as one of the most promising candidates as hosts for bioproduction (Knoot et al., 2018). Because cyanobacteria utilize solar energy to fix carbon dioxide, a greenhouse gas, and can convert these reduced carbon products into valuable metabolites (Lau et al., 2015), they are especially attractive in an era where sustainable biotechnological processes are of increasing importance (Ruffing, 2011). Additionally, cyanobacteria possess a number of advantageous features relative to other photosynthetic organisms. In comparison to eukaryotic algae and plants, cyanobacteria are more genetically tractable (Parmar et al., 2011; Lau et al., 2015), grow more rapidly, and can achieve higher efficiencies of solar energy capture and conversion (Dismukes et al., 2008; National Academies of Sciences, 2018). Furthermore, cyanobacteria can be cultivated without the need for arable landmass or potable water supplies (Nozzi et al., 2013) and can potentially even degrade aquatic pollutants, such as aromatic hydrocarbons (Ellis, 1977; Cerniglia et al., 1979, 1980; Narro et al., 1992) and xenobiotics (Megharaj et al., 1987; Kuritz and Wolk, 1995) to remediate contaminated water supplies. Yet, relative to other photosynthetic organisms, especially plants, cyanobacteria are currently not used in many scaled agricultural or biotechnological applications. The underutilization of cyanobacteria stems partially from their relative novelty as crop species. Whereas, technologies for cultivating, harvesting, and breeding plants have been under extensive development for many millennia, comparable research to improve the prospects for cyanobacterial cultivation has largely been pursued only since the 1970s (Sheehan et al., 1998).

Cyanobacteria make a number of compounds that are comparable to food, fiber, and fuel products routinely acquired from plants, although cyanobacterial strains have not been as extensively modified to improve their compatibility for scaled cultivation. Like many genera of eubacteria, cyanobacteria can synthesize polyhydroxyalkanoates, a thermoplastic class of biodegradable polyesters that includes polyhydroxybutyrate (Quintana et al., 2011). Many cyanobacterial strains also produce a wide spectrum of secondary metabolites with high-value commercial properties, such as pigments, vitamins, amino acids, macrolides, fatty acids, lipopeptides, and amides (Lau et al., 2015). In total, cyanobacteria are estimated to have the capacity to produce around 1,100 secondary metabolites (specific cyanobacterial bioproducts are beyond the scope of this article, but are reviewed comprehensively in Dittmann et al., 2015; Salvador-Reyes and Luesch, 2015; Xiong et al., 2015, 2017). Beyond natural metabolites, engineering efforts have been used to redirect the metabolism of model cyanobacteria toward biosynthesis of heterologous bioproducts including alcohols, fatty acids, hydrocarbons, fatty alcohols, olefins, organic

acids, sugars, and polyols (engineered cyanobacterial metabolites reviewed in Lai and Lan, 2015; Knoot et al., 2018). Finally, biomass derived from cyanobacterial production processes could be used in animal feed supplements or converted into organic fertilizers, especially if cells are engineered for optimal nutritional and nutraceutical content (Singh et al., 2016).

Relative to heterotrophic microbial host species, cyanobacteria possess key limitations that have kept them from being as widely adopted as bioproduction chassis. First, in comparison to many heterotrophic workhorses, such as the bacterium *Escherichia coli* and the yeast *Saccharomyces cerevisiae*, cyanobacteria have relatively slow division rates. In part, this is due to the relative energy density that is contained in solar energy as compared to a rich medium (Zhang, 2011), since the carbohydrates and other organic carbon sources in rich media have high levels of potential energy stored in their bonds (Utschig et al., 2011). The low energy density of sunlight also impacts cyanobacterial autotrophic productivity, and the specific productivity of target metabolites tend to be lower than that of heterotrophic microbes (Nogales et al., 2013). Nevertheless, some species of cyanobacteria have division rates that compete with that of industrial yeasts (Yu et al., 2015; Jaiswal et al., 2018) (see below). Secondly, the number of genetic tools available for cyanobacterial hosts continues to be limited relative to leading heterotrophic model organisms. Additionally, most cyanobacterial species are polyploid (Griese et al., 2011), which can complicate acquisition of fully segregated strains (Kelly et al., 2018), especially in the presence of restriction-modification systems in cyanobacteria that can limit transformation efficiencies (Stucken et al., 2013). Importantly, the technology for growing cyanobacteria at large scales is underdeveloped and low-cost bioreactors or other cultivation platforms systems such as open ponds need to be improved (Knoot et al., 2018). Bioreactors design must contend with the conflicting demands of scaling in two dimensions (to capture sunlight) while minimizing liquid volumes and reactor cost so that operational and capital expenses can be economically viable (Chisti, 2013; Nozzi et al., 2013; Acién et al., 2017). Genetic instability of heterologous pathways can also decrease bioproduction of cyanobacterial strains (Jones, 2014) and is increased by the abundance of repeated DNA motifs that lead to increased homologous recombination in cyanobacteria (Cassier-Chauvat et al., 2016). As outlined below, a number of research efforts have been directed toward overcoming cyanobacterial host limitations in recent years.

Model Cyanobacterial Strains

While cyanobacteria are an extremely diverse phylum, a relatively small number of cyanobacterial strains have been selected as models, often because these strains have features that mitigate some of the limitations described above (**Table 1**). Notably, model workhorse cyanobacteria include *Synechococcus elongatus* PCC 7942 (hereafter *S. elongatus* PCC 7942), the first reported strain to be transformed through natural DNA uptake pathways (Shestakov and Khyen, 1970), and *Synechocystis* sp. PCC 6803 (*Synechocystis* PCC 6803), originally isolated in 1968 (Stanier et al., 1971), the first strain to have complex *in silico* models

built for genome-scale prediction of metabolism (Fu, 2009). *S. elongatus* has been extensively used as a model of the circadian clock (Ditty et al., 2003), while *Synechocystis* PCC 6803 has served as a useful species for the investigation of core photosynthetic complexes due to its capacity to be grown under photoautotrophic, mixotrophic, or heterotrophic conditions (Vermaas, 1996). *Nostoc* sp. PCC 7120 (*Nostoc* PCC 7120) is a filamentous freshwater cyanobacterial strain which has been used extensively as a model to investigate cellular differentiation (Kumar et al., 2010). *Nostoc* PCC 7120 is capable of fixing nitrogen by forming heterocysts (Cai and Wolk, 1997), which are differentiated cells that efficiently catalyze the reduction of dinitrogen (Herrero et al., 2001), or be exploited for hydrogen production (Tamagnini et al., 2002). These three model species are arguably the best studied strains, but all have relatively modest doubling times: *Nostoc* PCC 7120 (14–15 h) (Callahan and Buikema, 2001), while *S. elongatus* PCC 7942 and *Synechocystis* PCC 6803 have doubling times around 7–12 h (Vermass et al., 1988; Mori et al., 1996). Furthermore, these established models have relatively limited capacity to withstand high light intensities and elevated temperatures that are expected to be encountered in outdoor bioreactors (Yu et al., 2013).

Driven by the desire for cyanobacterial models that are more amenable for bioindustrial applications, other cyanobacterial strains with much more rapid division times have emerged as important models. Significant efforts have been focused on *Synechococcus* sp. PCC 7002 (*Synechococcus* PCC 7002) in recent years, a unicellular cyanobacterium with a faster doubling time of ∼2.6 h (Ludwig and Bryant, 2012). *Synechococcus* PCC 7002 is also capable of growth in a variety of salt, temperature, and light conditions (Sheng et al., 2011; Ruffing et al., 2016), enabling the possibility of utilizing saltwater resources for growth media. More recently, still faster-growing strains have been reported, including *Synechococcus elongatus* UTEX 2973 (*S. elongatus* UTEX 2973) (Yu et al., 2015) and *Synechococcus elongatus* PCC 11801 (*S. elongatus* PCC 11801) (Jaiswal et al., 2018). Interestingly, although the growth rates of these strains are substantially faster (1.5–3 h), genome sequence and proteomic approaches have shown that they are exceptionally closely-related to the much slower-growing model, *S. elongatus* PCC 7942 (Mueller et al., 2017). Indeed, the genome of *S. elongatus* UTEX 2973 is 99.8% identical to *S. elongatus* PCC 7942 and differential regulation of a relatively small subset of common pathways largely accounts for the substantial growth differences (Abernathy et al., 2017; Mueller et al., 2017; Tan et al., 2018). Another related strain, *S. elongatus* PCC 11801 is ∼83% identical to *S. elongatus* PCC 7942 and shares some key modifications with that of *S. elongatus* UTEX 2973 that are responsible for its higher growth rates and increased tolerance to certain environmental stresses (Jaiswal et al., 2018). In a recent publication, Ungerer et al. identified five single nucleotide polymorphisms (SNPs) in three genes (*atpA*, *ppnK*, and *rpaA*) as responsible for rapid growth in *S. elongatus* UTEX 2973 (Ungerer et al., 2018) and these SNPs were also present in *S. elongatus* 11801. The a*tpA* SNP yielded an ATP synthase with higher specific activity, the *ppnK* SNP encoded a NAD$^+$ kinase with significantly improved kinetics, and the *rpaA* SNPs caused broad changes

TABLE 1 | Common cyanobacterial model organisms.

Strain	Genome size	Endogenous plasmids	Lifestyle features	DNA transfer methods
Nostoc sp. PCC 7120	6.4 Mb	6 plasmids size ranging from 5.6 to 408 kb	Freshwater; filamentous	Conjugation (Wolk et al., 1984), electroporation (Thiel and Poo, 1989)
Synechococcus elongatus PCC 7942	2.7 Mb	46 kb plasmid (& 7.6 kb non-essential plasmid)	Freshwater; unicellular	Conjugation (Tsinoremas et al., 1994), natural transformation (Shestakov and Khyen, 1970), electroporation (Marraccini et al., 1993)
Synechococcus elongatus UTEX 2973	2.7 Mb	46 kb plasmid	Freshwater; unicellular	Conjugation (Yu et al., 2015)
Synechococcus sp. PCC 7002	3.0 Mb	6 plasmids size ranging from 4.8 to 186 kb	Euryhaline; unicellular	Conjugation (Kopka et al., 2017), natural transformation (Stevens and Porter, 1980)
Synechocystis sp. PCC 6803	3.6 Mb	7 plasmids size ranging from 2.3 to 120 kb	Freshwater; unicellular	Conjugation (Marraccini et al., 1993), natural transformation (Grigorieva and Shestakov, 1982), ultrasonic transformation (Zang et al., 2007), electroporation (Marraccini et al., 1993)

Features of the most frequently-used model organisms for studies on cyanobacterial physiology and metabolic engineering. Size of primary genome, endogenous extrachromosomal plasmids, osmotolerance, and routinely utilized genetic transformation methods are reported.

in the transcriptional profile. The filamentous cyanobacterium *Leptolyngbya* sp. strain BL0902 (*Leptolyngbya* BL0902) was also recently identified (Taton et al., 2012), can be transformed by conjugation (Taton et al., 2012) and some molecular tools have been characterized (Taton et al., 2014).

Curiously, while *S. elongatus* UTEX 2973 is closely related to *S. elongatus* PCC 7942, it can be only transformed by conjugation (Yu et al., 2015), while *S. elongatus* PCC 11801 is naturally competent for genetic transformation. Apart from *Nostoc* PCC 7120, the other models described above are naturally competent, facilitating the genetic modification of these strains (Porter, 1986; Koksharova and Wolk, 2002). Furthermore, the genomes of the model cyanobacteria described in **Table 1** have been sequenced and the information has been organized in the database, CyanoBase (Fujisawa et al., 2017). The access of genome information, together with the construction of metabolic models (see below), makes it possible to both understand the basic metabolism of cyanobacteria and to achieve higher levels of metabolic redirection and control (Lai and Lan, 2015).

NEW TOOLS FOR SYNTHETIC BIOLOGY IN CYANOBACTERIA

In comparison to the genetic toolboxes for work in popular heterotrophic chasses like *E. coli* and *S. cerevisiae*, relatively limited genetic tools have been developed in cyanobacteria (Sun et al., 2018). The design of standardized gene expression parts (e.g., promoters, terminators) that can be modularly recombined with other genetic elements with predictable outputs has catalyzed a revolution in the complexity of heterotrophic genetic circuits (Popp et al., 2017). Unfortunately, many of the synthetic biology tools and modular parts developed for model heterotrophs often do not perform as robustly in the context of cyanobacterial strains (Huang et al., 2010). For example, many characterized *E. coli* promoters do not display similar traits when used in cyanobacteria (Heidorn et al., 2011). These problems may arise in part because of species-dependent distinctions in

key features, such as RBS sequences or promoter recognition by endogenous transcription, leading to unpredictability in gene expression (Huang et al., 2010; Wang et al., 2012; Camsund and Lindblad, 2014). In recent years, significant efforts have been made in adapting molecular techniques from other organisms in order to extend cyanobacterial toolboxes and facilitate cellular reprogramming for increased production yields. Here, we outline the latest advances in cyanobacterial synthetic biology tools, including promoters, riboswitches, RBSs, reporters, modular vectors, and markerless selection systems.

Engineering Promoters to Enhance Protein Expression

Many of the most advanced synthetic biology circuits and pathways today are firmly rooted in an extensive, well-defined library of promoter "parts" that exhibit predictable behaviors when used to drive the expression of a range of target genes (Shetty et al., 2008). By contrast, the number of constitutive and inducible promoters that have been well-characterized in cyanobacteria has been historically small, and considerable variation is often observed in the expression level achieved for distinct heterologous genes. Some limiting factors have compromised the development of cyanobacterial inducible genetic systems, including toxicity of inducers, leaky expression in the absence of inducer, and inducer photolability. Constitutive promoters may also be useful when continuous gene expression is desired, although many endogenous cyanobacterial promoters that are used for this purpose are dynamically regulated by circadian rhythms.

A few recent studies have focused on expanding and characterizing the collection of foreign promoters that can be used to drive gene expression in cyanobacteria. Some inducible promoter elements have been commonly used in selected cyanobacterial models for a number of years, such as the nickel-inducible *nrsB* promoter (Englund et al., 2016; Santos-Merino et al., 2018) or the IPTG-responsive *trc* promoter (P_{trc}) in *S. elongatus* PCC 7942 or *Synechocystis* PCC 6803 (Geerts et al., 1995; Huang et al., 2010). P_{trc} has recently been adapted to

other model cyanobacteria, such as *Synechococcus* PCC 7002 (Ruffing, 2014) and *Leptolyngbya* BL0902 (Ma et al., 2014). The L-arabinose-inducible *araBAD* promoter (P_{BAD}) that has long been a staple induction system in *E. coli* has recently been introduced in *S. elongatus* PCC 7942 (Cao et al., 2017) and *Synechocystis* PCC 6803 (Immethun et al., 2017). This system is based on the arabinose utilization network, which positively regulates P_{BAD} through the AraC regulator protein (Schleif, 2010) (**Figure 1A**). Most recently, the use of the rhamnose-inducible *rhaBAD* promoter of *E. coli* has been implemented in the model freshwater cyanobacterium *Synechocystis* PCC 6803 (Kelly et al., 2018) (**Figure 1B**). Another orthogonal inducible promoter, P_{van} from *Corynebacterium glutamicum,* that relies upon vanillate-induced suppression of the repressor VanR has been tested in *S. elongatus* PCC 7942 (Taton et al., 2017) (**Figure 1C**). Some of these promoters have been translated to faster-growing cyanobacterial strains, such as P_{trc} in *Synechococcus* PCC 7002 (Ruffing, 2014) and the IPTG-induced promoter P_{lac} in *S. elongatus* UTEX 2973 (Song et al., 2016), although the number of inducible promoters remains relatively limited in the cyanobacterial strains that possess the most promising features for bioproduction.

In some model cyanobacteria, a limited number of constitutive promoters have been characterized relative to one another. For example, P_{A2520} and P_{A2579}, two native promoters of *Synechococcus* PCC 7002, have been found to drive strong expression of heterologous genes (Ruffing et al., 2016). Most recently, nine native promoters have been characterized for their expression in *Synechocystis* PCC 6803 to enrich the promoter toolboxes (Liu and Pakrasi, 2018). Other examples of publications that characterize endogenous promoters in cyanobacteria can be found here (Huang et al., 2010; Markley et al., 2015). Yet, in cyanobacteria, many endogenous promoters are strongly influenced by the circadian clock machinery, and therefore they may not be truly constitutive through a 24 h period (Liu et al., 1995; Markson et al., 2013; Camsund and Lindblad, 2014).

Development of synthetic promoters that have been modified for improved expression specifically in cyanobacteria has emerged as another promising strategy. In *S. elongatus* PCC 7942, a tandem promoter composed of a truncated native promoter PR from *rrnA* of *S. elongatus* PCC 7942, and the consensus-σ70 PS promoter from *E. coli* has been designed (Chungjatupornchai and Fa-Aroonsawat, 2014). Zhou et al. reported the development of a constitutive promoter P_{cpc560} in *Synechocystis* PCC 6803 that produced a high-level of gene expression (Zhou et al., 2014). It is based on a truncated native promoter (P_{cpcB}) and its strength resides in the presence of multiple transcription factor binding sites. In another study, P_{sca3-2}, a variant of P_{tac} of *E. coli* was found to act as constitutive promoter with high levels of expression (Albers et al., 2015). Within the same context of synthetic constitutive promoters, a truncated version of *psbA2* native promoter was developed in *Synechocystis* PCC 6803 (Englund et al., 2016). The expression of this derivate of P_{psbA2} increased 4-fold compared to the original. Similarly, inducible promoters can be tuned for better performance in cyanobacteria,

such as the anhydrotetracycline-activated variant of P_{R40} from *E. coli*, P_{L03} (Huang and Lindblad, 2013), the IPTG-inducible promoter P_{sca6-2} (Albers et al., 2015), or the T7 RNA Polymerase promoter (Ferreira et al., 2018). Because the promoters described above are modified and/or heterologous, it is possible that they could escape regulatory activities of the circadian clock, yet it is unusual for these promoters to be characterized in more than one circadian period, or at varied light intensities. To solve this problem, some promoters have been characterized for their capacity to drive heterologous gene expression under multiple conditions (e.g., light/dark, aerobic/anaerobic) such as the FNR-activated promoter, P_{O2}, described in *Synechocystis* PCC 6803 (Immethun et al., 2016).

Optimizing Ribosome Binding Sites for Biotechnological Applications

The rate of protein production from a mRNA transcript also depends on the strength of the RBS in recruiting ribosomes for translation. The position and sequence of a given RBS significantly influences translational efficiency. Although "RBS calculators" have long been under development for heterotrophic microbes (Salis, 2011), it is only relatively recently that such efforts extended toward development of RBS libraries for cyanobacteria. RBS sequences from the BioBrick Registry of standard biological parts have been characterized in *Synechocystis* PCC 6803 (Heidorn et al., 2011; Englund et al., 2016). In one recent example, 20 native RBS elements have been characterized in *Synechocystis* PCC 6803 (Liu and Pakrasi, 2018), including two previously described by Englund et al. (2016). These efforts are becoming increasingly coordinated with attempts to develop synthetic RBSs that are based on *in silico* modeling tools for *Synechococcus* PCC 7002 (Markley et al., 2015), *Synechocystis* PCC 6803 (Heidorn et al., 2011; Taton et al., 2014; Xiong et al., 2015; Thiel et al., 2018), or *S. elongatus* PCC 7942 (Taton et al., 2014); (Wang et al., 2018).

Efforts to expand the toolbox of characterized promoter elements are foundational contributions that will enable more sophisticated circuit design in cyanobacteria, yet the capacity to fully predict expression output from a given element remains elusive. Context-specific features of a given expression construct can alter the performance of a given promoter-RBS combination, decreasing the modularity of promoter elements. For example, it is well-established that secondary structure can arise between a specific heterologous gene that can interfere with transcription or translation, leading to variability when expressing different genes from an identical promoter-RBS cassette. Promoter elements that are based on a bicistronic design have been developed for *E. coli* that exhibit much more consistent performance, regardless of the downstream gene sequence that is being expressed (Mutalik et al., 2013). No such system has been described in cyanobacteria and few promoters have been as extensively characterized. Therefore, it often remains difficult to anticipate the likely expression level of a construct during the design phase.
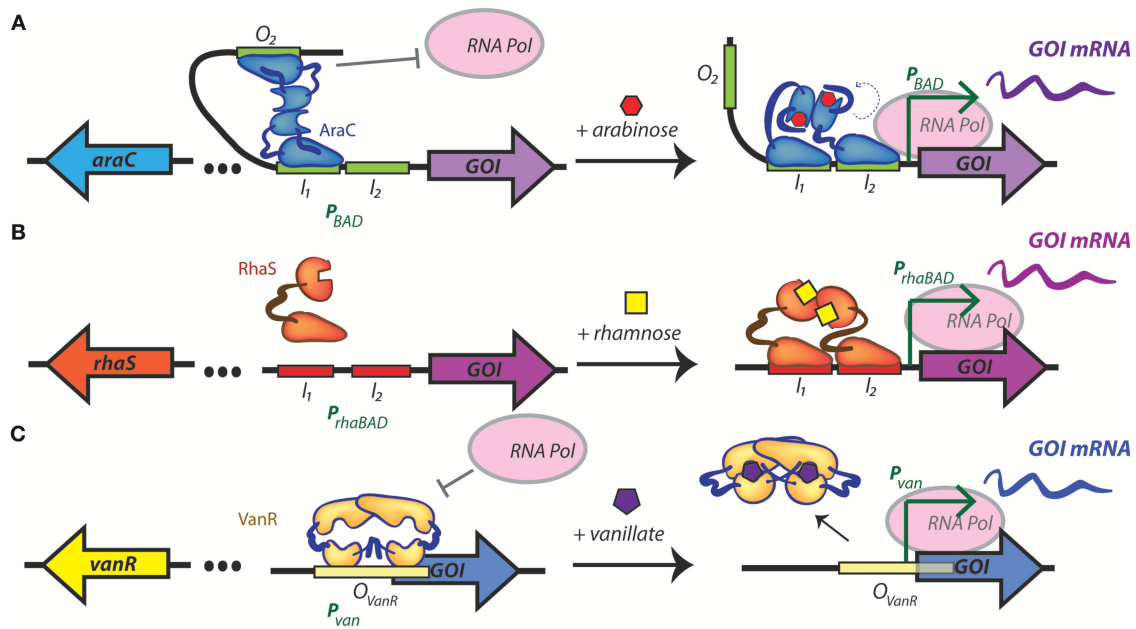
FIGURE 1 | Regulatory mechanisms of inducible promoters recently introduced into cyanobacteria. **(A)** Arabinose-inducible gene expression from P_{BAD}, which is both positively and negatively regulated by the transcriptional regulator AraC. In the absence of arabinose, an AraC dimer contacts the O_2 and I_1 half sites of the promoter favoring the formation of a DNA loop that is unfavorable for RNA polymerase binding. When arabinose is present, AraC undergoes a structural change that releases the DNA loop, and ideally positions an AraC activation domain that promotes gene transcription. **(B)** Rhamnose-inducible gene expression from P_{rhaBAD}. In the absence of rhamnose RhaS is unable to dimerize, limiting its capacity to bind to promoter elements. In the presence of rhamnose, a dimer of RhaS binds to the I_1 and I_2 repeat half-sites of the *rhaBAD* promoter, recruiting RNA polymerase and activating the transcription of the target gene. **(C)** The vanillate-inducible promoter, P_{van}, is repressed by a dimer of VanR bound to its operator. When vanillate is present, two molecules bind the VanR dimer, resulting in a conformation change and release of transcriptional repression. I_1, binding site 1 in P_{BAD} and P_{rhaBAD}; I_2, binding site 2 in P_{BAD} and P_{rhaBAD}; GOI, gene of interest; mRNA, messenger RNA; O_2, operator of P_{BAD}; O_{VanR}, operator in P_{van}; RNA pol, RNA polymerase. In all cases, left and right panel represent uninduced and induced promoters, respectively.

Riboswitches as Tools for Robust Control of Gene Expression

Riboswitches are versatile tools for genetic engineering and allow control of gene expression by manipulating secondary structure within a mRNA transcript. A riboswitch is composed of an aptamer sequence that imposes a secondary-structural conformation on the mRNA and which influences translational efficiency from the transcript. The aptamer sequence has cis-activating or cis-repressing effects on the mRNA on which it is encoded, while a trans-acting factor can bind and encourage the formation of an alternative secondary-structure conformation (Nudler, 2006). Frequently, the aptamer sequence is designed to form a stem-loop structure that prevents the attachment of ribosomes to a 5′ RBS. A trans-acting factor (which can be a small molecule metabolite or a non-coding regulatory RNA) binds to the aptamer in a manner that stimulates a new conformational state and increases accessibility of the RBS, improving translation of the encoded protein. Riboregulators have features that make them powerful tools for controlling gene expression when used in tandem with transcriptional-based approaches. These include the fact that they typically exhibit a high degree of modularity, can drive the expression of proteins over a large, physiologically-relevant range (thus avoiding issues such as toxicity or inclusion body formation), minimize "leaky" expression (protein production in the absence of inducer),

have fast response times (time between ligand binding and protein expression), are tunable (a large dynamic range of protein expression with increasing inducer), and can be used to regulate multiple genes simultaneously (Callura et al., 2010).

Relative to other microbial systems, the use of cyanobacterial riboregulators is just beginning to become more wide-spread and only a few riboswitch designs are in common use for metabolic engineering applications (Connor and Atsumi, 2010). A modified theophylline-dependent synthetic riboswitch was first reported in *S. elongatus* PCC 7942, allowing a strict regulation of protein production in this cyanobacterium (Nakahira et al., 2013). This effective riboswitch has since been implemented in other cyanobacteria such as *Synechocystis* PCC 6803, *Leptolyngbya* BL0902, *Nostoc* PCC 7120, and *Synechocystis* sp. strain WHSyn (*Synechocystis* WHSyn) (Ma et al., 2014; Armshaw et al., 2015; Ohbayashi et al., 2016). Similarly, Taton et al. have also developed NOT gate molecular circuits using transcriptional repressors controlled by theophylline-dependent synthetic riboswitches to downregulate gene expression in five diverse strains of cyanobacteria, including three model organisms, *Nostoc* PCC 7120, *Synechocystis* PCC 6803, and *S. elongatus* PCC 7942, as well as two recent isolates, *Leptolyngbya* BL0902 and *Synechocystis* WHSyn (Taton et al., 2017). A native cobalamin-dependent riboswitch has been reported in *Synechococcus* PCC 7002 (Pérez et al., 2016), although it remains unclear if this genetic

tool can be implemented in other cyanobacteria that are not cobalamin auxotrophs. More recently, another native riboswitch, a glutamine aptamer, has been described in *Synechocystis* PCC 6803 (Klähn et al., 2018). However, it exhibited poor ligand affinity compared to the aptamers of other riboswitch classes. These newly-described native regulatory sequences open up the possibility of discovering new cyanobacterial riboswitches still not identified. Future approaches that focus on the combination of *in silico* structural analysis and *in vivo* genetic tuning of riboswitches that can be modulated by a wide variety of ligands hold the promise of contributing significantly to metabolic engineering efforts (Berens and Suess, 2015).

Reporter Proteins

Reporters allow for easy quantitation of gene expression, visualization of subcellular localization, and interaction of proteins with other cellular components (Berla et al., 2013). Fluorescent reporter proteins do not require additional substrates to be detected *in vivo* and there are an ample range of colors. However, the use of fluorophores can be complicated in cyanobacteria due to the autofluorescence of photosynthetic pigments. The phycobilins and chlorophyll molecules that are major components of the photosynthetic electron transport chain can cause competitive absorbance of excitation light source, re-absorbance of fluorescent or bioluminescent reporter emission, and signal interference (Yokoo et al., 2015; Ruffing et al., 2016). Because of strong chlorophyll autofluorescence, use of red fluorophores is not recommended, but many other reporters including GFPmut3B (a mutant of green fluorescent protein) and EYFP (enhanced yellow fluorescent protein) have been routinely used (Pédelacq et al., 2006; Huang et al., 2010; Yang et al., 2010; Heidorn et al., 2011; Huang and Lindblad, 2013; Landry et al., 2013; Cohen et al., 2014). Despite the frequent use of fluorescent proteins in cyanobacteria and the continual development of improved variants, further improvements to available reporters could greatly increase their range of applications (Rodriguez et al., 2017). Recently developed fluorophores with superior brightness, photostability, and quantum yield have great promise, including mOrange, mTurquiose, mNeonGreen, or Ypet (Chen et al., 2012; Ruffing et al., 2016; Jordan et al., 2017). In addition to fluorescent proteins, luciferase-based bioluminescence assays are routinely used, especially for tracking gene expression patterns throughout the circadian cycle or under different environmental conditions (Fernández-Piñas et al., 2000; Cohen et al., 2015). Luciferase reporters are ideal for monitoring gene expression because of the short half-life of the enzymes, which provides a readout that is close to real-time (Ghim et al., 2010). Alternatively, fluorophores can be modified with protein degradation sequences that greatly reduce their half-life (Wang et al., 2012), increasing their suitability as a transcriptional readout (Noguchi and Golden, 2017).

Modular Vector Systems for Engineering Cyanobacteria

Vector systems based on the use of standard biological parts (http://parts.igem.org) and assembly schemes (e.g., BioBrick and BglBrick) were originally designed for organisms such as *E. coli*,

B. subtilis, or yeast in order to increase the modular assembly of many different parts. By contrast, most standard genetic tools and vectors were developed for a specific cyanobacterial strain and generally, these tools have not been designed to be modular. In recent years, a few vector systems have been designed specifically to work in diverse strains and/or to contain a modular organization, which could facilitate standardization and characterization of component parts.

A notable modular vector system was described in 2014, using a range of plasmids designed to be compatible with a broad host-range (Taton et al., 2014). These plasmids included both autonomously replicating plasmids and suicide plasmids for gene knockout and knockin, and were characterized in diverse cyanobacterial strains to ensure their proper functioning. As a part of this work, the authors also created a web server, CYANO-VECTOR, that can assist in the *in silico* design of plasmids and assembly strategies. Similarly, chromosomal integration vectors carrying standard prefix and suffix sequences suitable for BioBrick-based cloning have been designed for *Synechococcus* PCC 7002 (Vogel et al., 2017) and *S. elongatus* PCC 7942 (Kim et al., 2017). In a recent article, a versatile system called CyanoGate based on the Plant Golden Gate MoClo kit and the MoClo kit for the microalgae *Chlamydomonas reinhardtii* was developed (Vasudevan et al., 2018). Vasudevan et al. demonstrated that the functionality of this system was robust across different two cyanobacterial species, *Synechocystis* PCC 6803 and *S. elongatus* UTEX 2973.

Markerless Selection as a Tool to Facilitate Cyanobacterial Engineering

In cyanobacteria, as in other bacteria, metabolic engineering involving multiple genetic manipulations requires multiple selective markers. To make deletions, genes are normally replaced by antibiotic resistance markers to evaluate what phenotypic effects take place. However, the generation of a strain with numerous deletions is restricted using this method because the availability of resistance markers is limited. Alternatively, a range of markerless selection strategies has been developed to increase the number of modifications that may be performed to modify genetically these organisms.

The first markerless system described in the cyanobacterium *S. elongatus* PCC 7942 relies on a dominant streptomycin-sensitive *rps12* mutation (Matsuoka et al., 2001). The method is based on a double selection cassette composed of a kanamycin resistance gene and, as an alternative negative selection marker, a *rps12* wild-type copy that confers a dominant streptomycin sensitive phenotype. Streptomycin-resistant, kanamycin-sensitive markerless mutants can be recovered in a second transformation (Takahama et al., 2004). The main drawback of this method is the need to work in a genetic background that contains the appropriate *rps12* mutation. Moreover, this strategy requires two, time-consuming transformation events and cloning of two different suicide vectors. Recently, new time-saving alternatives have been developed for markerless gene replacement in cyanobacteria.

Begemann et al. described a counter-selection method for *Synechococcus* PCC 7002 based on organic acid toxicity (Begemann et al., 2013). The system was based on the use of the product of the *acsA* gene, an acetyl-CoA ligase. The loss of AcsA function was used to develop an acrylate counter-selection method. Another alternative counter-selection method was developed for *Synechocystis* PCC 6803, which involves the use of the endogenous nickel inducible promoter to drive an *E. coli* derived toxin gene known as *mazF* (Cheah et al., 2013). MazF is an endoribonuclease that acts as a global inhibitor for the synthesis of cellular proteins, because it cleaves mRNA at the ACA triplet sequence. A different markerless gene deletion system that only requires a single vector has also been described for *Synechocystis* PCC 6803 and uses a *nptI-sacB* double selection cassette (Viola et al., 2014). The *nptI* gene confers resistance to the antibiotic kanamycin, while expression of the *sacB* gene is toxic to bacteria grown on sucrose-containing media. Counter-selection based on *sacB* is not functional in *Synechococcus* PCC 7002 (Zhang and Song, 2018), possibly because *sacB* selection is sensitive to salt (Kunst and Rapoport, 1995) required for growth of this marine cyanobacterium. A few markerless gene deletion systems have been shown to work in multiple cyanobacterial strains, for example Kojima et al. developed an efficient method for generating knockouts in *Synechocystis* PCC 6803 and *Synechococcus* PCC 7002 (Kojima et al., 2016). This system is based on knocking out the *aas* gene, an acyl-acyl carrier protein synthetase, and selecting the mutants by their free fatty acid tolerance.

The most recent markerless systems are based on CRISPR-based technology (**Figure 2**) that does not require any counter-selection genes (Behler et al., 2018). In general, all CRISPR-Cas [clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein (Cas) system] technology relies on the capacity to target a protein (typically a nuclease) to a very precise genomic locus. This specificity is conferred by a single-guide RNA (sgRNA) is programmed to be a complement to the target genomic site, and which assembles with Cas9 or Cas12a (formerly known as Cpf1) into an effector complex (**Figure 2A**). Depending upon the CRISPR system, the effector complex may contain other RNA sequence: the CRISPR-Cas9 system requires two separate RNA strands, the CRISPR RNA (crRNA) that encodes the guide sequence, and trans-activating crRNA (tracrRNA), while CRISPR-Cas12a requires only a single crRNA (**Figure 2A**). Most cyanobacterial genomes naturally contain CRISPR-Cas repeat sequences (Cai et al., 2013) useful in defending the cell against foreign genetic material as the Cas nuclease can be directed to cleave sequences that are specific to an exogenous source (e.g., viral DNA).

The specificity of Cas9 and Cas12a targeting has been repurposed to design markerless genome editing systems, and other genetic control elements (Li et al., 2016; Ungerer and Pakrasi, 2016; Wendt et al., 2016; Niu et al., 2018; Ungerer et al., 2018; Xiao et al., 2018). When directed to a DNA target within the cyanobacterial genome, the double-stranded DNA breaks induced by the CRISPR effector are lethal unless they can be repaired in a manner that alters the DNA so that it is no longer recognized by the sgRNA (**Figure 2A**) (Behler et al., 2018).

Double-stranded DNA breaks engage cyanobacterial DNA repair machinery, which can resolve the damage by error-prone non-homologous end joining or by homologous recombination if a suitable template is available. To induce an inactivating mutation in a target gene it is often sufficient to express Cas9 or Cas12a along with a gene-specific sgRNA and to rely upon the error in the genetic repair systems to introduce point mutations and frameshifts that will inactivate the gene without the need for a selectable marker. More advanced genome editing (e.g., knockout of specific regions or insertion of new DNA at the target locus) can be accomplished through homologous recombination if a suitable template is also introduced into the cell (see **Figure 2B**). This markerless methodology has been used to introduce point mutations, knock-out large genomic regions, and "knock-in" genes in a range of cyanobacterial species, including *Synechocystis* PCC 6803 (Xiao et al., 2018), *S. elongatus* UTEX 2973 (Ungerer and Pakrasi, 2016; Wendt et al., 2016), *S. elongatus* 7942 (Li et al., 2016; Ungerer et al., 2018), and *Nostoc* 7120 (Ungerer and Pakrasi, 2016; Niu et al., 2018).

The CRISPR-Cas system is still under active development and some limitations may need to be overcome before it becomes a methodology that completely replaces more traditional cyanobacterial genome engineering techniques. One limitation is that colonies recovered from CRISPR-mediated transformations can have a low penetrance of the desired genomic alteration. For example, early reports of CRISPR editing in cyanobacteria have shown between 20 and 70% of recovered strains are the desired mutant (Li et al., 2016; Ungerer and Pakrasi, 2016; Wendt et al., 2016; Xiao et al., 2018). The low efficiency of some transformations requires screening and validation of a higher number of recovered colonies to obtain the correct strain. It may be possible to overcome this limitation by encoding two spacers that target the genomic region rather than one (see spacer depiction in **Figure 2B**), as shown in recent report in *Nostoc* PCC 7120 (Niu et al., 2018). The expression of Cas9 also appears to be toxic in a dose-dependent manner in some cyanobacteria, such as *S. elongatus* UTEX 2973 (Wendt et al., 2016), although this toxicity is not apparent in others (Xiao et al., 2018), even in some closely-related species (Li et al., 2016). While the mechanism by which Cas9 causes toxicity remains unclear, it is possible to substitute Cas12a for Cas9 to bypass the issue in many species (Ungerer and Pakrasi, 2016; Swarts and Jinek, 2018). Yet, such uncertainties also contribute to the concern that CRISPR-mediated techniques can lead to alter off-target genomic sites that could lead to misinterpretation of observed phenotypes (Fu et al., 2013). Strategies to minimize off-target CRISPR-Cas activity have been extensively explored in other organisms, but have not been rigorously evaluated in cyanobacteria. Finally, although genomic modifications introduced by CRISPR are themselves markerless, the sgRNA and nuclease themselves are often introduced on plasmids that require selectable markers (**Figure 2B**), and following genome editing, it is often desirable to cure the plasmid through some form of counter-selection (Xiao et al., 2018). This can make the process of genome editing by CRISPR-Cas considerably longer than traditional selection-based approaches.
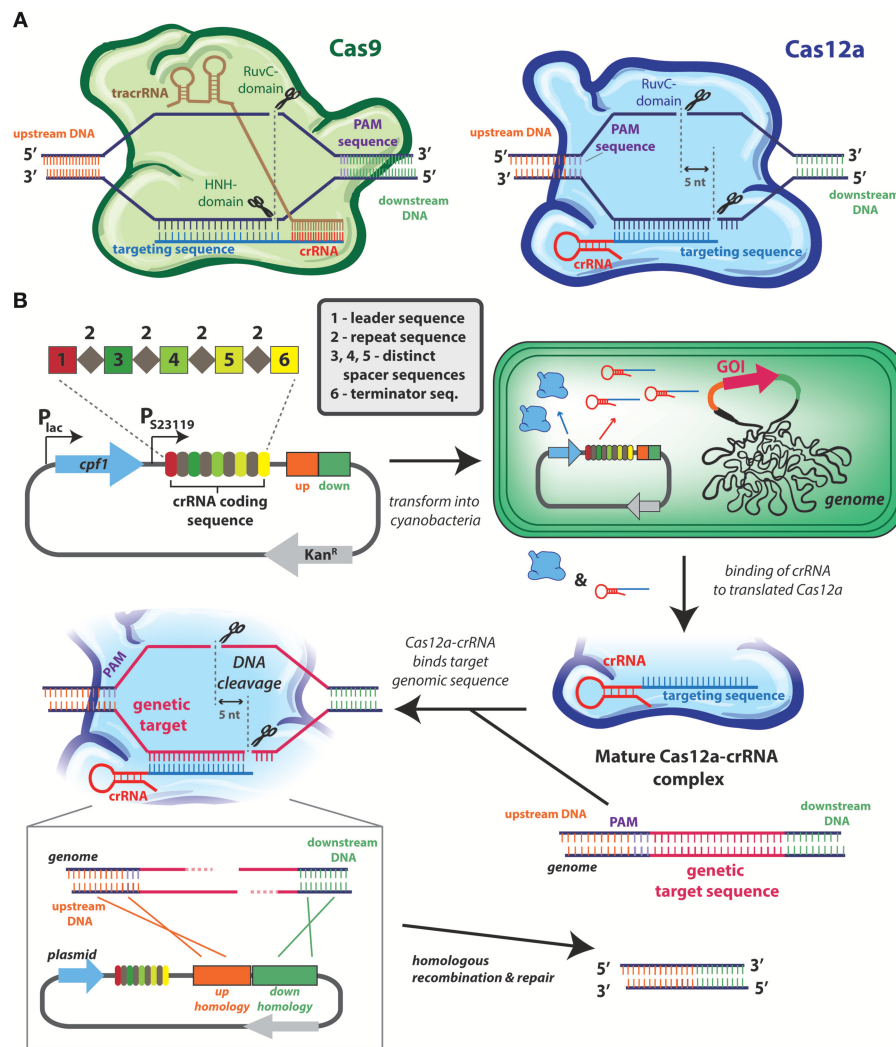
FIGURE 2 | CRISPR-*cas* technologies used for genome editing of cyanobacterial genome. **(A)** CRISPR-*cas* systems used in cyanobacteria: Cas9 (left) and Cas12a (right). In type II systems, Cas9 (green) forms a complex with the crRNA (red) and tracrRNA (brown), whereas in type V systems, Cas12a (blue) forms a complex with its crRNA (red). Both nucleases identify their target sites on the basis of complementarity of the guide sequence (blue) by heteroduplex formation, as well as the presence of a neighboring protospacer-adjacent motif (PAM; purple), although the positioning of the PAM is distinct between systems Cas9 cleaves the complementary DNA strand using its HNH domain, and a second RuvC domain to cleave the non-complementary strand. Cas12a cleaves target DNA using one RuvC endonuclease domain, and a second endonuclease activity is attributed to a novel nuclease domain. **(B)** Example application of CRISPR/Cas12a in *Synechococcus elongatus* PCC 7942. This system is based on the use of a single plasmid that contains: the *cas12a* gene under the *lac* promoter and a crRNA-encoding array (from *Francisella novicida*) under a constitutive (J23119) promoter. This plasmid allows for inducible, transient expression of *cas12a*. This crRNA-encoding array contains a leader sequence (1) followed by direct repeat (2) and spacer (3, 4, 5) sequences, and a terminator (6). In most cyanobacterial CRISPR-Cas reports a single spacer region (e.g., number 3) acts as the genome target spacer, while other spacers are not utilized. Multiple genomic regions can be targeted by encoding other spacers with complementary sequences, or the efficiency of modifying a single genomic loci can be improved by utilizing multiple spacers (Niu et al., 2018). Homology regions designed to promote recombination-mediated repair can also be encoded in the plasmid (represented by orange and green rectangles). Once inside the cell, an active and mature Cas12a-crRNA complex forms and is directed to the target sequence. After the introduction of the 5-nt staggered double-stranded break, the homology regions in the plasmid serve as a repair template to introduce the intended genomic changes. The transconjugants are selected on BG11 agar plates supplemented with kanamycin. The positive colonies are streaked on BG11 agar plates to cure the plasmid and they are assayed to check the loss by the inability to grow on kanamycin-containing media. Depicted system is modeled after the one described by Ungerer and Pakrasi (2016).

CRISPR-based applications reach beyond markerless genome engineering to a wider range of applications. Briefly, nuclease-dead variants of Cas9/Cas12a do not generate DNA breaks, but can still be targeted to specific DNA sequences via sgRNA can be fused to other functional domains (e.g., transcriptional activator or repressor domains). This creates hybrid Cas proteins that bind to endogenous sequences, but which can perform other functions, such as regulating gene expression. Recently, these approaches have been used to create synthetic transcription factors able to modulate the expression of essential

genes and key metabolic pathways in cyanobacteria. For a more comprehensive review of CRISPR-based applications in cyanobacteria, see (Behler et al., 2018).

IMPROVING CARBON CAPTURE AND DIRECTING CARBON FLUX THROUGH RATIONAL ENGINEERING APPROACHES

Engineering cyanobacterial strains for bioproduction requires that not only heterologous genes and pathways be expressed in new hosts, but also that carbon that is captured by cyanobacteria be efficiently directed toward the desired metabolic products (Stephanopoulos, 2012; Woo, 2017). Thus, it is a prerequisite for metabolic engineers to optimize pathways to increase the total carbon flux toward target product generation, and/or to enhance the pool size of rate-limiting metabolites (Kanno et al., 2017; Carroll et al., 2018). Although the regulation of carbon partitioning in cyanobacteria is not fully understood, it can be diverted under certain conditions like nutrient deprivation and irradiance stress, and therefore stress conditions are routinely used in studies that aim to increase product yield (Woo, 2017). New predictive tools that can assist in channeling fixed carbon to desired metabolic pathways have been in active development. These tools include models for metabolic analysis and flux balance analysis and will be described in greater detail below. Beyond channeling carbon flux, another strategy is to improve the total pool of available carbon, and therefore multiple efforts to improve carbon fixation rates of cyanobacteria have been pursued.

Improving Carbon Fixation Rates

Many efforts to engineer cyanobacteria for enhanced productivity have emphasized increasing the total rate of carbon fixation, most frequently by improving the activity or efficiency of RuBisCO. Cyanobacteria have highly efficient carbon concentration mechanisms (CCM) relative to other phototrophs such as algae and plants (Price et al., 2011). The cyanobacterial CCM is efficient in part because it uses bicarbonate transporters to actively transport bicarbonate into the cell, which effectively overcomes the slower (10^4-fold) diffusion rates of CO_2 in water compared to air (Price et al., 2011). Ultimately, the accumulated bicarbonate is transported across the carboxysome shell, and converted to CO_2 in the carboxysome lumen, where RuBisCO is concentrated (Badger et al., 2002; Price et al., 2008). Increasing the efficiency of the cyanobacterial CCM could not only increase the concentration of RuBisCO's substrate, but also reduce the production of energetically-costly photorespiratory byproducts. Recently, extra bicarbonate transporters were expressed in *Synechocystis* PCC 6803, leading to a 2-fold enhancement of the growth rate and a higher amount of biomass accumulation (Kamennaya et al., 2015).

Other efforts have focused upon improving carbon fixation rates by balancing the enzymatic activities in the Calvin-Benson-Bassham (CBB) cycle to improve its total metabolic flux. Liang and Lindblad have demonstrated that overexpression of any of four separate enzymes in the CBB cycle (RuBisCO,

Sedoheptulose bisphosphatase, fructose bisphosphate aldolase, or transketolase) can improve total carbon fixation rates in *Synechocystis* (Liang and Lindblad, 2016) (**Figure 3**). Furthermore, modifying the expression of these CBB enzymes could also enhance the rate of heterologous production of ethanol in *Synechocystis* PCC 6803 (Liang et al., 2018). Kanno et al. have recently described a novel strategy to improve carbon fixation rates by focusing upon importing a continuous supply of substrate for RuBisCO. By expressing glucose transporters, they successfully used the oxidative pentose phosphate pathway to convert some imported sugars for the synthesis of D-ribulose 1,5 bisphosphate. Using this approach, they demonstrated that an engineered strain was able to produce a targeted biochemical, 2,3-butanediol, under both light and dark conditions (Kanno et al., 2017).

Improving Carbon Fixation Through Sink Engineering

While it is frequently assumed that the rate limiting metabolic step of carbon fixation is directly related to the slow catalytic activity of RuBisCO, photosynthetic activity can be limited by other metabolic steps. Utilization of the primary products of the CBB cycle can limit step photosynthesis rather than the fixation of carbon dioxide itself. This is concept was first explored in plant models, where the mechanisms underlying triose phosphate utilization (TPU) limitation are more deeply characterized. In plants, there is a physical separation between where primary products of the CBB are generated (source tissues, e.g., leaves) and where much of these products will ultimately be utilized to support metabolic activity and growth (sink tissues; e.g., roots). Efficient operation of photosynthetic metabolism requires the balance of light energy ("source") that can be highly dynamic in the environment with an equivalent capacity to utilize/dissipate this energy using anabolic metabolism or quenching mechanisms ("sinks"). Without absorption of adequate photons (i.e., low "source"), a photosynthetic organism will starve; and without sufficient pathways to process or dissipate absorbed energy (low "sink"), end products of photosynthesis can accumulate leading to feedback inhibition and overreduction of the electron transport chain (ETC) (Gifford et al., 1984; Paul and Foyer, 2001). TPU limitation on photosynthesis can occur either because of experimental manipulation (e.g., exogenously supplied carbohydrates, or chemical inhibition of carbohydrate transport/catabolism), or because the sum of metabolic processes downstream the CBB are insufficient to remove triose phosphates at an equivalent rate as they are being generated (Sawada et al., 1986; Sharkey et al., 1986; Krapp et al., 1991; Paul and Foyer, 2001; Adams et al., 2013; Demmig-Adams et al., 2014).

While a number of published reports in plants have sought to enhance biomass accumulation by reducing TPU limitation, relatively few studies exist that suggest that cyanobacterial photosynthesis may also be limited by downstream metabolism. Much of our current knowledge on sink limitation in cyanobacteria is indirect, where researchers have found the activity of a heterologous metabolic pathway expressed in a cyanobacterial model leads to increased photosynthetic activity
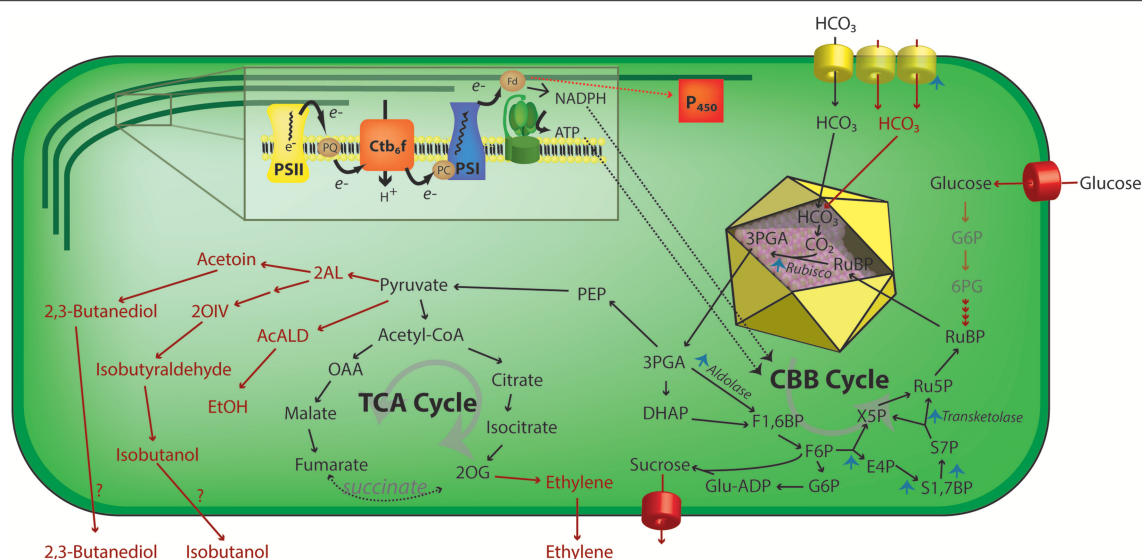
FIGURE 3 | Schematic representation of engineering strategies that increase photosynthetic activity in cyanobacteria. Using ATP and NADPH produced by the light reactions (inset), the CBB acts to assimilate inorganic carbon via a RuBisCO—mediated reaction in the carboxysome (yellow icosahedron) lumen. Upregulation of multiple steps of the carbon concentration mechanism or CBB cycle enzymes has been shown to increase total carbon uptake rates by cyanobacteria (blue arrowheads). Additionally, the regeneration of RuBisCO substrate, ribulose 1,5-bisphosphate, by feeding external carbohydrates has been used to increase CO_2 assimilation rates. Multiple metabolic pathways (red arrows) that have been heterologously expressed in cyanobacteria have also been demonstrated to increase the effective rate of carbon fixation, perhaps by relieving sink inhibition of photosynthesis. 2AL, 2-acetolactate; 2 OIV, 2-ketoisovalerate; 2OG, 2-oxoglutarate; 3PGA, 3-phosphoglycerate; 6PG, 6-phosphoglycerate; AcALD, acetaldehyde; Ctb_6f, cytochrome b_6f; E4P, erythrose 4-phosphate; EtOH, ethanol; F1,6BP, fructose-1,6-bisphosphate; F6P, fructose-6-phosphate; Fd, ferredoxin; G6P, glucose 6-phosphate; Glu-ADP, Glucose-ADP; OAA, oxalacetate; P450, cytochrome P450; PC, plastocyanin; PEP, phosphoenolpyruvate; PQ, plastoquinone pool; PSI, photosystem I; PSII, photosystem II; Ru5P, ribulose-5-phosphate; S1,7BP, sedoheptulose-1,7-bisphosphate; S7P, sedoheptulose-7-phosphate; TCA, tricarboxylic acid; X5P, xylulose-5-phosphate.

and/or quantum efficiency (Ducat et al., 2012; Oliver et al., 2013). For example, expression of transporters that allow export of a number of bioproducts in cyanobacteria can lead to increases in photosynthetic activity relative to a parental line that lacks the production pathway. Specifically, some degree of enhanced photosynthetic activity has been in diverse cyanobacterial strains engineered to export sucrose (Ducat et al., 2012), isobuteraldehyde (Li et al., 2014), 2,3-butanediol (Oliver et al., 2013), or ethylene (Ungerer et al., 2012) (**Figure 3**; red text and arrows).

Our group has recently reported a more comprehensive analysis on the photosynthetic effects that occur following activation of a heterologous carbon sink (i.e., sucrose production and secretion) in *S. elongatus* PCC 7942 (Abramson et al., 2016). In this analysis, we found that photosystem II and photosystem I activities were significantly increased within hours of activating the heterologous sucrose secretion pathway. The quantum efficiency of photosystem II transiently increased following sucrose export and photosystem I activity became less constrained by acceptor-side limitations (i.e., the ability of electron carriers such as ferredoxin to remove the excited electrons generated at the reactive chlorophyll pair of photosystem I; Abramson et al., 2016) and total CO_2 fixation rates increased (Ducat et al., 2012). Taken together, these results suggest that overall electron flux through the ETC is enhanced following activation of a heterologous export pathway, suggesting that the endogenous

metabolism of *S. elongatus* PCC 7942 can be insufficient to completely utilize the products of the light reactions under standard laboratory conditions. Similar studies have shown that heterologous electron sinks also have potential to enhance photosynthesis. When a mammalian cytochrome P450, CYP1A1, was expressed in *Synechococcus* PCC 7002 it was able to utilize reductant from the ETC to catalyze a desired monooxygenation reaction and introduction of this pathway was also associated with an improved photosynthetic efficiency and increased electron flow rate by up to ∼30% (Berepiki et al., 2016).

From the above discussion, it appears that at least slower-growing strains of cyanobacteria may be limited by their capacity to utilize products of the CBB cycle. However, it remains unknown if the fastest–growing strains of cyanobacteria (e.g., *S. elongatus* UTEX 2973 and *S. elongatus* PCC 11801) will also exhibit similar increases in photosynthetic flux when engineered to export bioproducts. Current evidence suggests that a greater flux of carbon is allocated away from storage products (e.g., glycogen) and instead invested in cell growth and light-harvesting/carbon-fixation machinery in fast-growing cyanobacteria (Mueller et al., 2017; Zhang et al., 2017; Jaiswal et al., 2018). It is possible that strains with naturally high growth rates will experience less photosynthetic limitations due to over-accumulation of CBB end products, and therefore not exhibit similar enhancements in photosynthesis upon activation of a heterologous metabolic sink.

PREDICTING AND ENGINEERING CYANOBACTERIAL METABOLISM VIA GENOME-SCALE MODELS

Genome-scale models (GSMs) are large-scale stoichiometric models that describe metabolic pathways as stoichiometric coefficients and mass balances of participating metabolites, and are simulated using numerical optimization (Kim et al., 2016). The ultimate goal of these metabolic reconstructions is to give a comprehensive explanation of all biochemical conversions taking place within a living cell or organism, including transport and non-enzymatic reactions (Steuer et al., 2012). Due to ease of implementation and relatively high predictive power, modeling approaches have been used as tools to assist metabolic engineering and production strain development (O'Brien et al., 2015). Computational modeling methods based on the use of GSMs complement experimental research and give a powerful tool to rapidly generate and prioritize testable hypotheses that can be used to guide subsequent experimentation (Dreyfuss et al., 2013). On the other hand, GSMs also provide potential mechanistic explanations for the results obtained in the laboratory (O'Brien et al., 2015).

Using the genome sequence of an organism, a draft GSM can be relatively easily compiled (Thiele and Palsson, 2010). Standard procedures have been detailed in the literature, to generate high-quality GSM reconstructions (Feist et al., 2009; Thiele and Palsson, 2010). Moreover, many steps of the reconstruction process have been successfully automated by several software programs (Hamilton and Reed, 2014). This progress has allowed the fast reconstruction of draft GSMs of multiple species (Kim et al., 2016) (**Figure 4**, Central panel). However, some manual evaluation and curation is required to ensure a high-quality reconstruction (Hamilton and Reed, 2014).

An Overview of Metabolic Models Developed for Cyanobacteria

A number of GSMs of phototrophic organisms have been published in the last decade, but they are still underrepresented in comparison to heterotrophic microorganisms (Gudmundsson et al., 2017). At the time of this writing, genomes for >270 cyanobacterial species have been sequenced (Fujisawa et al., 2017). However, a limited number of cyanobacterial reconstructions have been made available and preliminary models have only been refined for only a handful of species (**Table 2**). The earliest metabolic reconstructions were based on biochemical data, focusing mainly on central carbon metabolism and photosynthetic pathways (**Figure 4**, left panel). However, the recent advances in genome sequencing have allowed the generation of metabolic models at the genome scale. A general naming convention used to describe *in silico* models has been proposed with the form "*i*XXxxx"; where "*i*" refers to an *in silico* model, "XX" are the initials of the person who developed the model and "xxx" the number of genes included in the model (Reed et al., 2003). However, many of the published GSMs do not follow this rule. Moreover, most of them lack the universal metabolite and reaction conventions in the network

model. This lack of consistency impedes direct information extraction between different models; maintenance and adherence to a universal standard would greatly improve the updating and curation of GSMs. The inconsistent nomenclature is a key bottleneck in the speed of reconstruction of new high quality GSMs (Kumar et al., 2012).

Among cyanobacteria, *Synechocystis* PCC 6803, is the most extensively studied and well-modeled cyanobacterium, with a total of 12 GSMs (**Table 2**). Network reconstruction is an iterative process and the most robust models are generally created by gradually expanding and updating a prior draft when new data and tools are available (Gudmundsson et al., 2017). This is the case of GSMs: *i*Syn811 (Montagud et al., 2011), *i*Syn731 (Saha et al., 2012), *i*HK677 (Knoop et al., 2013), and *im*Syn716 (Gopalakrishnan et al., 2018). While distinct cyanobacterial species have unique characteristics, there are common pathways and core carbon metabolic processes that tend to be described in most GSM models. These common pathways can be schematically decomposed as: photosynthesis (to produce ATP and NADPH and fix inorganic carbon), glycolysis (to produce ATP, NADH, and generate precursor metabolites), the citric acid cycle (to produce other precursor metabolites), oxidative phosphorylation (to produce ATP), the pentose phosphate pathway (to produce reducing equivalents and precursor metabolites), carbohydrate synthesis and triacylglycerol synthesis (to build cell walls and store carbon), and inorganic nitrogen assimilation (to produce proteins, DNA, RNA, chlorophyll, and other secondary metabolites using the relevant pertinent precursor molecules) (Baroukh et al., 2015).

Unlike obligate heterotrophic microorganisms, cyanobacteria can utilize light and inorganic carbon, in addition to organic compounds, for the generation of energy and metabolic precursors. The complex mechanisms of light capture makes it difficult to represent it as a simple biochemical reaction in a metabolic networks (Baroukh et al., 2015). Accurate modeling of phototrophic metabolism requires a new level of detail, including modeling the process of light harvesting and electron transport through a variety of possible pathways. Some advances in this area have been achieved in recent years. Nogales et al. proposed a modeling approach for photosynthetic electron flow pathways in detail in *Synechocystis* PCC 6803, including many cyclic electron flow and accessory pathways, enabling the study of photosynthetic processes at the system level (Nogales et al., 2012). In an updated representation of the GSM of *Synechocystis* PCC 6803, the role of photorespiration in cellular growth and the peculiarities of photosynthetic reactions such as light-dependent oxidative stress were integrated (Knoop et al., 2013). Most recently, the development of an approach to incorporate light absorption that factors in the effects of cell shading was achieved in *S. elongatus* PCC 7942 by modeling light as a metabolite (Broddrick et al., 2016). On the other hand, Qian et al. incorporated a light-dependent PSI/PSII electron transport rate algorithm in a GSM of *Synechococcus* PCC 7002, which allowed simulations of photoautotrophic growth at different light intensities (Qian et al., 2017). In a new GSM of *Synechocystis* PCC 6803, an unconstrained photo-respiratory reaction and a mechanism to
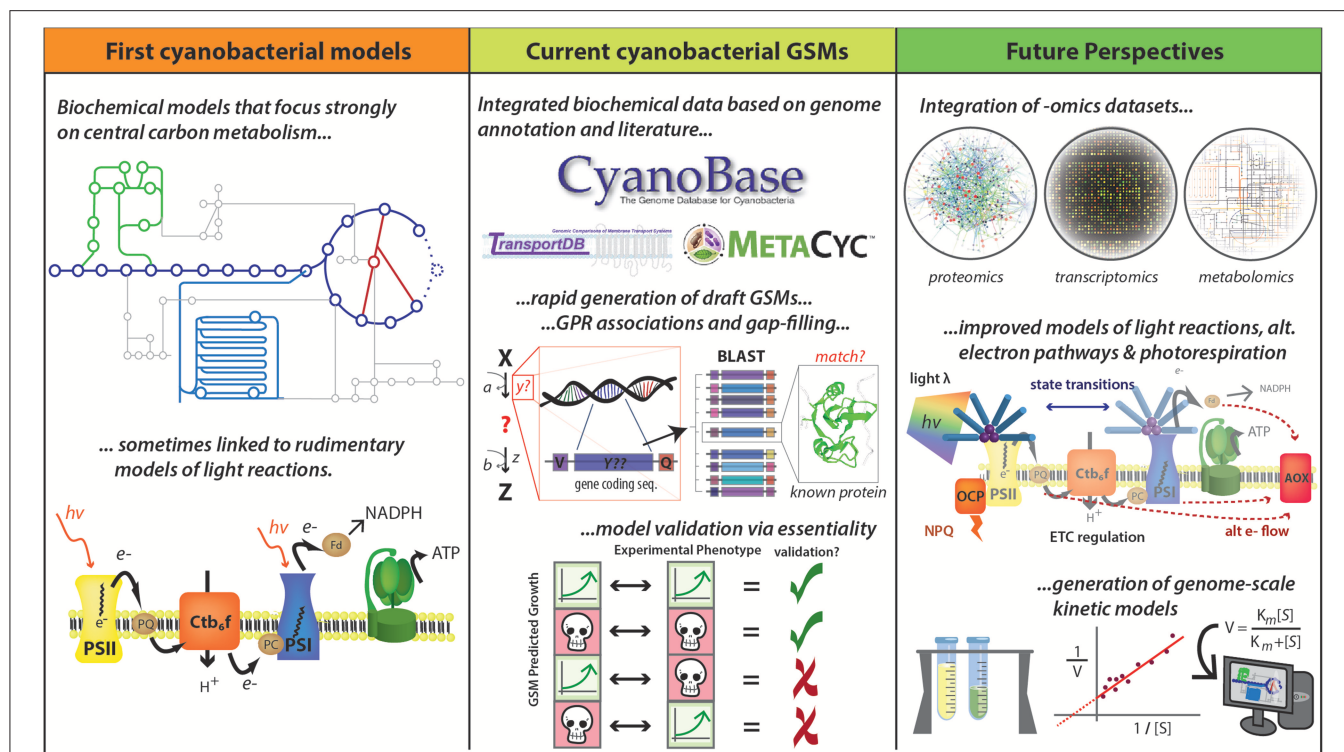
FIGURE 4 | Development of cyanobacterial GSMs. Early cyanobacterial GSMs were developed based on biochemical data and were focused mainly on central carbon metabolism and linked to basic representations of photosynthetic pathways (Cogne et al., 2003; Shastri and Morgan, 2005; Hong and Lee, 2007). Numerous advances have accelerated the reconstruction of GSMs, including improved quantity and quality of data described in public databases (e.g., CyanoBase, TransportDB, KEGG, and MetaCyc) and the development of tools for rapid generation of drafts and for gap-filling. The quality of GSMs has been substantially improved with the inclusion of gene-protein-reaction (GPR) associations, which has been assessed by essential gene prediction based on the availability of experimental data. This methodology allows identify four different kind of predictions: true positives, true negatives, false positives, and false negatives. True positives and true negatives are obtained when both prediction and experimental data indicate a gene is essential and non-essential, respectively (the first and the second case). False positives occur when the model says a gene is essential, but experiments suggest otherwise (the third case). False negatives are generated when a gene is predicted to be non-essential but in reality, it is essential (the fourth case). However, new improvements could further improve GSM predictive ability, including the integration of—omics datasets and the improvement of photosynthesis and photorespiration modeling. A long-term goal would be the development of genome-scale kinetic models, which might be expected to provide more accurate metabolic predictions. Logos for MetaCyc, TransportDB, and CyanoBase are used with permission.

account for changes in energy absorption from light at different wavelengths have been developed (Joshi et al., 2017). To facilitate a better understanding of respiratory and photosynthetic interactions, these authors included features to model known molecular mechanisms of the photosynthetic network around the thylakoid membrane.

A canonical test useful to benchmark and validate the accuracy of GSMs is to examine their capacity to predict essential genes of the metabolic network (Becker and Palsson, 2008) (**Figure 4**, central panel).Gene essentiality prediction has been successfully used in several bacteria such as *E. coli* (Suthers et al., 2009; Orth et al., 2011) and *Pseudomonas putida* (Nogales et al., 2008, 2017). The quality of two recent cyanobacterial GSMs has been assessing using gene essentiality datasets: *i*JB785 of *S. elongatus* PCC 7942 (Broddrick et al., 2016) and *i*SynCJ816 of *Synechocystis* PCC 6803 (Joshi et al., 2017). In the first case, 78% of genes were correctly assigned as either essential or non-essential based on data of essentiality *in vivo* obtained from previous dense-transposon mutagenesis experiments (Rubin et al., 2015). In the second case, the new GSM of *Synechocystis* PCC 6803 was able to predict gene deletions with 77% accuracy, based on a qualitative growth

comparison of 167 gene-deletion mutants with experimental studies obtained from online databases and a detailed literature search (Joshi et al., 2017).

There are several approaches to improve the accuracy of GSMs. First, the use of data derived from high-throughput growth phenotyping experiments (e.g., knockout mutant strains grown in various media conditions) is very useful for validating and refining metabolic network reconstructions (Gawand et al., 2013). Secondly, integrating of transcriptional regulation has proved to be a vital alternative to build improved models and to investigate the capabilities of reconstructed metabolic networks (Vivek-Ananth and Samal, 2016). Finally, the integration of multi-omics datasets have been used in a number of instances to improve the accuracy of a GSM (Kim and Reed, 2014).

Tools to Improve GSM Quality and Reconstruction

Manual curation is one the most important steps after the generation of the initial draft of a GSM. The process of manually

TABLE 2 | GSMs described for cyanobacterial strains.

Cyanobacteria	GSM name	Genes	Reactions	Metabolites	% of annotated genes	References
Arthrospira platensis PCC 8005	n/a	n/a	121	134	n/a	Cogne et al., 2003
Arthrospira platensis C1	*i*AK692	692	875	837	11	Klanchui et al., 2012
Arthrospira platensis NIES-39	n/a	620	746	673	24	Yoshikawa et al., 2015
Synechocystis sp. PCC 6803	n/a	n/a	70	46	n/a	Shastri and Morgan, 2005
	n/a	86	56	48	2.7	Hong and Lee, 2007
	n/a	633	831	704	20	Fu, 2009
	n/a	337	380	291	11	Knoop et al., 2010
	*i*Syn669	669	882	790	21	Montagud et al., 2010
	*i*Syn811	811	956	911	26	Montagud et al., 2011
	n/a	393	493	465	12	Yoshikawa et al., 2011
	*i*JN678	678	863	795	21	Nogales et al., 2012
	*i*Syn731	731	1,156	996	23	Saha et al., 2012
	*i*HK677	677	759	601	21	Knoop et al., 2013
	*i*SynCJ816	816	1,045	925	25.7	Joshi et al., 2017
	*im*Syn716	n/a	729	679	n/a	Gopalakrishnan et al., 2018
Cyanothece sp. ATCC 51142	*i*Cce806	806	667	587	16	Vu et al., 2012
	*i*Cyt773	773	946	811	15	Saha et al., 2012
Cyanothece sp. PCC 7424	*i*Cyc792	792	1.242	1.107	14	Mueller et al., 2013
Cyanothece sp. PCC 7425	*i*Cyn731	731	1.306	1.160	14	Mueller et al., 2013
Cyanothece sp. PCC 7822	*i*Cyj826	826	1.258	1.110	14	Mueller et al., 2013
Cyanothece sp. PCC 8801	*i*Cyp752	752	1.172	994	17	Mueller et al., 2013
Cyanothece sp. PCC 8802	*i*Cyh755	755	1.161	973	17	Mueller et al., 2013
Nostoc sp. PCC 7120	n/a	n/a	804	777	n/a	Malatinszky et al., 2017
Synechococcus elongatus PCC 7942	*i*Syf715	715	851	838	26	Triana et al., 2014
	*i*JB785	785	850	768	29	Broddrick et al., 2016
Synechococcus elongatus UTEX 2973	*i*Syu683	687	1,178	1,028	26	Mueller et al., 2017
Synechococcus sp. PCC 7002	*i*Syp611	611	552	542	19	Hamilton and Reed, 2012
	n/a	728	742	684	23	Hamilton and Reed, 2012
	*i*Syp708	708	602	581	22.2	Vu et al., 2013
	*i*Syp821	821	744	777	26	Qian et al., 2017

reconstructing GSMs is complex and requires arduous and time-consuming curation, without which the model quality remains low (Machado et al., 2018). The manual review is required to reflect the true metabolic capabilities of the target organism (Gudmundsson et al., 2017) and relies heavily on experimental, organism-specific information (Thiele and Palsson, 2010). This step involves the examination of the mass and charge balance of individual reactions, gene associations of reactions and reaction directionality (Gudmundsson et al., 2017). This task is typically addressed using the information stored in many public databases such as CyanoBase (Fujisawa et al., 2017), CyanoEXpress (Hernandez-Prieto and Futschik, 2012), KEGG (Kanehisa and Goto, 2000), MetaCyc (Caspi et al., 2006), SEED (Overbeek et al., 2014), BRENDA (Schomburg et al., 2002), and TransportDB (Ren et al., 2004) (**Figure 4**, central panel). To solve this bottleneck, several tools for rapid automated reconstruction

of GSMs are currently available, each offering different degrees of trade-off between automation and human intervention. Some of these tools haven been previously reviewed (Faria et al., 2018; Machado et al., 2018), but only a small number of them have been used in cyanobacteria.

In many cyanobacteria, a high percentage of proteins are annotated as "unknown function" or "hypothetical protein," e.g., in *Synechocystis* PCC 6803 up to 60% (Lv et al., 2015). Gene-protein-reaction (GPR) relationships define the association between genes, metabolic enzymes, and the biochemical transformations that they perform (Thomas et al., 2014) (**Figure 4**, central panel). GPR determines the set of metabolic reactions encoded in the genome and provides a mechanistic link between genotype and phenotype (Machado et al., 2016). The inclusion of GPRs within GSMs is essential to improve the quality of GSMs, with the aim of improving their phenotypic predictions

(Krishnakumar et al., 2013). New methodologies have been applied to automate the process of adding GPR associations to cyanobacterial GSMs, such as SHARP (Systematic, Homology-based Automated Re-annotation for Prokaryotes). SHARP is a novel PSI-BLAST-based methodology GPR association of the metabolic enzymes involved in prokaryotes, which has been used in cyanobacteria (Krishnakumar et al., 2013). These authors were able to predict 3,781 new GPR associations for the 10 prokaryotes considered, eight of which were cyanobacterial species. These new GPR associations allowed them to annotate gaps in metabolic networks, and to discover several pathways that may be active, thereby providing new directions for metabolic engineering of cyanobacteria.

Another bottleneck during GSM reconstruction is the problem of gap-filling. Metabolic reconstruction via functional genomics (MIRAGE) has been developed specifically to address this problem (Vitkin and Shlomi, 2012), and searches for reactions that are missing from reconstructions purely based on enzyme homology, but where functional genomic data suggests the reactions are present. MIRAGE performance was directly tested by applying it to the reconstruction of a network model for *Synechocystis* PCC 6803, and was then successfully validated against an existing, manually-curated model for this cyanobacterium. Vitkin and Shlomi compared the reconstructed network model for *Synechocystis* PCC 6803 generated by MIRAGE with the manually curated models of Knoop et al. (2010) and *i*Syn811 (Montagud et al., 2011), and found a predictive precision of 70% in the first case and 37.5% for *i*Syn811. MIRAGE was also applied to reconstruct GSMs for 36 sequenced cyanobacteria, including some model cyanobacteria such as *S. elongatus* PCC 7942 and *Synechococcus* PCC 7002 (Vitkin and Shlomi, 2012). However, it has been demonstrated that this methodology was not useful to develop a GSM for the cyanobacterium *Cyanothece* sp. PCC 7424 (Mueller et al., 2013). For example, MIRAGE analysis generated a GSM that contained menaquinone and ubiquinone, compounds shown to not exist within *Cyanothece* PCC 7424 (Collins and Jones, 1981). Moreover, the biomass composition of this GSM did not contain some metabolites that are known to be important components of this species, including lipids, pigments, and cyanophycin. Automated model development tools are useful when there is enough information in the training set of models that they can extract to develop the new one (Mueller et al., 2013). However, some cyanobacterial metabolites are unique, therefore they cannot be detected using this methodology, and require manual annotation.

Using Flux Balance Analysis for Systems Metabolic Engineering

The mathematical approach most widely used for studying the characteristics and capabilities of large-scale biochemical networks is flux balance analysis (FBA) (Orth et al., 2010). It depends on an assumption of steady-state growth and mass balance (influx equals efflux) (Orth et al., 2010; Qian et al., 2017). FBA calculates the flow of metabolites through a metabolic network, thereby making it possible to predict the growth rate

of an organism or the rate of production of a biotechnologically important metabolite (Orth et al., 2010). In FBA simulations, the biomass function is normally used to simulate cellular growth, because it is composed of all necessary compounds needed to create a new cell including DNA, amino acids, lipids, and polysaccharides (O'Brien et al., 2015). However, when FBA is used to simulate the expected rates of production for a metabolite of interest, instead of maximizing the growth rate, it is necessary to maximize the production of this metabolite by optimizing the output flux of the reaction that produces it (Lewis et al., 2012).

FBA and related constraint-based methods can be used to predict the optimal set of gene knockout and overexpression targets to increase the ability of one organism to produce a chemical of interest. With this aim, FBA has been successfully used in GSMs of cyanobacteria to increase the production of several compounds, as summarized in **Table 3**. This method is a useful alternative when GSMs have little or no kinetic data available for their metabolic enzymes (Orth et al., 2010). Readers may also find more details about FBA results getting in cyanobacteria in this reference (Gudmundsson et al., 2017) and in the references detailed in **Table 3**.

Algorithms Used *in silico* to Improve Cyanobacterial Productivity

Numerous constraint-based methods of GSMs are available to identify the phenotypic properties of an organism and to validate hypothesis-driven engineering of cellular functions toward specific objectives (Kim et al., 2015). In addition to the increasing refinement of higher-quality GSM reconstructions, other computational algorithms have been developed that have expanded the scope, accuracy, and applications for GSMs. Some of these algorithms have been used in cyanobacteria to predict promising gene deletion targets for increased production of target compounds, such as OptGene, minimization of metabolic adjustment (MOMA), OptKnock, and OptORF (Vu et al., 2013; Shabestary and Hudson, 2016). On the other hand, other algorithms have been developed to identify the possible interventions (e.g., up or downregulation of gene expression) that lead to overproduction of a target metabolite (e.g., OptForce) (Shabestary and Hudson, 2016; Lin et al., 2017).

The OptGene algorithm is based on the random implementation of reaction knockouts with the aim to create optimal knockout sets. This algorithm provides the advantage of a high computational speed, enabling solutions to be efficiently reached even for problems of larger size. Additionally, OptGene can optimize for non-linear objective functions, such as the productivity of one specific compound (Patil et al., 2005). This algorithm was recently used to identify knockouts in *Synechocystis* PCC 6803 that improve production of fermentation, fatty-acid, and terpene-derived biofuels (Shabestary and Hudson, 2016). The same authors also used MOMA, an algorithm that predicts the optimal flux distribution of altered metabolism that would require the smallest change from that of wild-type metabolism (Segrè et al., 2002). However, a primary drawback of the MOMA algorithm is that it does not assume optimality of growth or any other metabolic functions

TABLE 3 | Applications of FBA to maximize the production of different compounds in cyanobacteria.

	Compound	Organism	Reference
Alcohols	1-propanol	*Synechocystis* sp. PCC 6803	Yoshikawa et al., 2011
	Butanol	*Synechocystis* sp. PCC 6803	Kämäräinen et al., 2012
	Several alcohols	*Synechococcus* sp. PCC 7002	Vu et al., 2013
	Ethanol, isobutanol, 3-methyl-1-butanol, 2-methyl-1-butanol and propanol	*Synechocystis* sp. PCC 6803	Mohammadi et al., 2016
	Butanol, ethanol	*Synechococcus* sp. PCC 7002	Hendry et al., 2016
	1,3-propanediol, glycerol	*Synechococcus elongatus* PCC 7942	Hirokawa et al., 2017
	Ethanol	*Synechocystis* sp. PCC 6803	Yoshikawa et al., 2017
Bulk chemicals	Isoprene	*Synechocystis* sp. PCC 6803	Saha et al., 2012
	Ethylene	*Synechocystis* sp. PCC 6803	Zavrel et al., 2016
	Succinic acid	*Synechocystis* sp. PCC 6803	Shirai et al., 2016
	Terpenes	*Synechocystis* sp. PCC 6803	Englund et al., 2018
Gases	Hydrogen	*Synechocystis* sp. PCC 6803	Montagud et al., 2010
	Hydrogen	*Cyanothece* sp. ATCC 51142	Saha et al., 2012
	Methane	*Synechococcus* sp. PCC 7002	Comer et al., 2017

(Raman and Chandra, 2009). Despite of these issues, it has been successfully applied in *E. coli* to predict the metabolic phenotype of gene knockouts and to find solutions not detected with FBA (Segrè et al., 2002). Shabestary et al. used MOMA to find knockout strategies that could increase biofuel productivity in *Synechocystis* PCC 6803 (Shabestary and Hudson, 2016). MOMA has been used in *Synechococcus* PCC 7002 to identify knockout mutants to improve chemical production under photoautotrophic and/or dark anoxic conditions (Vu et al., 2013). Shabestary et al. used a third algorithm to find a set of knockouts that led to coupling between biofuel and growth (Shabestary and Hudson, 2016). OptKnock is a powerful algorithm that identifies and subsequently removes metabolic reactions that are capable of coupling cellular growth with chemical production (Burgard et al., 2003). OptORF is an algorithm very similar to OptKnock, but it identifies gene deletions (instead of reaction deletions) and regulatory changes needed to couple growth and chemical production (Kim and Reed, 2010). Nevertheless, the applicability of this algorithm relies heavily on the availability of integrated metabolic and regulatory models, which is not always possible, especially for cyanobacteria (Maia et al., 2016). It was used in *Synechococcus* PCC 7002 to predict metabolic engineering strategies that improve production of both native and non-native chemicals (Vu et al., 2013).

OptForce is an algorithm that identifies all possible engineering interventions by classifying reactions in the metabolic model depending upon whether their flux values must increase, decrease or become equal to zero to meet a pre-specified overproduction target (Ranganathan et al., 2010). This algorithm has been recently extended to include kinetic descriptions for some of the reaction steps (Chowdhury et al., 2014). The new algorithm, k-OptForce, can only be directly applied when kinetic models of cellular pathways are available and unfortunately detailed kinetic models are extremely limited in cyanobacteria (Steuer et al., 2012). OptForce algorithm was applied in

Synechocystis PCC 6803 to find a set of interventions to couple growth to 1-octanol and limonene production (Shabestary and Hudson, 2016). In the first case, OptForce predicted several reactions in alternative electron flow and required the upregulation of the ferredoxin:NADPH oxidoreductase reaction. In the case of limonene production-growth couple, a significant number of reaction knockouts were required, including some components of the cyclic electron flow and alternative electron flow. In a recent report, Lin et al. used this algorithm in *Synechocystis* PCC 6803 to enhance isoprenoid production (Lin et al., 2017). OptForce predicted several interventions: the up-regulation of two pentose phosphate pathway genes, ribose 5-phosphate isomerase and ribulose 5-phosphate 3-epimerase, and the overexpression of a geranyl diphosphate synthase involved in the limonene biosynthetic pathway. The optimized strain with these modifications demonstrated a 2.3-fold improvement in productivity.

CONCLUSIONS AND PERSPECTIVES

Cyanobacteria possess desirable characteristics as a chassis for biotechnological production and have demonstrated capacities to produce high-value bioproducts. However, tools from heterotrophs are not always transferrable to these microorganisms, leading to the delay in their advancement as industrial hosts. Accelerating the development of sophisticated genomic tools is likely to greatly assist in making scaled cyanobacterial bioproduction a commercially-viable endeavor. Many recent efforts have been focused on the characterization of new parts and development of libraries and standardized modular parts in cyanobacteria. However, characterization of such parts could likely be significantly accelerated by the development of robust and modular expression libraries, like those that have been established in *E. coli* and yeast. Moreover, the development of well-defined genetic libraries such as

genomic, expression, and knockout libraries might facilitate a better understanding of complex phenotypes in cyanobacteria (Ramey et al., 2015). On the other hand, the use of markerless modification systems as well as CRISPR/Cas based technologies could expand and improve the efficiency of genome editing in cyanobacteria. The iterative modification of the genome would allow the assembly of long and complex metabolic pathways, a capacity that is essential for many of the more advanced and ambitious metabolic engineering projects (Ungerer and Pakrasi, 2016; Behler et al., 2018). The development and/or optimization of high-throughput approaches that allow the introduction of several genetic modifications in a single step are needed (Ramey et al., 2015). In *E. coli*, techniques such as Multiplex Automated Genome Engineering (MAGE) (Wang et al., 2009) and trackable multiplex recombineering (Warner et al., 2010) have been successfully used to perform multiple genome engineering modifications in tandem. While some technical challenges would be complicate adoption of such techniques in cyanobacteria, it is notable that MAGE has been successfully applied in combination with CRISPR/Cas technology to engineer polyploid hosts, such as industrial yeast (Lian et al., 2018), and other organisms with more complex genomes, including plants (Sakuma et al., 2014; Hashimoto et al., 2018). The establishment of these approaches in higher organisms, together with the recent development of CRISPR/Cas systems in cyanobacteria, open up the possibility of the development of high-throughput genome engineering of cyanobacterial strains.

The increasing availability of GSMs for different cyanobacterial species have allowed for reiterative design-build-test cycles whereby the predictions from *in silico* models can be validated and improved from experimental outcomes *in vivo*. Recent efforts have focused on the establishment of GPRs to improve the quality of GSMs, and thus, their phenotypic predictions. However, there are still some challenges that need to be addressed to ensure the accuracy of GSM predictions. Improving the scope and accuracy of existing GSMs will likely involve the integration of -omic datasets in to improve *in silico* representation of metabolism and identify additional biological unknowns (**Figure 4**, right panel). In addition, it is necessary to improve the modeling of photosynthesis, photorespiration, and photodamage (**Figure 4**, right panel). The correct modeling of photosynthesis and respiratory electron transfer processes might provide some insight into their physiological role in cyanobacteria. Another critical element that is currently absent in GSMs is the incorporation of regulatory pathways that can dynamically alter enzymatic activity and pathway flux (**Figure 4**, right panel), although in many cases information about regulatory functions is poorly understood in cyanobacteria.

Although existing GSMs have been successfully applied to inform genetic engineering, it is clear that future efforts are being redirected to genome-scale kinetic models, which have still-greater promise. These new models overcome many of GSM shortcomings, such as the lack of representation of metabolite concentrations and enzymatic regulation, which are necessary for a complete physiologically relevant model. Moreover, they enable dynamic analysis of biological systems for enhanced *in silico* hypothesis generation (Srinivasan et al., 2015).

Most efforts in the genetic and metabolic engineering of cyanobacteria for the production of compounds with biotechnological applications have been focused on increasing product yield under laboratory conditions. A limited number of studies has been performed with the aim to optimize cyanobacterial cells to be more compatible with downstream scaled cultivation and processing (Singh et al., 2016; Johnson et al., 2017). The successful utilization of cyanobacterial species for industrial production depends on the development of accurate large-scale cultivation systems. However, most academic researchers lack access to large-scale production systems that are necessary to evaluate the potential of engineered strains under more realistic environmental conditions (Schoepp et al., 2014) and such information is important for the recursive design-build-test process for strain engineering. More communication between academic researchers and industrial partners could assist in the development of accurate equipment to mimic outdoor production conditions at the laboratory scale (as developed in Lucker et al., 2014). In addition to improving product yields, future research should also be addressed to the development of efficient and cost-effective photosynthetic bioreactors (Lau et al., 2015), as well as technologies to harvest the end products (Knoot et al., 2018) to minimize operation costs. Moreover, the high requirements of water and nutrients are other major challenges for economic profitability of large-scale cyanobacterial cultures (Pathak et al., 2018). Further investigation of the capacity of industrial strains of cyanobacteria to grow and remediate wastewater streams might help to alleviate some of these scalability concerns.

AUTHOR CONTRIBUTIONS

MS-M, AS, and DD contributed to the writing and editing of the manuscript and generation of figures and tables.

FUNDING

REFERENCES

Abernathy, M. H., Yu, J., Ma, F., Liberton, M., Ungerer, J., Hollinshead, W. D., et al. (2017). Deciphering cyanobacterial phenotypes for fast photoautotrophic growth via isotopically nonstationary metabolic flux analysis. *Biotechnol. Biofuels* 10:273. doi: 10.1186/s13068-017-0958-y

Abramson, B. W., Kachel, B., Kramer, D. M., and Ducat, D. C. (2016). Increased photochemical efficiency in cyanobacteria via an engineered sucrose sink. *Plant Cell Physiol.* 57, 2451–2460. doi: 10.1093/pcp/pcw169

Acién, F. G., Molina, E., Reis, A., Torzillo, G., Zittelli, G. C., Sepúlveda, C., et al. (2017). "Photobioreactors for the production of microalgae," in *Microalgae-Based Biofuels and Bioproducts,* eds C. Gonzalez-Fernandez and R. Muñoz (Kidlington: Woodhead Publishing), 1–44.

Adams, W. W. III., Muller, O., Cohu, C. M., and Demmig-Adams, B. (2013). May photoinhibition be a consequence, rather than a cause, of limited plant productivity? *Photosyn. Res.* 117, 31–44. doi: 10.1007/s11120-013-9849-7

Albers, S. C., Gallegos, V. A., and Peebles, C. A. (2015). Engineering of genetic control tools in *Synechocystis* sp. PCC 6803 using rational design techniques. *J. Biotechnol.* 216, 36–46. doi: 10.1016/j.jbiotec.2015.09.042

Angermayr, S. A., Gorchs Rovira, A., and Hellingwerf, K. J. (2015). Metabolic engineering of cyanobacteria for the synthesis of commodity products. *Trends Biotechnol.* 33, 352–361. doi: 10.1016/j.tibtech.2015.03.009

Armshaw, P., Carey, D., Sheahan, C., and Pembroke, J. T. (2015). Utilising the native plasmid, pCA2.4, from the cyanobacterium *Synechocystis* sp. strain PCC6803 as a cloning site for enhanced product production. *Biotechnol. Biofuels* 8:201. doi: 10.1186/s13068-015-0385-x

Badger, M. R., Hanson, D., and Price, G. D. (2002). Evolution and diversity of CO_2 concentrating mechanisms in cyanobacteria. *Funct. Plant Biol.* 29, 161–173. doi: 10.1071/PP01213

Baroukh, C., Muñoz-Tamayo, R., Steyer, J. P., and Bernard, O. (2015). A state of the art of metabolic networks of unicellular microalgae and cyanobacteria for biofuel production. *Metab. Eng.* 30, 49–60. doi: 10.1016/j.ymben.2015.03.019

Becker, S. A., and Palsson, B. O. (2008). Three factors underlying incorrect *in silico* predictions of essential metabolic genes. *BMC Syst. Biol.* 2:14. doi: 10.1186/1752-0509-2-14

Begemann, M. B., Zess, E. K., Walters, E. M., Schmitt, E. F., Markley, A. L., and Pfleger, B. F. (2013). An organic acid based counter selection system for cyanobacteria. *PLoS ONE* 8:e76594. doi: 10.1371/journal.pone.0076594

Behler, J., Vijay, D., Hess, W. R., and Akhtar, M. K. (2018). CRISPR-based technologies for metabolic engineering in cyanobacteria. *Trends Biotechnol.* 36, 996–1010. doi: 10.1016/j.tibtech.2018.05.011

Berens, C., and Suess, B. (2015). Riboswitch engineering - making the all-important second and third steps. *Curr. Opin. Biotechnol.* 31, 10–15. doi: 10.1016/j.copbio.2014.07.014

Berepiki, A., Hitchcock, A., Moore, C. M., and Bibby, T. S. (2016). Tapping the unused potential of photosynthesis with a heterologous electron sink. *ACS Synth. Biol.* 5, 1369–1375. doi: 10.1021/acssynbio.6b00100

Berla, B. M., Saha, R., Immethun, C. M., Maranas, C. D., Moon, T. S., and Pakrasi, H. B. (2013). Synthetic biology of cyanobacteria: unique challenges and opportunities. *Front. Microbiol.* 4:246. doi: 10.3389/fmicb.2013.00246

Broddrick, J. T., Rubin, B. E., Welkie, D. G., Du, N., Mih, N., Diamond, S., et al. (2016). Unique attributes of cyanobacterial metabolism revealed by improved genome-scale metabolic modeling and essential gene analysis. *Proc. Natl. Acad. Sci. U.S.A.* 113, E8344–E8353. doi: 10.1073/pnas.1613446113

Burgard, A. P., Pharkya, P., and Maranas, C. D. (2003). Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* 84, 647–657. doi: 10.1002/bit.10803

Cai, F., Axen, S. D., and Kerfeld, C. A. (2013). Evidence for the widespread distribution of CRISPR-Cas system in the Phylum Cyanobacteria. *RNA Biol.* 10, 687–693. doi: 10.4161/rna.24571

Cai, Y., and Wolk, C. P. (1997). *Anabaena* sp. strain PCC 7120 responds to nitrogen deprivation with a cascade-like sequence of transcriptional activations. *J. Bacteriol.* 179, 267–271. doi: 10.1128/jb.179.1.267-271.1997

Callahan, S. M., and Buikema, W. J. (2001). The role of HetN in maintenance of the heterocyst pattern in *Anabaena* sp. PCC 7120. *Mol. Microbiol.* 40, 941–950. doi: 10.1046/j.1365-2958.2001.02437.x

Callura, J. M., Dwyer, D. J., Isaacs, F. J., Cantor, C. R., and Collins, J. J. (2010). Tracking, tuning, and terminating microbial physiology using synthetic riboregulators. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15898–15903. doi: 10.1073/pnas.1009747107

Camsund, D., and Lindblad, P. (2014). Engineered transcriptional systems for cyanobacterial biotechnology. *Front. Bioeng. Biotechnol.* 2:40. doi: 10.3389/fbioe.2014.00040

Cao, Y. Q., Li, Q., Xia, P. F., Wei, L. J., Guo, N., Li, J. W., et al. (2017). AraBAD based toolkit for gene expression and metabolic robustness improvement in *Synechococcus elongatus. Sci. Rep.* 7:18059. doi: 10.1038/s41598-017-17035-4

Carroll, A. L., Case, A. E., Zhang, A., and Atsumi, S. (2018). Metabolic engineering tools in model cyanobacteria. *Metab. Eng.* 50, 47–56. doi: 10.1016/j.ymben.2018.03.014

Caspi, R., Foerster, H., Fulcher, C. A., Hopkinson, R., Ingraham, J., Kaipa, P., et al. (2006). MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.* 34(Database issue), D511–D516. doi: 10.1093/nar/gkj128

Cassier-Chauvat, C., Veaudor, T., and Chauvat, F. (2016). Comparative genomics of DNA recombination and repair in cyanobacteria: biotechnological implications. *Front. Microbiol.* 7:1809. doi: 10.3389/fmicb.2016.01809

Cerniglia, C. E., Gibson, D. T., and Baalen, C. (1979). Algal oxidation of aromatic hydrocarbons: formation of 1-naphthol from naphthalene by *Agmenellum quadruplicatum*, strain PR-6. *Biochem. Biophys. Res. Commun.* 88, 50–58. doi: 10.1016/0006-291X(79)91695-4

Cerniglia, C. E., Gibson, D. T., and Van Baalen, C. (1980). Oxidation of naphthalene by cyanobacteria and microalgae. *J. General Microbiol.* 116, 495–500. doi: 10.1099/00221287-116-2-495

Cheah, Y. E., Albers, S. C., and Peebles, C. A. (2013). A novel counter-selection method for markerless genetic modification in *Synechocystis* sp. PCC 6803. *Biotechnol. Prog.* 29, 23–30. doi: 10.1002/btpr.1661

Chen, A. H., Afonso, B., Silver, P. A., and Savage, D. F. (2012). Spatial and temporal organization of chromosome duplication and segregation in the cyanobacterium *Synechococcus elongatus* PCC 7942. *PLoS ONE* 7:e47837. doi: 10.1371/journal.pone.0047837

Cheng, A. A., and Lu, T. K. (2012). Synthetic biology: an emerging engineering discipline. *Ann. Rev. Biomed. Eng.* 14, 155–178. doi: 10.1146/annurev-bioeng-071811-150118

Chisti, Y. (2013). Constraints to commercialization of algal fuels. *J. Biotechnol.* 167, 201–214. doi: 10.1016/j.jbiotec.2013.07.020

Chowdhury, A., Zomorrodi, A. R., and Maranas, C. D. (2014). k-OptForce: integrating kinetics with flux balance analysis for strain design. *PLoS Comput. Biol.* 10:e1003487. doi: 10.1371/journal.pcbi.1003487

Chungjatupornchai, W., and Fa-Aroonsawat, S. (2014). The *rrnA* promoter as a tool for the improved expression of heterologous genes in cyanobacteria. *Microbiol. Res.* 169, 361–368. doi: 10.1016/j.micres.2013.09.010

Cogne, G., Gros, J. B., and Dussap, C. G. (2003). Identification of a metabolic network structure representative of *Arthrospira* (*Spirulina*) *platensis* metabolism. *Biotechnol. Bioeng.* 84, 667–676. doi: 10.1002/bit.10808

Cohen, S. E., Erb, M. L., Pogliano, J., and Golden, S. S. (2015). Best practices for fluorescence microscopy of the cyanobacterial circadian clock. *Meth. Enzymol.* 551, 211–221. doi: 10.1016/bs.mie.2014.10.014

Cohen, S. E., Erb, M. L., Selimkhanov, J., Dong, G., Hasty, J., Pogliano, J., et al. (2014). Dynamic localization of the cyanobacterial circadian clock proteins. *Curr. Biol.* 24, 1836–1844. doi: 10.1016/j.cub.2014.07.036

Collins, M. D., and Jones, D. (1981). Distribution of isoprenoid quinone structural types in bacteria and their taxonomic implication. *Microbiol. Rev.* 45, 316–354.

Comer, A. D., Long, M. R., Reed, J. L., and Pfleger, B. F. (2017). Flux balance analysis indicates that methane is the lowest cost feedstock for microbial cell factories. *Metabol. Eng. Commun.* 5, 26–33. doi: 10.1016/j.meteno.2017.07.002

Connor, M. R., and Atsumi, S. (2010). Synthetic biology guides biofuel production. *J. Biomed. Biotechnol.* 2010:541698. doi: 10.1155/2010/541698

Decoene, T., De Paepe, B., Maertens, J., Coussement, P., Peters, G., De Maeseneire, S. L., et al. (2018). Standardization in synthetic biology: an engineering discipline coming of age. *Crit. Rev. Biotechnol.* 38, 647–656. doi: 10.1080/07388551.2017.1380600

Demmig-Adams, B., Garab, G., Adams Iii, W., and Govindjee (2014). *Non-Photochemical Quenching and Energy Dissipation in Plants, Algae and Cyanobacteria*. Dordrecht: Springer.

Dismukes, G. C., Carrieri, D., Bennette, N., Ananyev, G. M., and Posewitz, M. C. (2008). Aquatic phototrophs: efficient alternatives to land-based crops for biofuels. *Curr. Opin. Biotechnol.* 19, 235–240. doi: 10.1016/j.copbio.2008.05.007

Dittmann, E., Gugger, M., Sivonen, K., and Fewer, D. P. (2015). Natural product biosynthetic diversity and comparative genomics of the cyanobacteria. *Trends Microbiol.* 23, 642–652. doi: 10.1016/j.tim.2015.07.008

Ditty, J. L., Williams, S. B., and Golden, S. S. (2003). A cyanobacterial circadian timing mechanism. *Annu. Rev. Genet.* 37, 513–543. doi: 10.1146/annurev.genet.37.110801.142716

Dreyfuss, J. M., Zucker, J. D., Hood, H. M., Ocasio, L. R., Sachs, M. S., and Galagan, J. E. (2013). Reconstruction and validation of a genome-scale metabolic model for the filamentous fungus *Neurospora crassa* using FARM. *PLoS Comput. Biol.* 9:e1003126. doi: 10.1371/journal.pcbi.1003126

Ducat, D. C., Avelar-Rivas, J. A., Way, J. C., and Silver, P. A. (2012). Rerouting carbon flux to enhance photosynthetic productivity. *Appl. Environ. Microbiol.* 78, 2660–2668. doi: 10.1128/AEM.07901-11

Ellis, B. E. (1977). Degradation of phenolic compounds by fresh-water algae. *Plant Sci. Lett.* 8, 213–216. doi: 10.1016/0304-4211(77)90183-3

Englund, E., Liang, F., and Lindberg, P. (2016). Evaluation of promoters and ribosome binding sites for biotechnological applications in the unicellular cyanobacterium *Synechocystis* sp. PCC 6803. *Sci. Rep.* 6:36640. doi: 10.1038/srep36640

Englund, E., Shabestary, K., Hudson, E. P., and Lindberg, P. (2018). Systematic overexpression study to find target enzymes enhancing production of terpenes in *Synechocystis* PCC 6803, using isoprene as a model compound. *Metab. Eng.* 49, 164–177. doi: 10.1016/j.ymben.2018.07.004

Faria, J. P., Rocha, M., Rocha, I., and Henry, C. S. (2018). Methods for automated genome-scale metabolic model reconstruction. *Biochem. Soc. Trans.* 46, 931–936. doi: 10.1042/BST20170246

Feist, A. M., Herrgård, M. J., Thiele, I., Reed, J. L., and Palsson BØFeist, B. O. (2009). Reconstruction of biochemical networks in microorganisms. *Nat. Rev. Microbiol.* 7, 129–143. doi: 10.1038/nrmicro1949

Fernández-Piñas, F., Leganés, F., and Wolk, C. P. (2000). Bacterial lux genes as reporters in cyanobacteria. *Method Enzymol.* 305, 513–527. doi: 10.1016/S0076-6879(00)05510-5

Ferreira, E. A., Pacheco, C. C., Pinto, F., Pereira, J., Lamosa, P., Oliveira, P., et al. (2018). Expanding the toolbox for *Synechocystis* sp. PCC 6803: validation of replicative vectors and characterization of a novel set of promoters. *Synthet. Biol.* 3:ysy014. doi: 10.1093/synbio/ysy014

Fu, P. (2009). Genome-scale modeling of *Synechocystis* sp. PCC 6803 and prediction of pathway insertion. *J. Chem. Technol. Biotechnol.* 84, 473–483. doi: 10.1002/jctb.2065

Fu, P. (2013). Grand challenges in synthetic biology to be accomplished. *Front. Bioeng. Biotechnol.* 1:2. doi: 10.3389/fbioe.2013.00002

Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., et al. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 31, 822–826. doi: 10.1038/nbt.2623

Fujisawa, T., Narikawa, R., Maeda, S. I., Watanabe, S., Kanesaki, Y., Kobayashi, K., et al. (2017). CyanoBase: a large-scale update on its 20th anniversary. *Nucleic Acids Res.* 45, D551–D554. doi: 10.1093/nar/gkw1131

Gawand, P., Yang, L., Cluett, W. R., and Mahadevan, R. (2013). Metabolic model refinement using phenotypic microarray data. *Methods Mol. Biol.* 985, 47–59. doi: 10.1007/978-1-62703-299-5_3

Geerts, D., Bovy, A., de Vrieze, G., Borrias, M., and Weisbeek, P. (1995). Inducible expression of heterologous genes targeted to a chromosomal platform in the cyanobacterium *Synechococcus* sp. PCC 7942. *Microbiology* 141(Pt 4), 831–841.

Ghim, C. M., Lee, S. K., Takayama, S., and Mitchell, R. J. (2010). The art of reporter proteins in science: past, present and future applications. *BMB Rep.* 43, 451–460. doi: 10.5483/BMBRep.2010.43.7.451

Gifford, R. M., Thorne, J. H., Hitz, W. D., and Giaquinta, R. T. (1984). Crop productivity and photoassimilate partitioning. *Science* 225, 801–808. doi: 10.1126/science.225.4664.801

Gopalakrishnan, S., Pakrasi, H. B., and Maranas, C. D. (2018). Elucidation of photoautotrophic carbon flux topology in *Synechocystis* PCC 6803 using genome-scale carbon mapping models. *Metab. Eng.* 47, 190–199. doi: 10.1016/j.ymben.2018.03.008

Gordon, G. C., Korosh, T. C., Cameron, J. C., Markley, A. L., Begemann, M. B., and Pfleger, B. F. (2016). CRISPR interference as a titratable, trans-acting regulatory tool for metabolic engineering in the cyanobacterium *Synechococcus* sp. strain PCC 7002. *Metab. Eng.* 38, 170–179. doi: 10.1016/j.ymben.2016.07.007

Griese, M., Lange, C., and Soppa, J. (2011). Ploidy in cyanobacteria. *FEMS Microbiol. Lett.* 323, 124–131. doi: 10.1111/j.1574-6968.2011.02368.x

Grigorieva, G., and Shestakov, S. (1982). Transformation in the cyanobacterium *Synechocystis* sp. 6803. *FEMS Microbiol. Lett.* 13, 367–370. doi: 10.1111/j.1574-6968.1982.tb08289.x

Gudmundsson, S., Agudo, L., and Nogales, J. (2017). "Applications of genome-scale metabolic models of microalgae and cyanobacteria in biotechnology," in *Microalgae-Based Biofuels and Bioproducts,* eds C. Gonzalez-Fernandez and R. Muñoz (Kidlington: Woodhead Publishing), 93–111.

Hamilton, J. J., and Reed, J. L. (2012). Identification of functional differences in metabolic networks using comparative genomics and constraint-based models. *PLoS ONE* 7:e34670. doi: 10.1371/journal.pone.0034670

Hamilton, J. J., and Reed, J. L. (2014). Software platforms to facilitate reconstructing genome-scale metabolic networks. *Environ. Microbiol.* 16, 49–59. doi: 10.1111/1462-2920.12312

Hashimoto, R., Ueta, R., Abe, C., Osakabe, Y., and Osakabe, K. (2018). Efficient multiplex genome editing induces precise, and self-ligated type mutations in tomato plants. *Front. Plant Sci.* 9:916. doi: 10.3389/fpls.2018.00916

Heidorn, T., Camsund, D., Huang, H. H., Lindberg, P., Oliveira, P., Stensjo, K., et al. (2011). Synthetic biology in cyanobacteria engineering and analyzing novel functions. *Method Enzymol.* 497, 539–579. doi: 10.1016/B978-0-12-385075-1.00024-X

Hendry, J. I., Prasannan, C. B., Joshi, A., Dasgupta, S., and Wangikar, P. P. (2016). Metabolic model of *Synechococcus* sp. PCC 7002: prediction of flux distribution and network modification for enhanced biofuel production. *Bioresour. Technol.* 213, 190–197. doi: 10.1016/j.biortech.2016.02.128

Hernandez-Prieto, M. A., and Futschik, M. E. (2012). CyanoEXpress: a web database for exploration and visualisation of the integrated transcriptome of cyanobacterium *Synechocystis* sp. PCC6803. *Bioinformation* 8, 634–638. doi: 10.6026/97320630008634

Herrero, A., Muro-Pastor, A. M., and Flores, E. (2001). Nitrogen control in cyanobacteria. *J. Bacteriol.* 183, 411–425. doi: 10.1128/JB.183.2.411-425.2001

Hirokawa, Y., Matsuo, S., Hamada, H., Matsuda, F., and Hanai, T. (2017). Metabolic engineering of *Synechococcus elongatus* PCC 7942 for improvement of 1,3-propanediol and glycerol production based on *in silico* simulation of metabolic flux distribution. *Microb. Cell Fact.* 16:212. doi: 10.1186/s12934-017-0824-4

Hong, S. J., and Lee, C. G. (2007). Evaluation of central metabolism based on a genomic database of *Synechocystis* PCC 6803. *Biotechnol. Bioprocess Eng.* 12, 165–173. doi: 10.1007/BF03028644

Huang, H. H., Camsund, D., Lindblad, P., and Heidorn, T. (2010). Design and characterization of molecular tools for a synthetic biology approach towards developing cyanobacterial biotechnology. *Nucleic Acids Res.* 38, 2577–2593. doi: 10.1093/nar/gkq164

Huang, H. H., and Lindblad, P. (2013). Wide-dynamic-range promoters engineered for cyanobacteria. *J. Biol. Eng.* 7:10. doi: 10.1186/1754-1611-7-10

Immethun, C. M., DeLorenzo, D. M., Focht, C. M., Gupta, D., Johnson, C. B., and Moon, T. S. (2017). Physical, chemical, and metabolic state sensors expand the synthetic biology toolbox for *Synechocystis* sp. PCC 6803. *Biotechnol. Bioeng.* 114, 1561–1569. doi: 10.1002/bit.26275

Immethun, C. M., Ng, K. M., DeLorenzo, D. M., Waldron-Feinstein, B., Lee, Y. C., and Moon, T. S. (2016). Oxygen-responsive genetic circuits constructed in *Synechocystis* sp. PCC 6803. *Biotechnol. Bioeng.* 113, 433–442. doi: 10.1002/bit.25722

Jaiswal, D., Sengupta, A., Sohoni, S., Sengupta, S., Phadnavis, A. G., Pakrasi, H. B., et al. (2018). Genome features and biochemical characteristics of a robust, fast growing and naturally transformable cyanobacterium *Synechococcus elongatus* PCC 11801 isolated from India. *Sci. Rep.* 8:16632. doi: 10.1038/s41598-018-34872-z

Johnson, T. J., Jahandideh, A., Isaac, I. C., Baldwin, E. L., Muthukumarappan, K., Zhou, R., et al. (2017). Determining the optimal nitrogen source for large-scale cultivation of filamentous cyanobacteria. *J. Appl. Phycol.* 29, 1–13. doi: 10.1007/s10811-016-0923-3

Jones, P. R. (2014). Genetic instability in cyanobacteria - an elephant in the room? *Front. Bioeng. Biotechnol.* 2:12. doi: 10.3389/fbioe.2014.00012

Jordan, A., Chandler, J., MacCready, J. S., Huang, J., Osteryoung, K. W., and Ducat, D. C. (2017). Engineering cyanobacterial cell morphology for enhanced recovery and processing of biomass. *Appl. Environ. Microbiol.* 83:AEM.00053-17. doi: 10.1128/AEM.00053-17

Joshi, C. J., Peebles, C. A. M., and Prasad, A. (2017). Modeling and analysis of flux distribution and bioproduct formation in *Synechocystis* sp. PCC 6803 using a new genome-scale metabolic reconstruction. *Algal Res.* 27, 295–310. doi: 10.1016/j.algal.2017.09.013

Kämäräinen, J., Knoop, H., Stanford, N. J., Guerrero, F., Akhtar, M. K., Aro, E. M., et al. (2012). Physiological tolerance and stoichiometric potential of cyanobacteria for hydrocarbon fuel production. *J. Biotechnol.* 162, 67–74. doi: 10.1016/j.jbiotec.2012.07.193

Kamennaya, N. A., Ahn, S. E., Park, H., Bartal, R., Sasaki, K. A., Holman, H. Y., et al. (2015). Installing extra bicarbonate transporters in the cyanobacterium

Synechocystis sp. PCC6803 enhances biomass production. *Metabol. Eng.* 29, 76–85. doi: 10.1016/j.ymben.2015.03.002

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27

Kanno, M., Carroll, A. L., and Atsumi, S. (2017). Global metabolic rewiring for improved CO_2 fixation and chemical production in cyanobacteria. *Nat. Commun.* 8, 1–11. doi: 10.1038/ncomms14724

Kelly, C. L., Taylor, G. M., Hitchcock, A., Torres-Mendez, A., and Heap, J. T. (2018). A rhamnose-inducible system for precise and temporal control of gene expression in cyanobacteria. *ACS Synth. Biol.* 7, 1056–1066. doi: 10.1021/acssynbio.7b00435

Kim, B., Kim, W. J., Kim, D. I., and Lee, S. Y. (2015). Applications of genome-scale metabolic network model in metabolic engineering. *J. Ind. Microbiol. Biotechnol.* 42, 339–348. doi: 10.1007/s10295-014-1554-9

Kim, H. U., Charusanti, P., Lee, S. Y., and Weber, T. (2016). Metabolic engineering with systems biology tools to optimize production of prokaryotic secondary metabolites. *Nat. Prod. Rep.* 33, 933–941. doi: 10.1039/C6NP00019C

Kim, J., and Reed, J. L. (2010). OptORF: optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Syst. Biol.* 4:53. doi: 10.1186/1752-0509-4-53

Kim, J., and Reed, J. L. (2014). Refining metabolic models and accounting for regulatory effects. *Curr. Opin. Biotechnol.* 29, 34–38. doi: 10.1016/j.copbio.2014.02.009

Kim, W. J., Lee, S. M., Um, Y., Sim, S. J., and Woo, H. M. (2017). Development of SyneBrick vectors as a synthetic biology platform for gene expression in *Synechococcus elongatus* PCC 7942. *Front. Plant Sci.* 8:293. doi: 10.3389/fpls.2017.00293

Klähn, S., Bolay, P., Wright, P. R., Atilho, R. M., Brewer, K. I., Hagemann, M., et al. (2018). A glutamine riboswitch is a key element for the regulation of glutamine synthetase in cyanobacteria. *Nucleic Acids Res.* 46, 10082–10094. doi: 10.1093/nar/gky709

Klanchui, A., Khannapho, C., Phodee, A., Cheevadhanarak, S., and Meechai, A. (2012). *i*AK692: a genome-scale metabolic model of *Spirulina platensis* C1. *BMC Syst. Biol.* 6:71. doi: 10.1186/1752-0509-6-71

Knoop, H., Gründel, M., Zilliges, Y., Lehmann, R., Hoffmann, S., Lockau, W., et al. (2013). Flux balance analysis of cyanobacterial metabolism: the metabolic network of *Synechocystis* sp. PCC 6803. *PLoS Comput. Biol.* 9:e1003081. doi: 10.1371/journal.pcbi.1003081

Knoop, H., Zilliges, Y., Lockau, W., and Steuer, R. (2010). The metabolic network of *Synechocystis* sp. PCC 6803: systemic properties of autotrophic growth. *Plant Physiol.* 154, 410–422. doi: 10.1104/pp.110.157198

Knoot, C. J., Ungerer, J., Wangikar, P. P., and Pakrasi, H. B. (2018). Cyanobacteria: promising biocatalysts for sustainable chemical production. *J. Biol. Chem.* 293, 5044–5052. doi: 10.1074/jbc.R117.815886

Kojima, K., Keta, S., Uesaka, K., Kato, A., Takatani, N., Ihara, K., et al. (2016). A simple method for isolation and construction of markerless cyanobacterial mutants defective in acyl-acyl carrier protein synthetase. *Appl. Microbiol. Biotechnol.* 100, 10107–10113. doi: 10.1007/s00253-016-7850-8

Koksharova, O. A., and Wolk, C. P. (2002). Genetic tools for cyanobacteria. *Appl. Microbiol. Biotechnol.* 58, 123–137. doi: 10.1007/s00253-001-0864-9

Kopka, J., Schmidt, S., Dethloff, F., Pade, N., Berendt, S., Schottkowski, M., et al. (2017). Systems analysis of ethanol production in the genetically engineered cyanobacterium *Synechococcus* sp. PCC 7002. *Biotechnol. Biofuels* 10:56. doi: 10.1186/s13068-017-0741-0

Krapp, A., Quick, W. P., and Stitt, M. (1991). Ribulose-1,5-bisphosphate carboxylase-oxygenase, other Calvin-cycle enzymes, and chlorophyll decrease when glucose is supplied to mature spinach leaves via the transpiration stream. *Planta* 186, 58–69. doi: 10.1007/BF00201498

Krishnakumar, S., Durai, D. A., Wangikar, P. P., and Viswanathan, G. A. (2013). SHARP: genome-scale identification of gene-protein-reaction associations in cyanobacteria. *Photosyn. Res.* 118, 181–190. doi: 10.1007/s11120-013-9910-6

Kumar, A., Suthers, P. F., and Maranas, C. D. (2012). MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases. *BMC Bioinformatics* 13:6. doi: 10.1186/1471-2105-13-6

Kumar, K., Mella-Herrera, R. A., and Golden, J. W. (2010). Cyanobacterial heterocysts. *Cold Spring Harb. Perspect. Biol.* 2:a000315. doi: 10.1101/cshperspect.a000315

Kumar, R. R., and Prasad, S. (2011). Metabolic engineering of bacteria. *Indian J. Microbiol.* 51, 403–409. doi: 10.1007/s12088-011-0172-8

Kunst, F., and Rapoport, G. (1995). Salt stress is an environmental signal affecting degradative enzyme synthesis in *Bacillus subtilis*. *J. Bacteriol.* 177, 2403–2407. doi: 10.1128/jb.177.9.2403-2407.1995

Kuritz, T., and Wolk, C. P. (1995). Use of filamentous cyanobacteria for biodegradation of organic pollutants. *Appl. Environ. Microbiol.* 61:1169.

Lai, M. C., and Lan, E. I. (2015). Advances in metabolic engineering of cyanobacteria for photosynthetic biochemical production. *Metabolites* 5, 636–658. doi: 10.3390/metabo5040636

Landry, B. P., Stöckel, J., and Pakrasi, H. B. (2013). Use of degradation tags to control protein levels in the *Cyanobacterium Synechocystis* sp. strain PCC 6803. *Appl. Environ. Microbiol.* 79, 2833–2835. doi: 10.1128/AEM.03741-12

Lau, N. S., Matsui, M., and Abdullah, A. A. (2015). Cyanobacteria: photoautotrophic microbial factories for the sustainable synthesis of industrial products. *Biomed Res. Int.* 2015:754934. doi: 10.1155/2015/754934

Lee, J. W., Kim, T. Y., Jang, Y. S., Choi, S., and Lee, S. Y. (2011). Systems metabolic engineering for chemicals and materials. *Trends Biotechnol.* 29, 370–378. doi: 10.1016/j.tibtech.2011.04.001

Lewis, N. E., Nagarajan, H., and Palsson, B. O. (2012). Constraining the metabolic genotype-phenotype relationship using a phylogeny of *in silico* methods. *Nat. Rev. Microbiol.* 10, 291–305. doi: 10.1038/nrmicro2737

Li, H., Shen, C. R., Huang, C. H., Sung, L. Y., Wu, M. Y., and Hu, Y. C. (2016). CRISPR-Cas9 for the genome engineering of cyanobacteria and succinate production. *Metab. Eng.* 38, 293–302. doi: 10.1016/j.ymben.2016.09.006

Li, X., Shen, C. R., and Liao, J. C. (2014). Isobutanol production as an alternative metabolic sink to rescue the growth deficiency of the glycogen mutant of *Synechococcus elongatus* PCC 7942. *Photosyn. Res.* 120, 301–310. doi: 10.1007/s11120-014-9987-6

Lian, J., Bao, Z., Hu, S., and Zhao, H. (2018). Engineered CRISPR/Cas9 system for multiplex genome engineering of polyploid industrial yeast strains. *Biotechnol. Bioeng.* 115, 1630–1635. doi: 10.1002/bit.26569

Liang, F., Lindberg, P., and Lindblad, P. (2018). Engineering photoautotrophic carbon fixation for enhanced growth and productivity. *Sustain. Ener. Fuels.* 2, 2583–2600. doi: 10.1039/C8SE00281A

Liang, F., and Lindblad, P. (2016). Effects of overexpressing photosynthetic carbon flux control enzymes in the cyanobacterium *Synechocystis* PCC 6803. *Metab. Eng.* 38, 56–64. doi: 10.1016/j.ymben.2016.06.005

Lin, P. C., Saha, R., Zhang, F., and Pakrasi, H. B. (2017). Metabolic engineering of the pentose phosphate pathway for enhanced limonene production in the cyanobacterium *Synechocystis* sp. PCC 6803. *Sci. Rep.* 7:17503. doi: 10.1038/s41598-017-17831-y

Liu, D., and Pakrasi, H. B. (2018). Exploring native genetic elements as plug-in tools for synthetic biology in the cyanobacterium *Synechocystis* sp. PCC 6803. *Microb. Cell Fact.* 17:48. doi: 10.1186/s12934-018-0897-8

Liu, Y., Tsinoremas, N. F., Johnson, C. H., Lebedeva, N. V., Golden, S. S., Ishiura, M., et al. (1995). Circadian orchestration of gene expression in cyanobacteria. *Genes Dev.* 9, 1469–1478. doi: 10.1101/gad.9.12.1469

Lucker, B. F., Hall, C. C., Zegarac, R., and Kramer, D. M. (2014). The environmental photobioreactor (ePBR): an algal culturing platform for simulating dynamic natural environments. *Algal Res.* 6, 242–249. doi: 10.1016/j.algal.2013.12.007

Ludwig, M., and Bryant, D. A. (2012). *Synechococcus* sp. strain PCC 7002 transcriptome: acclimation to temperature, salinity, oxidative stress, and mixotrophic growth conditions. *Front. Microbiol.* 3:354. doi: 10.3389/fmicb.2012.00354

Lv, Q., Ma, W., Liu, H., Li, J., Wang, H., Lu, F., et al. (2015). Genome-wide protein-protein interactions and protein function exploration in cyanobacteria. *Sci. Rep.* 5:15519. doi: 10.1038/srep15519

Ma, A. T., Schmidt, C. M., and Golden, J. W. (2014). Regulation of gene expression in diverse cyanobacterial species by using theophylline-responsive riboswitches. *Appl. Environ. Microb.* 80, 6704–6713. doi: 10.1128/AEM.01697-14

Machado, D., Andrejev, S., Tramontano, M., and Patil, K. R. (2018). Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.* 46, 7542–7553. doi: 10.1093/nar/gky537

Machado, D., Herrgård, M. J., and Rocha, I. (2016). Stoichiometric representation of gene-protein-reaction associations leverages constraint-based analysis from

reaction to gene-level phenotype prediction. *PLoS Comput. Biol.* 12:e1005140. doi: 10.1371/journal.pcbi.1005140

Maia, P., Rocha, M., and Rocha, I. (2016). *In silico* constraint-based strain optimization methods: the quest for optimal cell factories. *Microbiol. Mol. Biol. Rev.* 80, 45–67. doi: 10.1128/MMBR.00014-15

Malatinszky, D., Steuer, R., and Jones, P. R. (2017). A comprehensively curated genome-scale two-cell model for the heterocystous cyanobacterium *Anabaena* sp. PCC 7120. *Plant Physiol.* 173, 509–523. doi: 10.1104/pp.16.01487

Markley, A. L., Begemann, M. B., Clarke, R. E., Gordon, G. C., and Pfleger, B. F. (2015). Synthetic biology toolbox for controlling gene expression in the cyanobacterium *Synechococcus* sp. strain PCC 7002. *ACS Synth. Biol.* 4, 595–603. doi: 10.1021/sb500260k

Markson, J. S., Piechura, J. R., Puszynska, A. M., and O'Shea, E. K. (2013). Circadian control of global gene expression by the cyanobacterial master regulator RpaA. *Cell* 155, 1396–1408. doi: 10.1016/j.cell.2013.11.005

Marraccini, P., Bulteau, S., Cassier-Chauvat, C., Mermet-Bouvier, P., and Chauvat, F. (1993). A conjugative plasmid vector for promoter analysis in several cyanobacteria of the genera *Synechococcus* and *Synechocystis*. *Plant Mol. Biol.* 23, 905–909. doi: 10.1007/BF00021546

Matsuoka, M., Takahama, K., and Ogawa, T. (2001). Gene replacement in cyanobacteria mediated by a dominant streptomycin-sensitive *rps12* gene that allows selection of mutants free from drug resistance markers. *Microbiology* 147, 2077–2087. doi: 10.1099/00221287-147-8-2077

Megharaj, M., Venkateswarlu, K., and Rao, A. S. (1987). Metabolism of monocrotophos and quinalphos by algae isolated from soil. *Bull. Environ. Contam. Toxicol.* 39, 251–256. doi: 10.1007/BF01689414

Mohammadi, R., Fallah-Mehrabadi, J., Bidkhori, G., Zahiri, J., Javad Niroomand, M., and Masoudi-Nejad, A. (2016). A systems biology approach to reconcile metabolic network models with application to *Synechocystis* sp. PCC 6803 for biofuel production. *Mol. Biosyst.* 12, 2552–2561. doi: 10.1039/C6MB00119J

Montagud, A., Navarro, E., Fernández de Córdoba, P., Urchueguía, J. F., and Patil, K. R. (2010). Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium. *BMC Syst. Biol.* 4:156. doi: 10.1186/1752-0509-4-156

Montagud, A., Zelezniak, A., Navarro, E., de Cordoba, P. F., Urchueguia, J. F., and Patil, K. R. (2011). Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium *Synechocystis* sp. PCC 6803. *Biotechnol. J.* 6, 330–342. doi: 10.1002/biot.201000109

Mori, T., Binder, B., and Johnson, C. H. (1996). Circadian gating of cell division in cyanobacteria growing with average doubling times of less than 24 hours. *Proc. Natl. Acad. Sci. U.S.A.* 93, 10183–10188. doi: 10.1073/pnas.93.19.10183

Mueller, T. J., Berla, B. M., Pakrasi, H. B., and Maranas, C. D. (2013). Rapid construction of metabolic models for a family of Cyanobacteria using a multiple source annotation workflow. *BMC Syst. Biol.* 7:142. doi: 10.1186/1752-0509-7-142

Mueller, T. J., Ungerer, J. L., Pakrasi, H. B., and Maranas, C. D. (2017). Identifying the metabolic differences of a fast-growth phenotype in *Synechococcus* UTEX 2973. *Sci. Rep.* 7:41569. doi: 10.1038/srep41569

Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q. A., et al. (2013). Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* 10, 354–360. doi: 10.1038/nmeth.2404

Nakahira, Y., Ogawa, A., Asano, H., Oyama, T., and Tozawa, Y. (2013). Theophylline-dependent riboswitch as a novel genetic tool for strict regulation of protein expression in Cyanobacterium *Synechococcus elongatus* PCC 7942. *Plant Cell Physiol.* 54, 1724–1735. doi: 10.1093/pcp/pct115

Narro, M. L., Cerniglia, C. E., Van Baalen, C., and Gibson, D. T. (1992). Evidence for an NIH shift in oxidation of naphthalene by the marine cyanobacterium *Oscillatoria* sp. strain JCM. *Appl. Environ. Microbiol.* 58, 1360–1363.

National Academies of Sciences, Engineering and Medicine (2018). *Gaseous Carbon Waste Streams Utilization: Status and Research Needs*. Washington, DC: The National Academies Press.

Niu, T. C., Lin, G. M., Xie, L. R., Wang, Z. Q., Xing, W. Y., Zhang, J. Y., et al. (2018). Expanding the potential of CRISPR-Cpf1 based genome editing technology in the cyanobacterium *Anabaena* PCC 7120. *ACS Synth. Biol.* 8, 170–180. doi: 10.1021/acssynbio.8b00437

Nogales, J., Gudmundsson, S., Duque, E., Ramos, J. L., and Palsson, B. O. (2017). Expanding the computable reactome in *Pseudomonas putida* reveals metabolic cycles providing robustness. *bioRxiv*. doi: 10.1101/139121

Nogales, J., Gudmundsson, S., Knight, E. M., Palsson, B. O., and Thiele, I. (2012). Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. *Proc. Natl. Acad. Sci. U.S.A.* 109, 2678–2683. doi: 10.1073/pnas.1117907109

Nogales, J., Gudmundsson, S., and Thiele, I. (2013). Toward systems metabolic engineering in cyanobacteria: opportunities and bottlenecks. *Bioengineered* 4, 158–163. doi: 10.4161/bioe.22792

Nogales, J., Palsson, B. O., and Thiele, I. (2008). A genome-scale metabolic reconstruction of *Pseudomonas putida* KT2440: *i*JN746 as a cell factory. *BMC Syst. Biol.* 2:79. doi: 10.1186/1752-0509-2-79

Noguchi, T., and Golden, S. S. (2017). *Bioluminescent and Fluorescent Reporters in Circadian Rhythm Studies*, 1–24. Available online at: https://ccb.ucsd.edu/the-bioclock-studio/education-resources/reporter-review/ReporterReviewPDF.pdf (Accessed October 15, 2018).

Nozzi, N. E., Oliver, J. W., and Atsumi, S. (2013). Cyanobacteria as a platform for biofuel production. *Front. Bioeng. Biotechnol.* 1:7. doi: 10.3389/fbioe.2013.00007

Nudler, E. (2006). Flipping riboswitches. *Cell* 126, 19–22. doi: 10.1016/j.cell.2006.06.024

O'Brien, E. J., Monk, J. M., and Palsson, B. O. (2015). Using genome-scale models to predict biological capabilities. *Cell* 161, 971–987. doi: 10.1016/j.cell.2015.05.019

Ohbayashi, R., Akai, H., Yoshikawa, H., Hess, W. R., and Watanabe, S. (2016). A tightly inducible riboswitch system in *Synechocystis* sp. PCC 6803. *J. Gen. Appl. Microbiol.* 62, 154–159. doi: 10.2323/jgam.2016.02.002

Oliver, J. W., Machado, I. M., Yoneda, H., and Atsumi, S. (2013). Cyanobacterial conversion of carbon dioxide to 2,3-butanediol. *Proc. Natl. Acad. Sci. U.S.A.* 110, 1249–1254. doi: 10.1073/pnas.1213024110

Orth, J. D., Conrad, T. M., Na, J., Lerman, J. A., Nam, H., Feist, A. M., et al. (2011). A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism. *Mol. Syst. Biol.* 7:535. doi: 10.1038/msb.2011.65

Orth, J. D., Thiele, I., and Palsson, B. O. (2010). What is flux balance analysis? *Nat. Biotechnol.* 28, 245–248. doi: 10.1038/nbt.1614

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 42(Database issue), D206–D214. doi: 10.1093/nar/gkt1226

Parmar, A., Singh, N. K., Pandey, A., Gnansounou, E., and Madamwar, D. (2011). Cyanobacteria and microalgae: a positive prospect for biofuels. *Bioresour. Technol.* 102, 10163–10172. doi: 10.1016/j.biortech.2011.08.030

Pasotti, L., Politi, N., Zucca, S., Cusella De Angelis, M. G., and Magni, P. (2012). Bottom-up engineering of biological systems through standard bricks: a modularity study on basic parts and devices. *PLoS ONE* 7:e39407. doi: 10.1371/annotation/91e7d3a1-2f50-4f84-8b12-2c21f88438c3

Pathak, J., Rajneesh, M. P. K., Singh, S. P., Häder, D. P., and Sinha, R. P. (2018). Cyanobacterial farming for environment friendly sustainable agriculture practices: innovations and perspectives. *Front. Environ. Sci.* 6:7. doi: 10.3389/fenvs.2018.00007

Patil, K. R., Rocha, I., Förster, J., and Nielsen, J. (2005). Evolutionary programming as a platform for *in silico* metabolic engineering. *BMC Bioinformatics* 6:308. doi: 10.1186/1471-2105-6-308

Paul, M. J., and Foyer, C. H. (2001). Sink regulation of photosynthesis. *J. Exp. Bot.* 52, 1383–1400. doi: 10.1093/jexbot/52.360.1383

Pédelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006). Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* 24, 79–88. doi: 10.1038/nbt1172

Pérez, A. A., Rodionov, D. A., and Bryant, D. A. (2016). Identification and regulation of genes for cobalamin transport in the *Cyanobacterium Synechococcus* sp. strain PCC 7002. *J. Bacteriol.* 198, 2753–2761. doi: 10.1128/JB.00476-16

Popp, P. F., Dotzler, M., Radeck, J., Bartels, J., and Mascher, T. (2017). The *Bacillus* BioBrick Box 2.0: expanding the genetic toolbox for the standardized work with *Bacillus subtilis*. *Sci. Rep.* 7:15058. doi: 10.1038/s41598-017-15107-z

Porter, R. D. (1986). Transformation in cyanobacteria. *Crit. Rev. Microbiol.* 13, 111–132. doi: 10.3109/10408418609108736

Price, G. D., Badger, M. R., and Von Caemmerer, S. (2011). The prospect of using cyanobacterial bicarbonate transporters to improve leaf photosynthesis in C3 crop plants. *Plant Physiol.* 155, 20–26. doi: 10.1104/pp.110.164681

Price, G. D., Badger, M. R., Woodger, F. J., and Long, B. M. (2008). Advances in understanding the cyanobacterial CO2-concentrating-mechanism (CCM): functional components, Ci transporters, diversity, genetic regulation and prospects for engineering into plants. *J. Exp. Bot.* 59, 1441–1461. doi: 10.1093/jxb/erm112

Qian, X., Kim, M. K., Kumaraswamy, G. K., Agarwal, A., Lun, D. S., and Dismukes, G. C. (2017). Flux balance analysis of photoautotrophic metabolism: uncovering new biological details of subsystems involved in cyanobacterial photosynthesis. *Biochim. Biophys. Acta Bioenerg.* 1858, 276–287. doi: 10.1016/j.bbabio.2016.12.007

Quintana, N., Van der Kooy, F., Van de Rhee, M. D., Voshol, G. P., and Verpoorte, R. (2011). Renewable energy from Cyanobacteria: energy production optimization by metabolic pathway engineering. *Appl. Microbiol. Biotechnol.* 91, 471–490. doi: 10.1007/s00253-011-3394-0

Raman, K., and Chandra, N. (2009). Flux balance analysis of biological systems: applications and challenges. *Brief. Bioinformatics* 10, 435–449. doi: 10.1093/bib/bbp011

Ramey, C. J., Baron-Sola, A., Aucoin, H. R., and Boyle, N. R. (2015). Genome engineering in cyanobacteria: where we are and where we need to go. *ACS Synth. Biol.* 4, 1186–1196. doi: 10.1021/acssynbio.5b00043

Ranganathan, S., Suthers, P. F., and Maranas, C. D. (2010). OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput. Biol.* 6:e1000744. doi: 10.1371/journal.pcbi.1000744

Reed, J. L., Vo, T. D., Schilling, C. H., and Palsson, B. O. (2003). An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.* 4:R54. doi: 10.1186/gb-2003-4-9-r54

Ren, Q., Kang, K. H., and Paulsen, I. T. (2004). TransportDB: a relational database of cellular membrane transport systems. *Nucleic Acids Res.* 32(Database issue), D284–D288. doi: 10.1093/nar/gkh016

Rodriguez, E. A., Campbell, R. E., Lin, J. Y., Lin, M. Z., Miyawaki, A., Palmer, A. E., et al. (2017). The growing and glowing toolbox of fluorescent and photoactive proteins. *Trends Biochem. Sci.* 42, 111–129. doi: 10.1016/j.tibs.2016.09.010

Rubin, B. E., Wetmore, K. M., Price, M. N., Diamond, S., Shultzaberger, R. K., Lowe, L. C., et al. (2015). The essential gene set of a photosynthetic organism. *Proc. Natl. Acad. Sci. U.S.A.* 112, E6634–E6643. doi: 10.1073/pnas.1519220112

Ruffing, A. M. (2011). Engineered cyanobacteria: teaching an old bug new tricks. *Bioeng. Bugs* 2, 136–149. doi: 10.4161/bbug.2.3.15285

Ruffing, A. M. (2014). Improved free fatty acid production in cyanobacteria with *Synechococcus* sp. PCC 7002 as host. *Front. Bioeng. Biotechnol.* 2:17. doi: 10.3389/fbioe.2014.00017

Ruffing, A. M., Jensen, T. J., and Strickland, L. M. (2016). Genetic tools for advancement of *Synechococcus* sp. PCC 7002 as a cyanobacterial chassis. *Microb. Cell Fact.* 15:190. doi: 10.1186/s12934-016-0584-6

Saha, R., Verseput, A. T., Berla, B. M., Mueller, T. J., Pakrasi, H. B., and Maranas, C. D. (2012). Reconstruction and comparison of the metabolic potential of cyanobacteria *Cyanothece* sp. ATCC 51142 and *Synechocystis* sp. PCC 6803. *PLoS ONE* 7:e48285. doi: 10.1371/journal.pone.0048285

Sakuma, T., Nishikawa, A., Kume, S., Chayama, K., and Yamamoto, T. (2014). Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Sci. Rep.* 4:5400. doi: 10.1038/srep05400

Salis, H. M. (2011). The ribosome binding site calculator. *Method Enzymol.* 498, 19–42. doi: 10.1016/B978-0-12-385120-8.00002-4

Salvador-Reyes, L. A., and Luesch, H. (2015). Biological targets and mechanisms of action of natural products from marine cyanobacteria. *Nat. Prod. Rep.* 32, 478–503. doi: 10.1039/C4NP00104D

Santos-Merino, M., Garcillán-Barcia, M. P., and de la Cruz, F. (2018). Engineering the fatty acid synthesis pathway in *Synechococcus elongatus* PCC 7942 improves omega-3 fatty acid production. *Biotechnol. Biofuels* 11:239. doi: 10.1186/s13068-018-1243-4

Sawada, S., Hayakawa, T., Fukushi, K., and Kasai, M. (1986). Influence of carbohydrates on photosynthesis in single, rooted soybean leaves used as a source-sink model. *Plant Cell Physiol.* 27, 591–600.

Schleif, R. (2010). AraC protein, regulation of the l-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. *FEMS Microbiol. Rev.* 34, 779–796. doi: 10.1111/j.1574-6976.2010.00226.x

Schoepp, N. G., Stewart, R. L., Sun, V., Quigley, A. J., Mendola, D., Mayfield, S. P., et al. (2014). System and method for research-scale outdoor production of microalgae and cyanobacteria. *Bioresour. Technol.* 166, 273–281. doi: 10.1016/j.biortech.2014.05.046

Schomburg, I., Chang, A., and Schomburg, D. (2002). BRENDA, enzyme data and metabolic information. *Nucleic Acids Res.* 30, 47–49. doi: 10.1093/nar/30.1.47

Segrè, D., Vitkup, D., and Church, G. M. (2002). Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15112–15117. doi: 10.1073/pnas.232349399

Sengupta, A., Pakrasi, H. B., and Wangikar, P. P. (2018). Recent advances in synthetic biology of cyanobacteria. *Appl. Microbiol. Biotechnol.* 102, 5457–5471. doi: 10.1007/s00253-018-9046-x

Shabestary, K., and Hudson, E. P. (2016). Computational metabolic engineering strategies for growth-coupled biofuel production by *Synechocystis*. *Metab. Eng. Commun.* 3, 216–226. doi: 10.1016/j.meteno.2016.07.003

Sharkey, T. D., Stitt, M., Heineke, D., Gerhardt, R., Raschke, K., and Heldt, H. W. (1986). Limitation of photosynthesis by carbon metabolism: II. O(2)-insensitive CO(2) uptake results from limitation of triose phosphate utilization. *Plant Physiol.* 81, 1123–1129. doi: 10.1104/pp.81.4.1123

Shastri, A. A., and Morgan, J. A. (2005). Flux balance analysis of photoautotrophic metabolism. *Biotechnol. Prog.* 21, 1617–1626. doi: 10.1021/bp050246d

Sheehan, J., Dunahay, T., Benemann, J., and Roessler, P. (1998). *A Look Back at the U.S. Department of Energy's Aquatic Species Program: Biodiesel From Algae.* Available online at: https://www.nrel.gov/docs/legosti/fy98/24190.pdf (Accessed October 15, 2018).

Sheng, J., Kim, H. W., Badalamenti, J. P., Zhou, C., Sridharakrishnan, S., Krajmalnik-Brown, R., et al. (2011). Effects of temperature shifts on growth rate and lipid characteristics of *Synechocystis* sp. PCC6803 in a bench-top photobioreactor. *Bioresource Technology* 102, 11218–11225. doi: 10.1016/j.biortech.2011.09.083

Shestakov, S. V., and Khyen, N. T. (1970). Evidence for genetic transformation in blue-green alga *Anacystis nidulans*. *MGG Mol. Gen. Genet.* 107, 372–375. doi: 10.1007/BF00441199

Shetty, R. P., Endy, D., and Knight, T. F. Jr. (2008). Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.* 2:5. doi: 10.1186/1754-1611-2-5

Shih, P. M. (2015). Cyanobacterial evolution: fresh insight into ancient questions. *Curr. Biol.* 25, R192–R193. doi: 10.1016/j.cub.2014.12.046

Shirai, T., Osanai, T., and Kondo, A. (2016). Designing intracellular metabolism for production of target compounds by introducing a heterologous metabolic reaction based on a *Synechosystis* sp. 6803 genome-scale model. *Microb. Cell Fact.* 15:13. doi: 10.1186/s12934-016-0416-8

Singh, J. S., Kumar, A., Rai, A. N., and Singh, D. P. (2016). Cyanobacteria: a precious bio-resource in agriculture, ecosystem, and environmental sustainability. *Front. Microbiol.* 7:529. doi: 10.3389/fmicb.2016.00529

Song, K., Tan, X., Liang, Y., and Lu, X. (2016). The potential of *Synechococcus elongatus* UTEX 2973 for sugar feedstock production. *Appl. Microbiol. Biotechnol.* 100, 7865–7875. doi: 10.1007/s00253-016-7510-z

Srinivasan, S., Cluett, W. R., and Mahadevan, R. (2015). Constructing kinetic models of metabolism at genome-scales: a review. *Biotechnol. J.* 10, 1345–1359. doi: 10.1002/biot.201400522

Stanier, R. Y., Kunisawa, R., Mandel, M., and Cohen-Bazire, G. (1971). Purification and properties of unicellular blue-green algae (order *Chroococcales*). *Bacteriol. Rev.* 35, 171–205.

Stephanopoulos, G. (2012). Synthetic biology and metabolic engineering. *ACS Synth. Biol.* 1, 514–525. doi: 10.1021/sb300094q

Steuer, R., Knoop, H., and Machne, R. (2012). Modelling cyanobacteria: from metabolism to integrative models of phototrophic growth. *J. Exp. Bot.* 63, 2259–2274. doi: 10.1093/jxb/ers018

Stevens, S. E., and Porter, R. D. (1980). Transformation in *Agmenellum quadruplicatum*. *Proc. Natl. Acad. Sci. U.S.A.* 77, 6052–6056. doi: 10.1073/pnas.77.10.6052

Stucken, K., Koch, R., and Dagan, T. (2013). Cyanobacterial defense mechanisms against foreign DNA transfer and their impact on genetic engineering. *Biol. Res.* 46, 373–382. doi: 10.4067/S0716-97602013000400009

Sun, T., Li, S., Song, X., Diao, J., Chen, L., and Zhang, W. (2018). Toolboxes for cyanobacteria: recent advances and future direction. *Biotechnol. Adv.* 36, 1293–1307. doi: 10.1016/j.biotechadv.2018.04.007

Suthers, P. F., Zomorrodi, A., and Maranas, C. D. (2009). Genome-scale gene/reaction essentiality and synthetic lethality analysis. *Mol. Syst. Biol.* 5:301. doi: 10.1038/msb.2009.56

Swarts, D. C., and Jinek, M. (2018). Cas9 versus Cas12a/Cpf1: structure-function comparisons and implications for genome editing. *Wiley Interdiscip Rev. RNA* 9:e1481. doi: 10.1002/wrna.1481

Takahama, K., Matsuoka, M., Nagahama, K., and Ogawa, T. (2004). High-frequency gene replacement in cyanobacteria using a heterologous *rps12* gene. *Plant Cell Physiol.* 45, 333–339. doi: 10.1093/pcp/pch041

Tamagnini, P., Axelsson, R., Lindberg, P., Oxelfelt, F., Wunschiers, R., and Lindblad, P. (2002). Hydrogenases and hydrogen metabolism of cyanobacteria. *Microbiol. Mol. Biol. Rev.* 66, 1–20. doi: 10.1128/MMBR.66.1.1-20.2002

Tan, X., Hou, S., Song, K., Georg, J., Klähn, S., Lu, X., et al. (2018). The primary transcriptome of the fast-growing cyanobacterium *Synechococcus elongatus* UTEX 2973. *Biotechnol. Biofuels* 11:218. doi: 10.1186/s13068-018-1215-8

Taton, A., Lis, E., Adin, D. M., Dong, G., Cookson, S., Kay, S. A., et al. (2012). Gene transfer in *Leptolyngbya* sp. strain BL0902, a cyanobacterium suitable for production of biomass and bioproducts. *PLoS ONE* 7:e30901. doi: 10.1371/journal.pone.0030901

Taton, A., Ma, A. T., Ota, M., Golden, S. S., and Golden, J. W. (2017). NOT gate genetic circuits to control gene expression in cyanobacteria. *ACS Synth. Biol.* 6, 2175–2182. doi: 10.1021/acssynbio.7b00203

Taton, A., Unglaub, F., Wright, N. E., Zeng, W. Y., Paz-Yepes, J., Brahamsha, B., et al. (2014). Broad-host-range vector system for synthetic biology and biotechnology in cyanobacteria. *Nucleic Acids Res.* 42:e136. doi: 10.1093/nar/gku673

Thiel, K., Mulaku, E., Dandapani, H., Nagy, C., Aro, E. M., and Kallio, P. (2018). Translation efficiency of heterologous proteins is significantly affected by the genetic context of RBS sequences in engineered cyanobacterium *Synechocystis* sp. PCC 6803. *Microb. Cell Fact.* 17:34. doi: 10.1186/s12934-018-0882-2

Thiel, T., and Poo, H. (1989). Transformation of a filamentous cyanobacterium by electroporation. *J. Bacteriol.* 171, 5743–5746. doi: 10.1128/jb.171.10.5743-5746.1989

Thiele, I., and Palsson, B. O. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* 5, 93–121. doi: 10.1038/nprot.2009.203

Thomas, A., Rahmanian, S., Bordbar, A., Palsson, B. O., and Jamshidi, N. (2014). Network reconstruction of platelet metabolism identifies metabolic signature for aspirin resistance. *Sci. Rep.* 4:3925. doi: 10.1038/srep03925

Triana, J., Montagud, A., Siurana, M., Fuente, D., Urchueguia, A., Gamermann, D., et al. (2014). Generation and evaluation of a genome-scale metabolic network model of *Synechococcus elongatus* PCC 7942. *Metabolites* 4, 680–698. doi: 10.3390/metabo4030680

Tsinoremas, N. F., Kutach, A. K., Strayer, C. A., and Golden, S. S. (1994). Efficient gene transfer in *Synechococcus* sp. strains PCC 7942 and PCC 6301 by interspecies conjugation and chromosomal recombination. *J. Bacteriol.* 176, 6764–6768. doi: 10.1128/jb.176.21.6764-6768.1994

Ungerer, J., and Pakrasi, H. B. (2016). Cpf1 is a versatile tool for CRISPR genome editing across diverse species of cyanobacteria. *Sci. Rep.* 6:39681. doi: 10.1038/srep39681

Ungerer, J., Tao, L., Davis, M., Ghirardi, M., Maness, P. C., and Yu, J. (2012). Sustained photosynthetic conversion of CO_2 to ethylene in recombinant cyanobacterium *Synechocystis* 6803. *Energy Environ. Sci.* 5, 8998–9006. doi: 10.1039/c2ee22555g

Ungerer, J., Wendt, K. E., Hendry, J. I., Maranas, C. D., and Pakrasi, H. B. (2018). Comparative genomics reveals the molecular determinants of rapid growth of the cyanobacterium *Synechococcus elongatus* UTEX 2973. *Proc. Natl. Acad. Sci. U.S.A.* 115, E11761–E11770. doi: 10.1073/pnas.1814912115

Utschig, L. M., Silver, S. C., Mulfort, K. L., and Tiede, D. M. (2011). Nature-driven photochemistry for catalytic solar hydrogen production: a Photosystem I-transition metal catalyst hybrid. *J. Am. Chem. Soc.* 133, 16334–16337. doi: 10.1021/ja206012r

Vasudevan, R., Gale, G. A. R., Schiavon, A. A., Puzorjov, A., Malin, J., Gillespie, M. D., et al. (2018). CyanoGate: a golden gate modular cloning suite for engineering cyanobacteria based on the Plant MoClo syntax. bioRxiv. doi: 10.1101/426700

Vermaas, W. (1996). Molecular genetics of the cyanobacterium *Synechocystis* sp. PCC 6803: principles and possible biotechnology applications. *J. Appl. Phycol.* 8, 263–273. doi: 10.1007/BF02178569

Vermass, W. F., Rutherford, A. W., and Hansson, O. (1988). Site-directed mutagenesis in photosystem II of the cyanobacterium *Synechocystis* sp. PCC 6803: donor D is a tyrosine residue in the D2 protein. *Proc. Natl. Acad. Sci. U.S.A.* 85, 8477–8481. doi: 10.1073/pnas.85.22.8477

Viola, S., Rühle, T., and Leister, D. (2014). A single vector-based strategy for marker-less gene replacement in *Synechocystis* sp. PCC 6803. *Microb. Cell Fact.* 13:4. doi: 10.1186/1475-2859-13-4

Vitkin, E., and Shlomi, T. (2012). MIRAGE: a functional genomics-based approach for metabolic network model reconstruction and its application to cyanobacteria networks. *Genome Biol.* 13:R111. doi: 10.1186/gb-2012-13-11-r111

Vivek-Ananth, R. P., and Samal, A. (2016). Advances in the integration of transcriptional regulatory information into genome-scale metabolic models. *BioSystems* 147, 1–10. doi: 10.1016/j.biosystems.2016.06.001

Vogel, A. I. M., Lale, R., and Hohmann-Marriott, M. F. (2017). Streamlining recombination-mediated genetic engineering by validating three neutral integration sites in *Synechococcus* sp. PCC 7002. *J. Biol. Eng.* 11:19. doi: 10.1186/s13036-017-0061-8

Vu, T. T., Hill, E. A., Kucek, L. A., Konopka, A. E., Beliaev, A. S., and Reed, J. L. (2013). Computational evaluation of *Synechococcus* sp. PCC 7002 metabolism for chemical production. *Biotechnol. J.* 8, 619–630. doi: 10.1002/biot.201200315

Vu, T. T., Stolyar, S. M., Pinchuk, G. E., Hill, E. A., Kucek, L. A., Brown, R. N., et al. (2012). Genome-scale modeling of light-driven reductant partitioning and carbon fluxes in diazotrophic unicellular cyanobacterium *Cyanothece* sp. ATCC 51142. *PLoS Comput. Biol.* 8:e1002460. doi: 10.1371/journal.pcbi.1002460

Wang, B., Eckert, C., Maness, P. C., and Yu, J. (2018). A genetic toolbox for modulating the expression of heterologous genes in the Cyanobacterium *Synechocystis* sp. PCC 6803. *ACS Synth. Biol.* 7, 276–286. doi: 10.1021/acssynbio.7b00297

Wang, B., Wang, J., Zhang, W., and Meldrum, D. R. (2012). Application of synthetic biology in cyanobacteria and algae. *Front. Microbiol.* 3:344. doi: 10.3389/fmicb.2012.00344

Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., et al. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460, 894–898. doi: 10.1038/nature08187

Wang, Y. H., Wei, K. Y., and Smolke, C. D. (2013). Synthetic biology: advancing the design of diverse genetic systems. *Annu. Rev. Chem. Biomol. Eng.* 4, 69–102. doi: 10.1146/annurev-chembioeng-061312-103351

Warner, J. R., Reeder, P. J., Karimpour-Fard, A., Woodruff, L. B., and Gill, R. T. (2010). Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat. Biotechnol.* 28, 856–862. doi: 10.1038/nbt.1653

Wendt, K. E., Ungerer, J., Cobb, R. E., Zhao, H., and Pakrasi, H. B. (2016). CRISPR/Cas9 mediated targeted mutagenesis of the fast growing cyanobacterium *Synechococcus elongatus* UTEX 2973. *Microb. Cell Fact.* 15:115. doi: 10.1186/s12934-016-0514-7

Wolk, C. P., Vonshak, A., Kehoe, P., and Elhai, J. (1984). Construction of shuttle vectors capable of conjugative transfer from *Escherichia coli* to nitrogen-fixing filamentous cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A.* 81, 1561–1565. doi: 10.1073/pnas.81.5.1561

Woo, H. M. (2017). Solar-to-chemical and solar-to-fuel production from CO2 by metabolically engineered microorganisms. *Curr. Opin. Biotechnol.* 45, 1–7. doi: 10.1016/j.copbio.2016.11.017

Xiao, Y., Wang, S., Rommelfanger, S., Balassy, A., Barba-Ostria, C., Gu, P., et al. (2018). Developing a Cas9-based tool to engineer native plasmids in *Synechocystis* sp. PCC 6803. *Biotechnol. Bioeng.* 115, 2305–2314. doi: 10.1002/bit.26747

Xiong, W., Cano, M., Wang, B., Douchi, D., and Yu, J. (2017). The plasticity of cyanobacterial carbon metabolism. *Curr. Opin. Chem. Biol.* 41, 12–19. doi: 10.1016/j.cbpa.2017.09.004

Xiong, W., Morgan, J. A., Ungerer, J., Wang, B., Maness, P. C., and Yu, J. (2015). The plasticity of cyanobacterial metabolism supports direct CO_2 conversion to ethylene. *Nat. Plants* 1:15053. doi: 10.1038/nplants.2015.53

Yang, Q., Pando, B. F., Dong, G., Golden, S. S., and van Oudenaarden, A. (2010). Circadian gating of the cell cycle revealed in single cyanobacterial cells. *Science* 327, 1522–1526. doi: 10.1126/science.1181759

Yokoo, R., Hood, R. D., and Savage, D. F. (2015). Live-cell imaging of cyanobacteria. *Photosyn. Res.* 126, 33–46. doi: 10.1007/s11120-014-0049-x

Yoshikawa, K., Aikawa, S., Kojima, Y., Toya, Y., Furusawa, C., Kondo, A., et al. (2015). Construction of a genome-scale metabolic model of *Arthrospira platensis* NIES-39 and metabolic design for cyanobacterial bioproduction. *PLoS ONE* 10:e0144430. doi: 10.1371/journal.pone.0144430

Yoshikawa, K., Kojima, Y., Nakajima, T., Furusawa, C., Hirasawa, T., and Shimizu, H. (2011). Reconstruction and verification of a genome-scale metabolic model for *Synechocystis* sp. PCC 6803. *Appl. Microbiol. Biotechnol.* 92, 347–358. doi: 10.1007/s00253-011-3559-x

Yoshikawa, K., Toya, Y., and Shimizu, H. (2017). Metabolic engineering of *Synechocystis* sp. PCC 6803 for enhanced ethanol production based on flux balance analysis. *Bioprocess Biosyst. Eng.* 40, 791–796. doi: 10.1007/s00449-017-1744-8

Yu, J., Liberton, M., Cliften, P. F., Head, R. D., Jacobs, J. M., Smith, R. D., et al. (2015). *Synechococcus elongatus* UTEX 2973, a fast growing cyanobacterial chassis for biosynthesis using light and CO_2. *Sci. Rep.* 5:8132. doi: 10.1038/srep08132

Yu, Y., You, L., Liu, D., Hollinshead, W., Tang, Y. J., and Zhang, F. (2013). Development of *Synechocystis* sp. PCC 6803 as a phototrophic cell factory. *Mar. Drugs* 11, 2894–2916. doi: 10.3390/md11082894

Zang, X., Liu, B., Liu, S., Arunakumara, K. K., and Zhang, X. (2007). Optimum conditions for transformation of *Synechocystis* sp. PCC 6803. *J. Microbiol.* 45, 241–245.

Zavrel, T., Knoop, H., Steuer, R., Jones, P. R., Cerveny, J., and Trtílek, M. (2016). A quantitative evaluation of ethylene production in the recombinant cyanobacterium *Synechocystis* sp. PCC 6803 harboring the ethylene-forming enzyme by membrane inlet mass spectrometry. *Bioresour. Technol.* 202, 142–151. doi: 10.1016/j.biortech.2015.11.062

Zhang, A., Carroll, A. L., and Atsumi, S. (2017). Carbon recycling by cyanobacteria: improving CO_2 fixation through chemical production. *FEMS Microbiol. Lett.* 364, 1–7. doi: 10.1093/femsle/fnx165

Zhang, W., and Song, X. (2018). *Synthetic Biology of Cyanobacteria*. Singapore: Springer.

Zhang, Y. H. P. (2011). "Hydrogen production from carbohydrates: a mini-review," in *Sustainable Production of Fuels, Chemicals, and Fibers from Forest Biomass*, eds J. Zhu, X. Zhang, and X. Pan (Washington, DC: American Chemical Society), 203–216.

Zhou, J., Zhang, H., Meng, H., Zhu, Y., Bao, G., Zhang, Y., et al. (2014). Discovery of a super-strong promoter enables efficient production of heterologous proteins in cyanobacteria. *Sci. Rep.* 4:4500. doi: 10.1038/srep04500

Check for
updates

Metabolic Engineering and Synthetic Biology: Synergies, Future, and Challenges

Raúl García-Granados [1,2], Jordy Alexis Lerma-Escalera [1,2,3] and José R. Morones-Ramírez [1,2]*

[1] Facultad de Ciencias Químicas, Universidad Autónoma de Nuevo León, San Nicolás de los Garza, Mexico, [2] Centro de Investigación en Biotecnología y Nanotecnología, Facultad de Ciencias Químicas, Universidad Autónoma de Nuevo León, Apodaca, Mexico, [3] Facultad de Ciencias Biológicas, Universidad Autónoma de Nuevo León, San Nicolás de los Garza, Mexico

The "-omics" era has brought a new set of tools and methods that have created a significant impact on the development of Metabolic Engineering and Synthetic Biology. These fields, rather than working separately, depend on each other to prosper and achieve their individual goals. Synthetic Biology aims to design libraries of genetic components (promoters, coding sequences, terminators, transcriptional factors and their binding sequences, and more), the assembly of devices, genetic circuits and even organism; in addition to obtaining quantitative information for the creation of models that can predict the behavior of biological systems (Cameron et al., 2014). Metabolic engineering seeks for the optimization of cellular processes, endemic to a specific organism, to produce a compound of interest from a substrate, preferably cheap and simple. It uses different databases, libraries of components and conditions to generate the maximum production rate of a desired chemical compound and avoiding inhibitors and conditions that affect the growth rate and other vital functions in the specific organism to achieve these goals; metabolic fluxes manipulation represents an important alternative (Stephanopoulos, 2012).

Keywords: synthetic biology, metabolic engineering, biotechnology, bioengineering, omics analyses

While synthetic biology provides the components and information about different biological phenomena, metabolic engineering tries to apply all this information toward the optimization of the biological synthesis trajectory of a desired compound. We can even mention, that some examples of synthetic biology could be also classified as examples of genetic engineering. Despite this, these two areas are dependent on the advances in the actual methods, techniques and tools for DNA modification. The reason is that both areas of research seek to achieve basic requirements such as rational changes in DNA sequences, the generation of specific mutations, the assembly of parts or components in genetic circuits or biosynthetic pathways, the knockout of genes, the integration of DNA pieces in the genome of an organism of interest or in a plasmid (Boyle and Silver, 2012). Although, the PCR and its variants are one of the best tools to generate some of the necessary modifications (it is especially useful for the extraction of fragments of a specific region) it results to be ineffective for other purposes.

The creation of complementary new techniques and methods to achieve other objectives has been vital in the development of both areas, some examples are: Biobricks, Recombinase

technologies (integrases), Gibson Assembly, Gap-repair, Lambda-red, MAGE and CRISPR-Cas9, among others (Boyle and Silver, 2012; Stephanopoulos, 2012). During the "-omics" era, the sequencing of both, known and unknown organisms (metagenomics), allows obtaining more and better information from them; the catalog of enzymes and processes was expanded, and certain biological phenomena (for example viruses or phage infections) has been achieved (Goodwin et al., 2016).

Rational design is the introduction of desired mutations into a specific DNA sequence eliding the modification of proteins to improve their catalytic activity, stability or some other property (binding domain specificity); subsequently tools as the Gibson assembly, the biobricks or Golden Gate can assemble the different genetic components to form genetic circuits, expression cassettes or devices (Casini et al., 2015). Finally, the use of Lambda-red, CRISPR-Cas9 and other recombination techniques allow us to insert the different constructions into expression vectors or in strategic loci of the genome of a given organism (Esvelt and Wang, 2013; Liu et al., 2015; Wang et al., 2015).

Some of these technologies can be applied to insert genes of partial or complete metabolic pathways for the synthesis of a specific compound, they can also be used to eliminate genes from the organism that interfere with the synthesis. In addition, these methodologies allow carrying out point mutations that reduce the activity or expression of native proteins to modify the metabolic flux. However, to create all these modifications it is necessary to have certain information about the enzymes that participate in the reactions, the metabolic pathway that includes these reactions and information about the organism where the modifications will be made; because of this, sequencing genomes of organisms, characterization of proteins and metabolic studies provides extremely useful tools and information (**Figure 1**).

Therefore, one of the main issues is the lack of information; very few organisms have been sequenced and characterized at different levels (genome, transcriptome and metabolome); making the genetic modifications a difficult task to achieve. Added to this, the techniques and methods that currently exist are not effective or have low performance rates that depend on the organism. The creation of plasmids capable of being used in different organisms (shuttle vectors) will aid in the solution of this problem but sequencing remains a necessary step. With this information we could discover new proteins or metabolic pathways that could be useful for higher compound production titles (enzymes with better production rates or simpler metabolic pathways); we could even describe new biological phenomena that are useful for the modification of DNA sequences.

Another important point related to the information available, is the selection of the chassis (or microorganism on which we will working on). The basic information and techniques available, as well as the special qualities (specific metabolic pathways or certain resistant to conditions) are important criteria when we are choosing the chassis and could facilitate the development of a project (Brophy and Voigt, 2014; Khoury et al., 2014).

A problem related to the choice of the chassis and the modification of both the metabolic pathway and metabolic flux, is the production of toxic intermediates or the accumulation of an intermediate that could cause the inhibition of the route

(feedback). The production of certain compounds can result in toxicity for the microorganism that produce it, especially when is handled at large scales (Sopko et al., 2006; Förster and Gescher, 2014). In these cases, a correct design of the bioreactor is usually helpful, removing the toxic compound before it reaches a threshold concentration of damage, and therefore improving significantly the production titles. Here appears another challenge for both areas, to discover a microorganism that can resist a greater concentration of that compound, or to discover and adapt the mechanism of resistance from organism to another (Keasling, 2012). However, to avoid the accumulation of an intermediate in any of the steps from the metabolic pathway, it is necessary to determine the rates of synthesis and production of each of the compounds; once you have this information, tinkering with the strength of promoters, RBS, transcriptional factors, or in some cases, the growth conditions allow us to avoid this problem, but the main problem still relies in having these information available.

The most important challenge for both areas is the obtention of quantitative data from different biological phenomena (Le Novère, 2015). We need more projects that are responsible for characterizing biological components; like the strength of promoters, RBSs, terminators; as well as the behavior of certain phenomena (transcription and translation levels of genes, average life of transcripts and proteins, catalytic activity, repressive force). Also, the creation of genetic circuits and devices has helped with the description of more complex behaviors and the understanding of previously mentioned phenomena (Canton et al., 2008).

All these data allow the creation of mathematical and computational models that provides us with a better understanding of the behavior of the system; and helps to have a visualization of expected behaviors, approximate productions, selection of conditions, etc. (Hwang et al., 2009). There is a diversity of works focused on biological simulation: Karr et al. (2012) report a whole-cell computational model of *Mycoplasma genitalium*, including the molecular components and their interactions by combining different mathematical and computational approaches (ODEs with Boolean, probabilistic and constraint-based submodels). Moreover, O'Brien et al. (2015) mentions the different applications that the mathematical models and the computational simulation offers: simulating gene and reaction knockouts, comparing inputs and outputs, iterative improvement, gap-filling approaches, discovery of regulatory interactions, quantification of predictable phenotypes from optimality principles, model-based design of environmental and genetic perturbations, to mention some.

The application of the mathematical models could be a powerful tool to understand biological systems and the computational simulations could give us a better perspective of the processes and the changes that occur or of the predictions. Technologies such as the Machine Learning, the Big Data and the Artificial Intelligence are being of great help for the analysis of biological data. Different methods (Artificial Neural Nets, Bayesian, Decision Tree and Random Forrest, Multi-layer perceptron (MLP), Radial basis function (RBF) networks, Support Vector Machines, K-means, Farthest First,

FIGURE 1 | Different stages for the development of synthetic biology and metabolic engineering projects.
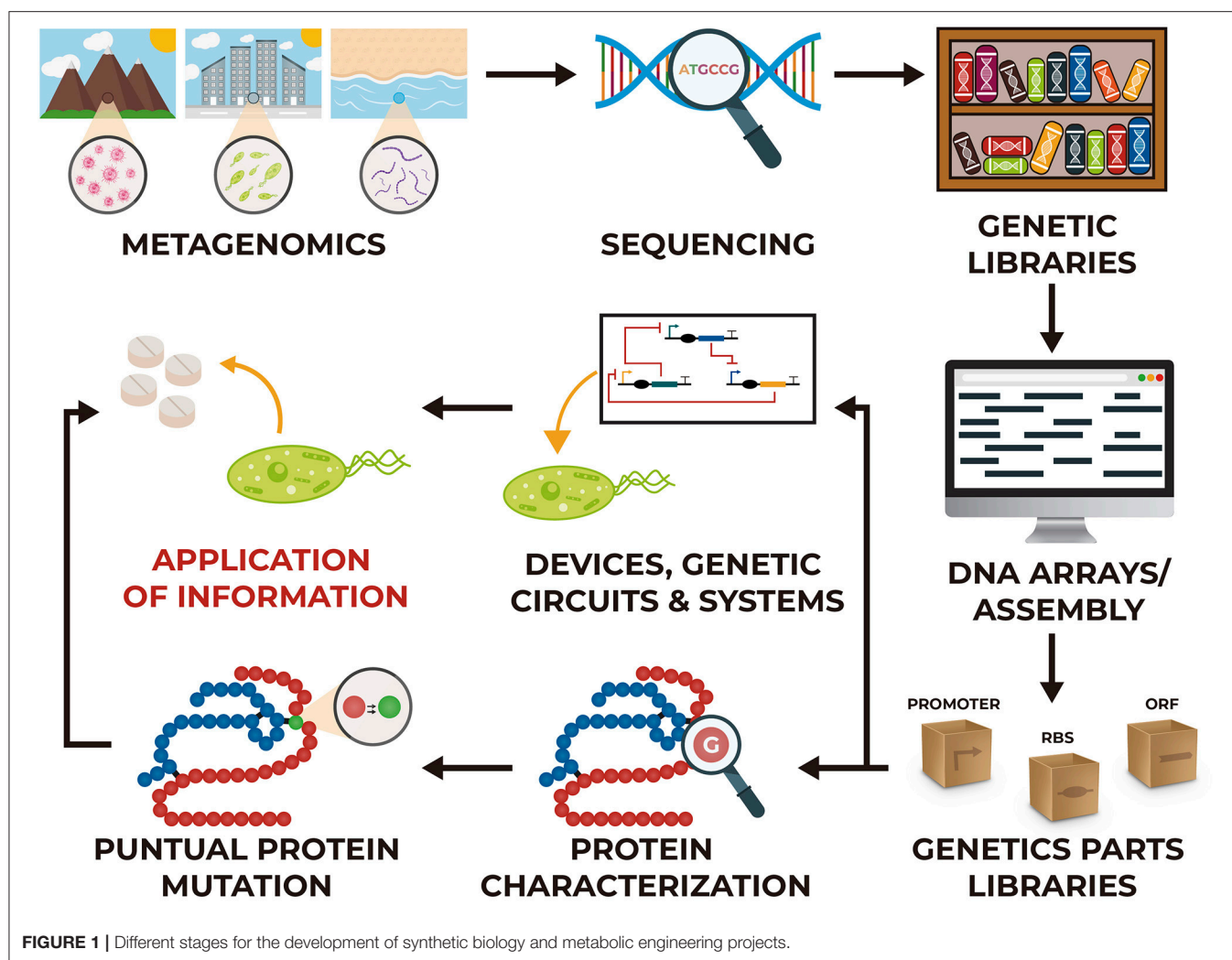
Density Base Clustering, etc.) are being used for microarrays, study of diseases, epidemiology; generating useful information for different researches (Pirooznia et al., 2008; Libbrecht and Noble, 2015). The use of quantum computers represents a very important tool for the development of these simulations and methods; these computers process the information in qubits (can represent either 1 or 0 or combinations of them, known as superposition) so it has a greater processing scope, something very valuable when you work with large amounts of information; the first big breakthrough has been the IBM to take public sale the first commercial quantum computer in history (the IBM Q System One). It is only a matter of time to discover the power that gives different researchers the analysis of information that provides the quantum computer in the biological and health area.

Both metabolic engineering and synthetic biology are two promising areas that have made great advances in biotechnology and have contributed significantly toward the resolution of problems in production of drugs, vaccines, chemical compounds, etc. (Khalil and Collins, 2010). In addition, these fields have advanced our knowledge regarding life function. Despite all these advances, it is still necessary to continue collecting information

regarding functioning of cells and living organisms and discover new species of microorganisms that could aid in the development of new techniques and methods. Finally, synthesizing long sequences is often problematic because of the margin of error that exists with current chemical synthesis techniques (approximately 1 error in every 1,000 base pairs); the search for new nucleotide synthesis techniques (such as TdT-dNTP or enzymatic synthesis) or the improvement of the current chemical synthesis of nucleotides are aspects to be improved and could be very useful for both areas, especially because it opens the opportunity for the synthesis of whole genomes, vectors or artificial chromosomes (Kosuri and Church, 2014; Hughes and Ellington, 2017).

With all these tools at our disposal, we will be able to optimize microorganisms as small factories that allow us to obtain higher rates of yield and production of a chemical compound, preferably using simple substrates.

AUTHOR CONTRIBUTIONS

RG-G, JL-E, and JM-R all contributed toward the writing and editing of this manuscript.

FUNDING

REFERENCES

Boyle, P. M., and Silver, P. A. (2012). Parts plus pipes: synthetic biology approaches to metabolic engineering. *Metab. Eng.* 14, 223–232. doi: 10.1016/j.ymben.2011.10.003

Brophy, J. A., and Voigt, C. A. (2014). Principles of genetic circuit design. *Nat. Methods* 11: 508. doi: 10.1038/nmeth.2926

Cameron, D. E., Bashor, C. J., and Collins, J. J. (2014). A brief history of synthetic biology. *Nat. Rev. Microbiol.* 12: 381. doi: 10.1038/nrmicro3239

Canton, B., Labno, A., and Endy, D. (2008). Refinement and standardization of synthetic biological parts and devices. *Nat. Biotechnol.* 26: 78. doi: 10.1038/nbt1413

Casini, A., Storch, M., Baldwin, G. S., and Ellis, T. (2015). Bricks and blueprints: methods and standards for DNA assembly. *Nat. Rev. Mol. Cell Biol.* 16: 568. doi: 10.1038/nrm4014

Esvelt, K. M., and Wang, H. H. (2013). Genome-scale engineering for systems and synthetic biology. *Mol. Syst. Biol.* 9: 641. doi: 10.1038/msb.2012.66

Förster, A. H., and Gescher, J. (2014). Metabolic engineering of *Escherichia coli* for production of mixed-acid fermentation end products. *Front. Bioeng. Biotechnol.* 2:16. doi: 10.3389/fbioe.2014.00016

Goodwin, S., McPherson, J. D., and McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* 17: 333. doi: 10.1038/nrg.2016.49

Hughes, R. A., and Ellington, A. D. (2017). Synthetic DNA synthesis and assembly: putting the synthetic in synthetic biology. *Cold Spring Harb. Perspect. Biol.* 9:a023812. doi: 10.1101/cshperspect.a023812

Hwang, M., Garbey, M., Berceli, S. A., and Tran-Son-Tay, R. (2009). Rule-based simulation of multi-cellular biological systems—a review of modeling techniques. *Cell. Mol. Bioeng.* 2, 285–294. doi: 10.1007/s12195-009-0078-2

Karr, J. R., Sanghvi, J. C., Macklin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival, B., et al. (2012). A whole-cell computational model predicts phenotype from genotype. *Cell* 150, 389–401. doi: 10.1016/j.cell.2012.05.044

Keasling, J. D. (2012). Synthetic biology and the development of tools for metabolic engineering. *Metab. Eng.* 14, 189–195. doi: 10.1016/j.ymben.2012.01.004

Khalil, A. S., and Collins, J. J. (2010). Synthetic biology: applications come of age. *Nat. Rev. Genet.* 11: 367. doi: 10.1038/nrg2775

Khoury, G. A., Smadbeck, J., Kieslich, C. A., and Floudas, C. A. (2014). Protein folding and *de novo* protein design for biotechnologicalapplications.

Trends Biotechnol. 32, 99–109. doi: 10.1016/j.tibtech.2013. 10.008

Kosuri, S., and Church, G. M. (2014). Large-scale *de novo* DNA synthesis: technologies and applications. *Nat. Methods* 11, 499. doi: 10.1038/nmeth.2918

Le Novère, N. (2015). Quantitative and logic modelling of molecular and gene networks. *Nat. Rev. Genet.* 16: 146. doi: 10.1038/nrg3885

Libbrecht, M. W., and Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* 16: 321. doi: 10.1038/nrg3920

Liu, R., Bassalo, M. C., Zeitoun, R. I., and Gill, R. T. (2015). Genome scale engineering techniques for metabolic engineering. *Metab. Eng.* 32, 143–154. doi: 10.1016/j.ymben.2015.09.013

O'Brien, E. J., Monk, J. M., and Palsson, B. O. (2015). Using genome-scale models to predict biological capabilities. *Cell* 161, 971–987. doi: 10.1016/j.cell.2015.05.019

Pirooznia, M., Yang, J. Y., Yang, M. Q., and Deng, Y. (2008). A comparative study of different machine learning methods on microarray gene expression data. *BMC Genomics* 9:S13. doi: 10.1186/1471-2164-9-S1-S13

Sopko, R., Huang, D., Preston, N., Chua, G., Papp, B., Kafadar, K., et al. (2006). Mapping pathways and phenotypes by systematic gene overexpression. *Mol. Cell* 21, 319–330. doi: 10.1016/j.molcel.2005.12.011

Stephanopoulos, G. (2012). Synthetic biology and metabolic engineering. *ACS Synth. Biol.* 1, 514–525. doi: 10.1021/sb300094q

Wang, J. W., Wang, A., Li, K., Wang, B., Jin, S., Reiser, M., et al. (2015). CRISPR/Cas9 nuclease cleavage combined with Gibson assembly for seamless cloning. *BioTechniques* 58, 161–170. doi: 10.2144/000 114261

frontiers
in Bioengineering and Biotechnology

High-Performance Biocomputing in Synthetic Biology–Integrated Transcriptional and Metabolic Circuits

Angel Goñi-Moreno[1]* and Pablo I. Nikel[2]

[1] School of Computing, Newcastle University, Newcastle upon Tyne, United Kingdom, [2] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby, Denmark

Biocomputing uses molecular biology parts as the hardware to implement computational devices. By following pre-defined rules, often hard-coded into biological systems, these devices are able to process inputs and return outputs—thus *computing* information. Key to the success of any biocomputing endeavor is the availability of a wealth of molecular tools and biological motifs from which functional devices can be assembled. Synthetic biology is a fabulous *playground* for such purpose, offering numerous genetic parts that allow for the rational engineering of genetic circuits that mimic the behavior of electronic functions, such as *logic gates*. A grand challenge, as far as biocomputing is concerned, is to expand the molecular hardware available beyond the realm of genetic parts by tapping into the host metabolism. This objective requires the formalization of the interplay of genetic constructs with the rest of the cellular machinery. Furthermore, the field of metabolic engineering has had little intersection with biocomputing thus far, which has led to a lack of definition of metabolic dynamics as computing basics. In this perspective article, we advocate the conceptualization of metabolism and its motifs as the way forward to achieve *whole-cell biocomputations*. The design of merged transcriptional and metabolic circuits will not only increase the amount and type of information being processed by a synthetic construct, but will also provide fundamental control mechanisms for increased reliability.

Keywords: biocomputing, synthetic biology, metabolic engineering, boolean logic, genetic circuits, metabolic networks

BIOCOMPUTING

Computation can be broadly defined as the formal procedure by which input information is processed according to pre-defined rules and turned into output data. Since this definition does not specify the type of *information* and *rules* involved in the process, it is applicable to electronic devices as well as to biological systems. In other words, biological systems *do* perform computations. While the computational ability of biological matter has been explicitly described a number of times along the twentieth century (Bennett, 1982), it was Leonard Adleman who showed the feasibility of implementing human-defined computations with molecular (i.e., genetic) hardware (Adleman, 1994). Although the discussion on what would be the equivalent of computer hardware and software in biological systems is still largely open (Danchin, 2009), the term *hardware* in this

article identifies any physical, tangible component (e.g., nucleic acids or metabolites) in a cell. On this first example of biocomputation, Adleman physically encoded an instance of the Hamiltonian path problem (a well-known mathematical problem in graph theory) in DNA strands, and solved it *in vitro* by using routine molecular biology methods. A bacterial computer (i.e., an *in vivo* computer), would solve an instance of the same problem 15 years later (Baumgardner et al., 2009). By the end of last century, Weiss et al. (2002) showed that synthetic regulatory networks could be conceptualized *in vivo* as a series of Boolean logic gates–the key device of cellular computers. This novel conceptual framework set the start of a frantic wave of electronic-inspired bioengineering in synthetic biology. Additionally, these seminal works also shifted the inspiration within the biocomputing community drastically, from mathematics and computer science to electronic engineering.

WHOLE-CELL BIOCOMPUTATIONS

Cells are able to process input information in many different and intricate ways. For the sake of clarity, in this article we propose to group the processing of information into two types of computing (i.e., *genetic* and *metabolic*) depending on the nature of the input and components thereof. To date, most of the biocomputing developments in synthetic biology dealt almost exclusively with genetic material and parts. This type of approach limits the scope of the potential synthetic biocomputations that can be executed, since a number of important resources are not being utilized. In a challenging paper entitled "It's the metabolism, stupid!," de Lorenzo (2015) suggested that "the interplay of DNA and metabolism is [...] akin to that of politics and economy. Both realms drive their own autonomous agendas and obviously influence each other." In a similar fashion, the field of *heterotic computing* (Kendon et al., 2015) advocates the use of various types of computing that merge the strengths of individual types into more powerful, *heterotic* devices.

Synthetic Biology as an Active Biocomputing Field

Boolean logic is central to the field of computing. Therefore, the design and implementation of Boolean logic functions in cells—typically encoded into genetic material (**Figure 1A**)—is key to the development of synthetic biology approaches rooted on biocomputing (Amos and Goñi-Moreno, 2018). The engineering of a genetic *toggle* switch (Gardner et al., 2000) and an *oscillator* (Elowitz and Leibler, 2000) in *Escherichia coli* at the onset of the twenty-first century had set the start of what is a very active field nowadays. Over the last (almost) 20 years, a number of circuits have been successfully engineered in living cells, such as logic gates mentioned above (Wang et al., 2011), counters (Friedland et al., 2009), multiplexers (Moon et al., 2011), adders (Ausländer et al., 2012), and memories (Bonnet et al., 2013). Inspired by computer science, distributed computations have also been designed and build in multicellular systems by modifying cell-cell communication programmes (Goñi-Moreno et al., 2011, 2019; Regot et al., 2011). From solving

relatively simple mathematical problems to compute intricate Boolean logic operations, biological systems have proved to be a powerful platform for tackling applications that are restricted to traditional "silicon-based" computer technologies, such as diagnosis, bioproduction, and bioremediation.

Synthetic biocomputing circuits are growingly more complex and accurate every day, mostly due to endless efforts in improving the genetic toolkit (Silva-Rocha et al., 2013; Martínez-García et al., 2014; Durante-Rodríguez et al., 2018), mathematical methods (Church et al., 2014; Goñi-Moreno and Amos, 2015) and design procedures (Goñi-Moreno and Amos, 2012; McLaughlin et al., 2018) for the so-called *design-build-test-learn* synthetic biology cycle (Goñi-Moreno et al., 2016). There are, nevertheless, major challenges on the genetic computing front (Goñi-Moreno, 2014; Manzoni et al., 2016), such as the urgent need for standardization of components, measurements, and information (Myers et al., 2017; Fabre and Sonnenschein, 2019). As long as synthetic biology claims to be a true engineering discipline, such a standardization problem must be tackled without delay to enable *bona fide* modularity and predictability of genetic circuits (Vilanova et al., 2015). Altogether, the implementation of biocomputations using genetic material is driven by an excellent scientific momentum at the present time.

Metabolic Engineering as a *Potential* Biocomputing Field

While there is a phenomenal potential for development, the metabolic aspect of computation has not been explored to the same degree as it has been implemented *via* genetic circuits (**Figure 1B**). This fact arises from a still-limited knowledge on the complexity of metabolic networks even in the so-called "model" organisms (Benedetti et al., 2016; Calero and Nikel, 2019). Nielsen and Keasling (2016) have recently stressed the presence of metabolic networks with hard wired, tightly regulated lines of communication in virtually all living cells–which are inherently difficult to manipulate but, as the very definition implies, offer an unique opportunity for engineering multi-level computations. In the same way synthetic biology uses genetic parts and devices to build complex systems with pre-defined behaviors, metabolic networks are characterized by some (more or less conserved) principles that can be used for re-purposing biochemical nodes. The bowtie model of central metabolism indicates that the core biochemistry of the cell includes the biochemical transformations necessary for the synthesis of the 12 known essential biomass precursors (Noor et al., 2010). This architecture requires a high level of regulation, especially at the level of gene transcription (Kochanowski et al., 2017). Is precisely at this intersection between cellular processes that biocomputing could play a role in re-programming the metabolic machinery of cells.

High-Performance Biocomputing

Natural cellular pathways are rarely based on genetic or metabolic activities alone. Thus, the concept of heterotic computing (i.e., the coordination between different types of computing), is intrinsic to biological systems. However, synthetic circuits are not often exploiting the full computational power of the cellular

FIGURE 1 | Interfacing genetic and metabolic processes for high-performance biocomputations. **(A)** Biocomputing circuits are typically encoded into genetic material. Synthetic biology provides an extensive toolkit of genetic parts and devices that are assembled to build combinatorial (and even sequential) logic circuits. The metabolic environment where the circuit *runs* is often overlooked when it comes to formalize logic motifs. **(B)** The expanding field of biocomputation intersects synthetic biology. Genetic logic circuits have been central to synthetic biology since the formal inception of the discipline. Thus, far, there is no obvious exploitation of this type of biocomputation for metabolic engineering–there is, however, enough synergy between the three disciplines to find an overlapping (sub)field. **(C)** Information processing flows in merged transcriptional and metabolic circuits. Both transcriptional and metabolic networks are able to sense external inputs and yield output responses; the feedback from one layer to the other can effectively communicate information.

machinery. Although the type of processes is very different, the cooperation between them could pave the way to a new generation of whole-cell circuits with enhanced abilities. This aspect is what we refer to as *high-performance biocomputing*.

Against this background, **Figure 1C** shows the flows of information in high-performance biocomputations. A first challenge would be to describe what in computing are called *primitives*, which are the simplest elements with which software programs are built upon. This will result in a set of well-characterized genetic and metabolic units (e.g., coding sequences and metabolic reactions) and motifs (e.g., oscillations and switches), including types of inputs and outputs for each computing end. Although current efforts are individually tackling this challenge in either the genetic (Nielsen et al., 2016) or metabolic (Sánchez-Pascuala et al., 2017) fronts, there is still the issue that genetic and metabolic units must be *plugged* together to allow information flow in both directions. This connectivity will enable the direct modification of genetic motifs by the action of their metabolic counterparts (and vice-versa). Depending on the specific process, and the type of information being computed, either of the two ends could return the desired output.

The increasing focus on the interplay between genetic and metabolic networks (Shlomi et al., 2007; Kumar et al., 2018)

is resulting in a revolution of metabolic engineering driven by the core principles of synthetic biology. Not only molecular tools are actively being developed (Keasling, 2012; Nikel and de Lorenzo, 2018), but also control strategies to engineer genetic circuits are being increasingly exploited for the regulation of metabolism in a pre-defined fashion (Oyarzún and Stan, 2013; Chen and Liu, 2018; Moser et al., 2018). The foundations of high-performance biocomputing are therefore established and ready to benefit from the input of the computing community. Yet, a solid representation framework is needed to fully realize this purpose.

A UNIFIED REPRESENTATION FRAMEWORK

Boolean logic is a way to abstract the underlying mechanistic details of a device into its high-level functional performance. By doing so, gene expression can be abstracted into *ON/OFF* states (i.e., either the gene is, or is not, being expressed under a given environmental condition) regardless of the particularities the gene of interest might have. Even in the case of radical analog fluctuations in gene expression, the *ON/OFF* abstraction still provides an useful conceptual framework (García-Betancur et al., 2017; Goñi-Moreno et al., 2017). However, when it
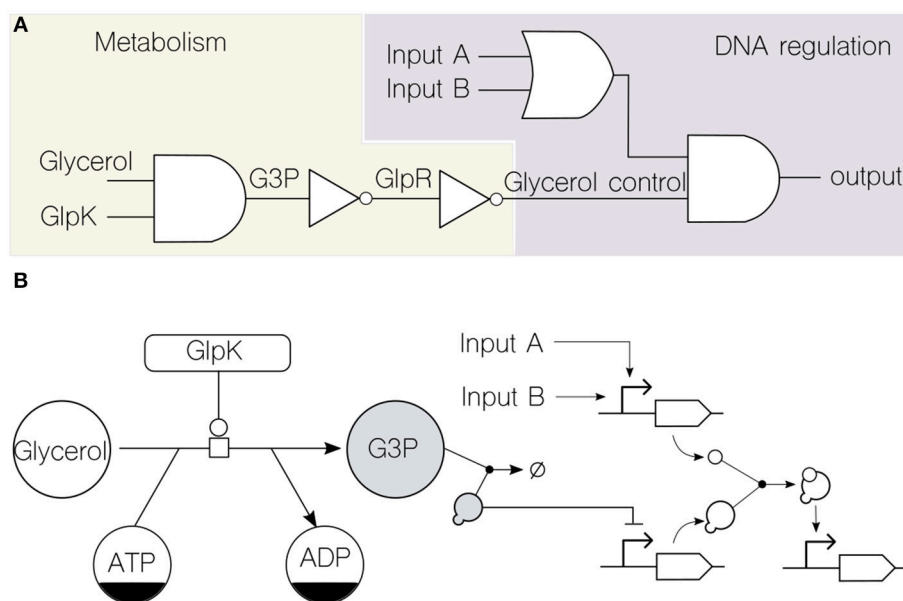
FIGURE 2 | Formal representation of merged transcriptional and metabolic circuits. **(A)** Circuit formalization into Boolean functions (i.e., logic gates) assists the combination of metabolism and DNA regulation. Using the known transcriptional and metabolic network that rules glycerol consumption in the soil bacterium *P. putida* KT2440 (Nikel et al., 2014) as an example, the circuit depicts the role of glycerol as an input along with other signals (*inputs A* and *B*) typically used in transcriptional logic. Note that the flow of information is bidirectional, since the metabolic GlpK input (a key enzyme involved in glycerol consumption) can be modified by the genetic circuit. This top-level logic design enables the abstraction of details about the type of substrate used by providing a unified computing framework. **(B)** The same glycerol circuit is formalized through the adoption of existing representation standards: the *Systems Biology Graphical Notation* (SBGN) for metabolic networks and the *Synthetic Biology Open Language* (SBOL) visual for genetic circuits. The two shaded components, the key metabolite glycerol-3-phosphate (G3P) and the transcriptional repressor GlpR, constitute the physical link that merge both computing layers.

comes to implementation, the Boolean abstraction needs to be complemented by a dynamical analysis of the components at stake. For example, the time-scales with which genetic and metabolic interactions occur can potentially be very different. Therefore, the dynamic analysis of individual reactions is as fundamental as the functional representation of the system as a whole.

To illustrate this point, we discuss a case of merged genetic and metabolic circuitry integration in a platform bacterium. **Figure 2A** shows a logic-gate representation of a simple merged transcriptional and metabolic circuit in the soil bacterium *Pseudomonas putida* KT2440. This device merges state-of-the-art DNA regulatory circuitry (Nielsen et al., 2016) with dynamics that are far beyond DNA reach: the metabolic ability of the cells to catabolize glycerol (Nikel et al., 2015). In this way, the circuit output depends not only in the upstream computation of typical genetic inputs (generic *inputs A* and *B* in the diagram) but also in the metabolic dynamics of glycerol uptake. The link that enables the functioning of the circuit is the transcriptional repressor GlpR, which somewhat encodes information about the metabolic state of the cell (the action of GlpR on the cognate *glp* gene cluster is relieved by the metabolite glycerol-3-phosphate, G3P) and acts on a specific promoter. Note that virtually any other signaling molecule or transcription factor that feeds the final genetic AND logic gate can be inserted downstream of this promoter. Moreover, any regulatory step

in the circuit can be connected back to, e.g., the key enzyme GlpK (essential for glycerol processing) thus providing feedback control from the genetic to the metabolic side of the device. As a result, the combinatorial genetic logic circuit is now linked to the physiological state of the cell concerning the dynamics of carbon source uptake, which can be both read and controlled.

Using a lower, more specific, layer of representation, the *Systems Biology Graphical Notation* (SBGN) (Le Novère et al., 2009) and the *Synthetic Biology Open Language* (SBOL) (Galdzicki et al., 2014) were used to formalize the circuit (**Figure 2B**) for both metabolic and genetic parts, respectively. This helps identifying the link, which in this case is formed by the interaction between a metabolite (G3P) and a transcriptional repressor (GlpR)—thereby merging the metabolic and genetic layers of regulation in the bacterial cell (**Figure 2B**).

We recently coined the term *metabolic widget* to refer to such merged circuits (Chavarría et al., 2016). The metabolic machinery of the cell, often referred to as the *context* when focusing on genetic logic, offers powerful resources that can greatly improve current biocomputations. Far from trying to avoid the context, the framework proposed herein is taking full advantage of it, which can lead to widgets that assist more complex and accurate pre-defined processes of information. The adoption of such a configuration will have a double impact by providing essential information about both the metabolic and genetic wiring of

the cell while taking full advantage of these interactions for re-programming core cellular functions.

From a broader perspective, evolution has shaped intricate cellular processes that merge both genetic and metabolic networks; yet, human-defined biocomputations rarely make use of both computing types. We advocate for taking this path into account in order to access and exploit the high-performance biocomputing power intrinsic to natural systems, empowering the design-build-test-learn cycle to entirely new directions.

AUTHOR CONTRIBUTIONS

Both authors have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

Adleman, L. M. (1994). Molecular computation of solutions to combinatorial problems. *Science* 266, 1021–1024. doi: 10.1126/science.7973651

Amos, M., and Goñi-Moreno, A. (2018). "Cellular computing and synthetic biology," in *Computational Matter,* eds S. Stepney, S. Rasmussen, and M. Amos (Berlin: Springer) 93–110.

Ausländer, S., Ausländer, D., Müller, M., Wieland, M., and Fussenegger, M. (2012). Programmable single-cell mammalian biocomputers. *Nature* 487, 123–127. doi: 10.1038/nature11149

Baumgardner, J., Acker, K., Adefuye, O., Crowley, S. T., Deloache, W., Dickson, J. O., et al. (2009). Solving a Hamiltonian path problem with a bacterial computer. *J. Biol. Eng.* 3:11. doi: 10.1186/1754-1611-3-11

Benedetti, I., de Lorenzo, V., and Nikel, P. I. (2016). Genetic programming of catalytic *Pseudomonas putida* biofilms for boosting biodegradation of haloalkanes. *Metab. Eng.* 33, 109–118. doi: 10.1016/j.ymben.2015.11.004

Bennett, C. H. (1982). The thermodynamics of computation—a review. *Int. J. Theor. Phys.* 21, 905–940. doi: 10.1007/BF02084158

Bonnet, J., Yin, P., Ortiz, M. E., Subsoontorn, P., and Endy, D. (2013). Amplifying genetic logic gates. *Science* 340, 599–603. doi: 10.1126/science.1232758

Calero, P., and Nikel, P. I. (2019). Chasing bacterial *chassis* for metabolic engineering: a perspective review from classical to non-traditional microorganisms. *Microb. Biotechnol.* 12, 98–124. doi: 10.1111/1751-7915.13292

Chavarría, M., Goñi-Moreno, Á., de Lorenzo, V., and Nikel, P. I. (2016). A metabolic widget adjusts the phosphoenolpyruvate-dependent fructose influx in *Pseudomonas putida*. *mSystems* 1, e00154–e00116. doi: 10.1128/mSystems.00154-16

Chen, X., and Liu, L. (2018). Gene circuits for dynamically regulating metabolism. *Trends Biotechnol.* 36, 751–754. doi: 10.1016/j.tibtech.2017.12.007

Church, G. M., Elowitz, M. B., Smolke, C. D., Voigt, C. A., and Weiss, R. (2014). Realizing the potential of synthetic biology. *Nat. Rev. Mol. Cell Biol.* 15, 289–294. doi: 10.1038/nrm3767

Danchin, A. (2009). Bacteria as computers making computers. *FEMS Microbiol. Rev.* 33, 3–26. doi: 10.1111/j.1574-6976.2008.00137.x

de Lorenzo, V. (2015). It's the metabolism, stupid! *Environ. Microbiol. Rep.* 7, 18–19. doi: 10.1111/1758-2229.12223

Durante-Rodríguez, G., de Lorenzo, V., and Nikel, P. I. (2018). A post-translational metabolic switch enables complete decoupling of bacterial growth from biopolymer production in engineered *Escherichia coli*. *ACS Synth. Biol.* 7, 2686–2697. doi: 10.1021/acssynbio.8b00345

Elowitz, M. B., and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 335–338. doi: 10.1038/35002125

Fabre, M. M., and Sonnenschein, N. (2019). Improving reproducibility in synthetic biology. *Front. Biotechnol. Bioeng.* 7:18. doi: 10.3389/fbioe.2019.00018

Friedland, A. E., Lu, T. K., Wang, X., Shi, D., Church, G., and Collins, J. J. (2009). Synthetic gene networks that count. *Science* 324, 1199–1202. doi: 10.1126/science.1172005

Galdzicki, M., Clancy, K. P., Oberortner, E., Pocock, M., Quinn, J. Y., Rodriguez, C. A., et al. (2014). The *Synthetic Biology Open Language* (SBOL) provides a community standard for communicating designs in synthetic biology. *Nat. Biotechnol.* 32, 545–550. doi: 10.1038/nbt.2891

García-Betancur, J. C., Goñi-Moreno, A., Horger, T., Schott, M., Sharan, M., Eikmeier, J., et al. (2017). Cell differentiation defines acute and chronic infection cell types in *Staphylococcus aureus*. *eLife* 6:e28023. doi: 10.7554/eLife.28023

Gardner, T. S., Cantor, C. R., and Collins, J. J. (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403, 339–342. doi: 10.1038/35002141

Goñi-Moreno, A. (2014). On genetic logic circuits: forcing digital electronics standards? *Memetic Comput.* 6, 149–155. doi: 10.1007/s12293-014-0136-8

Goñi-Moreno, A., and Amos, M. (2012). Continuous computation in engineered gene circuits. *BioSystems* 109, 52–56. doi: 10.1016/j.biosystems.2012.02.001

Goñi-Moreno, A., and Amos, M. (2015). "DiSCUS: a simulation platform for conjugation computing," in *International Conference on Unconventional Computation and Natural Computation*, (Auckland), 181–191.

Goñi-Moreno, A., Benedetti, I., Kim, J., and de Lorenzo, V. (2017). Deconvolution of gene expression noise into spatial dynamics of transcription factor–promoter interplay. *ACS Synth. Biol.* 6, 1359–1369. doi: 10.1021/acssynbio.6b00397

Goñi-Moreno, A., Carcajona, M., Kim, J., Martínez-García, E., Amos, M., and de Lorenzo, V. (2016). An implementation-focused bio/algorithmic workflow for synthetic biology. *ACS Synth. Biol.* 5, 1127–1135. doi: 10.1021/acssynbio.6b00029

Goñi-Moreno, A., de la Cruz, F., Rodríguez-Patón, A., and Amos, M. (2019). Dynamical task switching in cellular computers. *Life* 9:E14. doi: 10.3390/life9010014

Goñi-Moreno, A., Redondo-Nieto, M., Arroyo, F., and Castellanos, J. (2011). Biocircuit design through engineering bacterial logic gates. *Nat. Comput.* 10, 119–127. doi: 10.1007/s11047-010-9184-2

Keasling, J. D. (2012). Synthetic biology and the development of tools for metabolic engineering. *Metab. Eng.* 14, 189–195. doi: 10.1016/j.ymben.2012.01.004

Kendon, V., Sebald, A., and Stepney, S. (2015). Heterotic computing: past, present and future. *Philos. Trans. R. Soc. A* 373:20140225. doi: 10.1098/rsta.2014.0225

Kochanowski, K., Gerosa, L., Brunner, S. F., Christodoulou, D., Nikolaev, Y. V., and Sauer, U. (2017). Few regulatory metabolites coordinate expression of central metabolic genes in *Escherichia coli*. *Mol. Syst. Biol.* 13:903. doi: 10.15252/msb.20167402

Kumar, S., Mahajan, S., and Jain, S. (2018). Feedbacks from the metabolic network to the genetic network reveal regulatory modules in *E. coli* and *B. subtilis*. *PLoS ONE* 13:e0203311. doi: 10.1371/journal.pone.0203311

Le Novère, N., Hucka, M., Mi, H., Moodie, S., Schreiber, F., Sorokin, A., et al. (2009). The systems biology graphical notation. *Nat. Biotechnol.* 27, 735–741. doi: 10.1038/nbt.1558

Manzoni, R., Urrios, A., Velázquez-García, S., de Nadal, E., and Posas, F. (2016). Synthetic biology: insights into biological computation. *Integr. Biol.* 8, 518–532. doi: 10.1039/C5IB00274E

Martínez-García, E., Aparicio, T., Goñi-Moreno, A., Fraile, S., and de Lorenzo, V. (2014). SEVA 2.0: an update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities. *Nucleic Acids Res.* 43, D1183–D1189. doi: 10.1093/nar/gku1114

McLaughlin, J. A., Myers, C. J., Zundel, Z., Misirli, G., Zhang, M., et al. (2018). SynBioHub: a standards-enabled design repository for synthetic biology. *ACS Synth. Biol.* 7, 682–688. doi: 10.1021/acssynbio.7b00403

Moon, T. S., Clarke, E. J., Groban, E. S., Tamsir, A., Clark, R. M., Eames, M., et al. (2011). Construction of a genetic multiplexer to toggle between chemosensory pathways in *Escherichia coli. J. Mol. Biol.* 406, 215–227. doi: 10.1016/j.jmb.2010.12.019

Moser, F., Espah Borujeni, A., Ghodasara, A. N., Cameron, E., Park, Y., and Voigt, C. A. (2018). Dynamic control of endogenous metabolism with combinatorial logic circuits. *Mol. Syst. Biol.* 14:e8605. doi: 10.15252/msb.20188605

Myers, C. J., Beal, J., Gorochowski, T. E., Kuwahara, H., Madsen, C., McLaughlin, J. A., et al. (2017). A standard-enabled workflow for synthetic biology. *Biochem. Soc. Trans.* 45, 793–803. doi: 10.1042/BST20160347

Nielsen, A. A., Der, B. S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E. A., et al. (2016). Genetic circuit design automation. *Science* 352:aac7341. doi: 10.1126/science.aac7341

Nielsen, J., and Keasling, J. D. (2016). Engineering cellular metabolism. *Cell* 164, 1185–1197. doi: 10.1016/j.cell.2016.02.004

Nikel, P. I., and de Lorenzo, V. (2018). *Pseudomonas putida* as a functional *chassis* for industrial biocatalysis: from native biochemistry to *trans*-metabolism. *Metab. Eng.* 50, 142–155. doi: 10.1016/j.ymben.2018.05.005

Nikel, P. I., Kim, J., and de Lorenzo, V. (2014). Metabolic and regulatory rearrangements underlying glycerol metabolism in *Pseudomonas putida* KT2440. *Environ. Microbiol.* 16, 239–254. doi: 10.1111/1462-2920.12224

Nikel, P. I., Romero-Campero, F. J., Zeidman, J. A., Goñi-Moreno, A., and de Lorenzo, V. (2015). The glycerol-dependent metabolic persistence of *Pseudomonas putida* KT2440 reflects the regulatory logic of the GlpR repressor. *mBio* 6:e00340. doi: 10.1128/mBio.00340-15

Noor, E., Eden, E., Milo, R., and Alon, U. (2010). Central carbon metabolism as a minimal biochemical walk between precursors for biomass and energy. *Mol. Cell* 39, 809–820. doi: 10.1016/j.molcel.2010.08.031

Oyarzún, D. A., and Stan, G. B. V. (2013). Synthetic gene circuits for metabolic control: design trade-offs and constraints. *J. R. Soc. Interface* 10:20120671. doi: 10.1098/rsif.2012.0671

Regot, S., Macia, J., Conde, N., Furukawa, K., Kjellén, J., Peeters, T., et al. (2011). Distributed biological computation with multicellular engineered networks. *Nature* 469, 207–211. doi: 10.1038/nature09679

Sánchez-Pascuala, A., de Lorenzo, V., and Nikel, P. I. (2017). Refactoring the Embden–Meyerhof–Parnas pathway as a whole of portable *GlucoBricks* for implantation of glycolytic modules in Gram-negative bacteria. *ACS Synth. Biol.* 6, 793–805. doi: 10.1021/acssynbio.6b00230

Shlomi, T., Eisenberg, Y., Sharan, R., and Ruppin, E. (2007). A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol. Syst. Biol.* 3:101. doi: 10.1038/msb4100141

Silva-Rocha, R., Martínez-García, E., Calles, B., Chavarría, M., Arce-Rodríguez, A., de las Heras, A., et al. (2013). The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res.* 41, D666–D675. doi: 10.1093/nar/gks1119

Vilanova, C., Tanner, K., Dorado-Morales, P., Villaescusa, P., Chugani, D., Frías, A., et al. (2015). Standards not that standard. *J. Biol. Eng.* 9:17. doi: 10.1186/s13036-015-0017-9

Wang, B., Kitney, R. I., Joly, N., and Buck, M. (2011). Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nat. Commun.* 2:508. doi: 10.1038/ncomms1516

Weiss, R., Homsy, G. E., and Knight, T. F. (2002). "Toward *in vivo* digital circuits," in *Evolution as Computation*, eds L. F. Landweber and E. Winfree (Berlin, Heidelberg: Springer), 275–295.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Biological Parts for *Kluyveromyces marxianus* Synthetic Biology

Arun S. Rajkumar[1], Javier A. Varela[1], Hannes Juergens[2], Jean-Marc G. Daran[2] and John P. Morrissey[1]*

[1] School of Microbiology, Centre for Synthetic Biology and Biotechnology, Environmental Research Institute, APC Microbiome Institute, University College Cork, Cork, Ireland, [2] Department of Biotechnology, Delft University of Technology, Delft, Netherlands

Kluyveromyces marxianus is a non-conventional yeast whose physiology and metabolism lends itself to diverse biotechnological applications. While the wild-type yeast is already in use for producing fragrances and fermented products, the lack of standardised tools for its genetic and metabolic engineering prevent it from being used as a next-generation cell factory for bio-based chemicals. In this paper, we bring together and characterise a set of native *K. marxianus* parts for the expression of multiple genes for metabolic engineering and synthetic biology. All parts are cloned and stored according to the MoClo/Yeast Tool Kit standard for quick sharing and rapid construction. Using available genomic and transcriptomic data, we have selected promoters and terminators to fine-tune constitutive and inducible gene expression. The collection includes a number of known centromeres and autonomously replication sequences (ARS). We also provide a number of chromosomal integration sites selected for efficiency or visible phenotypes for rapid screening. Finally, we provide a single-plasmid CRISPR/Cas9 platform for genome engineering and facilitated gene targeting, and rationally create auxotrophic strains to expand the common range of selection markers available to *K. marxianus*. The curated and characterised tools we have provided in this kit will serve as a base to efficiently build next-generation cell factories from this alternative yeast. Plasmids containing all parts are available at Addgene for public distribution.

Keywords: *Kluyveromyces*, synthetic biology, metabolic engineering, genome engineering, yeast

INTRODUCTION

Cell factories can serve as the basis of a new bio-based economy based on the sustainable production of fine chemicals, pharmaceuticals, nutraceuticals, and biofuels from engineered or native microbes. As of now, the most commonly-used eukaryotic cell factory is baker's yeast *Saccharomyces cerevisiae*. An ease of genetic manipulation, along with a wealth of genomic, genetic and biochemical knowledge, have made it able to produce a diverse range of compounds with *S. cerevisiae* from simple or economical feedstocks. However, non-*Saccharomyces* yeasts can provide several advantages over *S. cerevisiae* for building cell factories as they often possess desirable tolerances or metabolic traits that would otherwise need to be extensively engineered into baker's yeast (Wagner and Alper, 2016). In general, such yeasts have had niche applications in biotechnology but are being developed to be next-generation cell factories. *Kluyveromyces marxianus* is one such alternative yeast. While thermotolerant, fast-growing and able to use various carbon and nitrogen sources, it broadly has the same nutritional requirements and culture

techniques as *S. cerevisiae*. Its natural strain diversity allows a wide number of phenotypes and gene variants to be exploited and combined to create an optimal cell factory chassis, and as a broadly Crabtree-negative species, it does not need to have metabolism routed away from ethanol production. While it has applications in the production of aroma compounds, fermented foods, and secreted enzymes (Morrissey et al., 2015; Gombert et al., 2016), *K. marxianus* is yet to be established as a metabolic engineering platform. Obstacles to such development include inefficient and random native gene targeting (hindering the stable expression of integrated heterologous genes), limited knowledge of its biochemistry and genetics, and a lack of standardised regulatory parts and expression systems on the level of baker's yeast. While such tools are starting to be developed, it still lacks the well-defined sets of regulatory elements to precisely control gene expression one uses in *S. cerevisiae*. As of now, individual parts are either selected from the native genome based on orthologues of genes of common parts from *S. cerevisiae* (Lee et al., 2013), or from other yeasts altogether (Chang et al., 2013), precluding the advantages of using *K. marxianus*' environmental triggers to fine-tune gene expression. Taken together, less than 20 native regulatory parts are currently in use for metabolic engineering (Bergkamp et al., 1993; Ball et al., 1999; Yang et al., 2015a; Gombert et al., 2016).

Kluyveromyces marxianus-specific techniques exist for the efficient *in vivo* assembly of large multigene constructs (Chang et al., 2012), and in conjunction with CRISPR/Cas9 (Löbs et al., 2017; Nambu-Nishida et al., 2017; Cernak et al., 2018) can allow us to specifically edit a genome, or efficiently target chromosmal integrations. Nonetheless *in vivo* assembly as it stands does not eliminate non-specific integrations of incomplete parts of the assembly. To further sidestep this problem, the MoClo standard, based on Golden Gate assembly, allows the efficient hierarchical *in vitro* assembly of multigene constructs either on episomal or integrative vectors for such purposes (Weber et al., 2011). It has been adapted for synthetic biology in diverse organisms, and is efficient enough to circumvent *in vivo* assembly. One variant of MoClo, the Yeast Toolkit (YTK), collects a number of well-characterised parts for *S. cerevisiae* (Lee et al., 2015). The YTK has 8 general classes of parts, defined by the 5' and 3' overhangs used for Golden Gate assembly, which allow directional cloning. Taken together, the parts of the original YTK and the system itself allow for the versatile construction of vectors for *S. cerevisiae* with several selection markers and integration sites of choice if needed.

The YTK's MoClo approach also sets up three tiers of plasmids for storage or use (**Figure 1B**). Level I plasmids correspond to part plasmids. A BsmBI site, and a BsaI site that after digestion will generate overhangs specific to that part type, flank each functional part. The BsmBI sites, in turn, are used to clone new parts into the entry vector YTK001 by Golden Gate assembly with that enzyme. At level II, level I plasmids are assembled together with BsaI to create gene expression cassettes or transcriptional units (TUs). Assembling TUs includes flanking them with synthetic and directional connector sequences which allow the construction of level III, plasmids. Here, multiple TUs are assembled together into a multigene expression or integrative

vector, again with BsmBI based on unique overhangs present in the connectors.

The YTK, either with its standard or simply by using its parts collection, has since been used in diverse metabolic engineering and synthetic biology applications (Feng et al., 2015; Awan et al., 2017; Denby et al., 2018), and has already been adopted when cloning parts for *Pichia pastoris* (Obst et al., 2017). In this paper we present a collection of *K. marxianus*-specific parts and vectors, cloned to the YTK standard, for the genetic and metabolic engineering of this yeast. Over 30 constitutive and inducible promoters, terminators and centromeres and autonomously replicating sequence (ARS) elements have been selected from existing gene expression data and characterised as part of this collection. We have also identified a number of integration sites for the integration of single- and multi-gene constructs and have tested the best means to eliminate random integration. Taken together with existing YTK parts, our parts collection is a valuable set of tools for researchers working with *K. marxianus* as a cell factory.

MATERIALS AND METHODS

Strains and Cultivation

Kluyveromyces marxianus strain CBS6556 (Westerdijk Fungal Biodiversity Institute, Utrecht, The Netherlands) was used for parts, unless specified otherwise. For part characterization and mutant creation, we used NBRC1777 [Biological Resource Centre, NITE (NBRC), Tokyo, Japan]. All strains used in this study are listed in **Table 1**. Yeast cultures were routinely grown in YPD broth (1% yeast extract, 2% peptone, 2% glucose) at 30°C. For promoter and terminator characterization, yeast was grown in synthetic drop-out medium without uracil [SD-ura; 0.17% yeast nitrogen base without amino acids or ammonium sulphate (Formedium, Hunstanton, UK), 0.5% ammonium sulphate and 0.19% SC-ura (Formedium)] containing 2% glucose or an alternate carbon source. For transformation of auxotrophic strains, yeast was grown in synthetic drop-out (SD) medium containing 2% glucose and lacking the appropriate nutrients (Formedium). For fermentation experiments, strains were cultured in minimal medium (MM) with 2% glucose (Fonseca et al., 2007). When needed, G418 or hygromycin (Fisher Scientific, Dublin, Ireland) was added to a concentration of $200 \, \text{mg} \, \text{L}^{-1}$ for selection or $150 \, \text{mg} \, \text{L}^{-1}$ for maintenance. Bacterial transformations used *E.coli* DH5α grown in LB medium (1% NaCl, 1% peptone, 0.5% yeast extract) supplemented with the appropriate antibiotics ($100 \, \text{mg} \, \text{L}^{-1}$ ampicillin, $50 \, \text{mg} \, \text{L}^{-1}$ chloramphenicol or $50 \, \text{mg} \, \text{L}^{-1}$ kanamycin).

Selection and Amplification of Parts

Native promoters, terminators, and replicating sequences were amplified from the genomic DNA of *K. marxianus* CBS6556 (Jeong et al., 2012) unless specified otherwise. The sequences in question were identified from the genome sequence data (assembly GCA 000299195.2) following in-house gene prediction (Varela et al., 2017). In general, a promoter was defined as the first 1,000 bases upstream of a gene's start codon, unless this overlapped with a neighbouring gene or other
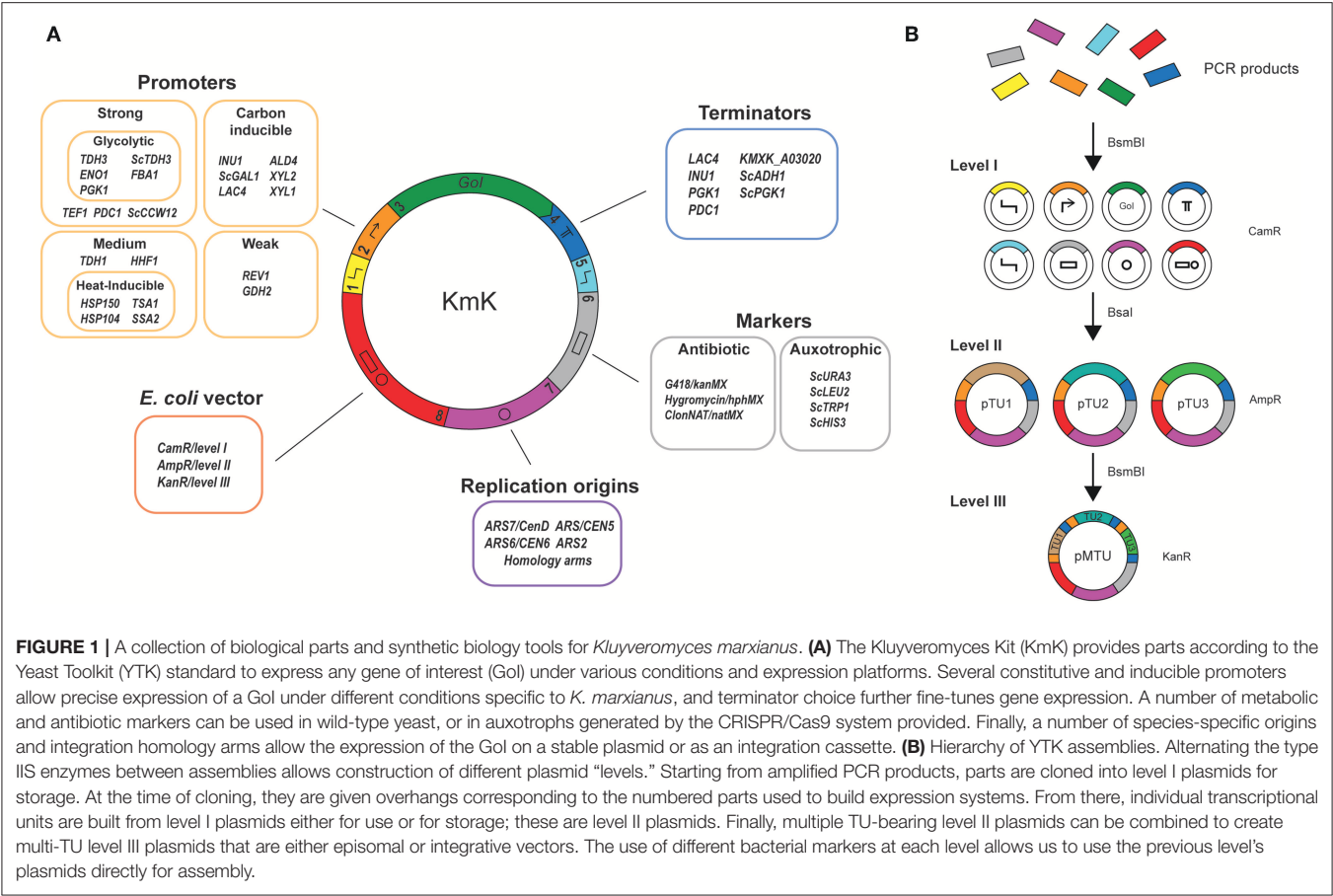
FIGURE 1 | A collection of biological parts and synthetic biology tools for *Kluyveromyces marxianus*. **(A)** The Kluyveromyces Kit (KmK) provides parts according to the Yeast Toolkit (YTK) standard to express any gene of interest (GoI) under various conditions and expression platforms. Several constitutive and inducible promoters allow precise expression of a GoI under different conditions specific to *K. marxianus*, and terminator choice further fine-tunes gene expression. A number of metabolic and antibiotic markers can be used in wild-type yeast, or in auxotrophs generated by the CRISPR/Cas9 system provided. Finally, a number of species-specific origins and integration homology arms allow the expression of the GoI on a stable plasmid or as an integration cassette. **(B)** Hierarchy of YTK assemblies. Alternating the type IIS enzymes between assemblies allows construction of different plasmid "levels." Starting from amplified PCR products, parts are cloned into level I plasmids for storage. At the time of cloning, they are given overhangs corresponding to the numbered parts used to build expression systems. From there, individual transcriptional units are built from level I plasmids either for use or for storage; these are level II plasmids. Finally, multiple TU-bearing level II plasmids can be combined to create multi-TU level III plasmids that are either episomal or integrative vectors. The use of different bacterial markers at each level allows us to use the previous level's plasmids directly for assembly.

TABLE 1 | List of *K. marxianus* strains used in this paper.

Strain name	Source	Genotype	Comments
CBS6556	Westerdijk Fungal Biodiversity Institute, Netherlands	Wild-type	
NBRC1777	NITE Biological Resource Center	Wild-type	
YBL001	This study	NBRC1777 *yku80-1*	Frameshift in codon 346 (double basepair deletion)
KmASR.005	This study	NBRC1777 *dnl4-1*	Frameshift in codon 11 (single base pair insertion)
YBL003	This study	NBRC1777 *nej1-1*	Frameshift in codon 61 (single base pair deletion)
KmASR.006	This study	NBRC1777 *ura3-1*	Frameshift in codon 139 (double base pair deletion)
KmASR.007	This study	NBRC1777 *his3-1*	Frameshift in codon 71 (single base pair insertion)
KmASR.008	This study	NBRC1777 *ura3-1 his3-2*	*HIS3:* Frameshift in codon 126 of *HIS3* (double base pair deletion)
KmASR.022	This study	NBRC1777 *leu2-1*	Frameshift across codons 124 and 125 double base pair deletion)
KmASR.023	This study	NBRC1777 *ura3-1 leu2-1*	*LEU2:* Frameshift across codons 124 and 125 (double base pair deletion)
KmASR.024	This study	NBRC1777 *his3-1 leu2-1*	*LEU2:* Frameshift across codons 124 and 125 (double base pair deletion)
KmASR.025	This study	NBRC1777 *ura3-1 his3-2 leu2-1*	*LEU2:* Frameshift across codons 124 and 125 (double base pair deletion)

features. Terminators were similarly defined as the first 250bp downstream of a gene's Stop codon (Curran et al., 2013). Parts were amplified from genomic DNA using Q5 High-Fidelity Polymerase (New England Biolabs, Ipswich, UK) and purified using a GeneJet PCR clean-up kit (Fisher Scientific) and eluted in sterile water prior to use. Yeast ARS elements and centromeres were selected from the literature, and either amplified from genomic DNA or from an appropriate plasmid. All assembled part plasmids with their Addgene plasmid IDs are listed in **Table S1**, and primers to amplify the parts are listed in **Table S2**.

Golden Gate Assembly

Golden Gate assemblies were carried out as recommended (Lee et al., 2015) with minor modifications. In a typical reaction, 40

fmol of each insert (either as an existing level I plasmid or PCR product) were combined with 20 fmol of the plasmid containing the backbone for the final product to be assembled along with 1 μL T4 DNA ligase 0.5 μL each of T7 DNA ligase (3,000U μL^{-1}) and BsmBI or BsaI, (10U μL^{-1}, NEB) and water to a total volume of 10 μL.

For Golden Gate assembly the protocol was as follows: 25 cycles of digestion and ligation (2 min at 42°C followed by 5 min at 16°C), followed by 10 min digestion at 60°C and 10 min inactivation at 80°C. For assemblies involving BsaI, the digestion step was carried out at 37°C for 3 min. When plasmids containing type IIS restriction sites that could not be removed, the final digestion and heat inactivation steps were eliminated.

Typically, half of a Golden Gate assembly was used to transform chemically competent *E.coli* DH5α. Transformants were screened by colony PCR with OneTaq Quick-Load DNA polymerase (NEB), using primers in the backbone vector flanking the insert if size permitted, or with one primer binding in the backbone and one in the insert. In the latter case, this was usually a primer used to amplify one of the component parts. For level I plasmids, colonies with the correct PCR product size had their plasmids extracted and inserts sequenced. For level II and III plasmids, PCR-positive colonies had their plasmids extracted and digested with NotI to verify the insert size All primers used to genotype strains are listed in **Table S3**, and maps of all the plasmids in **Table S1** are included as **Supplementary Material**.

Construction of Reporter Plasmids

20 fmol of a suitable "vector" level I plasmid from the Yeast Toolkit [pYTK083; (Lee et al., 2015)] were combined with 40 fmol each of the following "insert" level I plasmids: left and right connectors from the Yeast Toolkit (pYTK002 and pYTK067), mVenus (pYTK034), a *kanMX* expression marker for G418 resistance, and the appropriate *K. marxianus* promoters (P1-19), terminators (T1-5), and centromere (typically C1; **Table 2**). A Golden Gate assembly for a level II plasmid was then carried out as described above. In total, 8 parts were assembled and transformed into *E.coli* per plasmid. Assembly of the construct was verified by colony PCR with ASR_K001F and ASR_K005R as primers and subsequent restriction mapping of plasmids from PCR-positive colonies with NotI. All original YTK plasmids used for assemblies are listed in **Table S4**.

Construction of Integrative Reporter Constructs

For the construction of integrative reporter vectors to evaluate insertion sites, a similar 8-part Golden Gate assembly was carried out as above, except that pYTK002 was replaced with plasmids containing 850–900 bp left homology arms targeting an insertion site (I1L-I5L), and C1 was replaced with plasmids containing 850–900 bp right homology arms targeting the same insertion site (I1R-I5R). For the experiments carried out here, for the purpose of evaluating integration sites, mVenus was always under the control of the *PDC1* promoter (P2) and the *INU1* terminator (T1). Following assembly and transformation, transformants were screened by colony PCR with primers ASR_K001F and ASR_P002R, followed by NotI restriction digestion as above.

Promoter and Terminator Characterization

We constructed mVenus (YFP) reporter plasmids to characterise promoter and terminator strength. To minimise variations in expression due to copy number, we used centromeric plasmids with a *kanMX* cassette. While evaluating promoter strengths alone the reporter plasmids used the inulinase terminator in common (*INU1t*). In a similar manner, we used histone B promoter (*HHF1pr*) in common to regulate YFP expression when evaluating terminator strengths. Three hundred nanogram of reporter plasmids were transformed into *K. marxianus* by the LiOAc/PEG method (Gietz, 2014). After 48 h growth on selective medium, three transformant colonies were inoculated into 2 mL YPD with G418 and grown overnight at 30°C with 200 rpm agitation. The following day, the cultures were diluted 100-fold into 2 mL SD medium with 150mg L^{-1} G418 (approximately corresponding to a starting optical density of 0.1) and grown at 30°C for 24 h with 200 rpm shaking. For promoter inductions at high temperatures and xylose, overnight cultures were typically inoculated to a starting OD of 0.2–0.3 so that a comparable cell number would be present after 24 h growth. The cultures were then diluted 5 to 20-fold in identical SD medium on a 96-well microtitre plate, and YFP fluorescence measured on a Sirius HT platereader (MWG/BioTek, Winooski, USA) with excitation and emission set to 485 and 525 nm, respectively (bandwidth 20 nm). After correcting for the autofluorescence of wild-type *K. marxianus*, the signal was normalised to cell number by dividing by the OD at 600 nm. Differences between normalised YFP values under different conditions were tested for statistical significance by a paired t-test, with $p < 0.05$ taken to be significant.

For the characterization of integration sites, 2 μg of integrative plasmid containing an YFP expression cassette and *kanMX* marker was digested with SgsI/AscI (Fisher Scientific) and transformed into yeast. The amount corresponds to approximately 400 fmol of insertion cassette. G418-resistant colonies were screened for correct insertion at the intended locus by colony PCR and these alone were selected for YFP measurements. When the *LAC4* locus was targeted, transformation plates were replica-plated onto YPGal (2% galactose, 2% peptone, 1% yeast extract) containing 200 mg L^{-1} G418 and 40 mg L^{-1} X-Gal (Melford Laboratories, Ipswich, UK). The inability of disrupted *LAC4* to metabolise X-Gal—and not produce a blue dye—was made use of to pre-screen transformant colonies before genotyping them.

Genome Editing Using CRISPR/Cas9

The cross-yeast CRISPR/Cas9 plasmid pUDP002 (Juergens et al., 2018) was modified to pUCC001 to allow easy cloning of new guide RNA (gRNA) targets by Golden Gate assembly. The original gRNA expression cassette of pUDP002 consisted of a target gRNA and structural element flanked by self-cleaving hammerhead and HDV ribozymes at the 5' and 3' ends respectively (Ng and Dean, 2017). This was modified to contain a BsaI cloning site between the hammerhead ribozyme and gRNA structural element (Vyas et al., 2015). The new cassette was assembled from long oligonucleotides (Integrated DNA Technologies, USA) by annealing them in a thermocycler.

TABLE 2 | List of parts provided in the collection.

Part	Source	YTK part type	Gene/ systematic name/ description[a]	Addgene ID for plasmid containing part
REGULATORY PARTS				
PGK1pr/P1	*K. marxianus* CBS6556	2/Promoter	3-Phospho-glycerate kinase/KMXK_0A03010	125034
PDC1pr/P2	*K. marxianus* CBS6556	2/Promoter	Pyruvate decarboxylase/KMXK_0F05000	125035
ENO1pr/P3	*K. marxianus* CBS6556	2/Promoter	Enolase/KMXK_0A03750	125036
TDH1pr/P4	*K. marxianus* CBS6556	2/Promoter	Glyceraldehyde-3-phosphate dehydrogenase isozyme 1/KMXK_0D02420	125037
HSP150pr/P5	*K. marxianus* CBS6556	2/Promoter	Hypothetical cell wall manno-protein HSP150/KMXK_0E02570	125038
INU1pr/P6	*K. marxianus* CBS6556	2/Promoter	Inulinase/KMXK_0A03230	125039
TEF1pr/P7	*K. marxianus* CBS6556	2/Promoter	Translation elongation factor EF alpha-1/KMXK_0G03180	125040
REV1pr/P8	*K. marxianus* CBS6556	2/Promoter	Deoxycytidyl transferase/KMXK_0F03560	125041
ALD4pr/P9	*K. marxianus* CBS6556	2/Promoter	Aldehyde dehydrogenase/KMXK_0E00190	125042
GDH2pr/P10	*K. marxianus* CBS6556	2/Promoter	Glutamate dehydrogenase/KMXK_0B07490	125043
HHF1pr/P11	*K. marxianus* CBS6556	2/Promoter	Histone H4/KMXK_0G01260	125044
TSA1pr/P12	*K. marxianus* CBS6556	2/Promoter	Thioredoxin peroxidase/KMXK_0D06030	125045
HSP104pr/P13	*K. marxianus* CBS6556	2/Promoter	Heat shock protein 104/KMXK_0D01210	125046
SSA2pr/P14	*K. marxianus* CBS6556	2/Promoter	Heat shock protein SSA2/KMXK_0B04700	125047
TDH3pr/P15	*K. marxianus* NBRC1777	2/Promoter	Glyceraldehyde-3-phosphate dehydrogenase isoform 3/KMAR_80062[b]	125048
FBA1pr/P16	*K. marxianus* CBS6556	2/Promoter	Fructose 1,6-bisphosphate aldolase/KMXK_0D04110	125049
XYL1pr/P17	*K. marxianus* CBS6556	2/Promoter	Xylose reductase/KMXK_0A01570	125050
XYL2pr/P18	*K. marxianus* CBS6556	2/Promoter	Xylitol dehydrogenase/KMXK_0H00420	125051
LAC4pr/P19	*K. marxianus* CBS397	2/Promoter	Beta-galactosidase[c]	125052
INU1t/T1	*K. marxianus* CBS6556	4/Terminator	Inulinase	125053
LAC4t/T2	*K. marxianus* CBS6556	4/Terminator	Beta-galactosidase	125054
KMXK_A03020t/T3	*K. marxianus* CBS6556	4/Terminator	KMXK_A03020t	125055
PDC1t/T4	*K. marxianus* CBS6556	4/Terminator	Pyruvate decarboxylase	125056
PGK1t/T5	*K. marxianus* CBS6556	4/Terminator	3 Phospho-glycerate kinase	125057
ORIGINS OF REPLICATION				
KmARS7/CenD/C1	p11256/*K. marxianus* DMKU3-1042 (Yarimizu et al., 2013; Varela et al., 2017)	7/Origin	Minimal ARS and centromeric sequence	125059
KmARS/CEN5/C2	*K. marxianus* CBS6556	7/Origin	ARS and centromere from chromosome V	125060
KmARS/CEN6/C3	*K. marxianus* CBS6556	7/Origin	ARS and centromere from chromosome VI	125061
ARS2/C4	*K. marxianus* CBS6556	7/Origin	Minimal ARS from chromosome II	125063
HOMOLOGY ARMS FOR INSERTION VECTORS				
I1L/I1R	*K. marxianus* CBS6556	1, 7	Chromosome I:10187..11936 (*LAC4*)	125030/125063
I2L/I2R	*K. marxianus* CBS6556	1, 7	Chromosme V: 23743..21744 (Crick Strand) (downstream of *KmBDH1* and *KmBDH2*)	125031/125064
I3L/I3R	*K. marxianus* CBS6556	1, 7	Chromosome IV:388650..390345 (between *KmSWF1* and *KmARO1*)	125032/125065
I4L/I4R	*K. marxianus* CBS6556	1, 7	Chromosome IV:240042..241741 (downstream of *KmHSP104*)	125033/125066
OTHER PARTS				
ScTRP1/M1	*S. cerevisiae* CEN.PK 113-7D	6/Marker	Phosphoribosyl anthranilate isomerase expression cassette from *S. cerevisiae*	125058

[a] Based on the CBS6556 genome annotations generated in Varela et al. (2017); [b] Based on the NBRC annotations in (Inokuma et al., 2015); [c] Based on the genome sequence assembled in Ortiz-Merino et al. (2018).

A plasmid backbone was then amplified from the original pUDP002 using primers pUDP002-F and pUDP002-R, and Gibson assembly was used to create pUCC001 from the two parts.

When using CRISPR/Cas9 to edit a site in the genome, a gRNA sequence targeting the gene of interest was first predicted using the sgRNA software (Xie et al., 2014). Complementary

oligonucleotides comprising the gRNA sequence are designed with 5′ and 3′ overhangs (5′-CGTC-3′ and 5′-AAAC-3′,) on the sense and antisense oligonucleotides respectively, creating sticky ends to be ligated into pUCC001 cut by BsaI. One hundred pico mole oligos are phosphorylated with 1 μL of T4 polynucleotide kinase (10U μL^{-1}, NEB) in a total volume of 10 μL and then denatured and annealed. Fifty femto mole of the annealed gRNA insert is then used with 100 ng of pUCC001 in a Golden Gate reaction with BsaI. Following transformation, the correct insertion of the gRNA was subsequently verified by PCR using the primer Bsa-R and the sense oligo containing the gRNA target, and sequencing of the plasmid. gRNA targets used in this study are listed in **Table S5**.

Three hundred nanograms of a gRNA expression plasmid are used in a typical genome engineering experiment. As *K. marxianus* predominantly repairs double stranded breaks by non-homologous end-joining (Daley et al., 2005), mutations are created around the gRNA target site without providing a repair fragment (Cernak et al., 2018). After 48–72 h growth following a transformation, hygromycin-resistant colonies are then screened for mutations at the targeted locus by colony PCR and sequencing (Jakociunas et al., 2015). If the intended mutations create an observable phenotype (e.g., an auxotrophy), the transformed plate is replica-plated to appropriate medium to pre-screen colonies based on the phenotype. Colonies with frameshift mutations are then grown overnight in YPD and then passaged twice to fresh YPD cultures to ensure loss of the gRNA plasmid before being preserved. When multiple genes were to be edited, sequential mutation was performed; following each mutation, the gRNA plasmid was cured and the mutation confirmed before proceeding with transformation of the next gRNA expression plasmid.

RESULTS

A Collection of Parts and Assembly Pipeline for *Kluyveromyces marxianus* Using the Yeast Toolkit Standard

The biological parts collected and characterized here are intended to provide the same functions for *K. marxianus* in all the relevant categories of YTK parts (**Figure 1A**). A number of the parts have been described and identified elsewhere (Bergkamp et al., 1993; Yang et al., 2015a), but we also provide a more substantial number of characterised promoters, terminators and integration sites for expression cassettes. A number of high- and low-copy number origins have also been identified from *K. marxianus* genomes, or created from minimal elements. We have included three centromeric elements—one minimal (Yarimizu et al., 2013) and two genomic (Iborra and Ball, 1994; Ball et al., 1999)—and a minimal ARS element (Cernak et al., 2018) for the construction of expression vectors (**Table 2**).While the parts can be used to construct *K. marxianus* expression systems in general, they are optimally compatible with selection markers, bacterial vectors and synthetic connectors in the original YTK available from Addgene (#1000000061). They allow the *in vitro* construction of cloning and expression systems

for *K. marxianus* with the same flexibility one can for *S. cerevisiae*. Golden Gate assembly is an established *in vitro* assembly technique, and with it we were able to assemble up to 8 part-containing plasmids into reporter plasmids to characterize our parts.

A Set of Native Promoters to Fine-Tune Gene Expression

When selecting native yeast promoters to use in strain engineering, an important source is gene expression studies under conditions of interest. While a number of gene expression studies have been carried out on *K. marxianus* strains, few of them have focused on strains with the potential to be synthetic biology chassis, or on have instead focused on explicitly industrial conditions (Gao et al., 2015; Schabort et al., 2016; Diniz et al., 2017). To select promoters, we turned to two studies based on the strain DMKU3-1042, which also has the best quality publicly-available genome. One was TSS-seq transcriptome data published alongside its genome (Lertwattanasakul et al., 2015), as well as a small-scale gene expression study using fluorescent reporters of genes involved in carbon metabolism (Suzuki et al., 2015).The former gave us the opportunity to select not only promoters with distinct strengths, but also those induced by high temperature and xylose. Following this, the corresponding promoters were then identified from the CBS6556 genome. We excluded promoters with internal BsaI or BsmBI sites, as we lacked information on regulatory elements to justify removing the sites by mutagenesis.

This analysis allowed us to select 19 promoters: 10 constitutive and 9 inducible by heat, xylose, lactose and inulin, whose strengths we characterised using YFP reporter assays (**Table 2**, **Figure 2A**; **Figure S1**). While the promoter sequences came from CBS6556, we tested them in strain NBRC1777, due to its faster growth rate and superior thermotolerance compared to the former. Under standard conditions (30°C, glucose-rich medium), we provide a broad selection of promoter strengths for gene expression. From the weakest (*REV1pr*) to the strongest (*PDC1pr*) promoters, we can achieve a 40-fold range of gene expression in NBRC1777. The strongest promoters were those of genes involved in central carbon metabolism (*PDC1pr, FBA1pr, TDH3pr*), as well as that of the orthologue of the translation elongation factor EF-1α (*TEF1*) in *S. cerevisiae*. Interestingly, several of the same orthologous genes in *S. cerevisiae* also have strong promoters. However, the latter do not always achieve strong expression when used in *K. marxianus* (**Figure S2**)— a trade-off against the advantages of using an orthogonal yeast promoter.

Kluyveromyces marxianus's thermotolerance provides a unique induction signal for this yeast's promoters. With this in mind, we picked three heat-inducible promoters (*HSP104pr, SSA2pr, TSA1pr*) with different fold-induction at high temperatures, and one intended to be stable at ambient and high temperatures (*HSP150pr*) based on expression data for the strain DMKU3-1042 as earlier (Lertwattanasakul et al., 2015). All of our inducible promoters selected from expression data
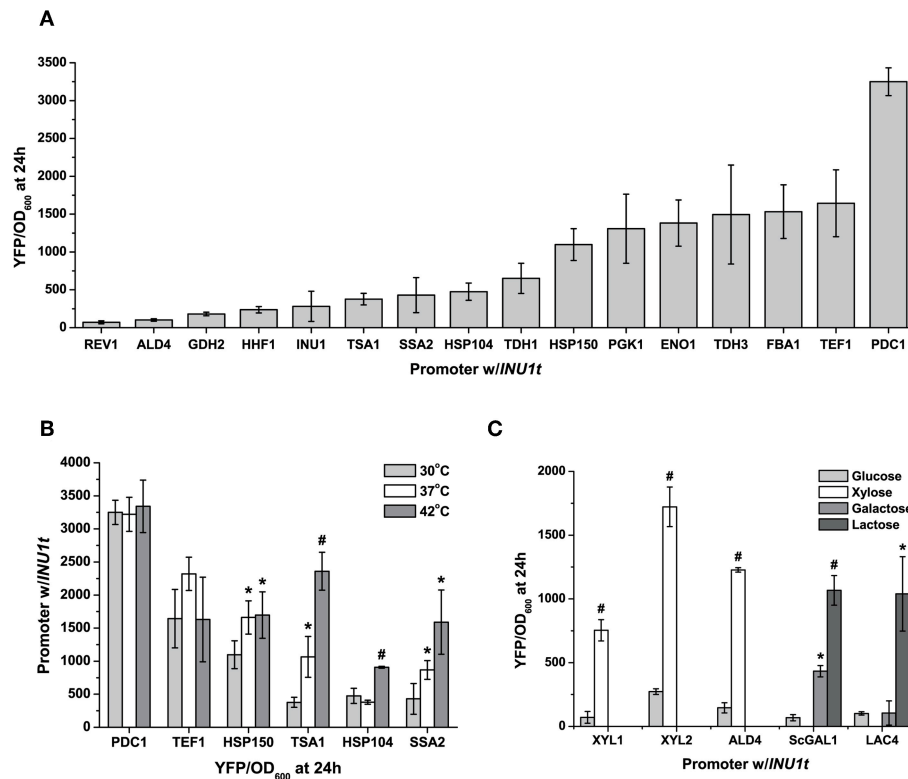
FIGURE 2 | Characterization of *K. marxianus* promoters. **(A)** A set of constitutive promoters, taken from genes with diverse functions, provide a wide range of expression levels—nearly two orders of magnitude—as seen from the YFP output under standard conditions (24 h at 30°C, 2% glucose in synthetic complete medium) **(B)** Expression of heat-inducible promoters at three different temperatures at which NBRC1777 can grow. Constitutive promoters as well (*PDC1pr*) have a stable output at high temperatures. **(C)** The diverse carbon source utilization of *K. marxianus* gives us a unique induction signal for this yeast, as seen by the induction of promoters by lactose, galactose, and xylose. All data are normalised to cell number using OD, and are plotted as the mean ± s.d. of at least three replicates. YFP values significantly different from those under baseline conditions (30 degrees in drop-out medium with 2% glucose) are marked with an asterisk (*p* < 0.05) or a hash (*p* < 0.001).

exhibited induction under the relevant conditions, though not all to the same extent (**Table S6**). When tested at 37 and 42°C, *TSA1pr*, *SSA2pr,* and *HSP104pr* promoters induced to different levels. Expression of YFP by *HSP150pr*, as predicted, remained relatively stable at higher temperatures, inducing weakly when exposed to 37°C but no further at 42°C (**Figure 2B**). The other three promoters offer a range of induction varying from 2 to 6.5-fold depending on the temperature. *TSA1pr* has the strongest fold-induction at 42°C and the highest measured fluorescence at high-temperature, whereas that of *SSA2pr* remains at increasing medium-to-high levels (relative to the constitutive promoters) as temperature increases (**Figure 2B**). In comparison, two strong promoters at 30°C—*PDC1pr* and *TEF1pr*—had their YFP expression relatively unchanged by increased temperature.

Another feature of *K. marxianus's* physiology that makes it an attractive cell factory chassis is its ability to utilize a wider range of carbon sources than *S. cerevisiae*, thus circumventing the need to engineer this yeast to consume the carbon sources in question. Wild-type *K. marxianus* can consume sugars found in plant (xylose and inulin) and dairy (lactose) waste, thus allowing reduced costs if using these as feedstocks for

fermentations. For xylose induction, we cloned and tested the induction of three xylose-inducible promoters: *XYL1pr, XYL2pr,* and *ALD4pr*. The xylose reductase *XYL1* reduces xylose to xylitol, this in turn converted to xylulose by the xylitol dehydrogenase *XYL2*. *XKS1* then phosphorylates it to xylulose-5-phosphate, which can then enter the non-oxidative branch of the pentose phosphate pathway. *ALD4pr* was chosen from a genome-wide expression study for its low background in glucose and high fold-induction by xylose (Lertwattanasakul et al., 2015). These promoters exhibited a 6-to-10-fold increase in YFP using xylose as a carbon source relative to glucose, with *XYL2pr* having the strongest fold-induction (**Figure 2C**). For a lactose-inducible promoter, we chose the promoter for the beta-galactosidase *LAC4*. Unlike the other promoters in this set, we chose *LAC4pr* from strain CBS397, known to grow well on lactose as a carbon source (Varela et al., 2017). It exhibited a 10-fold induction by lactose relative to glucose alone when used to express YFP in NBRC1777. In comparison, the *GAL1* promoter from *S. cerevisiae* was inducible by both galactose and lactose, but more strongly for the former (15-fold vs. 6-fold, **Figure 2C**).

Terminators Provide a Second Level of Control Over Gene Expression

Besides promoters, terminators can also control gene expression by affecting the lifetime of mRNA, and provide us with an extra means to fine-tune gene expression (Curran et al., 2013). While the terminators included with the original YTK aimed to keep gene expression output roughly constant (Lee et al., 2015), we have provided five native terminators to broaden the range of gene expression our parts can achieve (**Figure 3A**). Interestingly, two terminators from the YTK—those for *ScADH1* and *ScPGK1*—can change gene expression between them (as measured by YFP fluorescence) by nearly a factor of two in *K. marxianus*. The full ability of promoters to further optimise gene expression can be seen when a set of four terminators and three terminators are combinatorially used to express YFP. While the promoter of choice is still the dominant factor in determining the level of gene expression, choosing a "weak" or "strong" terminator can significantly affect expression as well (*TSA1pr* vs. *TDH1pr*, **Figure 3B**). In the case of inducible promoters, terminator effects are more pronounced under non-inducible conditions. Nonetheless, the increase in basal expression as seen in the inulinase promoter due to a change in a "stronger" terminator is enough to halve fold-induction by inulin (**Figure 3C**). In summary, depending on the promoter and the means of its induction, the terminators we provide could be used to minimise background, maintain a level of basal expression, or smooth out changes in expression during different conditions if needed when expressing a heterologous gene.

Efficient CRISPR/Cas9 Editing With pUCC001

Our CRISPR/Cas9 editing plasmid pUCC001 takes advantage of the cross-species pUDP002 system and makes it a more flexible and economical tool by introducing a BsaI cloning site for new gRNA targets (**Figure 4A**). In this way, annealed oligos containing overhangs matching those generated by the BsaI sites in pUCC001 can be easily cloned into a cut plasmid, or using Golden Gate assembly. As a proof of principle, we transformed *K. marxianus* with pUCC001 containing a gRNA targeting the *LAC4* locus. *lac4* mutants are unable to convert X-Gal to a blue dye. We observed that over 50% of the hygromycin-resistant colonies did not turn blue when grown on medium containing galactose and X-Gal, as opposed to ~10–20% when a deletion cassette is used without CRISPR/Cas9 (**Figure 4B**, **Table 3**). As a demonstration of more practical applications, we used pUCC001 to rapidly generate defined single, double and triple auxotrophs for uracil, leucine, and histidine (**Figures 4C,D**; **Table 1**). Using gRNA plasmids targeting the *K. marxianus* orthologues of *URA3*, *HIS3*, and *LEU2* we created frameshift mutations in these genes leading to loss of function. Given that no repair fragment was used, the efficiency of mutations was good (≥50%); in general, we were able to retrieve the frameshift mutants in the figure by screening fewer than 8 colonies per transformation. The auxotrophic mutants so created allow us to use the orthogonal metabolic markers

from the original YTK (*ScURA3, ScLEU2*, and *ScHIS3*), and show insignificant background when used in transformations with metabolic markers. Separately, the pUCC001 system was used to construct mutants in several other *K. marxianus* genes (Varela et al., 2019).

Inactivating Different Genes Involved in Non-homologous End-Joining Has Different Effects on Gene Targeting Efficiency

While metabolically engineering strains it is advantageous to integrate heterologous gene cassettes for stronger and more stable expression. In yeasts other than *S. cerevisiae*, the ability to efficiently target gene integration is hindered by their usage of non-homologous end-joining (NHEJ) as the dominant DNA repair mechanism (Daley et al., 2005). As a result, random and incomplete integrations are frequent, and much larger targeting homology sequences are also required (up to 1 kb) compared to *S. cerevisiae* (50 bp) for a successful integration (Baudin et al., 1993; Choo et al., 2014). While random integrations of multiple copies of a gene can be advantageous in some biotechnology applications (Lin et al., 2017), it is equally important to integrate single copies of heterologous genes or pathways to rationally construct a cell factory or evaluate different metabolic engineering strategies. Inactivating any of the key genes involved in NHEJ—*YKU70/80*, *NEJ1* or *DNL4* (Abdel-Banat et al., 2010; Choo et al., 2014; Nambu-Nishida et al., 2017)—has been shown to increase targeted integration in *K. marxianus* by forcing it to use homologous recombination (HR) alone for DNA repair. It remains unclear if the different mutations suppress random integration to the same extent. To decide which NHEJ mutant gave us the highest rate of targeted integration, we tested the integration of a YFP expression cassette at the *LAC4* locus in backgrounds that were either wild-type or had either *YKU80*, *DNL4*, or *NEJ1* previously inactivated by CRISPR/Cas9 (YBL001, KmASR.005, and YBL003, respectively; **Table 1**). The cassette was flanked by 880 bp on either end targeting *LAC4* (**Figure 5A**). As with inactivation by CRISPR/Cas9, correct integration would disrupt the gene and render the yeast unable to break down X-Gal, allowing us to pre-screen colonies by blue/white selection for genotyping (**Figure 4B**). The YFP cassette would also allow us to determine if transformants contained multiple integrants in a semi-quantitative manner. Sequencing around the insertion site revealed that integration of the cassette was "seamless" in all backgrounds; the sequencing region immediately downstream of the insertion sites revealed no mutations surrounding the insertion point (**Figure 5A**).

Our experiments revealed that inactivating NHEJ resulted in disruption of the *LAC4* locus in nearly all transformants, as opposed to only occurring for about ~10% of transformants in wild-type *K. marxianus*. Furthermore, all white colonies of the NHEJ-deficient strains had the YFP cassette correctly integrated, as opposed to <20% for the wild-type strain (**Table 3**). In spite of correct integration, wild-type transformants still had a wide spread of measured fluorescent intensities (**Figure 5B**). This
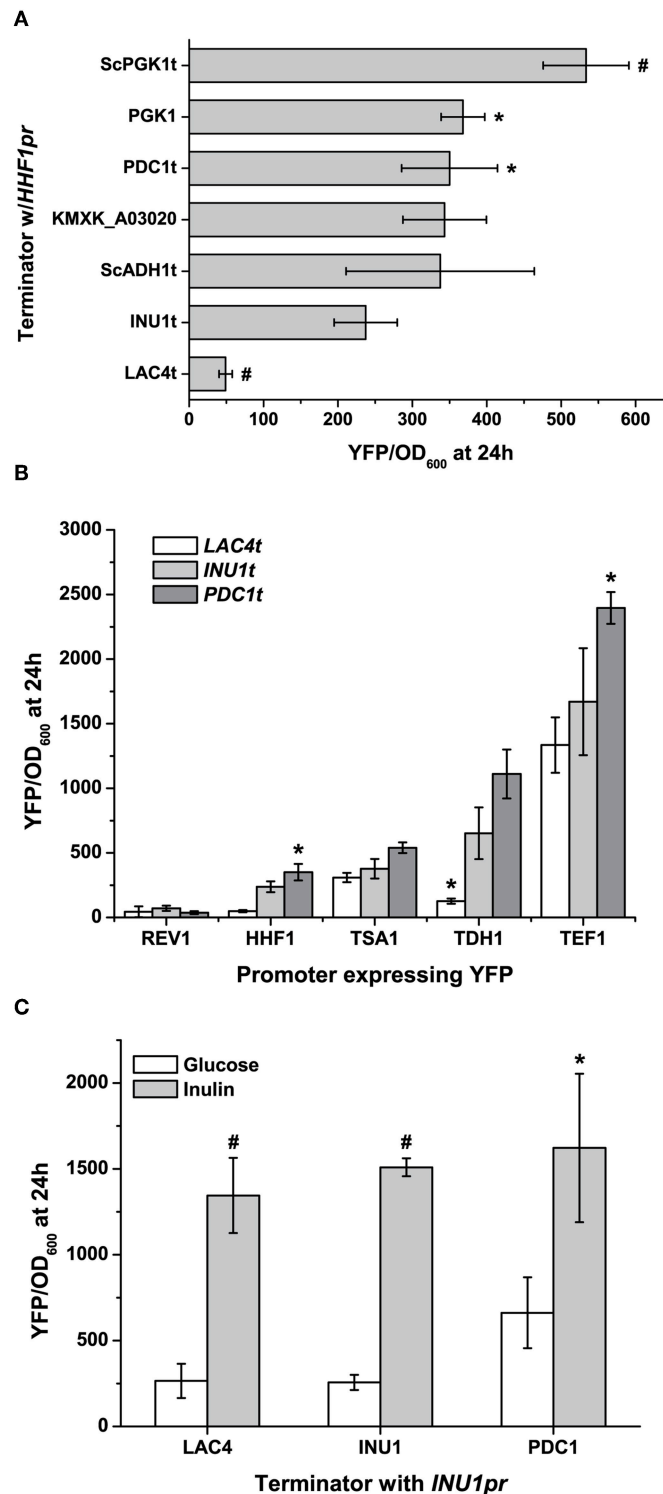
FIGURE 3 | Using terminators to fine-tune gene expression. **(A)** Terminator choice can change gene expression by a factor of nearly 5 using native terminators, and this range can be even further exchanged if terminators from *S. cerevisiae* are used. **(B)** This can further be used to control gene expression which promoters and terminators are used combinatorially, though the starting strength of the promoter can influence the range. **(C)** Inducible promoters' expression is largely altered under non-inducing conditions when used to express YFP with different terminators. Here the strong induction of the inulinase (*INU1*) promoter by inulin is not significantly affected by terminator choice, while it can affect leaky expression under non-inducing conditions. All data are plotted as the mean ± s.d. of at least three replicates. YFP values significantly different from those under baseline conditions (expression using *INU1t*) are marked with an asterisk ($p < 0.05$) or a hash ($p < 0.001$).
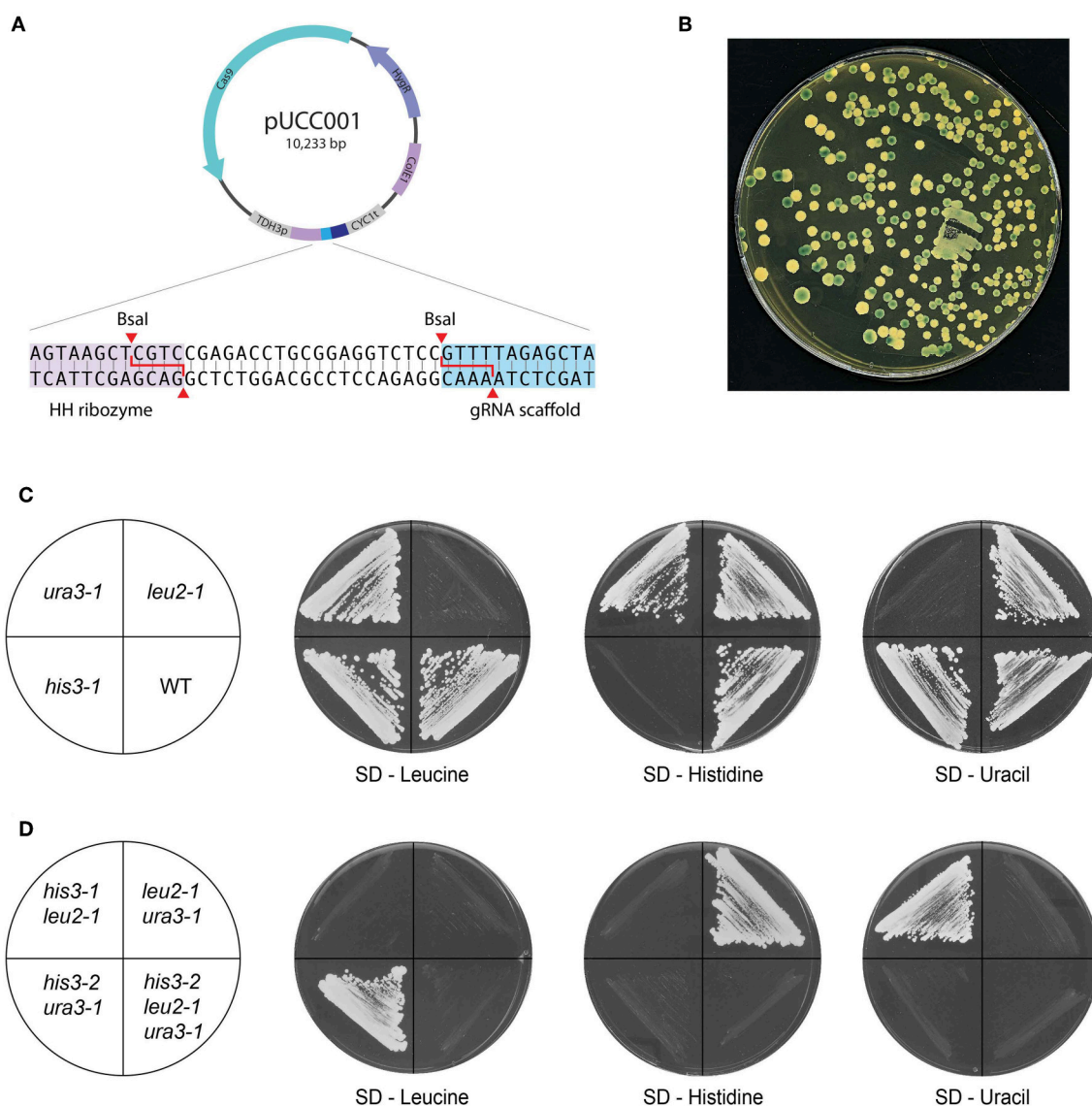
FIGURE 4 | CRISPR/Cas9 genome editing with pUCC001. **(A)** Map of the plasmid. Based on pUDP002, the plasmid contains Cas9 expressed using the *A.adeninivorans* promoter, a pangenomic ARS and a gRNA expression cassette under the control of the *S. cerevisiae TDH3* promoter. The inset shows the cloning site inserted between the ribozyme units; two BsaI sites allow the construction of new gRNA targets using Golden Gate cloning. **(B)** Proof-of-principle of pUCC001 function by inactivation of the *LAC4* locus. Colonies grown YPGal + X-Gal, with a functioning copy of *LAC4* should produce a blue colour on cleaving X-Gal; here, over half of the colonies remain white due to CRISPR-mediated inactivation of *LAC4*. **(C,D)** Construction of single and multiple auxotrophs. With pUCC001, we were able to rapidly construct genome editing plasmids targeting orthologues of genes commonly inactivated in *S. cerevisiae* laboratory strains: *URA3, LEU2* and *HIS3*, both as **(C)** single and **(D)** multiple mutants. Frameshift mutations in each gene created the desired auxotrophies, and the strains so generated could be targeted with other plasmids to create defined double and triple mutants.

led us to hypothesise that correct integration in a wild-type strain does not preclude random integration. The contrasting low variation in YFP fluorescence seen in *DNL4* and *YKU80* mutants suggests that these backgrounds incorporate a single copy of the YFP cassette. On the other hand, *NEJ1* mutants seemed to incorporate multiple copies of YFP. In the case of the latter, it is admittedly unclear if these extra copies are integrated sequentially at *LAC4*, or elsewhere in the genome. In summary, it is clear that *DNL4* or *YKU80* are the best NHEJ components to

inactivate to ensure targeted, single chromosomal integrations of DNA. Extrapolating from research in *S. cerevisiae,* our findings are given weight by the different roles in NHEJ of the proteins we targeted. *YKU80*, as part of the Ku complex, is the first protein to bind to and stabilise double-stranded breaks, followed by *DNL4* (Emerson and Bertuch, 2016). Therefore, the role of these "first responders" in DNA repair might make them better targets to thoroughly inactivate NHEJ. Furthermore, as other research has found that the binding of *DNL4* to DNA may not require *NEJ1*

TABLE 3 | Gene targeting and integration efficiencies at the *LAC4* locus in different backgrounds.

Background/strain	G418—resistant colonies	White resistant colonies	Gene disruption efficiency, %[a]	Gene targeting efficiency, %[b]
Wild-type/NBRC1777	863 ± 36	111 ± 13	12.9 ± 1.3	16.7 ± 5.9
yku80-1/YBL001	587 ± 71	587 ± 71	100	100
dnl4-1/KmASR.005	456 ± 44	455 ± 44	99.93	100
nej1-1/YBL003	469 ± 32	466 ± 32	99.36	100

The data are presented as the mean ± s.d of three replicates.
[a]*The percentage of white colonies with respect to all blue and white colonies;* [b]*Colony PCR was performed on eight white colonies per replicate to check for correct integration of the YFP cassette at LAC4.*

FIGURE 5 | Random and targeted integration in *K. marxianus*. **(A)** Using homology arms ∼880 bp in length, integration of a YFP reporter cassette (with marker) is seamless, with no sequence loss or modifications at the integration site irrespective of the genetic background (inset). **(B)** Inactivating *YKU80* or *DNL4* but not *NEJ1* is essential to eliminate random integrations. When targeting an YFP expression cassette at the *LAC4* locus, correct integration in a wild-type strain (as determined by genotyping) does not abolish random integration elsewhere, as seen by the spread in YFP fluorescence measured from five transformed colonies. YFP expression for the same cassette expressed on a centromeric plasmid is provided as a reference. **(C)** Genomic locations of insertion sites for gene expression cassettes, selected as described in the main text. **(D)** The effect of genetic context on expression levels of the same gene. The bar plots are plotted as the mean ± s.d. of at least three replicates.

(Wu et al., 2008), and the latter enhances but is not essential for NHEJ entirely (Yang et al., 2015b), this may explain why the inactivation of *NEJ1* in YBL003 was insufficient to ensure single-copy gene integration. In general, expression for a single integrated cassette was higher than that when the same cassette was expressed on a low-copy-number plasmid using the same antibiotic marker (**Figure 5B**); the same trend was seen across several promoters (**Figure S3**).

Evaluation of Sites for Chromosomal Integration

While selecting chromosomal integration sites in the *K. marxianus* genome, we considered three criteria: (i) providing a visible or easily detectable phenotype on integration, simplifying screening if necessary, (ii) near clusters of actively-transcribed genes, and (iii) near essential genes to ensure that only cells with correctly-inserted DNA would survive(Mikkelsen

et al., 2012). For insertion sites of the first type, we chose the *LAC4* locus (I1); insertion there has a visible phenotype that does not affect core metabolism (as opposed to *ADE2*, for example). For selecting sites of the second type, we examined existing TSS-seq data (Lertwattanasakul et al., 2015) and found two such sites where the gene clusters were on the same coding strand: one each on chromosome IV and V (I2 and I4) (**Figure 5C**; **Table 2**). We also selected an insertion site upstream of *ARO1*, the pentafunctional protein involved in aromatic amino acid biosynthesis (I3); an incorrect insertion would result in auxotrophy for the aromatic amino acids phenylalanine, tyrosine, and tryptophan. While cloning the 5' (left) homology arms into level I plasmids, we added the overhangs matching the 5' multigene connector ConLS' by PCR, turning it into a type 1 part. This included the nested connector sequence containing the BsmBI site with the appropriate overhang for multigene constructs, which we added by PCR. This way the part with the homology arm can also be used to construct integrative vectors for a single expression cassette, as it still possesses the appropriate BsaI overhang at its 3' end to clone a TU in, as described below. The 3' (right) homology arms, as in the original YTK, contain BsaI overhangs corresponding to a yeast replicative origin, or type 7 part (Lee et al., 2015). Finally, the homology arms have AscI restriction sites directly flanking them to linearise the insertion vector prior to transformation, to improve integration efficiency.

Using these arms in place of connectors provided for TUs in the original YTK, we constructed novel integrative reporter vectors, each expressing YFP under the control of the strongest promoter we identified (*PDC1pr;* P2) and targeting I1 to I4. We then evaluated the integration efficiency at each site in KmASR.005, in each case using the same YFP expression cassette. Insertion efficiency was 100% at all loci, and while variations in YFP expression were observed at all loci, none varied by more than 25% (**Figure 5D**). However, this difference in expression may be more significant when expressing multiple genes in a heterologous biosynthetic pathway as opposed to a single fluorescent protein. Integrating the gene at I3, upstream of *ARO1*, did not cause the auxotrophies for Phe, Trp, and Tyr expected if the gene was disrupted, suggesting that the integration there was seamless as well. However, insertion sites near essential genes may not always be this accessible; another site we tested near an orthologue of the translation initiation factor *TIF1* yielded <10% correct integrations in the wild-type strain and no colonies in an NHEJ-deficient background (data not shown).

DISCUSSION

Standardised biological parts for the expression and maintenance of genes, and means to assemble them, are a cornerstone of synthetic biology to build new biological systems. On a more practical level, they allow us to accelerate the design-build-test cycle which forms the core of applied synthetic biology and its sister discipline, metabolic engineering. A versatile collection of these parts is the Yeast Toolkit collected 3 years ago. Grouped into eight categories of parts, the standard set forth by the

YTK allows the hierarchical assembly of expression or storage vectors for expressing recombinant genes in *S. cerevisiae* with a wide range of selection markers, regulatory elements, and functional protein tags to fine-tune gene expression and engineer the genome as needed. Clearly, to move beyond being niche organisms in biotechnology, alternative yeasts require such part collections to make rapid metabolic engineering feasible. It is also advantageous to maintain a common standard for assembly to facilitate the exchange of parts between researchers and yeasts as needed. It was with this goal in mind that we selected and characterized the parts presented in this kit, while maintaining the YTK standard. Large part collections have been developed for other yeasts (Celinska et al., 2017; Prielhofer et al., 2017), and it is with such collections in mind that we have created ours. Using the established Golden Gate assembly protocols, we were able to assemble episomal and integrative reporter constructs from up to 8 component plasmids to characterize our parts.

Our extension of the YTK also includes the first collection of homology arms for insertion vectors targeting four loci in the *K. marxianus* genome, each with different characteristics. They are considered "full-length," but can easily be shortened using PCR or re-cloning in the case of their use in NHEJ-deficient strains to an optimal length to minimize construct size without compromising gene targeting efficiency. As better genomic and transcriptomic knowledge of *K. marxianus* is acquired, more insertion sites and parts will be identified to be added to the modest set we provide. We also foresee the set being expanded by synthetic promoters, engineered promoters (as has been done with *INU1pr*) and secretion tags (Zhou et al., 2018).

While the promoters we have characterized are largely selected from strain CBS6556, all but one of them had >90% sequence similarity with strain DMKU3-1042 and with strain NBRC1777, where the characterization was carried out. Within the observed differences, only a few promoters do exhibit significant sequence differences that could affect gene expression between strains based on orthologous transcription factor binding sites from *S. cerevisiae* (**Table S7**). Sequence differences between promoters for the same gene in different yeast strains can have implications in gene expression, and therefore should be taken into consideration for experimental and industrial applications (Liu et al., 2015; de Paiva et al., 2018); however, at this stage not enough is known about *K. marxianus'* native transcription factors and regulatory network to functionally dissect our promoter sequences. Their measured activity under the selected expression conditions, and in two different strains, demonstrate their practical usability.

Alongside parts collections, the existence of genome editing tools for allelic replacement and deletion speeds up the creation of strains with defined mutant genotypes and mating, similar to standard *S. cerevisiae* lab strains. This also opens up the possibility of creating the best *K. marxianus* strains for the laboratory and industry using classical genetics methods and synthetic biology side-by-side (Cernak et al., 2018; Lee et al., 2018). It is with this end in mind that our collection also provides pUCC001, a Cas9/gRNA genome editing plasmid derived from the broad-host platform pUDP002 (Juergens et al., 2018), into

which gRNA targets can be rapidly cloned by Golden Gate assembly simply as annealed and phosphorylated oligos. This saves the cost and time of cloning the entire gRNA expression cassette for every target as for the original plasmid. We have also cloned the *ScTRP1* expression cassette for use in future strains auxotrophic for tryptophan. The YTK standard makes it possible to multiplex gRNA expression using pUCC001. In theory, Cas9 and multiple gRNA cassettes could be separately cloned as level II plasmids and then reassembled into a level III "multi-TU" plasmid. The ease of assembly of both individual gRNA plasmids and the Golden Gate assembly would provide a credible, if not more efficient, alternative to existing multiplexing systems, (Löbs et al., 2018). However, further optimization of the gRNA expression system is necessary for an optimal *K. marxianus*—specific multiplexing system.

While investigating different NHEJ-deficient backgrounds, we found that inactivating *YKU80* or *DNL4*, but not *NEJ1,* was the best way to eliminate multiple or random integration. Identifying such a background is beneficial to improve the efficiency and specificity of *K. marxianus*-based *in vivo* assembly techniques such as PGASO (Chang et al., 2012), and to further define a genotype for a potential future "lab strain" for *K. marxianus*. As much as the versatility of the YTK standard is of relevance to synthetic biologists and metabolic engineers, the parts we have gathered may be of broader interest in the long term. *Kluyveromyces marxianus* is slowly emerging from its niche applications to become an alternative cell factory to *S. cerevisiae*. Several efforts have been, and are being made, to make it produce bio-based compounds of value (Cheon et al., 2014; Kim et al., 2014; Lin et al., 2017). The lack of standardised parts, and efficient synthetic biology tools and strategies has limited the scope or sophistication of these efforts. We believe this collection can enrich the existing synthetic biology landscape of *K. marxianus* and allow researchers to make more informed choices for the more efficient, predictable and practical design and testing of future cell factories for a bio-based economy.

AUTHOR CONTRIBUTIONS

AR and JV carried out the experimental work, interpreted the data and wrote the manuscript. HJ carried out the experimental work. J-MD and JM conceived the study, supervised the research, interpreted the data and contributed to writing the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2019.00097/full#supplementary-material

REFERENCES

Abdel-Banat, B. M. A., Nonklang, S., Hoshida, H., and Akada, R. (2010). Random and targeted gene integrations through the control of non-homologous end joining in the yeast *Kluyveromyces marxianus*. *Yeast* 27, 29–39. doi: 10.1002/yea.1729

Awan, A. R., Blount, B. A., Bell, D. J., Shaw, W. M., Ho, J. C. H., McKiernan, R. M., et al. (2017). Biosynthesis of the antibiotic nonribosomal peptide penicillin in baker's yeast. *Nat. Commun.* 8, 1–8. doi: 10.1038/ncomms15202

Ball, M. M., Raynala, A., Guérineau, M., and Iborra, F. (1999). Construction of efficient centromeric, multicopy and expression vectors for the yeast Kluyveromyces marxianus using homologous elements and the promoter of a purine-cytosine-like permease. *J. Mol. Microbiol. Biotechnol.* 1, 347–353.

Baudin, A., Ozier-Kalogeropoulus, O., Denouel, A., Lacroute, F., and Cullin, C. (1993). A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 21, 3329–3330. doi: 10.1093/nar/21.14.3329

Bergkamp, R. J. M., Bootsman, T. C., Toschka, H. Y., Mooren, A. T. A., Kox, L., Verbakel, J. M. A., et al. (1993). Applied microbiology biotechnology expression of an a-galactosidase gene under control of the homologous inulinase promoter in *Kluyveromyces marxianus*. *Appl. Microbiol.* 40, 309–317. doi: 10.1007/BF00170386

Celinska, E., Ledesma-Amaro, R., Larroude, M., Rossignol, T., Pauthenier, C., and Nicaud, J. M. (2017). Golden Gate Assembly system dedicated to complex pathway manipulation in *Yarrowia lipolytica*. *Microb. Biotechnol.* 10, 450–455. doi: 10.1111/1751-7915.12605

Cernak, R., Estrella, R., Poddar, S., Skerker, J. M., Cheng, Y.-F., Carlson, A. K., et al. (2018). Engineering *Kluyveromyces marxianus* as a robust synthetic biology platform host paul. *MBio* 9, 1–16. doi: 10.1128/mBio.01410-18

Chang, J. J., Ho, C. Y., Ho, F. J., Tsai, T. Y., Ke, H. M., Wang, C. H., et al. (2012). PGASO: a synthetic biology tool for engineering a cellulolytic yeast. *Biotechnol. Biofuels* 5, 1–12. doi: 10.1186/1754-6834-5-53

Chang, J. J., Ho, F. J., Ho, C. Y., Wu, Y. C., Hou, Y. H., Huang, C. C., et al. (2013). Assembling a cellulase cocktail and a cellodextrin transporter into a yeast host for CBP ethanol production. *Biotechnol. Biofuels* 6, 1–13. doi: 10.1186/1754-6834-6-19

Cheon, Y., Kim, J. S., Park, J. B., Heo, P., Lim, J. H., Jung, G. Y., et al. (2014). A biosynthetic pathway for hexanoic acid production in *Kluyveromyces marxianus*. *J. Biotechnol.* 182–183, 30–36. doi: 10.1016/j.jbiotec.2014.04.010

Choo, J. H., Han, C., Kim, J. Y., and Kang, H. A. (2014). Deletion of a KU80 homolog enhances homologous recombination in the thermotolerant yeast *Kluyveromyces marxianus*. *Biotechnol. Lett.* 36, 2059–2067. doi: 10.1007/s10529-014-1576-4

Curran, K. A., Karim, A. S., Gupta, A., and Alper, H. S. (2013). Use of expression-enhancing terminators in *Saccharomyces cerevisiae* to increase mRNA half-life and improve gene expression control for metabolic engineering applications. *Metab. Eng.* 19, 88–97. doi: 10.1016/j.ymben.2013.07.001

Daley, J. M., Palmbos, P. L., Wu, D., and Wilson, T. E. (2005). Nonhomologous end joining in yeast. *Annu. Rev. Genet.* 39, 431–451. doi: 10.1146/annurev.genet.39.073003.113340

de Paiva, D. P., Rocha, T. B., Rubini, M. R., Nicola, A. M., Reis, V. C. B., Torres, F. A. G., et al. (2018). A study on the use of strain-specific and homologous promoters for heterologous expression in industrial *Saccharomyces cerevisiae* strains. *AMB Express* 8, 1–11. doi: 10.1186/s13568-018-0613-4

Denby, C. M., Li, R. A., Vu, V. T., Costello, Z., Lin, W., Chan, L. J. G., et al. (2018). Industrial brewing yeast engineered for the production of primary flavor determinants in hopped beer. *Nat. Commun.* 9, 1–10. doi: 10.1038/s41467-018-03293-x

Diniz, R. H. S., Villada, J. C., Alvim, M. C. T., Vidigal, P. M. P., Vieira, N. M., Lamas-Maceiras, M., et al. (2017). Transcriptome analysis of the thermotolerant yeast *Kluyveromyces marxianus* CCT 7735 under ethanol stress. *Appl. Microbiol. Biotechnol.* 101, 6969–6980. doi: 10.1007/s00253-017-8432-0

Emerson, C. H., and Bertuch, A. A. (2016). Consider the workhorse: nonhomologous end-joining in budding yeast[1]. *Biochem. Cell Biol.* 94, 396–406. doi: 10.1139/bcb-2016-0001

Feng, J., Jester, B. W., Tinberg, C. E., Mandell, D. J., Antunes, M. S., Chari, R., et al. (2015). A general strategy to construct small molecule biosensors in eukaryotes. *Elife* 4, 1–23. doi: 10.7554/eLife.10606

Fonseca, G. G., Gombert, A. K., Heinzle, E., and Wittmann, C. (2007). Physiology of the yeast *Kluyveromyces marxianus* during batch and chemostat cultures with glucose as the sole carbon source. *FEMS Yeast Res.* 7, 422–435. doi: 10.1111/j.1567-1364.2006.00192.x

Gao, J., Yuan, W., Li, Y., Xiang, R., Hou, S., Zhong, S., et al. (2015). Transcriptional analysis of *Kluyveromyces marxianus* for ethanol production from inulin using consolidated bioprocessing technology. *Biotechnol. Biofuels* 8, 1–17. doi: 10.1186/s13068-015-0295-y

Gietz, R. D. (2014). Yeast transformation by the LiAc/SS carrier DNA/PEG method. *Methods Mol. Biol.* 1163, 33–44. doi: 10.1007/978-1-4939-0799-1_4

Gombert, A. K., Madeira, J. V., Cerdán, M.-E., and González-Siso, M.-I. (2016). *Kluyveromyces marxianus* as a host for heterologous protein synthesis. *Appl. Microbiol. Biotechnol.* 100, 6193–6208. doi: 10.1007/s00253-016-7645-y

Iborra, F., and Ball, M. M. (1994). *Kluyveromyces marxianus* small DNA fragments contain both autonomous replicative and centromeric elements that also function in *Kluyveromyces lactis*. *Yeast* 10, 1621–1629. doi: 10.1002/yea.320101211

Inokuma, K., Ishii, J., Hara, K. Y., Mochizuki, M., and Hasunuma, T. (2015). Complete genome sequence of *Kluyveromyces marxianus* NBRC1777, a nonconventional thermotolerant yeast. *Genome Announc.* 3, 2–3. doi: 10.1128/genomeA.00389-15

Jakociunas, T., Rajkumar, A. S., Zhang, J., Arsovska, D., Rodriguez, A., Jendresen, C. B., et al. (2015). CasEMBLR: Cas9-facilitated multiloci genomic integration of *in vivo* assembled DNA parts in *Saccharomyces cerevisiae*. *ACS Synth. Biol.* 4, 1126–1134. doi: 10.1021/acssynbio.5b00007

Jeong, H., Lee, D.-H., Kim, S. H., Kim, H.-J., Lee, K., Song, J. Y., et al. (2012). Genome sequence of the thermotolerant yeast *Kluyveromyces marxianus* var . marxianus KCTC 17555. *Eukaryot. Cell* 11, 1584–1585. doi: 10.1128/EC.00260-12

Juergens, H., Varela, J. A., Gorter de Vries, A. R., Perli, T., Gast, V. J. M., Gyurchev, N. Y., et al. (2018). Genome editing in *Kluyveromyces* and *Ogataea* yeasts using a broad-host-range Cas9/gRNA co-expression plasmid. *FEMS Yeast Res.* 18, 1–16. doi: 10.1093/femsyr/foy012

Kim, T. Y., Lee, S. W., and Oh, M. K. (2014). Biosynthesis of 2-phenylethanol from glucose with genetically engineered *Kluyveromyces marxianus*. *Enzyme Microb. Technol.* 61–62, 44–47. doi: 10.1016/j.enzmictec.2014.04.011

Lee, K. S., Kim, J. S., Heo, P., Yang, T. J., Sung, Y. J., Cheon, Y., et al. (2013). Characterization of Saccharomyces cerevisiae promoters for heterologous gene expression in *Kluyveromyces marxianus*. *Appl. Microbiol. Biotechnol.* 97, 2029–2041. doi: 10.1007/s00253-012-4306-7

Lee, M. E., DeLoache, W. C., Cervantes, B., and Dueber, J. E. (2015). A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth. Biol.* 4, 975–986. doi: 10.1021/sb500366v

Lee, M. H., Lin, J. J., Lin, Y. J., Chang, J. J., Ke, H. M., Fan, W. L., et al. (2018). Genome-wide prediction of CRISPR/Cas9 targets in *Kluyveromyces marxianus* and its application to obtain a stable haploid strain. *Sci. Rep.* 8, 1–10. doi: 10.1038/s41598-018-25366-z

Lertwattanasakul, N., Kosaka, T., Hosoyama, A., Suzuki, Y., Rodrussamee, N., Matsutani, M., et al. (2015). Genetic basis of the highly efficient yeast *Kluyveromyces marxianus*: complete genome sequence and transcriptome analyses. *Biotech. Biofuels* 8:47. doi: 10.1186/s13068-015-0227-x

Lin, Y. J., Chang, J. J., Lin, H. Y., Thia, C., Kao, Y. Y., Huang, C. C., et al. (2017). Metabolic engineering a yeast to produce astaxanthin. *Bioresour. Technol.* 245, 899–905. doi: 10.1016/j.biortech.2017.07.116

Liu, J., Martin-Yken, H., Bigey, F., Dequin, S., François, J. M., and Capp, J. P. (2015). Natural yeast promoter variants reveal epistasis in the generation of transcriptional-mediated noise and its potential benefit in stressful conditions. *Genome Biol. Evol.* 7, 969–984. doi: 10.1093/gbe/evv047

Löbs, A. K., Engel, R., Schwartz, C., Flores, A., and Wheeldon, I. (2017). CRISPR-Cas9-enabled genetic disruptions for understanding ethanol and ethyl acetate biosynthesis in *Kluyveromyces marxianus*. *Biotechnol. Biofuels* 10, 1–14. doi: 10.1186/s13068-017-0854-5

Löbs, A. K., Schwartz, C., Thorwall, S., and Wheeldon, I. (2018). Highly multiplexed CRISPRi repression of respiratory functions enhances mitochondrial localized ethyl acetate biosynthesis in *Kluyveromyces marxianus*. *ACS Synth. Biol.* 7, 2647–2655. doi: 10.1021/acssynbio.8b00331

Mikkelsen, M. D., Buron, L. D., Salomonsen, B., Olsen, C. E., Hansen, B. G., Mortensen, U. H., et al. (2012). Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. *Metab. Eng.* 14, 104–111. doi: 10.1016/j.ymben.2012.01.006

Morrissey, J. P., Etschmann, M. M., Schrader, J., and de Billerbeck, G. M. (2015). Cell factory applications of the yeast *Kluyveromyces marxianus* for the biotechnological production of natural flavour and fragrance molecules. *Yeast* 32, 3–16. doi: 10.1002/yea.3054

Nambu-Nishida, Y., Nishida, K., Hasunuma, T., and Kondo, A. (2017). Development of a comprehensive set of tools for genome engineering in a cold- and thermo-tolerant *Kluyveromyces marxianus* yeast strain. *Sci. Rep.* 7, 1–7. doi: 10.1038/s41598-017-08356-5

Ng, H., and Dean, N. (2017). Dramatic improvement of CRISPR/Cas9 editing in *Candida albicans* by increased single guide RNA expression. *Mol. Biol. Physiol.* 2, 1–17. doi: 10.1128/mSphere.00385-16

Obst, U., Lu, T. K., and Sieber, V. (2017). A modular toolkit for generating *Pichia pastoris* secretion libraries. *ACS Synth. Biol.* 6, 1016–1025. doi: 10.1021/acssynbio.6b00337

Ortiz-Merino, R. A., Varela, J. A., Coughlan, A. Y., Hoshida, H., da Silveira, W. B., Wilde, C., et al. (2018). Ploidy variation in *Kluyveromyces marxianus* separates dairy and non-dairy isolates. *Front. Genet.* 9, 1–16. doi: 10.3389/fgene.2018.00094

Prielhofer, R., Barrero, J. J., Steuer, S., Gassler, T., Zahrl, R., Baumann, K., et al. (2017). GoldenPiCS: a golden gate-derived modular cloning system for applied synthetic biology in the yeast *Pichia pastoris*. *BMC Syst. Biol.* 11, 1–14. doi: 10.1186/s12918-017-0492-3

Schabort, du. T. W., Letebele, P. K., Steyn, L., Kilian, S. G., and Du Preez, J. C. (2016). Differential RNA-seq, multi-network analysis and metabolic regulation analysis of *Kluyveromyces marxianus* reveals a compartmentalised response to xylose. *PLoS ONE* 11:e156242. doi: 10.1371/journal.pone.0156242

Suzuki, A., Fujii, H., Hoshida, H., and Akada, R. (2015). Gene expression analysis using strains constructed by NHEJ-mediated one-step promoter cloning in the yeast *Kluyveromyces marxianus*. *FEMS Yeast Res.* 15, 1–10. doi: 10.1093/femsyr/fov059

Varela, J. A., Montini, N., Scully, D., Van der Ploeg, R., Oreb, M., Boles, E., et al. (2017). Polymorphisms in the LAC12 gene explain lactose utilisation variability in *Kluyveromyces marxianus* strains. *FEMS Yeast Res.* 17, 1–13. doi: 10.1093/femsyr/fox021

Varela, J. A., Puricelli, M., Montini, N., and Morrissey, J. P. (2019). Expansion and Diversification of MFS Transporters in *Kluyveromyces marxianus*. *Front. Microbiol.* 9:3330. doi: 10.3389/fmicb.2018.03330

Vyas, V. K., Barrasa, M. I., and Fink, G. R. (2015). A *Candida albicans* CRISPR system permits genetic engineering of essential genes and gene families. *Sci. Adv.* 1:e1500248. doi: 10.1126/sciadv.1500248

Wagner, J. M., and Alper, H. S. (2016). Synthetic biology and molecular genetics in non-conventional yeasts: current tools and future advances. *Fungal Genet. Biol.* 89, 126–136. doi: 10.1016/j.fgb.2015.12.001

Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. *PLoS ONE* 6:e16765. doi: 10.1371/journal.pone.0016765

Wu, D., Topper, L. M., and Wilson, T. E. (2008). Recruitment and dissociation of nonhomologous end joining proteins at a DNA double-strand break in *Saccharomyces cerevisiae*. *Genetics* 178, 1237–1249. doi: 10.1534/genetics.107.083535

Xie, S., Shen, B., Zhang, C., Huang, X., and Zhang, Y. (2014). SgRNAcas9: a software package for designing CRISPR sgRNA and evaluating potential off-target cleavage sites. *PLoS ONE* 9:e100448. doi: 10.1371/journal.pone.0100448

Yang, C., Hu, S., Zhu, S., Wang, D., Gao, X., and Hong, J. (2015a). Characterizing yeast promoters used in *Kluyveromyces marxianus*. *World J. Microbiol. Biotechnol.* 31, 1641–1646. doi: 10.1007/s11274-015-1899-x

Yang, H., Matsumoto, Y., Trujillo, K. M., Lees-Miller, S. P., Osley, M. A., and Tomkinson, A. E. (2015b). Role of the yeast DNA repair protein Nej1 in end processing during the repair of DNA double strand breaks by non-homologous end joining. *DNA Repair* 31, 1–10. doi: 10.1016/j.dnarep.2015.04.003

Yarimizu, T., Nonklang, S., Nakamura, J., Tokuda, S., Nakagawa, T., Lorreungsil, S., et al. (2013). Identification of auxotrophic mutants of the yeast *Kluyveromyces marxianus* by non-homologous end joining-mediated integrative transformation with genes from *Saccharomyces cerevisiae*. *Yeast* 30, 485–500. doi: 10.1002/yea.2985

Zhou, J., Zhu, P., Hu, X., Lu, H., and Yu, Y. (2018). Improved secretory expression of lignocellulolytic enzymes in *Kluyveromyces marxianus* by promoter and signal sequence engineering. *Biotechnol. Biofuels* 11:235. doi: 10.1186/s13068-018-1232-7

Check for
updates

Production of 3-Hydroxypropanoic Acid From Glycerol by Metabolically Engineered Bacteria

Carsten Jers[1]*, Aida Kalantari[2], Abhroop Garg[1] and Ivan Mijakovic[1,3]

[1] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark, [2] Department of Biomedical Engineering, Duke University, Durham, NC, United States, [3] Systems and Synthetic Biology Division, Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden

3-hydroxypropanoic acid (3-HP) is a valuable platform chemical with a high demand in the global market. 3-HP can be produced from various renewable resources. It is used as a precursor in industrial production of a number of chemicals, such as acrylic acid and its many derivatives. In its polymerized form, 3-HP can be used in bioplastic production. Several microbes naturally possess the biosynthetic pathways for production of 3-HP, and a number of these pathways have been introduced in some widely used cell factories, such as *Escherichia coli* and *Saccharomyces cerevisiae*. Latest advances in the field of metabolic engineering and synthetic biology have led to more efficient methods for bio-production of 3-HP. These include new approaches for introducing heterologous pathways, precise control of gene expression, rational enzyme engineering, redirecting the carbon flux based on *in silico* predictions using genome scale metabolic models, as well as optimizing fermentation conditions. Despite the fact that the production of 3-HP has been extensively explored in established industrially relevant cell factories, the current production processes have not yet reached the levels required for industrial exploitation. In this review, we explore the state of the art in 3-HP bio-production, comparing the yields and titers achieved in different microbial cell factories and we discuss possible methodologies that could make the final step toward industrially relevant cell factories.

Keywords: 3-hydroxypropanoic acid, glycerol, biosynthesis, cell factory, synthetic biology, metabolic engineering

INTRODUCTION

The development of microbial cell factories is fueled by aspirations to develop sustainable processes based on renewable resources. The goal is mitigation of the negative environmental consequences of production of fuels, chemicals and other materials.

The development of microbial cell factories for production of the platform chemical 3-hydroxypropanoic acid (3-HP) has attracted much attention in the last decade. 3-HP is a non-chiral (optically inactive), small three-carbon molecule and a structural isomer of lactic acid. Specifically, 3-HP is a precursor for the production of a number of valuable chemicals including acrylic acid and bioplastics (**Figure 1**). Bio-based production of 3-HP also has the potential to "turn waste into a resource" since several metabolic pathways exist for converting glycerol, a by-product of biodiesel production, into 3-HP.

Here, we will briefly introduce the metabolic pathways that can be used to produce 3-HP from glycerol. Based on work accumulated over the last decade, we will present current knowledge pertaining to different production hosts, enzymes, and strategies for optimization of production pathway and host metabolism as well as process engineering for attaining high-level production of 3-HP. Finally, we will discuss how recent developments in synthetic biology and metabolic engineering might form the basis for further improvements of production strains and the eventual goal of industrially viable and sustainable 3-HP production.

METABOLIC PATHWAYS FOR SYNTHESIS OF 3-HP STARTING FROM GLYCEROL

Metabolic Pathways Found in Nature

A number of microorganisms have been reported to naturally produce 3-HP using various pathways and diverse substrates such as glycerol, glucose, CO_2, and uracil. Several reviews have described these in detail (Kumar et al., 2013a; de Fouchécour et al., 2018), and in this review we will focus only on the pathways for which glycerol is the substrate. Two pathways are known for conversion of glycerol into 3-HP: the CoA-dependent pathway and the CoA-independent pathway (**Figure 2**).

The CoA-dependent pathway has been most extensively studied in *Lactobacillus reuteri* and proceeds with conversion of glycerol to 3-hydroxypropanal (3-HPA) catalyzed by coenzyme B_{12}-dependent glycerol hydratase (PduCDE). 3-HPA is subsequently converted to 3-HP via 3-hydroxypropanoyl-CoA and 3-hydroxypropanoyl-phosphate, catalyzed by the enzymes propionaldehyde dehydrogenase (PduP), phosphotransacylase (PduL), and propionate kinase (PduW), respectively (Dishisha et al., 2014).

In the CoA-independent pathway, glycerol is converted to 3-HP in 2 steps. As in the case of the CoA-dependent pathway, glycerol is first converted to 3-HPA, while in the second step, 3-HPA is converted directly to 3-HP in a reaction catalyzed by aldehyde dehydrogenase (Kumar et al., 2013a). While the CoA-independent pathway is by far the most employed for production of 3-HP using engineered bacteria, it appears to have little relevance in nature. This is possibly due to low activity of aldehyde dehydrogenase in wild type bacteria (Zhu et al., 2009).

PRODUCTION HOSTS

A number of microorganisms have been used for the production of 3-HP, both natural isolates as well as engineered microorganisms (**Table 1**). For the selection of an appropriate production host, there are several parameters to consider. The microorganism should demonstrate tolerance to organic acids, specifically 3-HP, as well as potentially toxic impurities in crude glycerol. As one of the widely used glycerol dehydratases is coenzyme B_{12}-dependent, the production host should preferentially be capable of synthesizing coenzyme B_{12}. Addition to the production medium instead of *in situ* synthesis would significantly increase the overall production cost.

Lactobacillus reuteri is capable of naturally producing 3-HP from glycerol via the CoA-dependent pathway (Luo et al., 2011). In one study, three *L. reuteri* strains were evaluated, and all were found to produce 3-HP (Burgé et al., 2015). It also was shown that 3-HPA is toxic at a concentration of 5 g/L while 2.5 g/L 3-HP is not (as long as pH is maintained above 5). Notably, *L. reuteri* is capable of synthesizing coenzyme B_{12} (Burgé et al., 2015) and has been used for the bioconversion of biodiesel-derived glycerol into 3-HP (14 g/L) and 1,3-propanediol (1,3-PDO) (Dishisha et al., 2015). To our knowledge, there have been no attempts to engineer *L. reuteri* strains to improve production. Among lactic acid bacteria, production of 3-HP is not restricted to *L. reuteri*. Of 67 lactic acid bacteria isolates tested, 22 isolates belonging to *Lactobacillus diolivorans* and *Lactobacillus collinoides* were positive for 3-HP production from glycerol (Garai-Ibabe et al., 2008). Additionally, *L. diolivorans* is capable of using biodiesel-derived glycerol for the production of 3-HPA (Lindlbauer et al., 2017).

Klebsiella pneumoniae is a pathogenic bacterium that has been widely used for production of 3-HP. Along with *E. coli*, *K. pneumoniae* is one of the most frequently used hosts for strain development by genetic engineering for improved 3-HP production. In fact, the highest 3-HP titer reported, 83.8 g/L, was obtained in *K. pneumoniae* by the combining of optimized expression of aldehyde dehydrogenase (*K. pneumoniae* PuuC), blocking of lactic acid synthesis (*ldh1*, *ldh2*, and *pta* mutant), and optimization of the fermentation conditions (Li et al., 2016). An important asset of *K. pneumoniae* is its capability of producing coenzyme B_{12} (Luo et al., 2011).

E. coli is commonly used for metabolic engineering to produce a wide variety of compounds. It has also been widely used as a chassis for the production of 3-HP. The highest 3-HP titer reported in *E. coli* so far is 71.9 g/L, using a strain in which, besides introduction of glycerol dehydratase and aldehyde dehydrogenase, the central metabolism was modified to reduce by-product formation (Chu et al., 2015). It has been indicated that careful choice of the strain could be of importance. A comparison of nine *E. coli* strains in which a heterologous pathway for 3-HP synthesis was introduced, demonstrated differences in 3-HP production, as well as in enzyme level and activity (Sankaranarayanan et al., 2014). A drawback of using *E. coli* is the fact that it is not naturally capable of producing coenzyme B_{12}. However, insertion of the *Pseudomonas denitrificans* genes for coenzyme B_{12} synthesis (more than 25 genes in 6 operons) on three plasmids led to the production of coenzyme B_{12} under both anaerobic and aerobic conditions (Ko et al., 2014).

While *K. pneumoniae* and *E. coli* are by far the most widely used production hosts, a number of other bacteria have been engineered for 3-HP production as well. *P. denitrificans* synthesizes coenzyme B_{12} in aerobic conditions where $NAD(P)^+$ is efficiently regenerated (Zhou et al., 2013). Introduction of glycerol dehydratase and glycerol dehydratase reactivase from *K. pneumoniae* allowed for the production of 3.4 g/L 3-HP. By further introducing the *K. pneumoniae* aldehyde dehydrogenase, the yield increased to 4.9 g/L (Zhou et al., 2013).

FIGURE 1 | Potential industrial uses of 3-hydroxypropanoic acid. Global market of derived compounds are indicated; >1,000 M$ (blue), 100–1,000 M$ (green), <100 M$ (yellow).
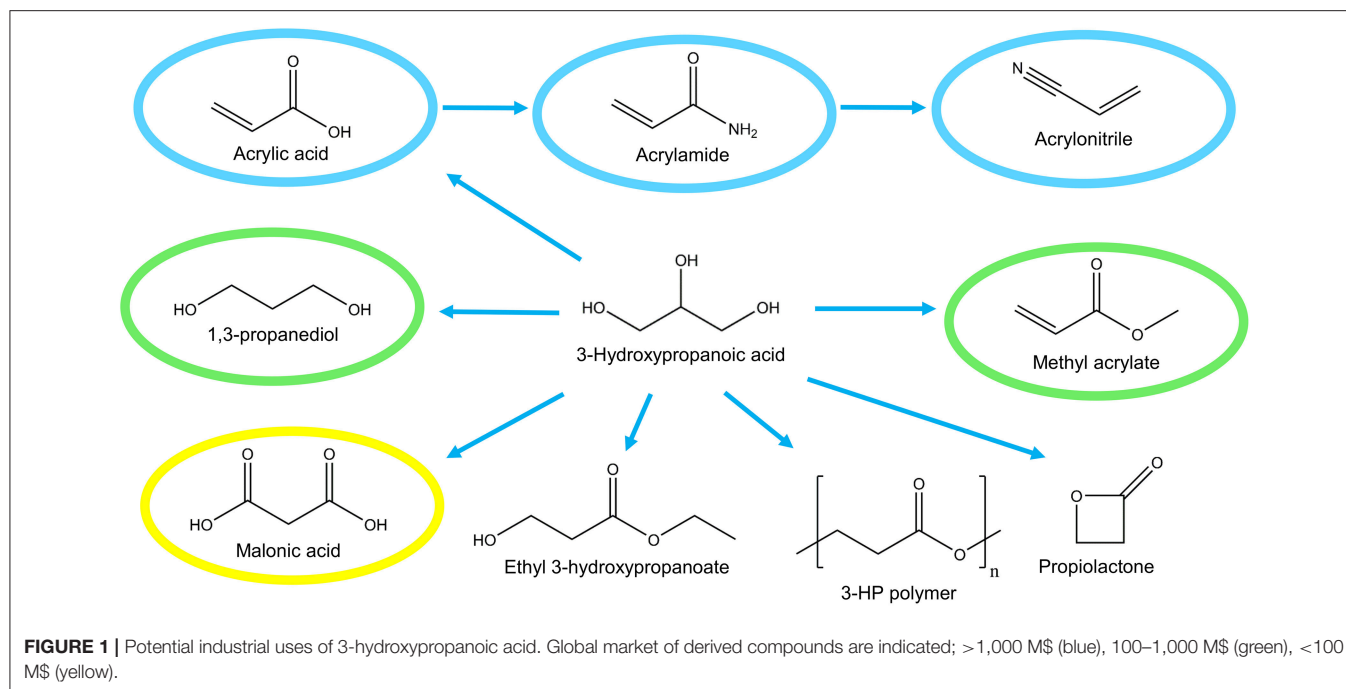
Another bacterium able to synthesize coenzyme B_{12}, *Shimwellia blattae*, is a 1,3-PDO producer using a native coenzyme B_{12}-dependent glycerol dehydratase. Introduction of various genes including aldehyde dehydrogenase from *Pseudomonas putida* KT2442 enabled the production of 0.26 g/L Poly-3HP using crude glycerol from biodiesel production as the substrate (Heinrich et al., 2013).

Corynebacterium glutamicum was engineered to produce 3-HP from glucose and xylose. By a combination of efforts, an impressive titer of 62.6 g/L 3-HP was obtained in a fed-batch fermentation (Chen et al., 2017). The steps taken to achieve this titer included enabling efficient production of glycerol from glucose, introduction of the *K. pneumoniae pduCDEGH* genes encoding diol dehydratase and its reactivase and *Cupriavidor necator* aldehyde dehydrogenase and modification in sugar uptake and glycolytic flux. In *Bacillus subtilis*, introduction of glycerol dehydratase along with its reactivases, and aldehyde dehydrogenase from *K. pneumoniae,* enabled the production of 3-HP from glycerol. Upon inactivation of glycerol kinase and optimization of growth conditions, a 3-HP titer of 10 g/L was obtained in a shake flask culture (Kalantari et al., 2017). A drawback of both of these bacteria is their inability to synthesize coenzyme B_{12}, which thus has to be supplemented to the production medium.

In an engineered cyanobacterium *Synechococcus elongatus* CO_2 was converted via glycerol to 3-HP, albeit at very low titer (31.7 mg/L). *Synechococcus elongatus* is unable to synthesize coenzyme B_{12}, but the problem was alleviated by producing 3-HP under anaerobic conditions using the oxygen-sensitive B_{12}-independent glycerol dehydratase from *Clostridium butyricum* (Wang et al., 2015). The use of glycerol as an additional carbon source in a strain engineered to assimilate glycerol has been suggested for increasing productivity (Kanno and Atsumi, 2017).

THE SUITE OF ENZYMES USED IN HETEROLOGOUS 3-HP SYNTHESIS PATHWAYS

As mentioned above, there are two principal pathways for production of 3-HP from glycerol. In both CoA-dependent and -independent pathways, the initial step is the conversion of glycerol to 3-HPA. In the CoA-dependent pathway, 3-HPA is converted via 3-hydroxypropanoyl-coenzyme A (3-HP-CoA) and 3-hydroxypropanoyl-phosphate (3-HP-P) to 3-HP, catalyzed by propionaldehyde dehydrogenase, phosphotransacylase, and propionate kinase, respectively (Dishisha et al., 2014). In the CoA-independent pathway, 3-HPA is converted directly to 3-HP by the action of aldehyde dehydrogenase (Kumar et al., 2013a). Both of the pathways have characteristics that are of importance for the design of cell factory and process conditions, for example, oxygen-sensitivity of enzymes and the need for co-factors such as coenzyme B_{12}, and NAD^+. In the following, we will describe the enzymes that make up these two pathways. Where available, kinetic data for the enzymes is presented in **Table 2**, to facilitate comparison.

Glycerol and Diol Dehydratase
Coenzyme B_{12}-Dependent Dehydratases
Coenzyme B_{12}-dependent glycerol dehydratase (EC 4.2.1.30) and diol dehydratase (EC 4.2.1.28) are isofunctional enzymes that catalyse the dehydration of 1,2-diols to the corresponding aldehyde (e.g., glycerol to 3-HPA). Both glycerol dehydratase and diol dehydratase are composed of three subunits that form a dimer of a heterotrimer ($\alpha_2\beta_2\gamma_2$; Shibata et al., 1999; Yamanishi et al., 2002). In mycobacteria, the α and β subunits are fused in a single polypeptide (Liu et al., 2010). Both of the enzymes

FIGURE 2 | Synthesis of 3-hydropropanoic acid from glycerol via the CoA-independent and -dependent pathways.

TABLE 1 | Overview of bacterial species applied for production of 3-HP.

Bacterial species	Natural producer	B$_{12}$ synthesis
Lactobacillus reuteri	Yes	Yes
Klebsiella pneumonia	Yes	Yes
Escherichia coli	No	No
Pseudomonas denitrificans	No	Yes
Shimwellia blattae	No[a]	Yes
Corynebacterium glutamicum	No	No
Bacillus subtilis	No	No
Synechococcus elongates	No	No

[a]*Naturally produces 3-HPA.*

hydratase, which in turn binds coenzyme B$_{12}$ to regenerate the active form of the enzyme (Toraya, 2003).

From the above, it can be deduced that the expression of five proteins is necessary for the efficient catalysis of glycerol to 3-HPA. In the context of engineering microorganisms for 3-HP production, the most widely used glycerol dehydratase is that of *K. pneumoniae* (*dhaB123*, and *gdrAB*), which has been used in *E. coli* (Rathnasingh et al., 2009), *K. pneumoniae* (Wang et al., 2013), *B. subtilis* (Kalantari et al., 2017), and *S. elongatus* (Wang et al., 2015). The use of other dehydratases has also been reported. In *E. coli*, the glycerol dehydratase from *Lactobacillus brevis* encoded by *dhaB123* and its reactivase *dhaR12* were used (Kwak et al., 2013). The diol dehydratase (*pduCDE*) and activator (*pduGH*) from *K. penumoniae* were used in *C. glutamicum* (Chen et al., 2017).

Considering the problem of enzyme instability, it is somewhat surprising that relatively few glycerol dehydratases have been tested. These enzymes have, by now, been discovered in many bacteria and this should provide a resource for enzyme discovery endeavors. In fact, for three selected glycerol dehydratases, the α subunit was systematically swapped, which led to identification of several combinations with improved stability and activity (Qi et al., 2006). Interestingly, fusion of the α and β subunit of *K. pneumoniae* glycerol dehydratase led to an increase in the k_{cat} (albeit with concomitant increase in K_M; Wang et al., 2009). As mentioned, mycobacteria contain a glycerol dehydratase variant in which the α and β subunits are fused, but to our knowledge they have not been characterized or used for 3-HP production. Enzyme engineering could also prove beneficial for generation of enzyme variants with improved properties such as increased stability and activity. As an example, by rational engineering, improved resistance to mechanism-based inactivation was conferred to a glycerol dehydratase from *Klebsiella oxytoga* (Yamanishi et al., 2012).

Coenzyme B$_{12}$-Independent Glycerol Dehydratase

This class of glycerol dehydratases is of particular interest since it negates the need for the rather costly coenzyme B$_{12}$. This enzyme also performs a radical catalysis, but instead of coenzyme B$_{12}$, it uses S-adenosylmethionine as a co-factor (Raynaud et al., 2003). This glycerol dehydratase is a homodimer and it is thus structurally simpler than its coenzyme B$_{12}$-dependent

catalyze a radical process in which an adenosyl radical is formed by homolytic cleavage of the Co-C bond in coenzyme B$_{12}$ (Daniel et al., 1998). The adenosyl radical abstracts a hydrogen from the substrate, glycerol, to generate a substrate radical. The substrate radical, upon rearrangement, re-abstracts a hydrogen to form the final product and regenerate coenzyme B$_{12}$ (Daniel et al., 1998). If a radical side reaction takes place, coenzyme B$_{12}$ is not regenerated, and instead, a catalytically inactive cobalamin species is formed, which binds tightly to the enzyme and thereby inactivates it (Toraya, 2003). Radical side reaction can be induced by both glycerol and oxygen, making the enzyme sensitive to oxygen (Wei et al., 2014). The inactive hydratase can be re-activated by the action of glycerol/diol dehydratase reactivase in the presence of ATP and intact coenzyme B$_{12}$ (Mori and Toraya, 1999). Glycerol/diol dehydratase reactivase consists of two subunits that form a heterotetramer ($\alpha_2\beta_2$; Liao et al., 2003). The reactivase binds ATP and hydrolyzes it to ADP. The ADP-bound reactivase forms a complex with the inactivated glycerol dehydratase. This leads to release of the damaged coenzyme B$_{12}$. Subsequently, reactivase binds ATP and is released from the

TABLE 2 | Summary of characterized enzymes of relevance in 3-HP production.

Source	gene / protein	K_M (mM)	V_{max} (U/mg)	k_{cat} (s^{-1})	k_{cat}/K_M x 10^3 (M^{-1}S^{-1})	Other characteristics	References
B12-dependent glycerol/diol dehydratase							
K. pneumoniae	*dhaB123*	n.a.	n.a.	n.a.	n.a.		
K. pneumoniae	*pduCDE*	n.a.	n.a.	n.a.	n.a.		
L. reuteri	*pduCDE*	0.60	130.1	n.a.	n.a.	pH$_{opt}$ 8.5	Qi et al., 2006
C. freundii	*gldABC*	0.59	142.9	n.a.	n.a.	pH$_{opt}$ 8.5	Qi et al., 2006
L. brevis	*dhaB*	n.a.	n.a.	n.a.	n.a.	pH$_{opt}$ 7.0	Kwak et al., 2013
B12-independent glycerol dehydratase							
C. butyricum		n.a.	n.a.	n.a.	n.a.	B$_{12}$-independent	
Propionaldehyde dehydrogenase							
L. reuteri	*pduP*	1.18	0.35	n.a.	n.a.	broad substrate specificity	Luo et al., 2011
		n.a.	n.a.	n.a.	n.a.	Inhibition by 3-HPA (from 7 mM). NAD$^+$ preference	Sabet-Azad et al., 2013
Aldehyde dehydrogenase							
E. coli K-12	*aldH*	n.a.	n.a.	28.5	58.6	NAD$^+$ preference	Jo et al., 2008
		n.a.	38.16	n.a.	57.28		Chu et al., 2015
		0.49	38.10	n.a.	n.a.		Ko et al., 2012
E. coli JM109	*aldA*	0.31	28.4	n.a.	n.a.	NAD$^+$ preference	Zhu et al., 2009
C. necator	*gabD4*	n.a.	55.12	n.a.	71.48		Chu et al., 2015
C. necator	*gabD4* E209Q/E269Q	n.a.	78.07	n.a.	162.34		Chu et al., 2015
K. pneumoniae DSM 2026	*puuC*	0.48	22.25	19.81	41.44	pH$_{opt}$ 8.0	Raj et al., 2010
		0.48	27.73	n.a.	n.a.		Ko et al., 2012
A. brasilense	KGSADH	1.6	n.a.	15	9)	NAD$^+$ preference	Park et al., 2017
A. brasilense	KGSADH 108-QR	0.17	n.a.	6	35	NAD$^+$ preference	Park et al., 2017
Aldehyde oxidase							
Pseudomonas sp. AIU 362	*aloD*	6.7	41.8	n.a.	n.a.	NAD$^+$-independent	Li et al., 2014

counterpart that is composed of three subunits (Raynaud et al., 2003). While it would seem the obvious enzyme of choice, it is important to note that it is extremely sensitive to oxygen and its use requires production under strict anaerobic conditions (Raynaud et al., 2003).

Considering the fact that coenzyme B$_{12}$ is not produced in most of the microorganisms that have been used as potential 3-HP production hosts, the use of a coenzyme B$_{12}$-independent dehydratase would be desirable. So far, its use has only been reported in a few studies. The *C. butyricum* glycerol dehydratase was used in the cyanobacterium *S. elongatus* for production of 3-HP under anaerobic conditions. As expected, it was not functional under aerobic conditions (Wang et al., 2015). In case of *E. coli*, the coenzyme B$_{12}$-independent dehydratase was used in the context of 1,3-PDO production. Although the reported 1,3-PDO production was relatively low, an accumulation of 3-HPA was observed, thus demonstrating functionality of the glycerol dehydratase (Dabrowski et al., 2012). Furthermore, it was used for engineering *E. coli* to produce poly (3-Hydroxypropanoate) (Andreeßen et al., 2010). The scarcity of reported use of this enzyme could be due to difficulties in engineering 3-HP producing strains using this dehydratase. Alternatively, and perhaps more likely, is that reports are scarce because of the requirement of strict anaerobic conditions for the enzyme to be functional, which is not compatible with growth of most production microorganisms.

Enzymes for Converting 3-HPA to 3HP
Enzymes in the CoA-Dependent Pathway

In the CoA-dependent pathway, 3-HPA is converted via 3-HP-CoA and 3-HP-P into 3-HP. This requires three enzymes: propionaldehyde dehydrogenase, phosphotransacylase, and propionate kinase, respectively (Dishisha et al., 2014). The propionaldehyde dehydrogenase requires the cofactor NAD$^+$, while one ATP is generated in the conversion of 3-HP-P to 3-HP. This pathway has been described in several organisms including *L. reuteri* and *K. pneumoniae* (Luo et al., 2012; Dishisha et al., 2014). The propionaldehyde dehydrogenase PduP from *L. reuteri* has been purified and characterized *in vitro* and shown to exhibit activity toward a broad spectrum of aldehydes, including 3-HPA, using either NAD$^+$ (preferred) or NADP$^+$ as a cofactor (Luo et al., 2011). *Lactobacillus reuteri* PduP was also reported to display substrate inhibition for 3-HPA (at 7 mM; Sabet-Azad et al., 2013).

The CoA-dependent pathway has been used in several engineered bacteria. The *K. pneumoniae pduPLW* genes were introduced in an *E. coli* strain harboring the *K. pneumoniae* glycerol dehydratase. Here they were shown to be functional (although the obtained titer was lower than that obtained using an aldehyde dehydrogenase from *Azospirillum brasilense*; Honjo et al., 2015). In *K. pneumoniae*, overexpression of the first gene in the pathway, *pduP*, led to a 4-fold (0.72 g/L) increase in the 3-HP titer (Luo et al., 2011). It could be speculated that the

native *K. pneumoniae* CoA-dependent pathway enzymes were responsible for the conversion of 3-HP-CoA to 3-HP, and that over-expression of all three genes in the pathway would further increase the titer.

It remains to be seen whether the CoA-dependent pathway could be an alternative to the more widely used aldehyde dehydrogenases for the conversion of 3-HPA to 3-HP. Although both pathways use an NAD(P)$^+$, ATP is generated in the CoA-dependent pathway. However, this should be balanced against the metabolic burden of over-expressing three enzymes instead of one. Further, it is unclear if the substrate inhibition exhibited by *L. reuteri* PduP at a relatively low concentration of 3-HPA is a general trait of propionaldehyde dehydrogenase, or whether a more robust enzyme can be discovered or engineered.

Aldehyde Dehydrogenase

Aldehyde dehydrogenase (EC 1.2.1.3) has decidedly been the enzyme of choice for conversion of 3-HPA to 3-HP in engineered cells. It catalyses the conversion of an aldehyde to its corresponding carboxylic acid using NAD(P)$^+$ as a co-factor. In the context of the 3-HP production pathway, it is of utmost importance that enzyme activity is sufficient to ensure that no accumulation of the toxic intermediate 3-HPA takes place. Aldehyde dehydrogenases are present in most organisms, but generally their activity is not high enough to sustain a high production of 3-HP (Raj et al., 2008; Zhu et al., 2009). Consequently, many studies have focused on discovering suitable aldehyde dehydrogenases. Earlier studies mainly employed the aldehyde dehydrogenase AldH from *E. coli* K-12 in both, *E. coli* and *K. pneumoniae* (Raj et al., 2008; Zhu et al., 2009). In a study by Chu and co-workers, an indirect comparison (based on 3-HP production in *E. coli*) was made amongst 17 aldehyde dehydrogenases from various organisms, benchmarked against *E. coli* AldH. One aldehyde dehydrogenase performing better than AldH, GabD4 from *C. necator*, was identified (Chu et al., 2015). The enzyme was further improved by enzyme engineering, substantially increasing its V_{max} and k_{cat}/K_M, and used for high-titer production of 3-HP in both *E. coli* (71.9 g/L) and *C. glutamicum* (62.6 g/L) (Chu et al., 2015; Chen et al., 2017).

The α-ketoglutaric semialdehyde dehydrogenase (KGSADH) from *A. brasilense* has also been used in several studies. KGSADH was evaluated against *E. coli* AldH and PuuC from *K. pneumoniae* (Ko et al., 2012). This study indicated higher activity (V_{max}) of KGSADH in extracts of the *K. pneumoniae* strain, while in contrast it had the lowest V_{max} and affinity for 3-HPA when enzymes purified from *E. coli* were evaluated (Ko et al., 2012). Whether this discrepancy is due to e.g., stability *in vivo* is unclear. The structure of KGSADH has been solved and has provided a basis for rational enzyme engineering. This has yielded enzyme variants with improved activity toward 3-HPA (Park et al., 2017; Son et al., 2017; Seok et al., 2018).

PuuC from *K. pneumoniae* has also been used in several production hosts, including *K. pneumoniae* and *B. subtilis* (Li et al., 2016; Kalantari et al., 2017). Substrate specificity was evaluated against 3-HPA and three other aldehydes indicating broad substrate specificity (and lowest activity when 3-HPA was used; Raj et al., 2010). Comparative studies indicated that *E.*

coli AldH and *C. necator* GabD4 E209Q/E269Q were better alternatives for converting 3-HPA to 3-HP (Huang et al., 2012; Chen et al., 2017). Nevertheless, PuuC was successfully used in an engineered *K. pneumoniae* strain that produced 3-HP with a titer of 83.8 g/L (highest reported) in a bioreactor (Li et al., 2016).

Several other aldehyde dehydrogenases have been tested in the context of 3-HP production. *E. coli* YneI was shown to selectively target 3-HPA over other aldehydes (Luo et al., 2013), although it was not tested against succinic semialdehyde which was previously reported as its primary substrate (Kurihara et al., 2010). The aldehyde dehydrogenase DhaS from *B. subtilis* has also been suggested as an aldehyde dehydrogenase that is specific for 3-HPA (Su et al., 2015).

Due to the oxygen-sensitivity of glycerol dehydratase, 3-HP production is often performed under anaerobic or microaerophilic conditions, where the regeneration of NAD$^+$ is diminished. Thus, efficient 3-HP production becomes a compromise between assuring optimal glycerol dehydratase activity (low oxygen level) and NAD$^+$-dependent aldehyde dehydrogenase activity (high oxygen level). An interesting strategy was suggested to overcome the limitation imposed by NAD$^+$, namely, the use of an aldehyde oxidase in place of dehydrogenase. Li and co-workers characterized the NAD$^+$-independent aldehyde oxidase from *Pseudomonas* sp. AIU 362 and found it to exhibit a broad substrate specificity but relatively low affinity for 3-HP. When expressed in *K. pneumoniae*, a relatively low 3-HP titer was obtained (Li et al., 2014). As can be seen in **Table 2**, the affinity of aldehyde oxidase to 3-HPA is several-fold lower than the described aldehyde dehydrogenases, with a K_M in the range where 3-HPA becomes toxic to the cell. Considering the importance of maintaining a low concentration of 3-HPA, better enzymes would likely be needed to successfully exploit aldehyde oxidase for 3-HP production. To the best of our knowledge, no other attempts have been made at identifying a more suitable aldehyde oxidase for the conversion of 3-HPA to 3-HP.

As outlined in this section, many enzyme variants have been used in the design of 3-HP-producing strains. While some of the enzymes have been characterized biochemically, it is obvious from **Table 2** that this is not a norm in the field. Nevertheless, a better understanding of the catalytic properties of these enzymes could provide a platform for more optimal selection of enzymes for constructing 3-HP pathways.

STRAIN ENGINEERING

There are several parameters to consider when engineering bacteria to produce 3-HP. As covered in sections Production Hosts and The Suite of Enzymes Used in Heterologous 3-HP Synthesis Pathways, a suitable host should be established and in most cases some or all of the enzymes needed for converting glycerol to 3-HP should be introduced. This provides a basis for 3-HP production, but is not enough to ensure the desired high titers. There are several hurdles that need to be overcome, and these include, but are not limited to, proper balancing of enzyme activities to prevent buildup of the toxic intermediate

3-HPA, efficient channeling of substrate into product, co-factor regeneration, prevention of by-product formation and countering of stress. In the following, we will explore challenges and strategies that have been applied in the engineering of strains for 3-HP production.

Improving 3-HP Production by Optimizing Expression of the Production Pathway

As illustrated above, a number of different enzymes have been used for the construction of synthetic operons to confer the ability to produce 3-HP to various bacteria. In a number of these studies, it was shown that further optimization of the expression of the pathway genes could enhance 3-HP production.

A critical parameter is to ensure that the intermediate product 3-HPA is kept at a low concentration. 3-HPA (also known as reuterin) is a broad range antimicrobial compound. Its minimal inhibitory concentrations for various bacteria are in the range of <1.9–50 mM, and it exerts its effect via modification of thiol groups of proteins and small molecules (Cleusix et al., 2007; Schaefer et al., 2010). Considering the fact that 3-HPA is toxic to bacteria even at minute concentrations, it is evident that balancing enzyme activities to prevent 3-HPA accumulation is critical. To this end, different approaches have been attempted to fine-tune the expression levels of glycerol dehydratase and aldehyde dehydrogenase. The order in which the genes encoding aldehyde dehydrogenase and glycerol dehydratase are arranged was shown to impact 3-HP production. The rationale here is that the gene adjacent to the promoter in an operon is normally more expressed. Specifically, the arrangement favoring aldehyde dehydrogenase expression led to a higher 3-HP production due to the diminished build-up of 3-HPA (Li et al., 2013a). Different promoters for driving the expression of aldehyde dehydrogenase gene has also been successfully evaluated (Li et al., 2016).

While these approaches are simple, they do not offer a high degree of tunability, and a number of more advanced approaches have been applied for optimization of expression levels. Optimization of expression level can be done by applying various $5'$ untranslated regions ($5'$UTRs) in the genetic constructs, which leads to differences in the translation rate. Using this approach, a better balance of the enzyme activities was obtained by fine-tuning expression of *dhaB1* encoding a glycerol dehydratase subunit. Testing only four different UTRs led to construction of a strain that enabled a 2.4-fold improvement of 3-HP titer in shake flask experiments (Lim et al., 2016). Several tools are available for the prediction of translation rate of $5'$UTRs, including the UTR Designer (Seo et al., 2013), and the RBS Calculator v2.0 (Espah Borujeni et al., 2014). To the same end, a study showed potential in tuning translation by modulation of the Shine-Dalgarno sequence for which an online tool, EMOPEC, is available (Bonde et al., 2016). For a selection of 106 Shine-Dalgarno sequences, the measured protein level was within 2-fold of the predicted level in 91 % of cases (Bonde et al., 2016). While *in silico* prediction of expression level is a very attractive venue, it should be noted that even for the good predictors, the number of false predictions appear to be significant (Bonde et al., 2016). Nevertheless, these computational approaches should

allow for design and evaluation of smaller, more focused libraries, compared to random mutagenesis libraries.

In most studies, researchers have taken advantage of plasmids for the expression of relevant genes in the production host. While these studies provide proof of principle, they might not be optimal for industrial production due to the well-known problem of plasmid instability in bioreactors (Gao et al., 2014). Consequently, the generation of genetically stable strains where expression cassettes are integrated into the genome will be a necessity. Additionally, the elimination of plasmids might improve the production via a reduced metabolic burden associated with plasmid maintenance and replication (Silva et al., 2012). To construct a plasmid-free strain, the *E. coli aldH* was inserted in the genome of *K. pneumoniae*. While 3-HP production was improved (over parent strain) the 3-HP titer upon production in shake flasks was low (Wang and Tian, 2017). In *B. subtilis*, a plasmid-free strain allowed the production of 3-HP in shake flasks with a titer of 10 g/L (Kalantari et al., 2017).

Considering that both multicopy vectors and strong promoters have been used for the expression of relevant genes, one foreseeable difficulty in shifting to a single (or a few) genome-integrated expression cassette will be to ensure sufficient protein synthesis in the cell. A recent study described an approach based on translational coupling between the gene of interest and an antibiotic resistance cassette, thus making a high rate of translation selectable via an increased resistance to an antibiotic (Rennig et al., 2018). With limited screening, the production of two selected proteins (a nanobody and an affibody) in *E. coli* was improved 2- and 10-fold, respectively (Rennig et al., 2018). The methodology was shown to also work in Gram-positive bacteria *B. subtilis* and *Lactococcus lactis* where production of a sialidase and a tyrosine ammonia lyase was improved 2- and 8-fold, respectively (Ferro et al., 2018).

Engineering Host Metabolism for Improving 3-HP Production

Besides integrating an efficient pathway for the conversion of glycerol to 3-HP, it is often necessary to further modulate the host metabolism, in order to direct the substrate into the production pathway more efficiently, eliminate unwanted by-products and/or reduce stress.

One of the more common approaches is to ensure that glycerol is converted more efficiently into 3-HP. To this end, targeting the first step in the glycerol utilization pathways is common. In *E. coli*, two pathways for glycerol utilization exist. Glycerol is converted to the glycolytic metabolite dihydroxyacetone phosphate either via glycerol-3-phosphate catalyzed by glycerol kinase under aerobic conditions or via dihydroxyacetone catalyzed by glycerol dehydrogenase under anaerobic conditions (Durnin et al., 2009; **Figure 3**). Inactivating the gene *glpK* (which encodes glycerol kinase in *E. coli*) increased the titer of 3-HP 1.6-fold in an aerobic fed batch fermentation when co-fed with glucose and glycerol. Under the conditions used (aerobic), further inactivating the gene encoding glycerol dehydrogenase had no effect (Kim et al., 2014). Similarly, the 3-HP titer increased in *B. subtilis* co-fed on glycerol and glucose upon inactivation of

FIGURE 3 | Overview of the glycerol metabolism in bacteria. Reactions consuming NADH/NAD$^+$ are indicated.

glpK (Kalantari et al., 2017). Inactivation of *glpK* is possible in process setups where glycerol is not the sole carbon source. When cells grow only on glycerol, targeting *glpK* becomes less trivial as the fluxes toward biomass accumulation and 3-HP production need to be balanced. In *E. coli*, the *glpK* was placed under control of an inducible promoter and fine-tuning of its expression led to an increase in both 3-HP titer and yield on glycerol (Jung et al., 2014). In that study, the gene *glpF* encoding glycerol facilitator was also overexpressed to increase the influx of glycerol in the cell, which led to a modest increase in 3-HP production (Jung et al., 2014). Interestingly, the opposite approach was evaluated by Su and co-workers. They overexpressed glycerol dehydrogenase in *K. pneumoniae* to stimulate growth, and found that this strain not only exhibited faster initial growth and similar final biomass yield while using less glycerol, but also exhibited increased 3-HP production (Su et al., 2014).

In natural 3-HP producers such as *K. pneumoniae* and *L. reuteri*, 3-HPA can be oxidized to 3-HP and reduced to 1,3-PDO (Zhu et al., 2009; Dishisha et al., 2014). The conversion of 3-HPA to 1,3-PDO is catalyzed by an oxidoreductase. In *K. pneumoniae*, the most important oxidoreductase appears to be the 1,3-propanediol reductase encoded by *dhaT*. However, significant amounts of 1,3-PDO are still produced in the *dhaT* mutant, indicating that other oxidoreductases in the cell can also catalyse this reaction (Ko et al., 2012). In a subsequent attempt to eliminate 1,3-PDO formation, four oxidoreductases were inactivated, but the quadruple mutant still retained the ability to produce 1,3-PDO (Ko et al., 2015). With respect to cofactor balance, 3-HP synthesis requires NAD$^+$ while 1,3-PDO formation requires NADH and thus regenerates NAD$^+$. Thus, the elimination of 1,3-PDO formation adversely affects the NAD$^+$/NADH balance, and in fact, it was reported that the *dhaT* mutation under 3-HP production conditions impeded

cell growth (Ko et al., 2012). In *E. coli*, 3-HPA is also converted to 1,3-PDO by the action of a broad-range aldehyde oxidoreductase encoded by *yqhD*. While the deletion of *yqhD* dramatically reduces 1,3-PDO formation, it is also here not fully abolished, likely due to the presence of alternate oxidoreductases (Tokuyama et al., 2014).

Other important by-products in 3-HP production are lactate and acetate. The deletion of *K. pneumoniae ldhA* encoding lactate dehydrogenase was reported to eliminate lactate formation (Kumar et al., 2013b). Another study reported a reduction in lactate formation upon inactivation of *ldh1* and *ldh2* in *K. pneumoniae* (Li et al., 2016). Likewise, to reduce the acetate formation, the synthesis gene *pta* was inactivated, leading to a reduction but not elimination of acetate formation in *K. pneuomoniae* (Li et al., 2016). To further optimize 3-HP producing cell factories, it is essential to assure that by-product formation is reduced to a minimum. Besides the obvious loss of carbon that could have otherwise been used for biomass and product formation, it also leads to a more complex fermentation broth thus potentially increasing costs associated with downstream purification.

The NAD$^+$/NADH balance is of importance for the 3-HP production due to the NAD$^+$-dependence of the employed aldehyde dehydrogenases. The synthesis of by-products 1,3-PDO and lactate also generates NAD$^+$. Elimination of respective pathways thus further skews the NAD$^+$/NADH balance (**Figure 3**). Under aerobic conditions, NAD$^+$ is also regenerated by the electron transport chain. However, due to the oxygen-sensitivity of glycerol dehydratase, 3-HP production is often performed under microaerophilic or anaerobic conditions where NAD$^+$ regeneration is reduced. To counter these effects, various enzymes can be used to modulate the NAD$^+$/NADH balance. In the context of 3-HP production, over-expression of the genes

encoding either NADH oxidase, NADH dehydrogenase, or glycerol-3-phosphate dehydrogenase *in K. pneumoniae* showed potential in stimulating 3-HP production via regenerating the NAD^+ pool in all cases (Li et al., 2013b).

From an economical point of view, it would be beneficial to use a production strain capable of synthesizing coenzyme B_{12}. However, even in *K. pneumoniae* which is capable thereof, it has been reported that addition of coenzyme B_{12} increases the glycerol dehydratase activity (Ashok et al., 2013). This could indicate that upregulation of coenzyme B_{12} synthesis is another parameter for increasing the glycerol dehydratase activity. Studies on *Bacillus megaterium* and more recently on *P. denitrificans* show good potential for increasing coenzyme B_{12} synthesis, e.g., by removing the riboswitch-based feedback inhibition system (Biedendieck et al., 2010; Nguyen-Vo et al., 2018).

The application of genome scale metabolic models is a common approach in the field of metabolic engineering and has been used successfully in multiple studies. Using a genome-scale metabolic model for *K. pneumoniae*, no single mutants that would improve 3-HP production were predicted. Instead, a double knockout of *tpi* (triose phosphate isomerase) and *zwf* (glucose-6-phosphate-1-dehydrogenase) involved in central metabolism was suggested. The introduction of these two mutations led to a 4.4-fold increase in the 3-HP yield compared to the parent strain (Tokuyama et al., 2014). Predictions were made of beneficial mutations in *B. subtilis*, but here only the more obvious candidate, glycerol kinase, was suggested as a beneficial mutation (Kalantari et al., 2017).

Another important aspect of optimizing the production strain is to improve the robustness of the strain, specifically, alleviating stress. Probably the most significant stressor is that imposed by the intermediate product 3-HPA and, to a lesser extent, the product 3-HP. Warnecke et al. developed a method to screen for regions and genes that could be involved in the tolerance to 3-HP. Starting with a genomic library of *E. coli* clones, each carrying a plasmid with a genomic DNA insert, they selected for increased tolerance to 3-HP and subsequently quantified the enriched plasmid DNA by microarray analysis (Warnecke et al., 2010). Using this method, they concluded that 3-HP inhibition is due to the limitations in the chorismate and threonine pathways. Surprisingly, they found that over-expression of almost any of the genes in the pathways dramatically alleviated the stress imposed by 3-HP (Warnecke et al., 2010). In a follow up study, the authors identified a 21-amino acid peptide that, when expressed, increased the 3-HP tolerance about 2.3-fold (Warnecke et al., 2012).

The potential of using omics data to identify the targets for improving 3-HP tolerance has been demonstrated. Using 2D gel-based proteomics, Liu and co-workers identified 46 up- and 23 down-regulated proteins upon challenging *E. coli* with 5 g/L of 3-HP. Over-expression of several of these proteins alleviated the stress imposed by 3-HP (Liu et al., 2016).

In recent years, there have been key improvements in the technology for generating targeted libraries of mutants, which is exemplified in a study by Liu and co-workers where a method termed iCREATE for iterative genome editing was used to generate 162,000 mutations in 115 genes (Liu et al., 2018). Using this approach, the production of 3-HP via the Malonyl-CoA pathway was increased 60-fold (Liu et al., 2018). However, for its successful employment, appropriate screening and/or selection techniques need to be in place. Small molecule biosensors based on transcription factors or riboswitches can couple metabolite concentration to a measurable output such as fluorescence for selection or selectable output such as antibiotic resistance (Liu et al., 2017). It is, thus, of particular importance that a 3-HP-inducible system was identified in *P. putida* recently and demonstrated to be functional when expressed in *E. coli* (Hanko et al., 2017). This biosensor has already found application for screening in the directed evolution of the aldehyde dehydrogenase KGSADH which improved the catalytic efficiency of the enzyme 2.8-fold (Seok et al., 2018).

Combining approaches such as iCREATE for generating diversity, and biosensors for detection or selection of improved variants should hold high promises for moving bacterial 3-HP production strains to industrially relevant levels of production.

PROCESS ENGINEERING FOR OPTIMIZING 3-HP PRODUCTION

In a majority of the studies reviewed for this paper, the ability of the recombinant strains to produce 3-HP was tested in 1.5–5 L bioreactors operated in fed-batch mode (**Table 3**; **Supplementary Table 1**). A fed-batch operation typically entails growing the cells until the end of the exponential phase. This is followed by the addition of feed medium containing the substrates, thus keeping the growth rate of the microorganism at a desired level (Liu, 2017). A fed-batch mode of operation has certain advantages over batch fermentation. For instance, a fed-batch operation has the ability to prevent overflow metabolism, that is, the phenomenon where fast-growing cells, in the presence of oxygen, utilize fermentation instead of respiration to produce energy (Basan et al., 2015). Another advantage of a fed-batch operation is its ability to maintain the substrate concentration at a low level (de Fouchécour et al., 2018). This is particularly important when the substrate is glycerol. It was reported that no growth of *K. pneumoniae* cells could be seen above 110 and 133 g/L glycerol (extrapolated values) in aerobic and anaerobic conditions, respectively, with glycerol concentrations higher than 40 g/L showing significant growth inhibition (Cheng et al., 2005). Consequently, high titers of 3-HP cannot be obtained from glycerol in a batch fermentation. In *L. reuteri*, the conversion of glycerol to 3-HPA was reported to be about 10 times faster than the subsequent conversion of 3-HPA to 3-HP and 1,3-PDO, resulting in the accumulation of the toxic intermediate 3-HPA. To circumvent this problem, a fed-batch fermentation process was established. It produced 3-HP and 1,3-PDO without any accumulation of 3-HPA. This was achieved by identifying a maximum value of specific glycerol consumption rate at which only 1,3-PDO and 3-HP are produced, and no accumulation of 3-HPA takes place. By maintaining the glycerol amount in the bioreactor such that its specific consumption rate stayed under this maximum value, a 3-HP titer of 14 g/L with a productivity

TABLE 3 | Summary of studies reporting the highest 3-HP titer, yield, and productivity in various production hosts.

Strain	Carbon source	Fermentation conditions	Titer[a]	Yield[a]	Produc-tivity[a]	Reference
K. pneumoniae Δ*ldh1*Δ*ldh2*Δ*pta*_pTAC-*puuC*	Glycerol	5-L bioreactor, fed-batch fermentation, pH 7.0, 37°C, 1.5 vvm air, 400 rpm	83.8	n.a.	1.16	Li et al., 2016
K. pneumoniae Kp4 Δ*ldhA*Δ*dhaT_ lac-aldH*	Glycerol	5-L bioreactor, fed-batch fermentation, pH 7.0, 37°C, 2.2 L/min air, 450 rpm	61.9	0.58	1.62[b]	Jiang et al., 2018
E. coli W3110 Δ*ackA-pta*Δ*yqhD_ T7-dhaB-gdrAB-gabD4*	Glycerol	5-L bioreactor, fed-batch fermentation, pH 7.0, 35°C, 1 vvm air, 500 rpm, vitamin B_{12} added externally	71.9	n.a.	1.8	Chu et al., 2015
E. coli W Δ*ackA-pta*Δ*yqhD_tac-dhaB-gdrAB-KGSADH*	Glycerol and glucose	5-L bioreactor, fed-batch fermentation, pH 7.0, 37°C, 1 vvm air, 500 rpm, coenzyme B_{12} added externally	40.5	0.97 g/g	1.35	Lim et al., 2016
L. reuteri RPRB3007[c]	Glycerol	3-L bioreactor, anaerobic fed-batch fermentation, pH 7.0, 37°C, 500 rpm	10.6	n.a.	1.08	Dishisha et al., 2014
B. subtilis 168 trp[+] Δ*glpK_pHyperspank-dhaB-gdrAB-puuC*	Glucose	Shake flask, 37°C, 200 rpm, coenzyme B_{12} added externally	10	0.79 g/g	n.a.	Kalantari et al., 2017
L. reuteri DSM 20016 and *G. oxydans* DSM 50049[d]	Glycerol	3-L bioreactor, Step 1: anaerobic fed-batch fermentation, pH 5.5, 37°C, 200 rpm. Step 2: aerobic batch fermentation, pH 5.5, 28°C, 1 L/min air, 800 rpm.	23.6	0.98	n.a.	Dishisha et al., 2015
K. pneumoniae and *G. oxydans*[e]	Glycerol	7-L bioreactor, fed-batch fermentation, Step 1: pH 7.0, 37°C, 0.2 vvm air, 150 rpm. Step 2: pH 5.5, 28°C, 0.5 vvm air, 600 rpm.	60.5	0.51	1.12	Zhao et al., 2015

[a] Unless otherwise mentioned, the units of titer, yield, and productivity are g/L, $mol_{3-HP}/mol_{Glycerol}$ and g/L.h, respectively

[b] Calculated based on the published data

[c] Mutated catabolite repression element (CRE) in the upstream region of the pdu operon

[d] L. reuteri converted glycerol to 3-HP and 1,3-propanediol followed by G. oxydans converting 1,3-propanediol to 3-HP

[e] K. pneumoniae converted glycerol to 1,3-propanediol followed by G. oxydans converting 1,3-propanediol to 3-HP

adhE, alcohol dehydrogenase; aldA, aldehyde dehydrogenase A (E. coli); aldH, γ-glutamyl-γ-aminobutyraldehyde dehydrogenase (E. coli); aldHk, NAD[+]-dependent homolog of E. coli aldH; CRE, catabolite repression element; dhaB, glycerol dehydratase; dhaR, dhaB reactivating factor in L. brevis; dhaS, putative aldehyde dehydrogenase from B. subtilis; dhaT, 1,3-propanediol oxidoreductase; DO, dissolved oxygen; frdA, succinate dehydrogenase; gabD4, aldehyde dehydrogenase (Cupriavidus necator); gdrA, gdrB, glycerol dehydratase reactivase; glpF, glycerol uptake facilitator protein; glpK, glycerol kinase; KGSADH, α-ketoglutaric semialdehyde dehydrogenase (A. brasilense); ldhA, lactate dehydrogenase; n.a. , not available; pduP, propionaldehyde dehydrogenase; puuC, γ-glutamyl-γ-aminobutyraldehyde dehydrogenase (K. pneumoniae); yqhD, NADPH-dependent aldehyde reductase/alcohol dehydrogenase.

of 0.25 g/L.h was obtained (Dishisha et al., 2014, 2015). This underlines how the knowledge of cell metabolism can be used to solve production problems by means of process engineering.

The challenge concerning the requirement of anaerobic conditions for coenzyme B_{12} generation and aerobic conditions for efficient NAD[+] regeneration for production of 3-HP from glycerol was tackled by electro-fermentation in a recently published study (Kim et al., 2017). Certain microorganisms possess the ability to deliver electrons to solid electrodes when grown under anaerobic conditions. These exoelectrogens, as these microorganisms are called, can be used to control the redox state in the cells independent of the electron transport chain or fermentation. Electro-fermentation utilizes this property of exoelectrogens to drive the unbalanced fermentations. Anodic electro-fermentation has been used for bioconversion of glycerol to 3-HP using recombinant *K. pneumoniae* over-expressing KGSADH from *A. brasilense*. An electrical potential of + 0.5 V vs. Ag/AgCl was applied to the anode, with 2-hydroxy-1,4-naphthoquinone used for shuttling the electrons between the bacteria and the anode. The transfer of electrons from the bacteria to the anode during anaerobic fermentation led to a decrease in

NADH/NAD[+] ratio in the cells, and hence to an enhanced 3-HP production as compared to the fermentative control (Kim et al., 2017). Although the 3-HP titer reached was low, the study nevertheless demonstrated that it is possible to improve NAD[+] regeneration under anaerobic conditions. With further optimization, this novel strategy might pave the way for providing new opportunities to enhance 3-HP production from glycerol. It might similarly enhance other bioconversions where there is a need to control the intracellular redox state of a cell factory (Kim et al., 2017; de Fouchécour et al., 2018).

In a fermentation operation, the process parameters such as aeration rate, medium composition, etc. play an important role in determining the final outcome. While there are ample reports on strain development, generally less focus has been put on the optimization of the production conditions. Conversion of glycerol to 3-HP being an oxidative reaction, the aeration rate does affect the final yield of the product (de Fouchécour et al., 2018). In one study, aerobic, micro-aerobic and anaerobic conditions were tested in batch fermentations performed to convert glycerol to 3-HP using recombinant *K. pneumoniae*. They reported that amongst the three conditions

tested, the micro-aerobic condition yielded the highest 3-HP titer (2.2 g/L; Zhu et al., 2009). In a more recent study, the transcript levels of the genes encoding the enzymes of the CoA-independent pathway of glycerol conversion to 3-HP and the genes coding for the enzymes of the formate hydrogen lyase pathway were analyzed to understand the intracellular response of *K. pneumoniae* under aerobic, micro-aerobic and anaerobic growth conditions (Huang et al., 2016). The transcription of the glycerol dehydratase operon was downregulated in the presence of oxygen, while aldehyde dehydrogenase-, hydrogenase- and formate dehydrogenase-coding genes were found to be upregulated. The authors suggested that in the presence of oxygen, the formate hydrogen lyase pathway consumed the excess NADH generated due to the over-expression of *aldH* (Huang et al., 2016). In another study aimed at enhancing the coproduction of 3-HP and 1,3-PDO in *K. pneumoniae*, the production of by-products such as lactate, ethanol, succinate, and acetate was reduced by disrupting their synthesis (Ko et al., 2017). However, the disruption of *pta-ackA* gene of the acetate pathway led to a reduction in cell growth, glycerol uptake, and 3-HP and 1,3-PDO production. The authors bypassed the disruption of this gene for reducing acetate production by testing various agitation speeds between 200 and 600 rpm. They reported an increase in 3-HP yield from 0.18 to 0.38 $mol_{3-HP}/mol_{glycerol}$, respectively, which could be attributed to an enhanced cell growth rate because of better oxygen transfer at 600 rpm. Although acetate production was not completely abolished, they did manage to reduce it from 163 mM (at 400 rpm) to only 80 mM (at 600 rpm; Ko et al., 2017). Recently, a production of 61.9 g/L 3-HP with a yield of 0.58 $mol_{3-HP}/mol_{glycerol}$ in 38 h in a 5 L bioreactor operated in fed-batch mode by using an engineered *K. pneumoniae* strain was reported (Jiang et al., 2018). The aeration rate was reduced to half of the initial rate once the cell biomass OD reached close to the maximum value. Furthermore, using the same strategy, they were able to scale-up the process in a 300 L bioreactor and reported a titer of 54.5 g/L for 3-HP with a yield of 0.58 $mol_{3-HP}/mol_{glycerol}$ in 51 h.

The growth medium used for cultivating a microorganism plays an important role in influencing the production process. The current record holder for the highest 3-HP titer, addressed this issue by testing different media for the production of 3-HP in *K. pneumoniae* in shake flasks (Li et al., 2016). A gradient concentration of each component of the media (except CaCl₂) was individually analyzed. The ameliorated medium enabled an increase of 80.5% in the production of 3-HP. They also tested the effect of pH on 3-HP production and found pH 7.0 to be the most appropriate for 3-HP production (Li et al., 2016). To optimize the growth medium for 3-HP production using *L. reuteri*, 30 different media with variable amounts of sugar beet and wheat processing coproducts (used as the carbon source), yeast extract, tween 80 and vitamin B₁₂ were tested. The authors reported an increase in 3-HP production yield by 70 %, accompanied with a decreased 3-HPA titer (Couvreur et al., 2017).

Another way to solve the earlier-mentioned challenge of conflicting requirements for coenzyme B₁₂ generation and NAD⁺ regeneration, is to divide the conversion of glycerol to 3-HP into 2 steps. First, glycerol is converted into either 3-HPA or a mixture of 3-HP and 1,3-PDO. The second step involves the conversion of 3-HPA and 1,3-PDO to 3-HP. This two-step strategy was employed in a study where wild-type *L. reuteri* resting cells first converted glycerol to 1,3-PDO and 3-HP in equimolar amounts. In the subsequent step, the supernatant was fed to the resting cells of *Gluconobacter oxydans* which oxidized 1,3-PDO to 3-HP, yielding a 3-HP titer of 23.6 g/L (Dishisha et al., 2015). In a similar study, Zhao et al. employed *K. pneumoniae* in the first step to convert glycerol into 1,3-PDO. In the second step, the resting cells of *G. oxydans* were introduced in the same bioreactor after heat inactivating *K. pneumoniae*. This led to the conversion of 1,3-PDO to 3-HP with a yield of 0.52 mol/mol on glycerol and 0.94 mol/mol on 1,3-PDO, and a 3-HP titer of 60.5 g/L (Zhao et al., 2015). The advantage of using two bioconversion steps is that each uncoupled step can be optimized separately. As attractive as it may appear because of high yields, the industrial implementation of this two-step strategy may not be as cost-effective and economical as a single step fermentation because of higher associated costs. Nevertheless, this strategy could well be a first step for achieving an integrated continuous process for the production of 3-HP, which in addition would be highly desirable because of its low operating costs (de Fouchécour et al., 2018).

CONCLUSIONS AND PERSPECTIVES

As outlined in this review considering the case of 3-HP production in bacterial cell factories, there are many elements to the successful design of a suitable production strain, ranging from the choice of organism, pathway, and enzymes to engineering of the cell for improved flux through the production pathway and the ability to tolerate the stress imposed by intermediate, by-, and final products (**Figure 4**). Most of the studies described in this review have targeted some of these elements. The highest titer recorded so far is 83.8 g/L which was achieved in *K. pneumoniae* by optimizing the expression of aldehyde dehydrogenase in the production pathway and increasing flux through the production pathway by blocking synthesis of the by-products lactic acid and acetic acid (Li et al., 2016). This impressive feat was based on rational design and accomplished by combining some of the successful approaches outlined above in very few engineering steps.

For bacterial 3-HP production to reach a level of economic viability, further improvements in titer, and productivity will be necessary. By now, many different enzyme variants have been used for constructing 3-HP production pathways but in many cases the enzymes have not been characterized biochemically. A better understanding of the properties of the enzymes with respect to catalytic properties, pH range, and stability could prove instrumental in identifying the optimal enzymes and would also help to identify enzyme qualities to improve by enzyme engineering.

FIGURE 4 | Representation of targets in the bacterial cell for engineering a 3-hydroxypropanoic acid producing strain. Reactions and proteins in green represents targets that are introduced or over-expressed to confer/improve 3-HP production. Unwanted reactions and the corresponding enzymes are shown in red. Inactivation of these targets have been shown to improve 3-HP production and/or reduce/eliminate unwanted by-product formation.

The impressive body of research on strain engineering has identified numerous challenges and solutions such as balancing enzyme activities, preventing by-product formation and alleviating stresses. With this accumulating knowledge base pertaining to 3-HP cell factory design, it seems likely that it will be possible to further tweak production by combining all of the successful approaches described herein. To facilitate this work, powerful engineering techniques are continually emerging that allow the design of large combinatorial libraries of targeted mutations. Notably, the recent development of a 3-HP biosensor can be used in combination with adaptive laboratory evolution for identifying as-of-yet unknown mechanisms for further fine-tuning of the production strains. While it seems reasonable that further improvements in titer will be achieved in the coming years, there should be an increased focus to also develop industrially relevant process designs and subsequent upscaling in order to provide proof of concept.

In conclusion, the massive developments in synthetic biology and metabolic engineering could be expected to soon deliver the long-sought goal of converting the abundant by-product glycerol into the much desired value-added chemical 3-HP.

AUTHOR CONTRIBUTIONS

CJ, AG, and AK performed literature search and drafted the manuscript. All authors revised the manuscript and approved the final version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2019.00124/full#supplementary-material

REFERENCES

Andreeßen, B., Lange, A. B., Robenek, H., and Steinbüchel, A. (2010). Conversion of glycerol to poly(3-hydroxypropionate) in recombinant *Escherichia coli*. *Appl. Environ. Microbiol.* 76, 622–626. doi: 10.1128/AEM.02097-09

Ashok, S., Sankaranarayanan, M., Ko, Y., Jae, K. E., Ainala, S. K., Kumar, V., et al. (2013). Production of 3-hydroxypropionic acid from glycerol by recombinant *Klebsiella pneumoniae* Δ*dhaT*Δ*yqhD* which can produce vitamin B12 naturally. *Biotechnol. Bioeng.* 110, 511–524. doi: 10.1002/bit.24726

Basan, M., Hui, S., Okano, H., Zhang, Z., Shen, Y., Williamson, J. R., et al. (2015). Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. *Nature.* 528, 99–104. doi: 10.1038/nature15765

Biedendieck, R., Malten, M., Barg, H., Bunk, B., Martens, J. H., Deery, E., et al. (2010). Metabolic engineering of cobalamin (vitamin B12) production in *Bacillus megaterium*. *Microb. Biotechnol.* 3, 24–37. doi: 10.1111/j.1751-7915.2009.00125.x

Bonde, M. T., Pedersen, M., Klausen, M. S., Jensen, S. I., Wulff, T., Harrison, S., et al. (2016). Predictable tuning of protein expression in bacteria. *Nat. Methods.* 13, 233–236. doi: 10.1038/nmeth.3727

Burgé G., Saulou-Bérion, C., Moussa, M., Pollet, B., Flourat, A., Allais, F., et al. (2015). Diversity of *Lactobacillus reuteri* strains in converting glycerol into 3-Hydroxypropionic acid. *Appl. Biochem. Biotechnol.* 177, 923–939. doi: 10.1007/s12010-015-1787-8

Chen, Z., Huang, J., Wu, Y., Wu, W., Zhang, Y., and Liu, D. (2017). Metabolic engineering of *Corynebacterium glutamicum* for the production of 3-hydroxypropionic acid from glucose and xylose. *Metab. Eng.* 39, 151–158. doi: 10.1016/j.ymben.2016.11.009

Cheng, K. K., Liu, H. J., and Liu, D. H. (2005). Multiple growth inhibition of *Klebsiella pneumoniae* in 1,3-propanediol fermentation. *Biotechnol. Lett.* 27, 19–22 doi: 10.1007/s10529-004-6308-8

Chu, H. S., Kim, Y. S., Lee, C. M., Lee, J. H., Jung, W. S., Ahn, J. H., et al. (2015). Metabolic engineering of 3-hydroxypropionic acid biosynthesis in *Escherichia coli*. *Biotechnol. Bioeng.* 112, 356–364. doi: 10.1002/bit.25444

Cleusix, V., Lacroix, C., Vollenweider, S., Duboux, M., and Le Blay, G. (2007). Inhibitory activity spectrum of reuterin produced by *Lactobacillus reuteri* against intestinal bacteria. *BMC Microbiol.* 7:101. doi: 10.1186/1471-2180-7-101

Couvreur, J., Teixeira, A., Allais, F., Spinnler, H. E., Saulou-Bérion, C., and Clément, T. (2017). Wheat and sugar beet coproducts for the bioproduction of 3-Hydroxypropionic Acid by *Lactobacillus reuteri* DSM17938. *Fermentation* 3:32. doi: 10.3390/fermentation3030032

Dabrowski, S., Pietrewicz-Kubicz, D., Zabłotna, E., and Długołecka, A. (2012). 1,3-propanediol production by *Escherichia coli* expressing genes of dha operon from *Clostridium butyricum* 2CR371.5. *Acta Biochim. Pol.* 59, 357–361 doi: 10.18388/abp.2012_2121

Daniel, R., Bobik, T. A., and Gottschalk, G. (1998). Biochemistry of coenzyme B12-dependent glycerol and diol dehydratases and organization of the encoding genes. *FEMS Microbiol. Rev.* 22, 553–566 doi: 10.1111/j.1574-6976.1998.tb00387.x

de Fouchécour, F., Sánchez-Castañeda, A. K., Saulou-Bérion, C., and Spinnler, H. E. (2018). Process engineering for microbial production of 3-hydroxypropionic acid. *Biotechnol. Adv.* 36, 1207–1222 doi: 10.1016/j.biotechadv.2018.03.020

Dishisha, T., Pereyra, L. P., Pyo, S. H., Britton, R. A., and Hatti-Kaul, R. (2014). Flux analysis of the *Lactobacillus reuteri* propanediol-utilization pathway for production of 3-hydroxypropionaldehyde, 3-hydroxypropionic acid and 1,3-propanediol from glycerol. *Microb. Cell Fact.* 13:76. doi: 10.1186/1475-2859-13-76

Dishisha, T., Pyo, S. H., and Hatti-Kaul, R. (2015). Bio-based 3-hydroxypropionic- and acrylic acid production from biodiesel glycerol via integrated microbial and chemical catalysis. *Microb. Cell Fact.* 14:200. doi: 10.1186/s12934-015-0388-0

Durnin, G., Clomburg, J., Yeates, Z., Alvarez, P. J., Zygourakis, K., Campbell, P., et al. (2009). Understanding and harnessing the microaerobic metabolism of glycerol in *Escherichia coli*. *Biotechnol. Bioeng.* 103, 148–161. doi: 10.1002/bit.22246

Espah Borujeni, A., Channarasappa, A. S., and Salis, H. M. (2014). Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res.* 42, 2646–2659. doi: 10.1093/nar/gkt1139

Ferro, R., Rennig, M., Hernández-Rollán, C., Daley, D. O., and Nørholm, M. H. H. (2018). A synbio approach for selection of highly expressed gene variants in Gram-positive bacteria. *Microb. Cell Fact.* 17:37. doi: 10.1186/s12934-018-0886-y

Gao, Y., Liu, C., Ding, Y., Sun, C., Zhang, R., Xian, M., et al. (2014). Development of genetically stable *Escherichia coli* strains for poly(3-hydroxypropionate) production. *PLoS ONE.* 9:e97845. doi: 10.1371/journal.pone.0097845

Garai-Ibabe, G., Ibarburu, I., Berregi, I., Claisse, O., Lonvaud-Funel, A., Irastorza, A., et al. (2008). Glycerol metabolism and bitterness producing lactic acid bacteria in cidermaking. *Int. J. Food Microbiol.* 121, 253–261. doi: 10.1016/j.ijfoodmicro.2007.11.004

Hanko, E. K. R., Minton, N. P., and Malys, N. (2017). Characterisation of a 3-hydroxypropionic acid-inducible system from *Pseudomonas putida* for orthogonal gene expression control in *Escherichia coli* and *Cupriavidus necator*. *Sci. Rep.* 7:1724. doi: 10.1038/s41598-017-01850-w

Heinrich, D., Andreeßen, B., Madkour, M. H., Al-Ghamdi, M. A., Shabbaj, I. I., and Steinbüchel, A. (2013). From waste to plastic: synthesis of poly(3-hydroxypropionate) in *Shimwellia blattae*. *Appl. Environ. Microbiol.* 79, 3582–3589. doi: 10.1128/AEM.00161-13

Honjo, H., Tsuruno, K., Tatsuke, T., Sato, M., and Hanai, T. (2015). Dual synthetic pathway for 3-hydroxypropionic acid production in engineered *Escherichia coli*. *J. Biosci. Bioeng.* 120, 199–204. doi: 10.1016/j.jbiosc.2014.12.023

Huang, Y., Li, Z., Shimizu, K., and Ye, Q. (2012). Simultaneous production of 3-hydroxypropionic acid and 1,3-propanediol from glycerol by a recombinant strain of *Klebsiella pneumoniae*. *Bioresour. Technol.* 103, 351–359. doi: 10.1016/j.biortech.2011.10.022

Huang, Y., Li, Z., and Ye, Q. (2016). Transcriptional regulation of genes involved in 3-hydroxypropionic acid production in response to aeration of recombinant *Klebsiella pneumoniae*. *Appl. Biochem. Biotechnol.* 178, 1129–1140 doi: 10.1007/s12010-015-1933-3

Jiang, J., Huang, B., Wu, H., Li, Z., and Ye, Q. (2018). Efficient 3-hydroxypropionic acid production from glycerol by metabolically engineered *Klebsiella pneumoniae*. *Bioresour. Bioprocess.* 5:34. doi: 10.1186/s40643-018-0218-4

Jo, J. E., Mohan Raj, S., Rathnasingh, C., Selvakumar, E., Jung, W. C., and Park, S. (2008). Cloning, expression, and characterization of an aldehyde dehydrogenase from *Escherichia coli* K-12 that utilizes 3-Hydroxypropionaldehyde as a substrate. *Appl. Microbiol. Biotechnol.* 81, 51–60. doi: 10.1007/s00253-008-1608-x

Jung, W. S., Kang, J. H., Chu, H. S., Choi, I. S., and Cho, K. M. (2014). Elevated production of 3-hydroxypropionic acid by metabolic engineering of the glycerol metabolism in *Escherichia coli*. *Metab. Eng.* 23, 116–122. doi: 10.1016/j.ymben.2014.03.001

Kalantari, A., Chen, T., Ji, B., Stancik, I. A., Ravikumar, V., Franjevic, D., et al. (2017). Conversion of glycerol to 3-Hydroxypropanoic acid by genetically engineered *Bacillus subtilis*. *Front. Microbiol.* 8:638. doi: 10.3389/fmicb.2017.00638

Kanno, M., and Atsumi, S. (2017). Engineering an obligate photoautotrophic cyanobacterium to utilize glycerol for growth and chemical production. *ACS Synth. Biol.* 6, 69–75. doi: 10.1021/acssynbio.6b00239

Kim, C., Kim, M. Y., Michie, I., Jeon, B. H., Premier, G. C., Park, S., et al. (2017). Anodic electro-fermentation of 3-hydroxypropionic acid from glycerol by recombinant *Klebsiella pneumoniae* L17 in a bioelectrochemical system. *Biotechnol. Biofuels.* 10:199. doi: 10.1186/s13068-017-0886-x

Kim, K., Kim, S. K., Park, Y. C., and Seo, J. H. (2014). Enhanced production of 3-hydroxypropionic acid from glycerol by modulation of glycerol metabolism in recombinant *Escherichia coli*. *Bioresour. Technol.* 156, 170–175. doi: 10.1016/j.biortech.2014.01.009

Ko, Y., Ashok, S., Ainala, S. K., Sankaranarayanan, M., Chun, A. Y., Jung, G. Y., et al. (2014). Coenzyme B12 can be produced by engineered *Escherichia coli* under both anaerobic and aerobic conditions. *Biotechnol. J.* 9, 1526–1535. doi: 10.1002/biot.201400221

Ko, Y., Ashok, S., Seol, E., Ainala, S. K., and Park, S. (2015). Deletion of putative oxidoreductases from *Klebsiella pneumoniae* J2B could reduce 1,3-propanediol during the production of 3-hydroxypropionic acid from glycerol. *Biotechnol Bioproc E.* 20, 834–843. doi: 10.1007/s12257-015-0166-1

Ko, Y., Ashok, S., Zhou, S., Kumar, V., and Park, S. (2012). Aldehyde dehydrogenase activity is important to the production of 3-hydroxypropionic

acid from glycerol by recombinant *Klebsiella pneumoniae*. *Process Biochem.* 47, 1135–1143 doi: 10.1016/j.procbio.2012.04.007

Ko, Y., Seol, E., Sundara Sekar, B., Kwon, S., Lee, J., and Park, S. (2017). Metabolic engineering of *Klebsiella pneumoniae* J2B for co-production of 3-hydroxypropionic acid and 1,3-propanediol from glycerol: Reduction of acetate and other by-products. *Bioresour. Technol.* 244, 1096–1103. doi: 10.1016/j.biortech.2017.08.099

Kumar, V., Ashok, S., and Park, S. (2013a). Recent advances in biological production of 3-hydroxypropionic acid. *Biotechnol. Adv.* 31, 945–961. doi: 10.1016/j.biotechadv.2013.02.008

Kumar, V., Sankaranarayanan, M., Durgapal, M., Zhou, S., Ko, Y., Ashok, S., et al. (2013b). Simultaneous production of 3-hydroxypropionic acid and 1,3-propanediol from glycerol using resting cells of the lactate dehydrogenase-deficient recombinant *Klebsiella pneumoniae* overexpressing an aldehyde dehydrogenase. *Bioresour. Technol.* 135, 555–563. doi: 10.1016/j.biortech.2012.11.018

Kurihara, S., Kato, K., Asada, K., Kumagai, H., and Suzuki, H. (2010). A putrescine-inducible pathway comprising PuuE-YneI in which γ-aminobutyrate is degraded into succinate in *Escherichia coli* K-12. *J. Bacteriol.* 192, 4582–4591. doi: 10.1128/JB.00308-10

Kwak, S., Park, Y. C., and Seo, J. H. (2013). Biosynthesis of 3-hydroxypropionic acid from glycerol in recombinant *Escherichia coli* expressing *Lactobacillus brevis dhaB* and *dhaR* gene clusters and *E. coli* K-12 *aldH*. *Bioresour Technol.* 135, 432–439. doi: 10.1016/j.biortech.2012.11.063

Li, Y., Ge, X., and Tian, P. (2013a). Gene arrangements in expression vector affect 3-hydroxypropionic acid production in *Klebsiella pneumoniae*. *Indian J. Microbiol.* 53, 418–424. doi: 10.1007/s12088-013-0390-3

Li, Y., Liu, L., and Tian, P. (2014). NAD⁺-independent aldehyde oxidase catalyzes cofactor balanced 3-hydroxypropionic acid production in *Klebsiella pneumoniae*. *Biotechnol. Lett.* 36, 2215–2221. doi: 10.1007/s10529-014-1590-6

Li, Y., Su, M., Ge, X., and Tian, P. (2013b). Enhanced aldehyde dehydrogenase activity by regenerating NAD⁺ in Klebsiella pneumoniae and implications for the glycerol dissimilation pathways. *Biotechnol Lett.* 35, 1609–1615. doi: 10.1007/s10529-013-1243-1

Li, Y., Wang, X., Ge, X., and Tian, P. (2016). High production of 3-hydroxypropionic acid in *Klebsiella pneumoniae* by systematic optimization of glycerol metabolism. *Sci. Rep.* 6:26932. doi: 10.1038/srep26932

Liao, D. I., Reiss, L., Turner, I., and Dotson, G. (2003). Structure of glycerol dehydratase reactivase: a new type of molecular chaperone. *Structure.* 11, 109–119 doi: 10.1016/S0969-2126(02)00935-8

Lim, H. G., Noh, M. H., Jeong, J. H., Park, S., and Jung, G. Y. (2016). Optimum rebalancing of the 3-Hydroxypropionic acid production pathway from glycerol in *Escherichia coli*. *ACS Synth. Biol.* 5, 1247–1255 doi: 10.1021/acssynbio.5b00303

Lindlbauer, K. A., Marx, H., and Sauer, M. (2017). 3-Hydroxypropionaldehyde production from crude glycerol by *Lactobacillus diolivorans* with enhanced glycerol uptake. *Biotechnol. Biofuels.* 10:295. doi: 10.1186/s13068-017-0982-y

Liu, M., Han, X., Xian, M., Ding, Y., Liu, H., and Zhao, G. (2016). Development of a 3-hydroxypropionate resistant *Escherichia coli* strain. *Bioengineered.* 7, 21–27. doi: 10.1080/21655979.2015.1122143

Liu, R., Liang, L., Choudhury, A., Bassalo, M. C., Garst, A. D., Tarasava, K., et al. (2018). Iterative genome editing of *Escherichia coli* for 3-hydroxypropionic acid production. *Metab. Eng.* 47, 303–313. doi: 10.1016/j.ymben.2018.04.007

Liu, S. (2017). "Cell cultivation" in *Bioprocess Engineering, 2nd Edn,* ed S. Liu, (Amsterdam: Elsevier), 699–782.

Liu, Y., Gallo, A. A., Bajpai, R. K., Chistoserdov, A., Nelson, A. T., Segura, L. N., et al. (2010). The diversity and molecular modelling analysis of B₁₂-dependent and B₁₂-independent glycerol dehydratases. *Int. J. Bioinform. Res. Appl.* 6, 484–507 doi: 10.1504/IJBRA.2010.037988

Liu, Y., Liu, Y., and Wang, M. (2017). Design, optimization and application of small molecule biosensor in metabolic engineering. *Front. Microbiol.* 8:2012. doi: 10.3389/fmicb.2017.02012

Luo, L. H., Kim, C. H., Heo, S. Y., Oh, B. R., Hong, W. K., Kim, S., et al. (2012). Production of 3-hydroxypropionic acid through propionaldehyde dehydrogenase PduP mediated biosynthetic pathway in *Klebsiella pneumoniae*. *Bioresour. Technol.* 103, 1–6. doi: 10.1016/j.biortech.2011.09.099

Luo, L. H., Seo, J. W., Baek, J. O., Oh, B. R., Heo, S. Y., Hong, W. K., et al. (2011). Identification and characterization of the propanediol utilization protein PduP

of *Lactobacillus reuteri* for 3-hydroxypropionic acid production from glycerol. *Appl. Microbiol. Biotechnol.* 89, 697–703. doi: 10.1007/s00253-010-2887-6

Luo, L. H., Seo, J. W., Heo, S. Y., Oh, B. R., Kim, D. H., and Kim, C. H. (2013). Identification and characterization of *Klebsiella pneumoniae* aldehyde dehydrogenases increasing production of 3-hydroxypropionic acid from glycerol. *Bioprocess Biosyst. Eng.* 36, 1319–1326. doi: 10.1007/s00449-012-0880-4

Mori, K., and Toraya, T. (1999). Mechanism of reactivation of coenzyme B₁₂-dependent diol dehydratase by a molecular chaperone-like reactivating factor. *Biochemistry.* 38, 13170–13178 doi: 10.1021/bi9911738

Nguyen-Vo, T. P., Ainala, S. K., Kim, J. R., and Park, S. (2018). Analysis, characterization of coenzyme B₁₂ biosynthetic gene clusters and improvement of B₁₂ biosynthesis in *Pseudomonas denitrificans* ATCC 13867. *FEMS Microbiol. Lett.* 1:365. doi: 10.1093/femsle/fny211

Park, Y. S., Choi, U. J., Nam, N. H., Choi, S. J., Nasir, A., Lee, S. G., et al. (2017). Engineering an aldehyde dehydrogenase toward its substrates, 3-hydroxypropanal and NAD⁺, for enhancing the production of 3-hydroxypropionic acid. *Sci. Rep.* 7:17155. doi: 10.1038/s41598-017-15400-x

Qi, X., Sun, L., Luo, Z., Wu, J., Meng, X., Tang, Y., et al. (2006). Rational design of glycerol dehydratase: Swapping the genes encoding the subunits of glycerol dehydratase to improve enzymatic properties. *Chinese Sci. Bull.* 51:2977. doi: 10.1007/s11434-006-2219-5

Raj, S. M., Rathnasingh, C., Jo, J. E., and Park, S. (2008). Production of 3-hydroxypropionic acid from glycerol by a novel recombinant *Escherichia coli* BL21 strain. *Process Biochem.* 43, 1440–1446 doi: 10.1016/j.procbio.2008.04.027

Raj, S. M., Rathnasingh, C., Jung, W. C., Selvakumar, E., and Park, S. (2010). A Novel NAD⁺-dependent aldehyde dehydrogenase encoded by the *puuC* Gene of *Klebsiella pneumoniae* DSM 2026 that Utilizes 3-Hydroxypropionaldehyde as a Substrate. *Biotechnol. Bioproc. E.* 15, 131–138 doi: 10.1007/s12257-010-0030-2

Rathnasingh, C., Raj, S. M., Jo, J. E., and Park, S. (2009). Development and evaluation of efficient recombinant *Escherichia coli* strains for the production of 3-hydroxypropionic acid from glycerol. *Biotechnol. Bioeng.* 104, 729–739. doi: 10.1002/bit.22429.

Raynaud, C., Sarçabal, P., Meynial-Salles, I., Croux, C., and Soucaille, P. (2003). Molecular characterization of the 1,3-propanediol (1,3-PD) operon of *Clostridium butyricum*. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5010–5015 doi: 10.1073/pnas.0734105100

Rennig, M., Martinez, V., Mirzadeh, K., Dunas, F., Röjsäter, B., Daley, D. O., et al. (2018). TARSyn: Tunable antibiotic resistance devices enabling bacterial synthetic evolution and protein production. *ACS Synth. Biol.* 7, 432–442. doi: 10.1021/acssynbio.7b00200

Sabet-Azad, R., Linares-Pastén, J. A., Torkelson, L., Sardari, R. R., and Hatti-Kaul, R. (2013). Coenzyme A-acylating propionaldehyde dehydrogenase (PduP) from *Lactobacillus reuteri*: kinetic characterization and molecular modeling. *Enzyme Microb. Technol.* 53, 235–242. doi: 10.1016/j.enzmictec.2013.05.007

Sankaranarayanan, M., Ashok, S., and Park, S. (2014). Production of 3-hydroxypropionic acid from glycerol by acid tolerant *Escherichia coli*. *J. Ind. Microbiol. Biotechnol.* 41, 1039–1050. doi: 10.1007/s10295-014-1451-2

Schaefer, L., Auchtung, T. A., Hermans, K. E., Whitehead, D., Borhan, B., and Britton, R. A. (2010). The antimicrobial compound reuterin (3-hydroxypropionaldehyde) induces oxidative stress via interaction with thiol groups. *Microbiology* 156, 1589–1599. doi: 10.1099/mic.0.035642-0

Seo, S. W., Yang, J. S., Kim, I., Yang, J., Min, B. E., Kim, S., et al. (2013). Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency. *Metab. Eng.* 15, 67–74. doi: 10.1016/j.ymben.2012.10.006

Seok, J. Y., Yang, J., Choi, S. J., Lim, H. G., Choi, U. J., Kim, K. J., et al. (2018). Directed evolution of the 3-hydroxypropionic acid production pathway by engineering aldehyde dehydrogenase using a synthetic selection device. *Metab. Eng.* 47, 113–120. doi: 10.1016/j.ymben.2018.03.009

Shibata, N., Masuda, J., Tobimatsu, T., Toraya, T., Suto, K., Morimoto, Y., et al. (1999). A new mode of B₁₂ binding and the direct participation of a potassium ion in enzyme catalysis: X-ray structure of diol dehydratase. *Structure* 7, 997–1008 doi: 10.1016/S0969-2126(99)80126-9

Silva, F., Queiroz, J. A., and Domingues, F. C. (2012). Evaluating metabolic stress and plasmid stability in plasmid DNA production by *Escherichia coli*. *Biotechnol. Adv.* 30, 691–708. doi: 10.1016/j.biotechadv.2011.12.005

Son, H. F., Park, S., Yoo, T. H., Jung, G. Y., and Kim, K. J. (2017). Structural insights into the production of 3-hydroxypropionic acid by aldehyde dehydrogenase from *Azospirillum brasilense*. *Sci. Rep.* 7:46005. doi: 10.1038/srep46005

Su, M., Li, Y., Ge, X., and Tian, P. (2015). 3-Hydroxypropionaldehyde-specific aldehyde dehydrogenase from *Bacillus subtilis* catalyzes 3-hydroxypropionic acid production in *Klebsiella pneumoniae*. *Biotechnol. Lett.* 37, 717–724. doi: 10.1007/s10529-014-1730-z

Su, M. Y., Li, Y., Ge, X. Z., and Tian, P. F. (2014). Insights into 3-hydroxypropionic acid biosynthesis revealed by overexpressing native glycerol dehydrogenase in *Klebsiella pneumoniae*. *Biotechnol Biotechnol Equip.* 28, 762–768 doi: 10.1080/13102818.2014.944419

Tokuyama, K., Ohno, S., Yoshikawa, K., Hirasawa, T., Tanaka, S., Furusawa, C., et al. (2014). Increased 3-hydroxypropionic acid production from glycerol, by modification of central metabolism in *Escherichia coli*. *Microb. Cell Fact.* 13:64. doi: 10.1186/1475-2859-13-64

Toraya, T. (2003). Radical catalysis in coenzyme B_{12}-dependent isomerization (eliminating) reactions. *Chem. Rev.* 103, 2095–2127 doi: 10.1021/cr020428b

Wang, K., and Tian, P. (2017). Engineering Plasmid-Free *Klebsiella Pneumoniae* for Production of 3-Hydroxypropionic Acid. *Curr. Microbiol.* 74, 55–58. doi: 10.1007/s00284-016-1153-2

Wang, X., Sa, N., Wang, F. H., and Tian, P. F. (2013). Engineered constitutive pathway in *Klebsiella pneumoniae* for 3-hydroxypropionic acid production and implications for decoupling glycerol dissimilation pathways. *Curr. Microbiol.* 66, 293–299. doi: 10.1007/s00284-012-0271-8

Wang, X. D., Zhang, X. E., Guo, Y. C., Zhang, Z. P., Cao, Z. A., and Zhou, Y. F. (2009). Characterization of glycerol dehydratase expressed by fusing its α- and β-subunits. *Biotechnol. Lett.* 31, 711–717. doi: 10.1007/s10529-009-9911-x

Wang, Y., Tao, F., Ni, J., Lia, C., and Xu, P. (2015). Production of C3 platform chemicals from CO_2 by genetically engineered cyanobacteria. *Green Chem.* 17, 3100–3110 doi: 10.1039/C5GC00129C

Warnecke, T. E., Lynch, M. D., Karimpour-Fard, A., Lipscomb, M. L., Handke, P., Mills, T., et al. (2010). Rapid dissection of a complex phenotype through genomic-scale mapping of fitness altering genes. *Metab. Eng.* 12, 241–250. doi: 10.1016/j.ymben.2009.12.002

Warnecke, T. E., Lynch, M. D., Lipscomb, M. L., and Gill, R. T. (2012). Identification of a 21 amino acid peptide conferring 3-hydroxypropionic acid stress-tolerance to *Escherichia coli*. *Biotechnol. Bioeng.* 109, 1347–1352. doi: 10.1002/bit.24398

Wei, X., Meng, X., Chen, Y., Wei, Y., Du, L., and Huang, R. (2014). Cloning, expression, and characterization of coenzyme-B_{12}-dependent diol dehydratase from *Lactobacillus diolivorans*. *Biotechnol. Lett.* 36, 159–165 doi: 10.1007/s10529-013-1346-8

Yamanishi, M., Kinoshita, K., Fukuoka, M., Saito, T., Tanokuchi, A., Ikeda, Y., et al. (2012). Redesign of coenzyme B_{12} dependent diol dehydratase to be resistant to the mechanism-based inactivation by glycerol and act on longer chain 1,2-diols. *FEBS J.* 279, 793–804. doi: 10.1111/j.1742-4658.2012.08470.x

Yamanishi, M., Yunoki, M., Tobimatsu, T., Sato, H., Matsui, J., Dokiya, A., et al. (2002). The crystal structure of coenzyme B_{12}-dependent glycerol dehydratase in complex with cobalamin and propane-1,2-diol. *Eur. J. Biochem.* 269, 4484–4494 doi: 10.1046/j.1432-1033.2002.03151.x

Zhao, L., Lin, J., Wang, H., Xie, J., and Wei, D. (2015). Development of a two-step process for production of 3-hydroxypropionic acid from glycerol using *Klebsiella pneumoniae* and *Gluconobacter oxydans*. *Bioprocess Biosyst. Eng.* 38, 2487–2495 doi: 10.1007/s00449-015-1486-4

Zhou, S., Catherine, C., Rathnasingh, C., Somasundar, A., and Park, S. (2013). Production of 3-hydroxypropionic acid from glycerol by recombinant *Pseudomonas denitrificans*. *Biotechnol. Bioeng.* 110, 3177–3187. doi: 10.1002/bit.24980

Zhu, J. G., Ji, X. J., Huang, H., Du, J., Li, S., and Ding, Y. Y. (2009). Production of 3-hydroxypropionic acid by recombinant *Klebsiella pneumoniae* based on aeration and ORP controlled strategy. *Korean J. Chem. Eng.* 26:1679. doi: 10.1007/s11814-009-0240-5

Build Your Bioprocess on a Solid Strain—β-Carotene Production in Recombinant *Saccharomyces cerevisiae*

OPEN ACCESS

Edited by:
Rodrigo Ledesma-Amaro,
Imperial College London,
United Kingdom

Reviewed by:
Mattheos Koffas,
Rensselaer Polytechnic Institute,
United States
Mingfeng Cao,
University of Illinois at
Urbana-Champaign, United States

*Correspondence:
Eduardo Agosin
agosin@ing.puc.cl

†*These authors have contributed*
equally to this work

‡**Present Address:**
Maximiliano Ibaceta,
Advanced Biotechnology Inc., Totowa,
NJ, United States

Javiera López[1,2†], **Vicente F. Cataldo**[2†], **Manuel Peña**[2], **Pedro A. Saa**[2], **Francisco Saitua**[1], **Maximiliano Ibaceta**[2‡] **and Eduardo Agosin**[1,2*]

[1] *Centro de Aromas and Sabores, DICTUC S.A., Santiago, Chile,* [2] *Department of Chemical and Bioprocess Engineering, School of Engineering, Pontificia Universidad Católica de Chile, Santiago, Chile*

Robust fermentation performance of microbial cell factories is critical for successful scaling of a biotechnological process. From shake flask cultivations to industrial-scale bioreactors, consistent strain behavior is fundamental to achieve the production targets. To assert the importance of this feature, we evaluated the impact of the yeast strain design and construction method on process scalability -from shake flasks to bench-scale fed-batch fermentations- using two recombinant *Saccharomyces cerevisiae* strains capable of producing β-carotene; SM14 and βcar1.2 strains. SM14 strain, obtained previously from adaptive evolution experiments, was capable to accumulate up to 21 mg/g_{DCW} of β-carotene in 72 h shake flask cultures; while the βcar1.2, constructed by overexpression of carotenogenic genes, only accumulated 5.8 mg/g_{DCW} of carotene. Surprisingly, fed-batch cultivation of these strains in 1L bioreactors resulted in opposite performances. βcar1.2 strain reached much higher biomass and β-carotene productivities (1.57 g/L/h and 10.9 mg/L/h, respectively) than SM14 strain (0.48 g/L/h and 3.1 mg/L/h, respectively). Final β-carotene titers were 210 and 750 mg/L after 80 h cultivation for SM14 and βcar1.2 strains, respectively. Our results indicate that these substantial differences in fermentation parameters are mainly a consequence of the exacerbated Crabtree effect of the SM14 strain. We also found that the strategy used to integrate the carotenogenic genes into the chromosomes affected the genetic stability of strains, although the impact was significantly minor. Overall, our results indicate that shake flasks fermentation parameters are poor predictors of the fermentation performance under industrial-like conditions, and that appropriate construction designs and performance tests must be conducted to properly assess the scalability of the strain and the bioprocess.

Keywords: bioprocess, scale up, fermentation, *Saccharomyces cerevisiae*, β-carotene

INTRODUCTION

The improvement of experimental and analytical techniques, together with emerging synthetic biology tools, have enabled rapid and precise genetic manipulation of microorganisms for the industrial production of diverse compounds (Ajikumar et al., 2010; Paddon and Keasling, 2014; Meadows et al., 2016; Lian et al., 2018). These advances have greatly accelerated the construction and evaluation of promising microbial cell factories replacing traditional chemical synthesis processes, however, there are still many obstacles that difficult scaling lab-scale experiments up to economically attractive industrial bioprocesses (Yadav et al., 2012). In addition to low production titers, yields, and productivities, the often unpredictable physiology of microbial cell factories renders the scale-up process an expensive, time-consuming, labor-intensive task (Woolston et al., 2013; Wu et al., 2016). While much attention has been paid to the development of upstream operations (i.e., strain construction), seldom assessed under production conditions (i.e., fed-batch cultivations), the production performance is typically only evaluated in laboratory-scale batch cultures, which may greatly differ from the actual process behavior (Lee and Kim, 2015; Petzold et al., 2015; Gustavsson and Lee, 2016).

A family of interesting compounds with seemingly attractive scalability potential in yeast cell factories are carotenoids, particularly β-carotene (Mata-Gómez et al., 2014; Larroude et al., 2018). This C_{40} isoprenoid is widely used in the food and health industries as feed additive and/or nutraceutical (Mata-Gómez et al., 2014; Niu et al., 2017). A number of studies have reported high heterologous β-carotene production in shake flasks (Yamano et al., 1994; Verwaal et al., 2007; Li et al., 2013, 2017; Zhao et al., 2015) and bench-scale batch fermentations (Reyes et al., 2014; Olson et al., 2016). However, there is scarce data about the scalability of the bioprocess under more realistic production conditions (i.e., fed-batch mode). To the best of our knowledge, there is only one study reporting high β-carotene production in a recombinant yeast strain during fed-batch fermentations (achieving up to 20.8 mg/g_{DCW} total carotenoid content and 9.6 mg/L/h volumetric productivity) (Xie et al., 2015). However, analysis of the overall scalability of the generated strain and proposed bioprocess are still lacking.

In this work, we evaluated the impact of the design and construction method on the production performance and scalability potential of two engineered yeast strains. As a case study, we chose the heterologous production of β-carotene in S. cerevisiae. The evaluated strains were built following two radically different approaches; one transformed and evolved by adaptive evolution using oxidative stress as selective pressure for accumulating high amounts of β-carotene (SM14)—to date the highest carotenoid-accumulating yeast strain in test tubes (18 mg/g_{DCW}) (Reyes et al., 2014) and batch fermenters (25 mg/g_{DCW}) (Olson et al., 2016); and the other, constructed using the industrial CEN.PK2-1c strain as metabolic chassis and transformed employing state-of-the-art molecular biology tools (βcar1.2). In addition to the usual fermentation characterization in shake flasks, genomic stability assays and fed-batch culture experiments in bench-scale bioreactors were performed to assess the suitability of the strains for high and robust β-carotene production. Our results show that conclusions drawn from preliminary characterizations performed under settings different than the actual production conditions can be misleading and that rigorous evaluation of the producer strains should be conducted to properly assess the scalability of both, the strain and the bioprocess.

MATERIALS AND METHODS

Plasmids Construction

Integrative plasmids needed for heterologous β-carotene production in the S. cerevisiae CEN.PK2-1c strains were constructed using Gibson assembly (Gibson, 2009). The carotenogenic enzyme genes (crtE, crtYB, and crtI) from Xanthophyllomyces dendrorhous were amplified by PCR, using genomic DNA from the S. cerevisiae SM14 strain as template. The catalytic domain of the truncated HMG-CoA reductase gene (tHMG1) was amplified by PCR using genomic DNA from the CEN.PK2-1c strain of S. cerevisiae as template. Backbone vectors were amplified by PCR, using the plasmid library developed by Mikkelsen et al. (2012) with primer pair homology to either TEF1 or PGK1 promoters, and to either ADH1 or tCYC1 terminators. All PCR products were gel-extracted to eliminate original vector residues. Purified PCR amplified vectors (100 ng) were mixed in a molar ratio depending on their length, following the manufacturer's instruction. DNA fragments were mixed with home-made Gibson master mix (5X isothermal mix buffer, T5 exonuclease 1 U/μL, Phusion DNA polymerase 2 U/μL, Taq DNA ligase 40 U/μL and Milli-Q purified water) until reaching 10 μL working volume. The mixture was incubated for 60 min at 50°C. Finally, the reaction mix was used to transform chemically competent E. coli TOP10 cells (ThermoFisher, USA). All vectors contained one marker gene (URA3, TRP1 or LEU2) depending on the parental strain auxotrophy.

Strains

Two S. cerevisiae strains, SM14 and βcar1.2, were employed to compare their production performance and evaluate their genetic stability. The SM14 is a β-carotene hyper-producer strain derived from adaptive evolution experiments (Reyes et al., 2014). For the genomic stability evaluation, the SM14 was transformed with a 120-bp PCR product containing a 60-base-pair homology with the flanking regions of the URA3 gene, yielding a yeast strain with uracil auxotrophy (SM14-ΔURA3). In the case of βcar1.2 strain, the CEN.PK2-1c strain was the chassis employed for its incremental construction through βcar1 and βcar1.1 strains. A summary of the strains employed in this study and their genotype is shown in **Table 1**.

Transformations were performed using lithium acetate/single-stranded DNA carrier/PEG procedure (Gietz and Woods, 2002) and SC proper plates for transformants selection. Finally, correct cassette integration into the specific loci was tested by colony PCR, and carotenoid production was evaluated in YPD medium at 30°C after 72 h (see *Carotenoid extraction and analysis*).

TABLE 1 | Strains used in this study.

Strain	Parental strain	Genotype	References
SM14	S288c	MATa, P_{TDH3}-crtYB-T_{CYC1} P_{TDH3}-crtI-T_{CYC1} P_{TDH3}-crtE-T_{CYC1} $\Delta CTT1$ $URA3$	Reyes et al., 2014
SM14-ΔURA3	SM14	$\Delta URA3$	This study
CEN.PK2-1c		MATa, ura3-52, trp1-289, leu2-3, 112 his3Δ, can1Δ::cas9-natNT2	Euroscarf IMX672
βcar1	CEN.PK2-1c	(XI-5) P_{TEF1}-tHMG1-T_{ADH1} P_{PGK1}-crtYB-T_{CYC1} (XI-3) P_{TEF1}-crtI-T_{ADH1} P_{PGK1}-crtE-T_{CYC1} $URA3$	This study
βcar1.1	βcar1	(X-2) P_{TEF1}-tHMG1-T_{ADH1} P_{PGK1}-tHMG1X.d.-T_{CYC1} $LEU2$	This study
βcar1.2	βcar1.1	(XI-1) P_{PGK1}-crtYB-T_{CYC1} P_{TEF1}-crtI-T_{ADH1} $TRP1$	This study

Shake Flask Cultures and Genetic Stability Evaluation

For ethanol, glucose, acetate, biomass and total carotenoids quantification, a single colony was picked from YPD or CSM (with or without auxotrophy) agar plates, subcultured in tubes overnight at 30°C and 160 rpm in a rotary shaker incubator in 3 mL of YPD medium. On the next day, the optical density at 600 nm (OD_{600}) of each culture tube was measured (see *Biomass determination*). The content of the tubes was then transferred to a 250 mL baffled shake flask with 50 mL final culture volume at an initial OD600 of 0.1 in YPD medium. Culture samples were periodically collected for biomass, extracellular metabolites and carotenoids quantification.

Genetic stability was evaluated in 72-h batch cultures, following the same shake-flask cultivation protocol. After 72 h, an aliquot of each culture (previously diluted to 1 mL at OD600 = 10) was diluted 10,000-fold and 100 μL were plated in YPD. Furthermore, the kinetics of carotenogenic gene loss after several generations was determined in exponential phase cultures. For this purpose, the cultures were started at OD600 of 0.1 in YPD until late exponential phase (12 h) and then diluted again to an OD600 of 0.1 (twice per day). Samples at different cultivation times were diluted and plated in YPD. After 4 days at 30°C, orange and white colonies were counted from all plates. The number of generations was calculated according to the duplication time of the cultures.

Fermentation Conditions and Culture Media

Fed-batch cultures were performed in 1-L in-house bioreactors equipped with a condenser, a stirrer and two Rushton turbines operated with brushless DC motors (Oriental Motor, Japan). A SIMATIC PCS7 control system (Siemens, Germany) was used to monitor and control the cultivations at 30°C, pH = 5.0

and dissolved oxygen above 2.8 mg/L. These culture conditions were employed throughout the entire study for all fermentations. Aerobiosis was maintained with a modified split-range control scheme varying the agitation, air and pure oxygen gas flows (Cárcamo et al., 2014). Briefly, as the oxygen demand increases, the control scheme first increases the agitation from 200 to 500 rpm, then the air flow from 0.3 to 1 L/min, and finally, if needed, pure oxygen gas flow from 0.05 to 1 L/min with the concomitant decrease in air flow, thereby maintaining the total gas inflow constant.

The batch medium used in the bioreactor cultivations of both strains contained (per liter): 20 g glucose, 15 g ammonium sulfate, 500 mg leucine, 160 mg histidine, 100 mg tryptophan, 4 g KH_2PO_4, 1.2 g $MgSO_4 \cdot 7H_2O$, and 150 mg NaCl. The medium was supplemented with 15 mL/L of a vitamin solution, 3 mL/L of a trace solution, 0.75 mL/L of a $CaCl_2 \cdot 2H_2O$ solution at 40 g/L, and 0.75 mL/L of a $FeSO_4 \cdot 7H_2O$ solution at 4.2 g/L. The trace solution is composed of 3.3 g/L of zinc sulfate heptahydrate, 2 g/L of cobalt chloride hexahydrate, 3.3 g/L of manganese sulfate monohydrate, 4.67 g/L of copper sulfate pentahydrate, 2 g/L of boric acid, 0.2 g/L of potassium iodide, 0.46 g/L of molybdic acid sodium salt dihydrate and 8 g/L of EDTA. In addition, the vitamin solution consists of 0.05 g/L of D-Biotin, 5 g/L of calcium pantothenate, 3.75 g/L of nicotinic acid, 40 g/L of myo-inositol, 1 g/L of thiamine- HCl, 2.5 g/L of Pyridoxine-HCl, 0.02 g/L of p-aminobenzoic acid, 1 g/L of riboflavin and 0.02 g/L of folic acid. Finally, in the case of the fed-batch cultures, the fed-batch feeding contained (per liter): 450 g glucose, 15 g KH_2PO_4, 5.5 g $MgSO_4 \cdot 7H_2O$, and 15 mL/L casamino acids solution at 75 g/L. The fed-batch feeding was supplemented with 15 mL/L and 9 mL/L of the previous vitamin and trace solutions, respectively, 1.35 mL/L of $CaCl_2 \cdot 2H_2O$ solution at 400 g/L, and 1.35 mL/L of a $FeSO_4 \cdot 7H_2O$ solution at 84 g /L. Both vitamin and trace solutions were filter-sterilized before their use in all the mentioned media.

Batch Cultures

A single colony from a working plate was cultured in 3 mL YPD medium in a pre-inoculum tube at 30°C for 10–14 h (overnight). On the next morning, 1 mL was cultured in a shake flask with 20 mL YPD medium for 8 h. Batch cultures in bioreactors were inoculated to a final concentration of OD600 of 0.1. Fermentation conditions were the same as the previous section. Culture samples were periodically collected every 2–3 h for biomass, extracellular metabolites and carotenoids quantification. Batch fermentations were stopped when all major carbon sources were exhausted (i.e., glucose, ethanol and acetic acid).

Fed-Batch Cultures

Fed-batch cultures were fed following an exponential feeding (Equation 1, 2) with an exponentially decreasing specific growth rate set (μ_{set}) (Equation 3). Both *S. cerevisiae* strains (SM14 and βcar1.2) followed the same feeding profile, albeit with different initial parameters, depending on the fermentation profile observed in the shake flask cultivations. As both strains behaved differently in these cultures (refer to ethanol and β-carotene production, **Figures 1C,D**), and in order to provide a fair evaluation, we employed slightly different feeding parameters

so that the strains could be more easily compared. For both strains, the feeding started after both glucose and ethanol were depleted using an initial fixed specific growth rate. The exponentially decreasing specific growth rate feeding strategy was started once the biomass concentration reached approx. an OD_{600} of 60. The general feeding profile is described by the conventional exponential formula (Villadsen and Patil, 2007),

$$F(t) = F_{in} \cdot \exp\{\mu_{set}(t) \cdot t\} \tag{1}$$

$$F_{in} = \mu_{set}(XV)_{in}/(Y_{sx} \cdot S_{in}) \tag{2}$$

where F_0 describes the initial feeding rate. The parameters of the latter formula correspond to the total amount of biomass in the reactor at the beginning of the feeding $(XV)_{in}$, the biomass yield on glucose (Y_{sx}), and the glucose concentration in the feed (Sin). In the case of $\mu_{set}(t)$, this parameter is strain-dependent and was defined by Equation 3 with values between 0.1–0.13 h^{-1} for μ_{init} and 0.03 h^{-1} for μ_{end}.

$$\mu_{set}(t) = (\mu_{init} - \mu_{end}) \cdot \exp\{-\mu_{end}(t) \cdot t\} + \mu_{end} \tag{3}$$

Finally, culture samples were periodically collected for biomass, extracellular metabolites and carotenoids quantification.

Biomass Determination

Biomass concentration was determined by optical density (OD600) using an UV-160 UV-visible spectrophotometer (Shimadzu, Japan). Biomass concentration was estimated using the linear relationship: 1 OD_{600} = 0.4 g/L determined experimentally.

Extracellular Metabolite Quantification

Culture samples were centrifuged at 10,000 rpm for 3 min and the supernatant stored at −80°C for metabolite analysis. Extracellular glucose, ethanol and acetic acid concentrations were quantified in duplicate by High-Performance Liquid Chromatography (HPLC) as detailed in Sánchez et al. (2014).

Carotenoid Extraction and Analysis

For each sample, 8 mg of biomass were pelleted into 2 mL Eppendorf tubes and the supernatant was discarded. Four hundred microliter of acid-washed glass beads (Sigma Aldrich, USA) and 1 mL of hexane were then added for cell disruption and carotenoid extraction from the cell membranes. Cells were disrupted at room temperature in a BeadBug6 cell homogenizer (Benchmark Scientific, USA) using a program consisting in 4 cycles of 90 s of disruption at 3,700 rpm, followed by a 10-s rest. Cell lysate was then centrifuged and the supernatant (hexane) was stored at −80°C until further analysis. Carotenoid quantification was performed by measuring the absorbance at 453 nm of the hexane extracts, and then converted into concentrations using a standard curve of β-carotene ranging from 0.5 to 10 mg/L.

RESULTS

Construction of βCar Yeast Strains

A series of β-carotene-producing yeast strains (βcar) were constructed using a CEN.PK strain as host cell. Integration of carotenogenic genes in stable constructs was achieved using two different promoters (in a bidirectional arrangement) and two different terminators. The integration of the three carotenogenic genes (*crtE*, *crtYB* and *crtI*), together with the tHMG1 gene, resulted in a strain (βcar1) that generated faint orange colonies. This strain was capable of accumulating 0.034 mg/g_{DCW} β-carotene in shake flasks after 72 h of incubation. This initial βcar strain was further optimized by integrating two extra copies of the tHMG1 gene (βcar1.1 strain) and, then, by adding an extra copy of the *crtYB* and *crtI* genes (final βcar1.2 strain) (**Figures 1A,B**). The βcar1.1 and βcar1.2 strains accumulated, respectively, 0.3 mg/g_{DCW} and 5.8 mg/g_{DCW} β-carotene in shake flasks after 72 h of incubation (**Figure 1A**).

Strain Performance in Shake Flask Cultures

Shake flask batch cultures of the βcar1.2 and SM14 displayed similar fermentation profiles during the first 24 h incubation (**Figures 1C,D**). Both, glucose consumption and ethanol production behaved similarly during the first 12 h, and there were no substantial differences in β-carotene levels before 24 h. In addition, the maximum specific growth rate reached similar values during this time (0.43 h^{-1} for SM14 and 0.40 h^{-1} for βcar1.2), consistent with the profiles of the main fermentation substrates and products. However, this trend was not maintained throughout the cultivation. While βcar1.2 showed an almost constant β-carotene concentration after the first 24 h of cultivation (∼5 mg/g_{DCW}), SM14 quadrupled this value, rising from 7.6 mg/g_{DCW} at 24 h, to 21 mg/g_{DCW} after 72 h cultivation (**Figures 1C,D**). Interestingly, the SM14 strain displayed an important increase in the specific β-carotene production rate upon reaching the stationary phase, which was not replicated by the βcar1.2 strain. Altogether, the final β-carotene titer of the SM14 strain reached 159.6 mg/L by the end of the shake flask fermentation, roughly four times higher than the βcar1.2 titers.

Strain Performance in Fed-Batch Cultures

SM14 strain was grown in fed-batch mode to evaluate its performance under production conditions (**Figure 2A**). Once the glucose from the batch phase was depleted (after approx. 26 h of cultivation), the feeding was started according to Equation 1 with a constant specific growth rate set of 0.1 h^{-1}. After 65 h of cultivation and upon reaching 60 OD_{600}, the exponentially decreasing specific growth rate set protocol was initiated using an initial μ_{init} of 0.1 h^{-1} and a final μ_{end} of 0.03 h^{-1} to be reached within the next 24 h. This change in the feeding policy was performed to enable reaching higher cell densities using a more conservative feeding strategy.

After 68 h cultivation, biomass growth stopped, which was consistent with the glucose (2.4 g/L) accumulation (**Figure 2A**). At this point, biomass concentration reached 32.3 g/L (72.6 OD_{600}, biomass volumetric productivity q_X = 0.475 g_{DCW}/L/h)
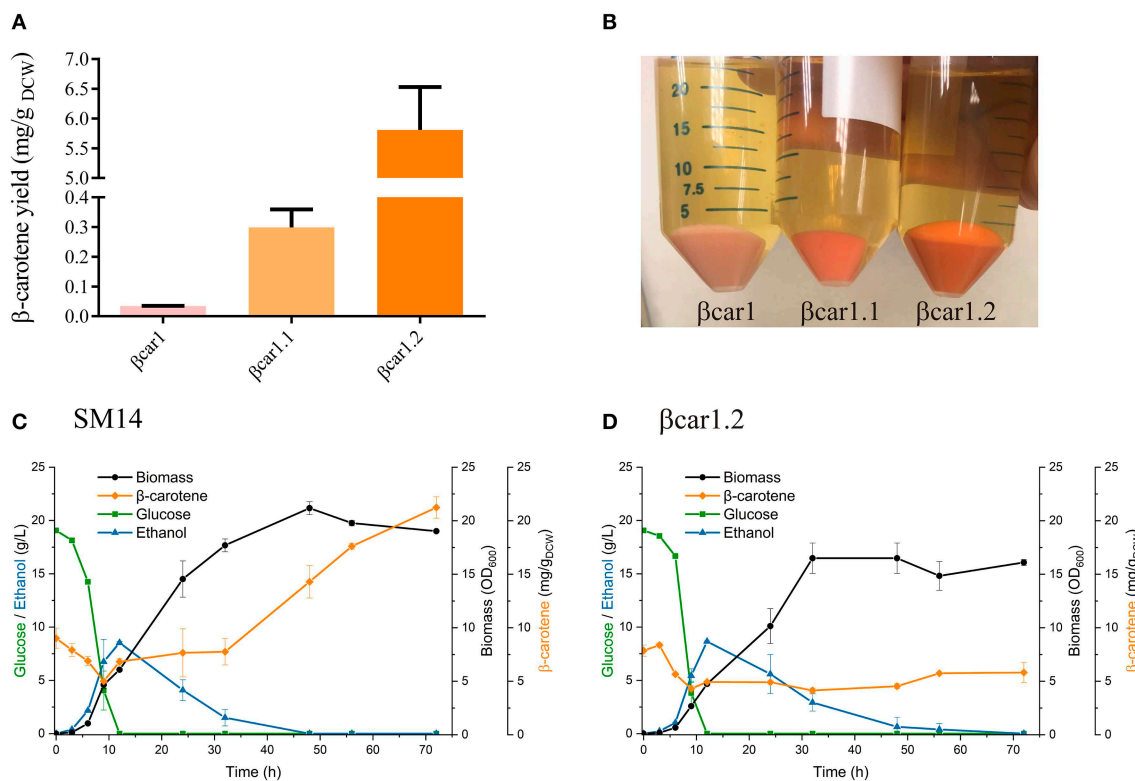
FIGURE 1 | Shake flask performance of *Saccharomyces cerevisiae* βcar and SM14 strains. **(A)** β-carotene to biomass yield for βcar strains after 72 h shake flask cultures. **(B)** Pellets of βcar strains in shake flask cultures after 72 h. **(C)** SM14 strain performance in shake flask culture after 72 h. **(D)** βcar1.2 strain performance in 72 h shake flask cultures.

and the carotenoid content of the cells was 6.48 mg/g_{DCW}, yielding a total carotenoid titer of 209 mg/L (carotenoid volumetric productivity $q_C = 3.07$ mg/L/h) (**Table 2**).

Both biomass and carotenoid concentrations plateaued thereafter. For instance, after 77 h of cultivation, 30.5 g/L of biomass (76.2 OD_{600}, $q_X = 0.396$ g_{DCW}/L/h) with a carotenoid content of 7.18 mg/g_{DCW} were achieved, yielding a total carotenoid titer of 218 mg/L (carotenoid volumetric productivity $q_C = 2.83$ mg/L/h). Finally, ethanol accumulation started as soon as the constant exponential feed was initiated, reaching concentrations of 12.3 g/L at 68 h cultivation and >18 g/L after 77 h.

Similar to SM14 strain, the engineered βcar1.2 strain, was evaluated in fed-batch fermentations under production conditions (**Figure 2B**). Once the glucose and ethanol from the batch phase were depleted, the feeding was started according to Equation 1 with a constant specific growth rate set of 0.13 h^{-1}. A slightly higher μ_{set} was employed in this case as this strain showed a higher $\mu_{critical}$, as determined in preliminary fermentations. Again, upon reaching 60 OD_{600} (approx. after 33 h cultivation), the feeding policy was changed to the exponentially decreasing policy with an initial μ_{init} of 0.13 h^{-1} and a final μ_{end} of 0.03 h^{-1}.

After 68 h of fed-batch cultivation, the βcar1.2 strain reached a biomass concentration of 107.1 g_{DCW}/L (267.9 OD_{600}, biomass volumetric productivity $q_X = 1.576$ g_{DCW}/L/h) with a carotenoid content of 6.91 mg/g_{DCW}, overall yielding a total carotenoid titer of 739.6 mg/L (carotenoid volumetric productivity $q_C = 10.88$ mg/L/h). Later fermentation results were consistent with this data. After 77 h of fermentation, 103.8 g/L of biomass were produced, yielding a volumetric productivity of 1,34 g_{DCW}/L/h. Importantly, the biomass productivity of the βcar1.2 strain significantly outperformed (∼3.5-fold higher) the previous results for the SM14 strain (**Table 2**). Likewise, the carotene productivity of the βcar1.2 strain was 3.3-fold higher than that reached by the SM14 strain. After 77 h cultivation, a total carotenoid titer of 729 mg/L (7 mg/g_{DCW} carotene yield) was achieved with this strain (**Table 2**).

Strain Stability Analysis

In order to determine if genetic instability had a significant impact on β-carotene productivity, we evaluated the rate of generation of white cells in the yeast population, which is indicative of carotenogenic genes loss. The SM14 strain showed a 3.9% of white colonies when 72-h shake flask cultures were plated, indicating an intrinsic genetic instability of the carotenogenic construct (**Figure 3C**). Since this strain contained repeated URA3 loci (URA3 and ura3-52) - flanking the carotenogenic genes -, and the same TDH3 promoters and CYC1 terminators for the latter genes, homologous recombination

FIGURE 2 | Fed-batch fermentation performance of (A) SM14 and (B) βcar1.2 strains. Exponential decreasing feeding started when the culture reached a biomass concentration of $OD_{600} = 60$.

TABLE 2 | Fermentation parameters in fed-batch cultivations at 68 h cultivation for the two *S. cerevisiae* strains of this study.

Parameter	SM14 strain	βcar1.2 strain
Biomass titer (g/L)	32.3	107.1
Biomass productivity (g/L/h)	0.475	1.576
β-carotene titer (mg/L)	209.0	739.6
β-carotene productivity (mg/L/h)	3.07	10.88
β-carotene to biomass yield (mg/g)	6.48	6.91

DISCUSSION

Design and Construction Strategy of the Engineered Yeast Strains Determines Fermentation Performance in Fed-Batch Cultures

Depending on fermentation phase, Reyes et al. (2014) reported different β-carotene to biomass yields for the SM14 strain in batch cultures: 6 mg/g_{DCW} during the glucose consumption phase, and 15 mg/g_{DCW} during the ethanol consumption phase. These results were in line with our findings for the SM14 strain behavior in shake flasks (5-7 mg/g_{DCW} during the glucose consumption phase, and 21 mg/g_{DCW} after the ethanol consumption phase, **Figure 1C**). However, these results were not scalable in fed-batch cultures for this strain (**Figure 2A**). In spite of employing a (conservative) μ decreasing strategy, the SM14 strain was unable to assimilate glucose without producing ethanol, even at low specific growth rates. In fact, ethanol accumulated to such high levels that growth was completely arrested (**Figure 2A**).

Previous DNA microarrays analysis for the SM14 strain showed that several genes involved in mitochondrial respiration and electron transport were downregulated, relative to its parental strain (e.g., SDH1, COX4, QCR9 and SDH3; Reyes et al., 2014). Apparently, β-carotene accumulation was not sufficient to overcome the oxidative stress from the oxygen peroxide shocks used to evolve the SM14 strain; thus, SM14 cells might reduce their mitochondrial oxidative capacity to lower the generation of radical oxygen species (ROS). These transcriptional changes are consistent with the enhanced Crabtree effect observed in our fed-batch fermentations. The latter suggests that impaired mitochondrial respiration is the main cause for the poor fermentation performance of the SM14 strain.

Since the SM14 strain has a poor oxidative capacity, other feeding policies could be explored in order to increase the biomass productivity. Nevertheless, the lack of knowledge of the complete genetic background of SM14 renders this strain unsuitable for its transfer to larger scales. In contrast to SM14, the βcar1.2 strain exhibited a more robust and satisfactory performance in the fed-batch cultivations (**Figure 2B**). Despite accumulating a quarter of the concentration of β-carotene compared to SM14 in shake flask cultures after 72 h (**Figures 1C,D**), the βcar1.2 strain greatly surpassed the production performance of SM14 in bioreactors (**Table 2**).

between direct repeat sequences, e.g., URA3 and ura3-52, had likely occurred (**Figure 3A**). In order to avoid possible recombination of the construct and improve the genetic stability of this strain, the URA3 marker was deleted and a new strain, SM14-ΔURA3, was generated. Targeted deletion of the marker decreased the white colony number in YPD plates from 3.9% to 1.3% (**Figure 3C**). However, the SM14-ΔURA3 still showed carotenogenic gene loss, suggesting an inter-promoter or inter-terminator homologous recombination. Genomic PCR results further supported the loss of all carotenogenic genes in the white phenotype of the SM14 and the loss of two genes (*crtI* and *crtYB*) in the case of the white phenotype of the SM14-ΔURA3 (**Figure 3D**). These results are in concordance with the recombination scheme proposed, i.e., URA3 loci recombination in SM14 and inter-promoter recombination in SM14-ΔURA3, although other recombinations cannot be discarded (e.g., inter-promoter recombination in SM14).

Figure 3E shows a more detailed analysis of the genetic stability of the SM14 strain. We observed an increase in the proportion of white cells as the number of generations progressed, consistent with a rate of appearance of 0,06% white cells/generation. Notably, after 15 generations (close to the duration of a fed-batch cultivation), the proportion of white cells reached ∼1.7 %.
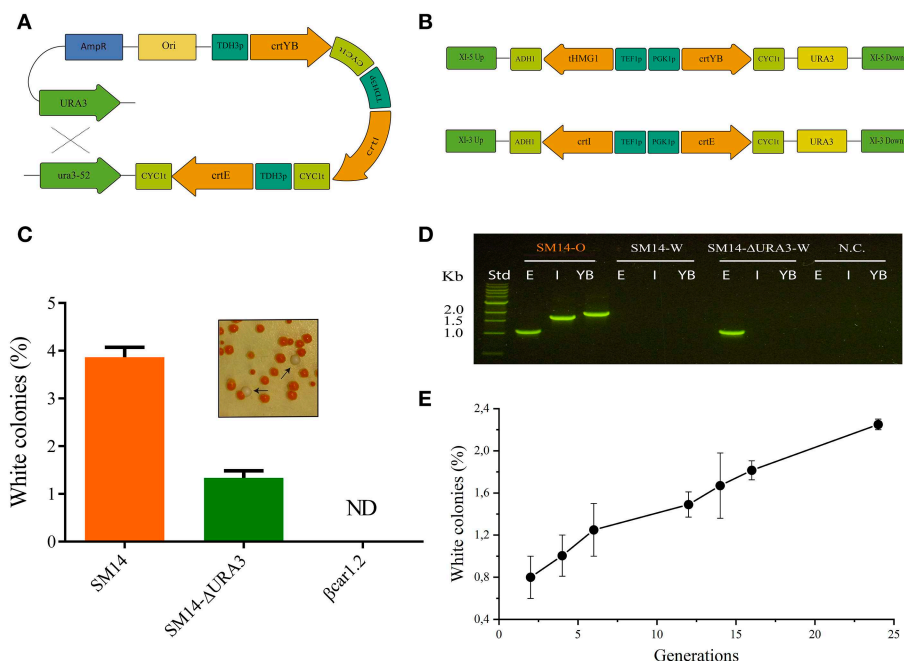
FIGURE 3 | Stability of βcar and SM14 strains. **(A)** General scheme of the possible homologous recombination of the SM14 carotenogenic expressing construct.
(B) Architecture for the expression of the carotenogenic genes in βcar strains. Integration stability is based in two recombination sites (UP and DOWN) and the use of two different promoters and terminators. **(C)** Percentage of white colonies observed after 72 h liquid cultures in strains with the original construct (SM14) and with the deletion of the URA3 gene (SM14-ΔURA3). ND: not detected. **(D)** Detection of carotenogenic genes by agarose gel of genomic PCR products from SM14 orange phenotype (SM14-O), SM14 white phenotype (SM14-W) and SM14-ΔURA3 white phenotype (SM14-ΔURA3-W). E, I, YB refers to *crtE*, *crtI* and *crtYB* respectively. Std: standard DNA ladder. NC: negative control. **(E)** Percentage of white colonies observed during exponential cultures of SM14.

In fed-batch cultures, βcar1.2 exhibited a high oxidative rate, consistent with a fully oxidative metabolism on glucose and the high biomass and β-carotene productivities achieved. These results clearly illustrate that the strain performance must be evaluated in a proper setting, such that initially modest but more robust producers are not discarded early.

Gene Integration Architecture Affects Strain Stability

SM14 strain was constructed using the YIPlac211 YBIE plasmid reported by Verwaal et al. (2007), which has the classic features of an integrating plasmid for yeast, i.e., it possesses a URA3 marker that also serves as recombination site for integration into the ura3-52 locus of the auxotrophic strain. As a consequence of a unique recombination event, the vector containing the carotenogenic construct is integrated and flanked by the URA3 loci (URA3 and ura3-52). Since these genes have sufficient sequence identity, direct repeat recombination and loss of the whole construct may occur. This can even happen in a selective medium (drop-out without uracil), as the recombination event can leave the URA3 allele instead of the ura3-52 (**Figures 3A,B**). Based on the time-course stability analysis up to 24 generations (**Figure 3E**), we note, however, that genetic instability of the SM14 strain does not heavily impact the fed-batch culture performance in the relevant time-scale (ca. 15 generations). This is also supported by the β-carotene production

profile in the fed-batch cultivation that shows a proportional increase of carotene with biomass (**Figure 2A**). However, the genetic instability is still a disadvantage in pre-bioreactor stages considering that it introduces practical difficulties (e.g., selection of pigmented colonies) when handling the strain before bioreactor cultures. Lange and Steinbüchel (2011) reported that episomal expression of the same carotenogenic construct led to an entire plasmid loss due to segregational and structural instabilities, even when grown in selective medium. In this sense, integrating plasmids with one recombination site or repeated sequences are better choices over episomal counterparts. Even so, the stability of classical integrating plasmids like YIp was shown to be unsatisfactory to arrive at a robust strain for bioproduction (**Figure 3**).

To avoid the above stability issues, we built a new β-carotene strain producer, using the plasmid set developed by Mikkelsen et al. (2012). In these plasmids, the target genes are flanked by two different sequences for integration, preventing excision of the construct by homologous recombination. Moreover, the use of two different promoters and two different terminators further decreases the probability of gene loss. Consistent with these features, white colonies were not observed neither in βcar1.2 shake flask cultivations nor in stability studies, even in the absence of a selective pressure (medium with uracil), confirming the high stability of the genomic construct (**Figures 3B,C**).

CONCLUSION

Robust fermentation performance is critical for the development and satisfactory scale up of biotechnological processes. Strain evaluation under realistic production conditions is critical to ensure appropriate process behavior. Here, we have compared the β-carotene production of two yeast strains built following radically different strategies. Our results indicated that the initially most promising evolved strain performed poorly under fed-batch production conditions compared to the conventionally-built strain. These results highlight the impact of the methodology employed for constructing and screening superior strains and its subsequent scale up. Particularly, the adaptive laboratory evolution impacted on the general microbial physiology, while the construct architecture affected the genetic stability of the strain, rendering a poorly scalable producer under production conditions. In contrast, the conventionally-built strain performed robustly, achieving higher biomass and β-carotene productivity in fed-batch cultivations. Overall, this work underscores the importance of carefully choosing the strain construction strategy and its optimization method taking into consideration the end goal. While some strains may perform well in batch cultures at small scales, and could be very useful for gene screening purposes, they may not necessarily perform adequately under production conditions. To this end, evaluation of a satisfactory production performance is a must and by no means can be extrapolated.

REFERENCES

Ajikumar, P. K., Xiao, W. H., Tyo, K. E., Wang, Y., Simeon, F., Leonard, E., et al. (2010). Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli. Science* 330, 70–74. doi: 10.1126/science.1191652

Cárcamo, M., Saa, P. A., Torres, J., Torres, S., Mandujano, P., Correa, J. R. P., et al. (2014). Effective dissolved oxygen control strategy for high-cell-density cultures. *IEEE Lat. Am. Trans.* 12, 389–394. doi: 10.1109/TLA.2014.6827863

Gibson, D. (2009). Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Res.* 37, 6984–6990. doi: 10.1093/nar/gkp687

Gietz, B. R. D., and Woods, R. A. (2002). Transformation of yeast by lithium acetate/single-stranded carrier DNA /polyethylene glycol method. *Methods Enzymol.* 350, 87–96. doi: 10.1016/S0076-6879(02)50957-5

Gustavsson, M., and Lee, S. Y. (2016). Prospects of microbial cell factories developed through systems metabolic engineering. *Microb. Biotechnol.* 9, 610–617. doi: 10.1111/1751-7915.12385

Lange, N., and Steinbüchel, A. (2011). β-Carotene production by *Saccharomyces cerevisiae* with regard to plasmid stability and culture media. *Appl. Microbiol. Biotechnol.* 91, 1611–1622. doi: 10.1007/s00253-011-3315-2

Larroude, M., Celinska, E., Back, A., Thomas, S., Nicaud, J. M., and Ledesma-Amaro, R. (2018). A synthetic biology approach to transform *Yarrowia lipolytica* into a competitive biotechnological producer of β-carotene. *Biotechnol. Bioeng.* 115, 464–472. doi: 10.1002/bit.26473

Lee, S. Y., and Kim, H. U. (2015). Systems strategies for developing industrial microbial strains. *Nat. Biotechnol.* 33, 1061–1072. doi: 10.1038/nbt.3365.

Li, J., Shen, J., Sun, Z., Li, J., Li, C., Li, X., et al. (2017). Discovery of several novel targets that enhance β-carotene production in *Saccharomyces cerevisiae. Front. Microbiol.* 8:1116. doi: 10.3389/fmicb.2017.01116

Li, Q., Sun, Z., Li, J., and Zhang, Y. (2013). Enhancing beta-carotene production in *Saccharomyces cerevisiae* by metabolic engineering. *FEMS Microbiol. Lett.* 345, 94–101. doi: 10.1111/1574-6968.12187

Lian, J., Mishra, S., and Zhao, H. (2018). Recent advances in metabolic engineering of *Saccharomyces cerevisiae*: new tools and their applications. *Metab. Eng.* 50, 85–108. doi: 10.1016/j.ymben.2018.04.011

Mata-Gómez, L. C., Montañez, J. C., Méndez-Zavala, A., and Aguilar, C. N. (2014): Biotechnological production of carotenoids by yeasts: an overview. *Microb. Cell. Fact.* 13, 12. doi: 10.1186/1475-2859-13-12

Meadows, A. L., Hawkins, K. M., Tsegaye, Y., Antipov, E., Kim, Y., Raetz, L., et al. (2016). Rewriting yeast central carbon metabolism for industrial isoprenoid production. *Nature.* 537, 694–697. doi: 10.1038/nature19769

Mikkelsen, M. D., Buron, L. D., Salomonsen, B., Olsen, C. E., Hansen, B. G., Mortensen, U. H., et al. (2012). Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. *Metab. Eng.* 14, 104–111. doi: 10.1016/j.ymben.2012.01.006

Niu, F. X., Lu, Q., Bu, Y. B., and Liu, J. Z. (2017). Metabolic engineering for the microbial production of isoprenoids: carotenoids and isoprenoid-based biofuels. *Synth. Syst. Biotechnol.* 2, 167–175. doi: 10.1016/j.synbio.2017.08.001

Olson, M. L., Johnson, J., Carswell, W. F., Reyes, L. H., Senger, R. S., and Kao, K. C. (2016). Characterization of an evolved carotenoids hyper-producer of *Saccharomyces Cerevisiae* through Bioreactor parameter optimization and raman spectroscopy. *J. Ind. Microbiol. Biotechnol.* 43, 1355–1363. doi: 10.1007/s10295-016-1808-9

Paddon, C. J., and Keasling, J. D. (2014). Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nature Rev. Microbiol.* 12, 355–367. doi: 10.1038/nrmicro3240

Petzold, C. J., Chan, L. J. G., Nhan, M., and Adams, P. D. (2015). Analytics for metabolic engineering. *Front. Bioeng. Biotechnol.* 3, 135. doi: 10.3389/fbioe.2015.00135

Reyes, L. H., Gomez, J. M., and Kao, K. C. (2014). Improving carotenoids production in yeast via adaptive laboratory evolution. *Metab. Eng.* 21, 26–33. doi: 10.1016/j.ymben.2013.11.002

Sánchez, B. J., Pérez-Correa, J. R., and Agosin, E. (2014). Construction of robust dynamic genome-scale metabolic model structures of *Saccharomyces cerevisiae* through iterative re-parameterization. *Metab. Eng.* 25, 159–173. doi: 10.1016/j.ymben.2014.07.004.

Verwaal, R., Wang, J., Meijnen, J. P., Visser, H., Sandmann, G., van den Berg, J. A., et al. (2007). High-level production of beta-carotene in Saccharomyces cerevisiae by successive transformation with carotenogenic genes from Xanthophyllomyces dendrorhous. *Appl. Environ. Microbiol.* 73, 4342–4350. doi: 10.1128/AEM.02759-06

Villadsen, J., and Patil, K. (2007). Optimal fed-batch cultivation when mass transfer becomes limiting. *Biotechnol. Bioeng.* 98, 706–710. doi: 10.1002/bit.21451

Woolston, B. M., Edgar, S., and Stephanopoulos, G. (2013). Metabolic engineering: past and future. *Annu. Rev. Chem. Biomol. Eng.* 4, 259–288. doi: 10.1146/annurev-chembioeng-061312-103312.

Wu, G., Yang, Q., Jones, J. A., Tang, Y. J., Fong, S. S., and Koffas, M. (2016). Metabolic burden: cornerstones in synthetic biology and metabolic engineering applications. *Trends Biotechnol.* 34, 652–664. doi: 10.1016/j.tibtech.2016.02.010

Xie, W., Ye, L., Lv, X., Xu, H., and Yu, H. (2015). Sequential control of biosynthetic pathways for balanced utilization of metabolic intermediates in *Saccharomyces cerevisiae*. *Metab. Eng.* 28, 8–18. doi: 10.1016/j.ymben.2014.11.007

Yadav, V. G., De Mey, M., Lim, C. G., Ajikumar, P. K., and Stephanopoulos, G. (2012). The future of metabolic engineering and synthetic biology: towards a systematic practice. *Metab. Eng.* 14, 233–241. doi: 10.1016/j.ymben.2012.02.001

Yamano, S., Ishii, T., Nakagawa, M., Ikenaga, H., and Misawa, N. (1994). Metabolic engineering for production of beta-carotene and lycopene in *Saccharomyces cerevisiae*. *Biosci. Biotechnol. Biochem.* 58, 1112–1114.

Zhao, X., Shi, F., and Zhan, W. (2015). Overexpression of ZWF1 and POS5 improves carotenoid biosynthesis in recombinant *Saccharomyces cerevisiae*. *Lett. Appl. Microbiol.* 345, 94–101. doi: 10.1111/lam.12463

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership