Recent advances in metabolic engineering of Saccharomyces cerevisiae: New tools and their applications

Jiazhang Lian\textsuperscript{a,b}, Shekhar Mishra\textsuperscript{b}, Huimin Zhao\textsuperscript{b,c,*}

\textsuperscript{a} Key Laboratory of Biomass Chemical Engineering of Ministry of Education, College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310027, China
\textsuperscript{b} Department of Chemical and Biomolecular Engineering, Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, United States
\textsuperscript{c} Key Laboratory of Biomass Chemical Engineering of Ministry of Education, College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310027, China
\textsuperscript{d} Departments of Chemistry, Biochemistry, and Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, United States

\section*{ARTICLE INFO}

\textbf{Keywords:}
Metabolic engineering
Pathway optimization
Genome engineering
Saccharomyces cerevisiae

\section*{ABSTRACT}

Metabolic engineering aims to develop efficient cell factories by rewiring cellular metabolism. As one of the most commonly used cell factories, \textit{Saccharomyces cerevisiae} has been extensively engineered to produce a wide variety of products at high levels from various feedstocks. In this review, we summarize the recent development of metabolic engineering approaches to modulate yeast metabolism with representative examples. Particularly, we highlight new tools for biosynthetic pathway optimization (i.e. combinatorial transcriptional engineering and dynamic metabolic flux control) and genome engineering (i.e. clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR associated (Cas) system based genome engineering and RNA interference assisted genome evolution) to advance metabolic engineering in yeast. We also discuss the challenges and perspectives for high throughput metabolic engineering.

\section{1. Introduction}

Due to the concerns on climate change and sustainability, there is an increasing interest in converting renewable resources into fuels and chemicals using microbial cell factories. Because microorganisms are evolved to maintain metabolic homeostasis under various environmental conditions, their metabolisms must be intensively rewired to achieve high titer, rate, and yield (TRY) for commercial production. Metabolic engineering is the science of rewiring cellular metabolism and has found broad applications in cell factory development, including \begin{enumerate*}[i] 
\item extended substrate scopes, such as xylose and arabinose;
\item increased production of the desired compounds;
\item enabled production of novel compounds, such as taxol and opioids in yeast;
\item improved cellular properties, such as tolerance to harsh industrial conditions (Du et al., 2011; Hong and Nielsen, 2012; Nevoigt, 2008; Nielsen and Keasling, 2016; Ostergaard et al., 2000).
\end{enumerate*}

\textit{Saccharomyces cerevisiae} has been widely used in the biotechnology industry because its generally regarded as safe (GRAS) status is suitable for large-scale operation. As a model eukaryotic system, molecular and cell biology of \textit{S. cerevisiae} has been studied in-depth with ample genetic engineering tools available. Unlike prokaryotes, \textit{S. cerevisiae} has multiple organelles providing different environments and compartments for biosynthesis. In addition, \textit{S. cerevisiae} exhibits high tolerance against harsh industrial conditions (Hong and Nielsen, 2012; Lian and Zhao, 2015; Nevoigt, 2008). Therefore, \textit{S. cerevisiae} has been developed as a platform microorganism for metabolic engineering.

Engineered microbial biosynthesis generally involves the conversion of some precursor metabolites into products of interest via metabolic pathways with part of or all the steps catalyzed by heterologous enzymes (Du et al., 2012). Thus, the major objectives of metabolic engineering include the enhancement of precursor supplies by genome engineering and the optimization of metabolic pathways by pathway engineering (Nielsen and Keasling, 2016). In this review, we mainly focus on the principles and tools for deregulation of endogenous metabolism and introduction and optimization of heterologous pathways in \textit{S. cerevisiae}. Readers are directed to other recent reviews on the use of yeast cell factories to produce specific fuels, chemicals, and pharmaceuticals, such as fatty acids and derivatives (Fernandez-Moya and Da Silva, 2017; Lian and Zhao, 2015), terpenoids (Zhang et al., 2017), butanol isomers (Generoso et al., 2015), and pharmaceutical proteins (Wang et al., 2017b). Advanced DNA assembly techniques that are required to build genetic circuits and heterologous pathways for metabolic engineering have also been reviewed previously (Chao et al., 2015) and will not be covered here. Specifically, we review metabolic engineering approaches at hierarchical levels of biological parts (i.e. protein engineering and promoter engineering), pathways (i.e.
combinatorial transcriptional engineering), organelles (i.e. mitochondrial compartmentalization), and systems (i.e. genome-scale modeling and the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR associated (Cas) system based multiplex genome engineering). Both experimental and computational tools will be discussed. We also include some representative examples on how these metabolic engineering approaches are applied to develop efficient yeast cell factories. Finally, future perspectives on metabolic engineering of yeast cell factories will be provided.

2. Metabolic engineering at the biological parts level

As the basic genetic elements of biosynthetic pathways and genomes, biological parts should be well-characterized and optimized for metabolic engineering applications. Based on the central dogma, we summarize the engineering of biological parts at DNA level (i.e. plasmid copy number engineering (Lian et al., 2016)), RNA level (i.e. promoter engineering (Alper et al., 2005) and terminator engineering (Curran et al., 2013a)), protein level (i.e. transporter engineering (Hu et al., 2018) and regulatory protein (Feng et al., 2015) engineering), and metabolite level (i.e. co-factor engineering (Guadalupe Medina et al., 2010)).

2.1. Copy number engineering

In S. cerevisiae, the CEN/ARS-based low-copy (~1–4 copies/cell) and 2μ-based high-copy plasmids (~20–30 copies/cell) have been commonly used for metabolic engineering. Notably, the copy number of the 2μ-based plasmids can be as low as CEN/ARS-based plasmids when antibiotic markers are included (Karim et al., 2013). Thus, the relative low plasmid copy numbers (PCNs) often become the bottleneck for metabolic engineering, where high level expression of heterologous genes is required in many cases. Interestingly, it was found that PCNs could be enhanced by using truncated or partially defective promoters to drive the expression of selection marker genes, i.e. LEU2 (Khart and Hollenberg, 1983) and URA3 (Oikkels, 1996). In a recent study, Lian et al. systematically evaluated the effect of the expression levels of the selection marker proteins on PCNs. A series of plasmids with gradually increased copy numbers were constructed, as high as ~80 copies/cell for plasmids with both auxotrophic and antibiotic markers (Lian et al., 2016). A similar strategy to enhance PCNs is to destabilize the marker proteins by fusion with a degradation tag (Chen et al., 2012a). The use of plasmids with higher copy numbers increased the expression of a lipase gene (Oikkels, 1996) and the production of patchoulol (Chen et al., 2012a) and artemisinin precursors (Ro et al., 2008). Plasmids with increased and tunable copy numbers will be invaluable tools to identify and address the bottleneck(s) of a metabolic pathway, and those with antibiotic markers are expected to find applications in industrial yeast strain engineering.

Genome integration, particularly multi-copy integration (Lee and Da Silva, 1997), is preferred for metabolic engineering, as plasmids suffer from low genetic stability and high maintenance cost. One strategy to increase the integration copy numbers is to select the clones under high selection pressure. For example, 6 copies of the wax syn-thase expression cassette was integrated into the δ sites by selection on 20 g/L G418, which increased the production of fatty acid ethyl esters (FAEEs, biodiesel) by 6-fold in S. cerevisiae (Shi et al., 2014b). Similar to the plasmid system, high copy numbers of genome integration can be achieved by using defective marker expression cassettes. For example, 30 copies of a fluorescent protein expression cassette (with a defective G418 resistance gene expression cassette) were integrated by selection on 200 mg/L of G418 (Lian et al., 2016). In another study, a malonyl-CoA reductase gene (MCR) was integrated into the yeast genome using a defective URA3 marker, which increased the production of 3-hydroxypropionic acid (3-HP) by more than 3-fold. Recently, the CRISPR/Cas system has been adopted to achieve multi-copy δ-integration, particularly for large DNA fragments (Shi et al., 2016). It is expected that the combination of the CRISPR/Cas system and the defective markers will make multi-copy genome integration of large DNA fragments much easier.

Notably, although several studies have reported positive results (Chen et al., 2012a; Lian et al., 2016; Ro et al., 2008), higher DNA copy numbers do not necessarily result in better performance of the cell factories (Yamada et al., 2010). For example, the cellobiose utilization pathway demonstrated a much higher efficiency when put on a CEN/ARS-based low copy number plasmid than that on a high copy number plasmid (Lian et al., 2014b). In other words, the effect of gene dosage on metabolic engineering should be evaluated case by case.

2.2. Promoter engineering

As the first step of gene expression, successful transcription using an appropriate promoter is essential for metabolic engineering. The expression level of the rate-limiting or branch-point enzymes can be precisely controlled by using promoters with different strengths. For example, Kim et al. used different promoters (YCp1, GPD2p, and TDH3p) to fine-tune the expression level of a pyruvate decarboxylase (PDC) gene for 2,3-butanediol (BDO) production. Complete removal of the PDC activity in yeast resulted in the inability of yeast to grow with glucose as the sole carbon source, while high expression of PDC led to the production of ethanol as the major product. Minimal ethanol production and high growth rate (BDO productivity) was achieved by fine-tuning PDC expression (Kim et al., 2016). In another study, Xu et al. used a similar strategy to evaluate the effect of TKL1 expression levels on xylose utilization rate (Xu et al., 2016a).

Yeast promoters are much more complicated than their prokaryotic counterparts because the promoter elements are not conserved or well-characterized. Therefore, yeast promoter engineering studies are mainly based on known constitutive or inducible promoters with different strengths, such as CYC1p, ADH1p, and TEF1p. Lee et al. characterized a panel of native promoters with different strength for yeast metabolic engineering, including 19 constitutive, 2 mating-type specific, and 2 inducible promoters (Lee et al., 2015). By cloning 700 bp DNA sequences directly upstream of the native start codon, promoters with 1000-fold difference in strengths were obtained. Sun et al. also characterized a panel of constitutive promoters for pathway engineering applications in S. cerevisiae (Sun et al., 2012a). In addition, random mutagenesis followed by fluorescence activated cell sorting (FACS) was employed to create an array of promoters with different strengths. For example, Alper et al. created a library of yeast TEF1 promoters with a wide dynamic range (Alper et al., 2005), which was used to modulate the expression of GPD1 (glycerol-3-phosphate dehydrogenase) for glycerol production in S. cerevisiae (Nevoigt et al., 2006). A similar approach was applied to engineer the oxygen-responsive DAN1 promoter, resulting in the creation of two promoter variants that could be induced under less-stringent anaerobic conditions. In this case, gene expression in yeast fermentations can be induced simply by oxygen depletion during cell growth. Therefore, the undetectable leaky expression and the low cost of the inducer make the DAN1 promoter mutants a promising alternative for yeast metabolic engineering (Nevoigt et al., 2007).

Significant work has also been done in the field of designing strong synthetic yeast promoters. A synthetic hybrid promoter library consisting of enhancing elements along with a core promoter element showed an expression level stronger by more than 2.5-fold than that of the strongest native promoter commonly used in yeast metabolic engineering (Blazeck et al., 2012). In another study, a hidden Markov model that predicted nucleosome occupancy on a promoter sequence was used to inform the design of both endogenous as well as synthetic promoters (Curran et al., 2014). The search for yeast promoters of shorter lengths led to a study in which minimal upstream activating sequences (UAS) were combined with minimal core promoter elements
to yield nine short synthetic promoters (Redden and Alper, 2015). Minimal promoters can reduce the DNA cargo load, thus aiding the efficiency of molecular biology experiments. Plant-based transcription factors, specifically Arabidopsis thaliana, along with synthetic promoters have also been employed to achieve orthogonal regulation of gene expression in S. cerevisiae (Naseri et al., 2017). Several combinations studied by the authors showed a 6- to 10-fold increase in expression strength over the strong TDH3 promoter. Nevertheless, the use of these synthetic promoters for metabolic engineering applications is yet to be explored.

2.3. Terminator engineering

The control of gene expression is mainly achieved via promoter engineering, while the role of terminators has been less explored in yeast. Curran et al. characterized over 30 terminators in the context of multiple promoters (Curran et al., 2013a). With a strong promoter, the difference in transcript level was determined to be as high as 6.5-fold. When coupled with a low-expression promoter, the difference in gene expression levels was even larger: 11-fold between an expression-enhancing terminator and the parent plasmid terminator and 35-fold higher than the no-terminator baseline. The changes in mRNA half-life were determined to be the major reason for varied gene expression levels. As a proof-of-concept of metabolic engineering applications, different terminators were used to control the expression of a mutant xylose isomerase gene. In another study, Yamanishi et al. comprehensively evaluated the activity of 5302 terminator regions from a total of 5880 genes (terminatome). Using GFP under the control of TDH3 promoter as the reporter, terminator region activities relative to that of the PGK1 terminator ranged from 0.036 to 2.52 (Yamanishi et al., 2013). Both studies demonstrate that terminators can be used in the same way as promoters to fine-tune gene expression levels for metabolic engineering applications in yeast.

Similar to the design of synthetic yeast promoters, synthetic terminators have also been designed and tested – by constructing a terminator from a minimal set of elements and ranging from 38 – 75 bp, the best performer was found to result in a 3.7-fold increase in fluorescent protein expression compared to the CYC1 terminator (Curran et al., 2015). A hidden Markov model was also used to test a hypothesis of nucleosome depletion in the terminator regions correlating with improved termination efficiency, aiding in the design of synthetic terminators that showed a 4-fold increase in protein expression (Morse et al., 2017).

2.4. Pathway enzyme engineering

In many cases, the catalytic activities of one or a few enzymes are the bottleneck for overall metabolic fluxes. To alleviate the rate-limiting steps, protein engineering, particularly directed evolution, has been applied to increase enzyme activities (Abatemarco et al., 2013). For example, although there is a growing interest in producing biofuels and chemicals from lignocolleucosic biomass, xylose utilization represents a rate-limiting step for a viable biotechnological process. Directed evolution was used to improve the catalytic efficiency of xylose isomerase, resulting in 61- and 8-fold improvement in the growth rate and ethanol productivity from xylose, respectively (Lee et al., 2012b). Similarly, the production of lycopene (Xie et al., 2015a) and isoprene (Wang et al., 2017a) have been improved by directed evolution of the rate-limiting enzymes in the corresponding metabolic pathways.

As directed evolution strategies have been extensively reviewed (Arnold, 2017; Cobb et al., 2013), we highlight two recent studies for continuous directed evolution of proteins and pathways in yeast. For continuous evolution, mutagenesis and screening/selection are coupled into an integrated process, which accelerates the discovery of improved variants. The first strategy took advantage of the retrotransposon Ty1, where an error-prone reverse transcriptase was used for diversity generation (Crook et al., 2016). Such an in vivo continuous evolution approach was used to evolve single enzymes, global transcriptional regulators, and multi-gene pathways. Another continuous evolution system was developed using an extrachromosomal orthogonal error-prone replication system in yeast. By introducing the cytoplasmic plasmid system of Kluyveromyces lactis into S. cerevisiae, the heterologous DNA polymerase-plasmid pair, which is orthogonal to host genomic DNA replication, can be used to mutate and evolve the gene of interest rapidly and exclusively (Ravikumar et al., 2014).

In addition to directed evolution, rational protein design has also been used to engineer enzymes in heterologous pathways in yeast. For example, the thermostability of a cellubiohydrolase from the thermophilic fungus Talaromyces emersonii expressed in S. cerevisiae was increased by introducing a disulfide bridge forming mutations using site-directed mutagenesis. The best mutant, carrying three disulfide bridges, exhibited a 9°C increase in unfolding temperature over the wild type enzyme and improved cellulose hydrolysis at 80°C (Voutilainen et al., 2010). In addition, Ding et al. used a computer-aided enzyme-substrate docking strategy to predict the catalytic efficiency among variants of the geranylgeranyl diphosphate (GGPP) synthase from 6 different organisms in the engineered taxadiene biosynthetic pathway in S. cerevisiae (Ding et al., 2014). Experimental data confirmed the predictions from this protein modeling strategy.

2.5. Transporter engineering

Although often overlooked, the transport of small molecules into and out of the cell can be the biggest limiting factor in metabolic engineering. Transporters can maximize the potential productivity by improving substrate availability or product secretion. For example, the transport of exogenous sugars (i.e. xylose and cellulbiose) is the first and rate-limiting step in the conversion of biomass sugars to fuels and chemicals. Therefore, there have been several efforts in directed evolution of transporter proteins to significantly improve biofuel productivity. Wang et al. evolved a xylose-specific transporter with significantly improved activities, whose overexpression enabled efficient xylose utilization and to some extent glucose/xylose co-fermentation (Wang et al., 2016b). Similarly, Lian et al. evolved the cellobextrin transporter 2 (CDT2) and addressed the challenges in anaerobic cellulbiose fermentation. As a facilitator with potential energetic benefits under anaerobic conditions, the engineered CDT2 increased the cellulbiose consumption rate and ethanol productivity by 4.0- and 4.4-fold, respectively (Lian et al., 2014b).

Product secretion can be equally important as substrate uptake. The removal of product from intracellular space will not only provide a driving force for biosynthesis, but also lower product toxicity and extraction cost (Hu et al., 2018; Zelle et al., 2008). For example, fermentative production of malate using an ethanol non-producing yeast strain was attempted by combined overexpression of a pyruvate carboxylase gene (PYC2), a cytosolically located malate dehydrogenase gene (MDH3ASKL), and a malate transporter gene from Schizosaccharomyces pombe (SpMAE1) (Zelle et al., 2008). Because malate is not efficiently secreted in yeast, introduction of SpMAE1 into the PYC2 and MDH3ASKL overexpressing strain further increased malate production by ~6-fold, with the titer and yield as high as 59 g/L and 0.42 mol/mol glucose, respectively. Most recently, transporter engineering was also demonstrated as a promising strategy to increase fatty alcohol production by alleviating growth inhibition. The introduction of human FATP1 resulted in 5-fold increase in fatty alcohol secretion and improved cell fitness (Hu et al., 2018). Notably, in both cases, the introduction of the heterologous transporter also increased the secretion of some by-products, such as succinate in the malate transporter expression strain (Zelle et al., 2008) and free fatty acids in the fatty alcohol exporter overexpression strain (Hu et al., 2018). Therefore, the design of substrate specific transporters using protein engineering approaches will be highly desirable to prevent the loss of carbon fluxes.
and minimize product separation cost.

2.6. Regulator and regulated protein engineering

Biological systems are highly regulated to maintain the concentrations of intermediate metabolites or final products at appropriate levels for optimal growth (Jackson et al., 1974). Thus, the presence of regulated enzymes (i.e. feedback inhibition) often becomes the bottleneck for efficient production in metabolic engineering. Control of enzyme activity occurs mainly at two levels, transcriptional regulation and protein level modification.

For post-translationally regulated proteins, an effective strategy is to mutate the inhibitor or regulator binding sites. To increase the activity of acetyl-CoA carboxylase (ACC1) and intracellular malonyl-CoA supply, Shi et al. mutated Ser659 and Ser1157 (phosphorylation sites by SNF1 protein kinase) to abolish post-translational regulation. ACC1 is responsible for the carboxylation of acetyl-CoA to produce malonyl-CoA, a universal precursor for a variety of high-value compounds, such as fatty acids, polyketides, and 3-HP. The introduction of the deregulated ACC1 enhanced the supply of malonyl-CoA and increased the production of two industrially important products derived from malonyl-CoA, FAEs (biodiesel) and 3-HP (a platform chemical). The elimination of feedback inhibition can also be achieved by simply removing the regulatory domain. For example, the truncated version of 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR) was introduced to increase the metabolic flux of the mevalonate pathway and the titer of the corresponding products. HMGR, containing a transmembrane regulatory domain and a cytosol catalytic domain, is the rate-limiting and flux controlling enzyme of the mevalonate pathway (Donald et al., 1997). By truncating the transmembrane domain, thMGR was expressed as a soluble protein in the cytosol and the regulatory features were not observed. Accordingly, the production of terpenoids, such as squalene (Donald et al., 1997; Rico et al., 2010) and artemisinin (Ro et al., 2006) and taxol precursors (Engels et al., 2008), was significantly improved.

Besides the modification of the regulated proteins, the metabolic fluxes can be increased by engineering the regulator proteins, such as the removal of negative regulators or overexpression of the positive regulators. For example, Feng et al. evaluated the deletion of a panel of negative regulators of lipid biosynthesis (i.e. RPD3, OP11, and SNF1) to increase fatty alcohol production. Yeast lipid metabolism is difficult to engineer because a broad number of reactions are tightly regulated at multiple levels (Fernandez-Moya and Da Silva, 2017). Nevertheless, the knockout of negative regulators was found to be an effective strategy to increase the metabolic flux of fatty acid biosynthesis, with the highest titer observed in the RPD3 deletion strain. By combining with ACC1 overexpression, acetyl-CoA availability engineering, and high cell density fermentation, fatty alcohol production as high as 1.1 g/L was achieved (Feng et al., 2015). In some cases, overexpression of positive regulators can be as effective for metabolic engineering, with the overexpression of UPC2-1 as a representative example. UPC2 is a transcription factor (TF) involved in the expression of sterol uptake and mevalonate pathway genes. UPC2-1 is the constitutively active mutant, whose overexpression resulted in significantly improved production of artemisinin precursors (Ro et al., 2006; Westfall et al., 2012) and bisabolene (Peralta-Yahya et al., 2011).

Another important strategy is global Transcription Machinery Engineering (gTME), where the expression levels of multiple genes are modulated simultaneously. Different from catalytic protein and transporter engineering, gTME applies directed evolution to transcription factors. Alper at al. performed directed evolution of SPT15, the TATA-box-binding protein, an essential general TF involved in directing the transcription of many genes by all three RNA polymerases (Alper et al., 2006). By enriching yeast strains containing mutated SPT15 genes in 6% ethanol, an SPT15 variant was isolated and found to significantly improve ethanol tolerance and productivity. Because directed evolution of regulatory machinery can change the expression level of multiple genes simultaneously for optimal performance of yeast cell factories, gTME is a powerful tool for metabolic engineering.

Regulatory interactions in eukaryotic systems including yeast are still being explored, and as a result, few studies have utilized regulatory engineering to the full capacity possible for achieving metabolic engineering targets. Ellis et al. introduced a model-guided framework to improve the predictability of synthetic gene circuit performance. The model-guided design was used to introduce a time-varying network in S. cerevisiae that induced flocculation in the culture (Ellis et al., 2009). This model-guided strategy for designing a gene regulatory network is a significant step towards engineering S. cerevisiae and other organisms’ native regulatory networks. In the context of designing or engineering regulatory networks, often the first step is decoding these hidden interactions, also called reverse engineering of regulatory networks. Reverse engineering algorithms employ gene expression data to identify statistically significant interactions and build a correlation network from that. These networks can then be used to inform new strain engineering strategies that rewire host regulation. One of the challenges when exploring reverse engineering algorithms is the process of benchmarking, or evaluating their predictive performance. To address this, a synthetic five-gene network was expressed in S. cerevisiae, and used to benchmark different mathematical models that capture the regulatory interactions. Such a platform can prove to be very useful for assessing algorithms that can later inform on the regulatory landscape of a host organism (Cantone et al., 2009).

2.7. Co-factor engineering

Co-factors, such as NADH and NADPH, are involved in many catabolic and biosynthetic pathways. Besides low enzymatic activities, co-factor imbalance is also a common bottleneck for maximizing the metabolic fluxes and bioprocess efficiencies. For cellulose biofuel production, the fungal xylose utilization pathway represents one of such examples, where xylose reductase (XR) utilizes NADPH to convert xylose into xylitol, and xylitol dehydrogenase (XDH) prefers NAD+ instead of glycerol formation as the electron sink during anaerobic fermentation (Guadalupe Medina et al., 2010). Because acetate is consumed of NADH, resulting in xylitol accumulation (Kim et al., 2013c; Wei et al., 2013). By adjusting the XR cofactor preference to NADH or XDH to NADP+, decreased xylitol accumulation and improved ethanol production can be achieved (Bengtsson et al., 2009; Matsushika et al., 2008).

In addition to co-factor specificities, co-factor excess can result in co-factor imbalance as well. In anaerobic cultures of S. cerevisiae, glycerol production serves as an electron sink to reoxidize excess NADH produced in biosynthetic processes. Consequently, glycerol is generated as the major by-product. To minimize glycerol formation and increase ethanol yield, one strategy is to express a water-forming NADH oxidase gene to decrease the intracellular NADH/NAD+ ratio (Kim et al., 2016). Nevertheless, the requirement for O2 as the electron acceptor may limit its large-scale industrial applications. A more promising strategy is to use acetate reduction to reoxidize excess NADH. By introducing an acetylating NAD+ -dependent acetaldehyde dehydrogenase gene (i.e. mhpF from E. coli), acetate is reduced to ethanol accompanied with the consumption of NADH. In other words, acetate reduction can be used instead of glycerol formation as the electron sink during anaerobic fermentation (Guadalupe Medina et al., 2010). Because acetate is available at significant amounts in lignocellulosic hydrolysates, this strategy not only increased ethanol yield, but also addressed the acetate toxicity concerns during cellulosic biofuel production. Similarly, such a redox balancing strategy was adopted to enable efficient xylose fermentation and in situ detoxification of acetate simultaneously (refer to example 5.1 for details) (Wei et al., 2013).

3. Metabolic engineering at the pathway level

Introduction of a biosynthetic pathway containing multiple
heterologous genes is generally the first step for microbial production of biofuels, chemicals, and pharmaceuticals. To develop a commercially viable process with maximal product TRY, metabolic fluxes of the multi-gene pathways must be carefully balanced (Eriksen et al., 2014; Martin et al., 2009). Based on the composition of biological parts, a multi-gene pathway can be optimized at DNA level (i.e. copy number modulation (Lian et al., 2016)), transcription and mRNA level (i.e. combinatorial transcriptional engineering (Du et al., 2012)), translation and protein level (i.e. protein co-localization (Tippmann et al., 2017)), and metabolite level (i.e. dynamic metabolic flux control (Scalcinati et al., 2012)) (Fig. 1). Here, we summarize different approaches to balance and optimize the multi-gene metabolic pathways, highlighting some of the most recently developed synthetic biology tools.

3.1. Metabolic pathway optimization at the DNA level

DNA copy number modulation is probably the most straightforward strategy for pathway optimization. As mentioned above, the relatively low copy numbers and narrow dynamic range of the commonly used yeast plasmid system (Fig. 1A) limit their applications in pathway optimization and metabolic engineering. To address the challenge, a series of plasmids with gradually increased and tunable copy numbers were constructed. Based on the finding that the copy number of the plasmids with engineered dominant markers showed a positive correlation with the concentration of antibiotics supplemented to the growth media, Lian et al. developed Pathway Optimization by Tuning Antibiotic Concentrations (POTAC), a simple yet highly efficient strategy to rapidly balance the flux of multi-gene pathways. This DNA level pathway optimization tool was used to optimize the lycopene and n-butanol biosynthetic pathways, increasing the production of lycopene and n-butanol by 10- and 100-fold, respectively (Lian et al., 2016).

For metabolic engineering applications, genome integration to maintain the heterologous pathways without selection pressure is preferred. Using the recently developed CRISPR system, multi-copy δ-
integration of large DNA fragments (> 10 kb) became possible (Shi et al., 2016). A bigger challenge is to precisely control the copy numbers of the integrated cassettes for metabolic pathway balancing. A few studies demonstrated that the genome integrated copy numbers could be controlled by the antibiotic concentration supplemented to the growth medium (Lian et al., 2016; Shi et al., 2014b; Yuan and Ching, 2015). For example, Yuan and Ching developed an antibiotic selection based approach for combinatorial assembly of large pathways. A 5-gene isobutanol pathway and an 8-gene mevalonate pathway were successfully assembled into yeast chromosomes and the optimal strains with proper DNA fragment copy numbers were screened and characterized for synthesis of the desired product (Yuan and Ching, 2015). In another study, Lian et al. found that the copy numbers of the integrated cassettes can be precisely tuned using antibiotic concentration when truncated promoters were used to drive the expression of the marker genes. By testing the combination of different antibiotic concentrations, the gene dosage of the upstream pathway and the downstream pathway can be pre-determined and balanced for optimal performance (Lian et al., 2016).

3.2. Metabolic pathway optimization at the RNA level

Transcriptional control plays important roles in fine-tuning the expression of multiple genes simultaneously to maximize product TRY. To achieve this, promoter libraries were used to drive the expression of each pathway gene, and the resultant combinatorial pathway libraries were further evaluated (Fig. 1B). Promoter libraries may comprise of different endogenous promoters (Latimer et al., 2014; Lee et al., 2013) or promoter mutants (generated by error-prone PCR) (Du et al., 2012) with varying strengths. Lee et al. used a library of endogenous promoters with orders of magnitude difference in strength (TDH3p, TEF1p, RPL11Bp, RNR2p, and REV1p) to fine-tune the expression of every individual gene of the bacterial violacin pathway and precisely controlled the metabolic fluxes for producing violacin and its intermediates (Lee et al., 2013). Latimer et al. used the same set of promoters to regulate the expression of each gene of the xylose utilization pathway and the resulting combinatorial library of pathways was screened under different conditions, i.e. aerobic and anaerobic conditions. The authors could not only obtain faster xylose utilization strains, but also identify important factors in xylose utilization. Interestingly, the optimized xylose pathways were determined to be screening condition dependent, where the aerobically-enriched strains under-performed under anaerobic conditions (Latimer et al., 2014).

Instead of native promoters, Du et al. created 3 libraries of promoter mutants (TEF1p mutants, ENO2p mutants, and PDC1p mutants), which were used to develop a simple, efficient, and programmable approach named COMPACTER, customized optimization of metabolic pathways by combinatorial transcriptional engineering, for rapid tuning and optimization of metabolic pathways. COMPACTER was applied to optimize a xylose utilization pathway and a cellulose utilization pathway in both laboratory and industrial yeast strains. Interestingly, the engineered xylose and cellulose utilizing pathways were found to be host-specific. In other words, COMPACTER is capable of tailor-making and optimizing metabolic pathways for different strain backgrounds (Du et al., 2012). In a following up study, Yuan et al. used random mutagenesis generated promoter libraries (ENO2p library and PDC1p library) rather than pre-characterized promoters (i.e. endogenous promoters or promoter mutants with well-defined strength) to optimize a cellulose utilizing pathway in an industrial yeast strain. For the first time, directed evolution and promoter engineering were combined for rapid and efficient multi-gene pathway optimization. After three rounds of pathway directed evolution, cellulose consumption rate and ethanol productivity were increased by 6.41-fold and 6.36-fold, respectively (Yuan and Zhao, 2013c).

3.3. Metabolic pathway optimization at the protein level

Besides the promoter based combinatorial pathway engineering approach, a multi-gene pathway can also be balanced and optimized by exploring various combinations of enzyme analogues with different properties (Fig. 1C). Considering the enzyme diversities, the protein based combinatorial pathway engineering method can be readily applied for rapid optimization of multi-gene pathways in a highly customized manner for metabolic engineering. For example, Kim et al. created a library of more than 8000 xylose utilization pathways with all possible combinations of 20 xylose reductase genes (XR), 22 xylitol dehydrogenase genes (XDH), and 19 xylulose kinase genes (XKS) to engineer a highly efficient xylose-utilizing pathway for cellulosic ethanol production. In the yeast strain with an optimal enzyme combination, the by-product yield (i.e. xylitol, glycerol, and acetate) was determined to be as low as 0.06 g/g xylose without compromising cell growth (Kim et al., 2013a). In another study, Chen et al. screened carotenogenic enzymes (CrtE, CrtB, and CrtI) from diverse species and the optimal combination produced lycopene at the highest yield (55.56 mg/g DCW) ever reported in yeast, when combined with host engineering and fed-batch fermentation (Chen et al., 2016b).

Instead of using enzyme analogues with different origins, Eriksen et al. explored the combination of random mutagenesis-generated protein mutants to optimize a cellobiose utilization pathway. Through a new directed evolution strategy by simultaneously engineering multiple proteins of the pathway, key mutations in different proteins synergistically improved the overall cellulose utilization and ethanol productivity by 49% and 64%, respectively. The improvement resulted from a more active celloextrin transporter (CDT) mutant and a β-glucosidase (BGL) mutant with increased specificity (Eriksen et al., 2013). Although the authors demonstrated the significance of evolving pathway proteins simultaneously, the biggest challenge is to create a high quality library with enough coverage.

The performance of a biosynthetic pathway can also be balanced and optimized by protein colocalization (Fig. 1D), which can largely increase the local concentrations of enzymes and metabolites. In addition, facilitated substrate channeling and decreased release of toxic intermediates are other possible reasons for enhanced pathway efficiency. The simplest way to colocalize proteins is to fuse metabolic enzymes together. For example, the fusion of the endogenous farnesyl diphosphate (FPP) synthase (FPPS, ERG20) with bisabolene synthase (AgBIS) increased the production of bisabolene by 2-fold, when compared with the individually expressed genes (Ozaydin et al., 2012). Albertsen et al. fused ERG20 with plant patchoulol synthase to increase the production of patchoulol by ~2-fold (Albertsen et al., 2011). A more flexible method for pathway protein colocalization is to use protein scaffolds, docking enzymes in close proximity to one another (Lee et al., 2012a). Tippmann et al. reported the use of affibodies as the protein scaffolds for enzyme colocalization, which is based on the recognition of affibodies to their anti-idiotypic partners in vivo. Through colocalization of FPPS and farnesene synthase (FS) in S. cerevisiae, the yield of farnesene on glucose could be improved by 135% in fed-batch cultivations (Tippmann et al., 2017). Lin et al. develop a protein-based scaffold for colocalizing multiprotein complexes on the membranes of intracellular lipid droplets for ester bioconversion in yeast. Because the last step enzyme, alcohol-0-acetyltransferase (ATF1), was localized in the lipid droplets, the plant lipid droplet protein oleosin and cohesion-dockerin interaction based scaffold was used to recruit precursor supply enzymes (ALD6 and ACS1 for acetyl-CoA generation) to the same location. Metabolic pathway colocalization increased the production of ethyl acetate by ~2-fold. Another successful example of protein colocalization is the display of an artificial cellulolusome (Wen et al., 2010) or hemicellulolusome (Sun et al., 2012b) on yeast cell walls for consolidated bioprocessing (CBP). For the assembly of a mini-cellulosome, three component enzymes, endoglucanase (EG), cellulohydrolase (CBH), and β-glucosidase (BGL), were co-localized onto the yeast surface.
associated scaffoldin through high affinity interaction between cohesins and dockersins. The tri-functional mini-cellulosome resulted in the production of 1.8 g/L ethanol from 10 g/L phosphoric acid-swollen cellulose (PASC). In a following study, Liang et al. integrated two additional cellulosytic enzymes, a lytic polysaccharide monooxygenase (LPMO) and a cellobiose dehydrogenase (CDH), to construct a penta-functional mini-cellulosome displaying yeast strain. The resultant strain produced 2.7 g/L ethanol from 10 g/L PASC and grew on cellulose as the sole carbon source (Liang et al., 2014).

3.4. Dynamic flux control for metabolic pathway optimization

Although the above mentioned metabolic pathway optimization strategies have been extensively used in metabolic engineering, the static nature makes them non-optimal in some cases. The ideal flux control should be able to respond to internal and/or external conditions in a real-time manner (Fig. 1E). In the dynamic flux control system, the expression levels of the pathway genes (particularly the flux control gene(s)) are dependent on the intracellular metabolite concentration. In other words, the pathways enzymes are only expressed when needed. Therefore, dynamic flux control can not only reduce metabolic burdens, but also minimize by-product formation.

The most widely used dynamic regulation system in yeast is the sugar-responsive element. For example, the strength of HXT1 promoter is dependent on glucose concentration. In the scenario of metabolic engineering, in the beginning of fermentation, cell density is low and gene expression is high due to abundant glucose; when cell density becomes higher, glucose is depleted to repress gene expression. Therefore, HXT1p can be used for dynamic regulation of the competing pathway genes, whose expression is essential for cell survival, with ERG9 serving as the most representative example. The expression level of ERG9 (squalene synthase) controls the distribution of metabolic fluxes through the endogenous ergosterol formation or the heterologous terpenoid biosynthesis. Scalcinati et al. used the HXT1 promoter to control ERG9 expression, diverting the carbon flux from sterol synthesis towards α-santalene, and improved the productivity by 3.4-fold (Scalcinati et al., 2012). Xie et al. followed the same strategy to control ERG9 expression with HXT1p for carotenoid production. Through sequential control of the pathways and balanced utilization of FPP, carotenoid was produced as high as 1156 mg/L (20.79 mg/g DCW) using high-cell-density fermentation (Xie et al., 2015b). In another study, Teixeira et al. used HTXIp to control the expression of FAA1 for dynamic regulation of the interconversion of fatty acyl-CoA and free fatty acid. Dynamic flux control resulted in an increased accumulation of fatty alcohols up to 41% while free fatty acid levels were decreased by 63% compared with the control strain (Teixeira et al., 2017).

Besides sugar responsive promoters, transcriptional factor based biosensors and also been used for dynamic flux control. David et al. developed a FapR based malonyl-CoA biosensor to dynamically regulate the expression of malonyl-CoA reductase, which significantly increased the production of 3-HP (David et al., 2016). A synthetic quorum-sensing system has been developed in yeast, which was combined with RNA interference (RNAi) for dynamic metabolic pathway control. The system autonomously triggered gene expression at high cell density, and was linked with an RNAi module to repress the expression of target genes. As proof-of-concept, the system was used to control flux through the shikimate pathway for producing para-hydroxybenzoic acid (PHBA). Dynamic RNAi resulted in the highest reported PHBA titer (1.1 mM) in yeast (Williams et al., 2015).

4. Metabolic engineering at the organelle level

The presence of a variety of organelles with various conditions (i.e. pH and oxygen levels) and precursor metabolite availabilities (i.e. acetyl-CoA and FPP) opens new opportunities for metabolic engineering in S. cerevisiae (Hammer and Avalos, 2017). Similar to the scaffold based protein colocalization strategy, compartmentalization of metabolic pathways increases local enzyme and metabolite concentrations for more efficient biocatalysis, prevents the release of toxic intermediates, and eliminates the bottleneck in substrate diffusion. In addition, some proteins have been found to function better inside specific organelles than those in the cytosol. Nevertheless, the release of the final products from organelles can be one of the biggest challenges for compartmentalization engineering and should be addressed to achieve high-level production. Based on the metabolic engineering needs, the whole metabolic pathways have been compartmentalized into various organelles, such as mitochondria, peroxisomes, endoplasmic reticulum (ER), and vacuoles (Table 1).

### Table 1: Compartment engineering strategies for yeast metabolic engineering.

<table>
<thead>
<tr>
<th>Compartment</th>
<th>Products</th>
<th>Advantages</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mitochondrion</td>
<td>Hydrocortisone</td>
<td>Electron transfer</td>
<td>(Szczepanska et al., 2003)</td>
</tr>
<tr>
<td></td>
<td>Valencene and amorphadiene</td>
<td>Availability of FPP</td>
<td>(Farhi et al., 2011)</td>
</tr>
<tr>
<td></td>
<td>Amorphadiene</td>
<td>Availability of acetyl-CoA</td>
<td>(Yuan and Ching, 2016)</td>
</tr>
<tr>
<td></td>
<td>Isoprene</td>
<td>Supply of acetyl-CoA and less competition</td>
<td>(Lo et al., 2016)</td>
</tr>
<tr>
<td></td>
<td>Branch-chain alcohols</td>
<td>Supply of precursor (KIV) and cofactor (NADPH) and better folding environment for ILV3</td>
<td>(Avalos et al., 2013)</td>
</tr>
<tr>
<td>Peroxisome</td>
<td>PHAs</td>
<td>Acyl-CoAs from β-oxidation</td>
<td>(Poier et al., 2001; Zhang et al., 2006)</td>
</tr>
<tr>
<td></td>
<td>Medium-chain fatty acids</td>
<td>Acyl-CoAs from β-oxidation</td>
<td>(Chen et al., 2014a)</td>
</tr>
<tr>
<td></td>
<td>Medium-chain fatty alcohols</td>
<td>Acyl-CoAs from β-oxidation</td>
<td>(Shen et al., 2016a)</td>
</tr>
<tr>
<td></td>
<td>Alkanes and olefins</td>
<td>Less competition and supply of precursor and NADPH</td>
<td>(Zhou et al., 2016a)</td>
</tr>
<tr>
<td></td>
<td>Terpenoids</td>
<td>Functional expression of P450s</td>
<td>(Arendt et al., 2017)</td>
</tr>
<tr>
<td></td>
<td>Opioids</td>
<td>Minimal by-product formation</td>
<td>(Thodey et al., 2014)</td>
</tr>
<tr>
<td>Vacuole</td>
<td>Methyl halides</td>
<td>Availability of SAM</td>
<td>(Bayer et al., 2009)</td>
</tr>
</tbody>
</table>

FPP: farnesyl diphosphate; KIV: α-ketoisovalerate; SAM: S-Adenosyl Methionine.

a General advantages of compartmentalization such as higher local enzyme concentration and better substrate channeling are not included.

b ILV3 (dihydroxyacid dehydratase) contains an iron–sulfur cluster (ISC), which is exclusively synthesized in mitochondria.
targeting oxidoreductase ARH1 and six additional mammalian proteins in the cytosol, Szczepanak et al. constructed a S. cerevisiae strain capable of producing hydrocortisone as the major steroid (~70%).

Avalos et al. systematically evaluated the mitochondrial compartmentalization for branched-chain alcohol (i.e. isobutanol) production (Avalos et al., 2013). In S. cerevisiae, the upstream enzymes (ILV2, ILV5, and ILV3) are mitochondrial, while the downstream enzymes (α-ketoacid decarboxylases (KDCs) and alcohol dehydrogenases (ADHs)) are cytosolic. Therefore, complete compartmentalization of the whole pathway addressed the bottleneck in α-ketoisovalerate (α-KIV) export from mitochondria and minimizes the loss of α-KIV to competing reactions in the cytosol. In addition, the availability of an iron–sulfur cluster for ILV3 and reducing power for ILV5 and ADHs makes mitochondria a better location for isobutanol production. Therefore, the mitochondrial pathway resulted in the production of 635 mg/L of isobutanol, 2.6-fold higher than the cytosolic pathway. In addition, the production of isopentanol (130 mg/L) and 2-methyl-1-butanol (113 mg/L) were increased by 3.7-fold and 5-fold, respectively.

Farhi et al. took advantage of the availability of FPP in mitochondria for metabolic engineering of plant terpenoid production. By targeting heterologous sesquiterpene synthases to mitochondria, the production of valencene and amorphadiene were increased by 3-fold and 7-fold, respectively. By targeting both the FPPS and a sesquiterpene synthase to mitochondria, the production of valencene and amorphadiene was further increased to 1.2 mg/L and 20 mg/L, which represented 8-fold and 20-fold higher titers than the cytosolic pathways, respectively (Farhi et al., 2011).

Recently, a growing interest is to utilize the acetyl-CoA pool in mitochondria for biosynthesis of value-added compounds. Yuan and Ching targeted the whole FPP pathway (ERG10, ERG13, thMG, ERG12, ERG8, ERG19, ID1, and ERG20) together with the amorphadiene synthase (ADS) to the mitochondria. The production of amorphadiene as high as 427 mg/L confirmed a large pool of acetyl-CoA in the mitochondria (Yuan and Ching, 2016). In another study, Lv et al. targeted a similar mevalonate pathway (ERG10, ERG13, thMG, ERG12, ERG8, ERG19, and ID1) together with a heterologous isoprene synthase (ISPS) to mitochondria for isoprene production. This mitochondrial targeting strain produced 108 mg/L of isoprene, 1.7-fold higher than the control strain with a cytosolic pathway. As expected, squalene was produced at a much lower level, demonstrating the significance of compartmentalization for reducing by-product formation. In addition, the authors showed that simultaneous utilization of mitochondrial and cytosolic acetyl-CoA increased the production of isoprene even further. Dual regulation of cytoplasmic and mitochondrial acetyl-CoA utilization enabled the production of ~2.5 g/L isoprene in fed-batch fermentation, the highest titer ever reported in eukaryotes (Lv et al., 2016).

4.2. Peroxisome engineering

As the peroxisomes are the natural places for fatty acid degradation, they have been harnessed to produce fatty-acid-derived chemicals, including polyhydroxyalkanoates (PHAs), medium-chain fatty acids (MCFAs), fatty alcohols, alkanes, and olefins (DeLoache et al., 2016). Peroxisome compartmentalization was firstly demonstrated for producing PHAs, a family of biodegradable polymers. Based on the PHA synthases introduced, yeast strains have been engineered to accumulate both medium-chain-length (mcl) PHAs (Poirier et al., 2001) and short-chain-length (scl) PHAs (Zhang et al., 2006) in the peroxisomes. By targeting the scl-PHA synthase to the peroxisomes, scl-PHA was accumulated as high as 0.8% DCW, which was about 40-fold higher than the cytosolic synthase expressing strain (Zhang et al., 2006).

More extensive engineering efforts were devoted to producing fatty-acid-derived chemicals by hijacking the β-oxidation cycle intermediates. Chen et al. attempted to produce MCFAs from low-value long-chain fatty acids. To prevent extensive MCFA degradation, the only acyl-CoA oxidase in S. cerevisiae (POX1) was replaced by POX2 from Yarrowia lipolytica, which shows a preference for long-chain acyl-CoAs. Expression of a peroxisomal carnitine O-octanoyltransferase enabled product export from the peroxisomes and increased MCFA production by 3.34-fold (Chen et al., 2014a). In another study, Sheng et al. targeted a heterologous fatty acyl-CoA reductase (FAR) to the peroxisome with a PTS2 signal peptide, which converted β-oxidation intermediates to medium-chain fatty alcohols (C10 and C12). Fed-batch fermentations in nitrogen-limited medium produced 1.3 g/L fatty alcohols, containing 34.4% C10 and C12 alcohols (Sheng et al., 2016).

Recently, peroxisomes have also been harnessed for alkane and olefin production. To minimize the loss of aldehyde intermediates to byproducts (fatty acids and/or fatty alcohols), Zhou et al. targeted a bacterial aldehyde-deformylating oxygenase (ADO) and a carboxyl acid reductase (CAR) to the peroxisomes, which led to increased alkane production by 90% compared to the cytosolic pathway. By engineering peroxisomal biogenesis factors to increase the number of peroxisomes, alkane production was further enhanced. Similarly, peroxisomal targeting of a bacterial P450 fatty acid decarboxylase (OleT) and the associated electron transfer system increased olefin production by 40% (Zhou et al., 2016a).

4.3. Compartment engineering of ER

Although it is mainly known as the major compartment for protein folding, ER including the ER membrane is also home to several endogenous (i.e. ergosterol biosynthesis enzymes) and heterologous (i.e. P450 enzymes) enzymes. Therefore, ER can be an important compartment for metabolic engineering applications. Arendt et al. constructed an ER-engineered platform yeast strain for producing plant natural products such as terpenoids. The authors found that the disruption of PAH1, encoding a phosphatidic acid phosphatase, resulted in significant morphological changes such as ER proliferation. The ER proliferated strain could be employed for functional overproduction of ER-localized proteins, such as the plant terpenoid biosynthesis enzymes. Compared to the wild-type strain, accumulation of β-amyrin, medicagenic acid, and medicagenic-28-O-glucoside was increased by 8-, 6-, and 16-fold, respectively. PAH1 deletion increased the production of artemisic acid by 2-fold as well (Arendt et al., 2017). In another study, ER was employed to reconstruct the opioid biosynthetic pathway from the poppy plant. Expression of dioxygenases T6ODM and C0DM and aldonolactoneductase COR, which are required for biosynthesis of morphine from thebaine, in the cytosol resulted in the production of 2.5 mg/L of morphine and 3.2 mg/L of neomorphine as the byproduct. By targeting COR to the ER membrane, morphine production was increased to 3.1 mg/L, while the by-product neomorphine was decreased to 0.5 mg/L (Thodey et al., 2014).

As ER is the location for triacylglycerides (TAGs) biosynthesis, it has the potential to be engineered for producing fatty acid derivatives. By targeting of the terminal enzymes (i.e. wax ester synthases) to the ER, the production of FAEEs and alkanes were significantly improved in Y. lipolytica (Xu et al., 2016b). Nevertheless, ER compartmentalization for producing fatty acid derivatives has yet to be explored in S. cerevisiae. Notably, fatty acid metabolism occurs in the cytosol, peroxisomes, ER, and mitochondria. Therefore, combined regulation for fatty acid precursor utilization may maximize the production of fatty acid derived compounds.

4.4. Vacuole compartmentalization

The vacuole, known as the digestive organelle, is not only the place for protein degradation, but also serves as a storage compartment. For example, S-Adenosyl Methionine (SAM) is mainly stored in this compartment. By targeting methyl halide transferase (MHT), which transfers a methyl group from SAM to a halide ion to synthesize methyl halides, to the vacuole, methyl iodide was produced at a rate of
190 mg/L/h, 1.5-fold higher than the cytosolic MHT expressing strain. When methionine was supplemented to boost SAM accumulation, the production of methyl iodide was further increased to 860 mg/L/h (Bayer et al., 2009).

5. Metabolic engineering at the systems level

Although yeast has been engineered to produce a wide range of fuels, chemicals, and drugs, the development of efficient cell factories for industrial application is still time and cost intensive. The biggest challenge for metabolic engineering lies in the robust metabolic and regulatory networks, i.e. to maintain cellular homeostasis even under varying environmental conditions. We summarize the principles (i.e. platform strains) and tools (i.e. genome engineering and systems biology tools) to perturb metabolic and regulatory networks for metabolic engineering applications.

5.1. Platform strains with enhanced supply of precursor metabolites

Although cellular metabolism is rather complicated and highly regulated, it is naturally organized into a “bowtie” structure, where the carbon and energy sources are converted through central metabolism into a few precursor metabolites, such as acetyl-CoA (Nielsen and Keasling, 2016). If we can construct yeast strains with enhanced supply of these general precursor metabolites, such platform strains can be converted to efficient cell factories with minimal modifications. Because a major bottleneck for metabolic engineering is the deregulation of cellular metabolism to enhance precursor supply, the constructed platform strains will enable the development of efficient cell factories much easier and faster.

5.1.1. Pyruvate platform strain

In S. cerevisiae, most of the glycolytic fluxes go through ethanol fermentation even under aerobic conditions. Thus, PDC is often chosen as the major target for metabolic engineering. Unfortunately, the PDC deficient strain (Pdc−, pdcΔ pdc5Δ pdc6Δ) cannot grow on glucose as the sole carbon source (Flikweert et al., 1999). Inverse metabolic engineering revealed that an internal deletion in the MTH1 coding sequence enabled the growth of Pdc− strain on glucose (Oud et al., 2012; van Maris et al., 2004). In other studies, a mutation in the MTH1 coding sequence (Ala81Pro) (Kim et al., 2013b) or overexpression of MTH1 gene on a multi-copy plasmid (Lian et al., 2014a) also rescued the van Maris et al., 2004). The pyruvate accumulating platform strain can be converted to efficient cell factories with minimal modifications. Because a major bottleneck for metabolic engineering is the deregulation of cellular metabolism to enhance precursor supply, the constructed platform strains will enable the development of efficient cell factories much easier and faster.

5.1.1.2. Acetyl-CoA platform strain

Acetyl-CoA is a central metabolite and key precursor in the biosynthesis of cellular components and many natural products. Unfortunately, acetyl-CoA biosynthesis in S. cerevisiae suffers from low metabolic fluxes, feedback inhibition, and high energy input requirement. Therefore, several metabolic engineering strategies have been attempted to construct acetyl-CoA overproducing strains, such as to eliminate competing pathways, to inactivate acetyl-CoA consuming pathways, and to introduce more efficient acetyl-CoA biosynthetic pathways (Fig. 2).

Acetyl-CoA synthetase (ACS) is subject to feedback inhibition and rate-limiting for acetyl-CoA biosynthesis. By introducing a point mutation at the acetylation position (i.e. ACSΔ614 from Salmonella enterica), the feedback inhibition insensitive ACS mutant could enhance acetyl-CoA supply. When combined with the overexpression of the upstream genes including aldehyde dehydrogenase (ALD6) and/or ethanol dehydrogenase (ADH2), the introduction of the ACS mutant increased the production of α-santalene (Chen et al., 2013), amorphadiene (Shiba et al., 2007), 3-HP (Wang et al., 2014), PHB (Kochhar et al., 2012), and n-butanol (Krivoruchko et al., 2013) by 1.75-, 1.22-, 1.5-, 18-, and 3.10-fold, respectively.

Considering the significance of acetyl-CoA in cellular metabolism, nature has evolved various routes to synthesize acetyl-CoA under different conditions. Several acetyl-CoA biosynthetic pathways with less energy input requirement have been heterologously expressed in yeast and their energetic benefits have been evaluated for metabolic engineering applications (Fig. 2). Pyruvate dehydrogenase (PDH) catalyzes the oxidative degradation of pyruvate for acetyl-CoA synthesis in an ATP-free manner. Functional expression of PDH in yeast has been achieved by co-expressing PDH genes and protein-lipoate ligase genes from both Enterococcus faecalis (pdhA pdhB aceF lpd lplA lplA2) (Kozak et al., 2014) and E. coli (lpd aceF aceE lplA) (Lian et al., 2014c; Lian and Zhao, 2016). With the supplementation of lipoic acid, functional PDH complex could fully replace the endogenous pathway for cytosolic acetyl-CoA synthesis, resulting in a similar growth rate as the wild-type. In addition, Lian and Zhao introduced the de novo lipoylation machinery for functional expression of PDH in absence of lipoic acid, although the growth rate was much lower (Lian and Zhao, 2016). ATP-dependent citrate lyase (ACL) converts citrate to acetyl-CoA and oxaloacetate at the cost of one ATP molecule, which is the major source of acetyl-CoA generation in oleaginous yeasts such as Y. lipolytica (Vorapreeda et al., 2012). ACLs from oleaginous yeasts (Lian et al., 2014c), plants (Fatland et al., 2002), and mammalian cells (Tang et al., 2013) have been functionally expressed in S. cerevisiae and were able to increase the production of fatty acids (Tang et al., 2013), n-butanol (Lian et al., 2014c), fatty alcohols (Feng et al., 2015), by 1.17-, 2-, and 2.4-fold, respectively. Phosphoketolase (PK) catalyzes the cleavage of xylulose-5-phosphate for acetyl-CoA generation (Panagiotou et al., 2008). By introducing the phosphoketolase pathway, the production of PHB (Kochhar et al., 2013) and FAEE (de Jong et al., 2014) was improved by 7- and 5.7-fold, respectively. Pyruvate: formate lyase (PFL) and acetylating aldehyde dehydrogenase (A-ALD) were also found to function as alternative routes for acetyl-CoA biosynthesis in yeast (Kozak et al., 2013).

As acetyl-CoA is generally involved in the synthesis of essential cellular components, glyoxylate shunt, a pathway for acetyl-CoA consumption, was mainly chosen as the metabolic engineering targets (Chen et al., 2012b) (Fig. 2). The disruption of CIT2 or MLS1 could increase the production of α-santalene (Chen et al., 2013), 3-HP (Wang et al., 2014), and n-butanol (Krivoruchko et al., 2013). Interestingly, the disruption of the glyoxylate shunt impaired the production of PHB (Kochhar et al., 2012).

As mentioned above, most of the metabolic fluxes go through ethanol fermentation during glucose fermentation, indicating a need to rewrite yeast central metabolism to enhance acetyl-CoA supply. Because acetaldehyde is the branch point to control the flux distribution between ethanol and acetyl-CoA, alcohol dehydrogenases (ADHs) were chosen as the metabolic engineering targets (Fig. 2). The deletion of ADH1, a major ADH responsible for ethanol fermentation, resulted in 1.9-fold improvement in fatty acid production (Li et al., 2014). Lian et al. knocked out ADH1 and ADH4 for ethanol formation and GPD1 and GPD2 for glycerol biosynthesis, the intracellular acetyl-CoA concentration and n-butanol production were increased around 2- and 4-fold, respectively. Due to the redundancy of ADHs in the yeast genome, ethanol is still produced as the major product and a promising strategy is to inactivate the upstream enzymes, PDCs. Considering the
accumulation of pyruvate to high levels and the functional expression of PDH to convert pyruvate to acetyl-CoA in yeast, it seems to be straightforward to further engineer the pyruvate platform strain into the acetyl-CoA platform strain. Nevertheless, initial attempts to introduce functional PDH into PdcΔ strain resulted in limited success (Lian and Zhao, 2016), indicating that more thorough studies at the systems level are required to further engineer the acetyl-CoA platform strain. Notably, the above mentioned metabolic engineering strategies should be combined to construct acetyl-CoA overproducing platform yeast strains. Chen et al. developed an acetyl-CoA platform strain by combining the overexpression of ADH2, ALD6, and AceS641P and disruption of the glyoxylate shunt (MLS1), which increased the production of a-santalene by 4-fold (Chen et al., 2013). Lian et al. combined the disruption of competing pathways (adh1Δ adh4Δ gpd1Δ gpd2Δ) and introduction of alternative acetyl-CoA pathways (PDH and ACL), increasing the intracellular acetyl-CoA concentration and n-butanol production by 3- and 12-fold, respectively (Lian et al., 2014c). The combination of A-ALD and PK for high titer (> 150 g/L) and yield (< 1% by-product formation) production of farnesene is described in example 5.3 (Meadows et al., 2016).

### 5.1.3. Shikimate and p-coumaric acid

Aromatic amino acids are precursors of numerous plant secondary metabolites with diverse biological functions. Due to the tight regulation of amino acid metabolism, Suastegui et al. adopted a multilevel approach to establish a shikimate (a precursor of the aromatic amino acid biosynthesis pathway) platform strain (Suastegui et al., 2017). Negative regulators were removed (i.e. ric1Δ) to alleviate transcriptional repression, competing pathways were eliminated (i.e. aro1Δ) to redirect the metabolic fluxes to shikimate, and central metabolism was rewired to enhance precursor supply (i.e. overexpression of TKI1 and RKL1). A combination of these modifications increased the production of shikimate by 7-fold, with the titer as high as 2.5 g/L. Notably, the authors incorporated the computational tool YEASTRACT and OptForce to identify novel targets for metabolic engineering. As a case study, the shikimate platform strain was demonstrated for the development of the muconic acid cell factory, and more importantly can be readily engineered toward the production of aromatic secondary metabolites.

p-Coumaric acid, an aromatic amino acid derivative, is another precursor metabolite for biosynthesis of plant secondary metabolites, such as polyphenols, flavonoids, and polyketides. Rodriguez et al. established a p-coumaric acid-overproducing yeast platform strain (Li et al., 2015a) by (i) reducing by-product formation via knockout competing pathways (i.e. phenylpyruvate decarboxylase ARO10 and pyruvate decarboxylase PDC5), (ii) eliminating feedback inhibition of pathway enzymes (i.e. DAHP (3-deoxy-D-arabino-heptulosonic acid 7-phosphate) synthase and chorismate mutase), (iii) overexpressing flux controlling enzymes of the aromatic amino acid pathway (i.e. the pentafunctional enzyme ARO1 and bifunctional chorismite synthase-flavin reductase ARO2 as well as their E. coli analogues, arob, aroD, aroO, ydiB, aroK, aroL, aroA, and aroC). Specifically, the highest production of p-coumaric acid (~2 g/L) was achieved by overexpressing tyrosine ammonia-lyase (TAL) from Flavobacterium johnsoniae, DAHP synthase mutant ARQ4220, chorismate mutase mutant ARQ701415, and E. coli shikimate kinase II (aroI) in pdc5Δaro10ΔA yeast strain. In a following study, the metabolic response to over-production of p-coumaric acid was studied in depth by comparing the physiology, transcriptome, and metabolome of the low and high producers. Further engineering of omics-guided targets in the high producing strain resulted in an additional 50% improvement in the final titer (Rodriguez et al., 2017). The p-coumaric acid overproducing strains was established as a good platform strain for the discovery and engineering of p-coumaric acid derived secondary metabolites in S. cerevisiae.

### 5.2. Genome engineering tools

As mentioned above, deregulation of cellular metabolism, i.e. to enhance the supply of precursor metabolites for biosynthesis, represents the biggest challenge for metabolic engineering. To perturb the extensive regulation and complex interactions between metabolic pathways, many metabolic engineering targets should be modified in different modes of regulation, such as introducing heterologous genes and pathways (knock in), increasing expression of rate-limiting enzymes (overexpression), decreasing expression of essential genes (knock down), and removing expression of competing pathways (knock out). Recently, the development of novel genome engineering tools, particularly the CRISPR/Cas based genome editing technology (Cong et al., 2013; Hsu et al., 2014; Mali et al., 2013), makes it possible to modify the yeast genome in a modular, parallel, and high-throughput manner (Lian et al., 2018b) (Fig. 3), which will greatly reduce the time and cost for yeast cell factory development.

#### 5.2.1. Genome editing and transcriptional regulation

Efficient genome editing in yeast can be achieved by combining the CRISPR/Cas induced double strand break and the endogenous homology directed repair (HDR) (Fig. 3A). By introducing the CRISPR/Cas system together with a double-strand oligonucleotide (dsOligo) as the HDR template, DiCarlo et al. achieved 100% knock out efficiency of the CAN1 gene without selection (DiCarlo et al., 2013). Thanks to the modularity and high efficiency of the CRISPR/Cas system, Jakociunas et al. reported multiple gene knock-out by co-transforming gRNAs and dsOligos for editing 5 metabolic engineering targets simultaneously.
with 100% efficiency. The resultant strain accumulated 41-fold more mevalonate (Jakociunas et al., 2015). Recently, Lian et al. attempted multiplex genome editing in polyploid industrial yeast strain, with up to 4 genes disrupted in both a diploid (8 alleles) and triploid (12 alleles) industrial yeast strains in a single step with 100% efficiency (Lian et al., 2018a). Such a multi-copy and multi-loci gene deletion confirmed the power of CRISPR/Cas system for multiplex genome editing.

Similarly, marker-free integration of multiple genes or metabolic pathways can be achieved by simply integrating genes or pathways of interest into the HDR donors. For example, Ronda et al. integrated three 5.1–6.6 kb fragments containing the carotene biosynthesis genes into three pre-characterized loci simultaneously in a single step with an efficiency of 85% without the use of any selection marker (Ronda et al., 2015). CRISPR mediated multiplex genome integration was also demonstrated by multi-copy integration of metabolic pathways (Lian et al., 2016; Shi et al., 2016). A combined xylose utilization and BDO biosynthetic pathway (a total size of 24 kb) was integrated into the yeast genome in up to 18 copies, leading to direct BDO production from xylose.

Most recently, Barbieri et al. reported eMAGE (eukaryotic multiplex automated genome engineering) for multiplex and precise genome editing in S. cerevisiae by engineering the DNA replication fork (Fig. 3B). During DNA replication, synthetic oligonucleotides were annealed at the lagging strand to achieve precise chromosome modifications with high efficiency (~40%). Up to 12 oligonucleotides could be incorporated simultaneously in one transformation. Through iterative transformations of a well-defined pool of oligonucleotides, eMAGE could generate large genomic diversity rapidly (Barbieri et al., 2017). As a proof-of-concept, eMAGE was used to engineer the β-carotene biosynthetic pathway by introducing precise mutations in both structural genes and regulatory elements (promoters and terminators).

Besides genome editing, transcriptional regulation is also a powerful approach and often adopted for metabolic engineering. Flexible and targeted gene knock-down can be achieved by CRISPR interference (CRISPRi) (Lian et al., 2017; Sheng et al., 2017; Smith et al., 2016; Zalatan et al., 2015) (Fig. 3C) and RNA interference (RNAi) (Crook et al., 2014; Si et al., 2015; Williams et al., 2015) (Fig. 3D). CRISPRi is based on the targeted binding of the nuclease deficient CRISPR protein to function as a physical block of transcription initiation and elongation. Because a repressor domain has been found to be necessary for maximal CRISPRi, Lian et al. screened and engineered a panel of repressor domains for improved CRISPRi efficiency in yeast (Lian et al., 2017). The engineered CRISPRi was adopted to repress the expression of ERG9 and MNN9 for improved lycopene production and heterologous protein secretion, respectively. Although RNAi is naturally present in most eukaryotes, it has only been recently reconstituted in S. cerevisiae via heterologous expression of Argonaute and Dicer and further optimized and adopted for metabolic engineering applications (Crook et al., 2014; Si et al., 2015). For example, the optimized synthetic RNAi was used for rapid prototyping of metabolic engineering strategies for itaconic acid production in yeast (Crook et al., 2014).

Besides cDNA overexpression, CRISPR activation (CRISPRa) has been introduced to achieve gene overexpression as well. CRISPRa can be achieved by recruiting an activator domain to the nuclease deficient CRISPR complex (Gilbert et al., 2013; Zalatan et al., 2015) (Fig. 3C). Several transcription activator domains have been tested and the rationally designed tripartite activator, VP64-p65-Rta (VPR), resulted in dramatically improved CRISPRa efficiency in a broad range of organisms, including S. cerevisiae (Chavez et al., 2015). In addition, Lian et al. optimized the combinations of 4 nuclease-deficient CRISPR proteins and 3 activation domains for maximal CRISPRa activity and more flexible metabolic engineering applications (Lian et al., 2017).

5.2.2. Multiple modulation and combinatorial metabolic engineering

As metabolic engineering involves different modes of genetic manipulation of many targets, there is a growing interest in combining gain- and loss-of-function genome engineering modules in the same cell to enable combinatorial optimization. For example, Vanegas et al. developed the SWITCH system to alternate between a genetic engineering state (Cas9 based genome editing) and a pathway control state (dCas9 based transcriptional interference) (Vanegas et al., 2017). Deaneer and Alper established a rapid method named STEPS (Systematically Test Enzyme Perturbation Sensitivities) for fine-tuned and graded expression of pathway enzymes via dCas9-VPR or dCas9-MXI1 regulation (Deaneer and Alper, 2017). STEPS was used to quantify rate limiting enzymes in glycerol biosynthesis and 3-dehydroshikimate production, and the subsequent addressing of the pathway bottlenecks increased the production of glycerol and 3-dehydroshikimate by 5.7-fold and 7.8-fold, respectively. Nevertheless, due to the possible cross-talk between Cas9 and dCas9, only one mode of modification is possible at a time for both SWITCH and STEP. To achieve simultaneous activation and repression within a cell, Deaneer at al repurposed the dCas9-VPR to act as a dual-mode activator/repressor (Deaneer et al., 2017). The dual-mode transcriptional reprogramming was solely dependent on the target position at gene expression cassette. Multiplex modulation of 4 native genes (2 for activation and the other 2 for repression) was demonstrated using dCas9-VPR. Zalatan et al. developed the scaffold RNA (scRNA) system to achieve orthogonal transcriptional activation and interference, which was demonstrated for redirecting the metabolic fluxes of the violacein biosynthetic pathway in yeast (Zalatan et al., 2015).
Recently, Lian et al. developed an orthogonal tri-functional CRISPR system that combines transcriptional activation, transcriptional interference, and gene deletion (CRISPR-AID) in the same cell (Lian et al., 2017). As a proof-of-concept metabolic engineering study, CRISPR-AID increased β-carotene production via simultaneous upregulation of HMG1, downregulation of ERG9, and deletion of ROX1. CRISPR-AID was also applied for combinatorial optimization of several metabolic engineering targets (14 targets for CRISPRa, 17 targets for CRISPRi, and 5 targets for CRISPRd) to enhance the expression and display of a recombinant protein on the yeast surface. Besides the CRISPR system, Si et al. combined cDNA overexpression and RNAi for simultaneous upregulation and downregulation of metabolic engineering targets (Si et al., 2017). Moreover, based on the design of the synthetic oligonucleotides, eMAGE could introduce multiple mutations into promoters, coding sequences, and terminators simultaneously and create combinatorial genomic diversity for metabolic engineering (Barbieri et al., 2017).

5.2.3. Genome-scale metabolic engineering

Although yeast is one of the most well studied microorganisms, there is still a lack of a clear and thorough understanding of the whole metabolic and regulatory networks. In previous metabolic engineering efforts, some unknown or unrelated targets increased the desired phenotype the most (Caspeta et al., 2014; Kim et al., 2013c). Therefore, it is highly desirable to use genome-scale metabolic engineering to cover all the possible modifications. Genome-scale cDNA overexpression library was firstly constructed (Liu et al., 1992) and used to study genes involved in lethality (Liu et al., 1992), cell cycle progression (Stevenson et al., 2001), and improved protein secretion and display levels (Wentz and Shusta, 2007).

For the genome-scale loss-of-function libraries, the Yeast Knock-out (YKO) collection (Giaever et al., 2002) was created for non-essential genes while the decreased abundance by the mRNA perturbation (DAmP) library (Breslow et al., 2008) was constructed for essential genes. The YKO collection contains a yeast strain library with every nonessential gene deleted individually, while the Damp library covers ~82% essential genes with mRNA abundance decreased by disrupting the 3′-untranslated region (UTR). Although yeast strain libraries have been widely used for functional genomics and provided invaluable knowledge about numerous important biological processes, the Synthetic Genetic Array approach requires multiple steps of manipulation (Tong et al., 2001), which makes it rather challenging to introduce genome-wide perturbation iteratively for metabolic engineering applications. To address these challenges, Si et al. developed RNAi assisted genome evolution (RAGE) to iteratively introduce genome-wide modifications in customized genetic backgrounds (Si et al., 2015). RAGE has been successfully used to increase the tolerance to both acetate and furfural (Si et al., 2015) and furfural (Xiao and Zhao, 2014). Iterative rounds of genome-scale screening identified knockdown targets (PTC6, YPR084W, and RNAV1A2C) that acted synergistically to improve acetate acid tolerance (Si et al., 2015).

RAGE was further developed by combining RNAi and cDNA overexpression libraries, in which both gain- and loss-of-function modifications required for metabolic engineering can be created simultaneously at the whole genome scale (Si et al., 2017). This combinatorial metabolic engineering approach has been proved to be effective in engineering various complex phenotypes, such as protein and chemical production, substrate utilization, and inhibitor tolerance. More importantly, by taking advantage of the standardized modulation parts used for genome-wide perturbation, the process of creating and screening genome-scale libraries was automated using an integrated robotic system. Such automated multiplex genome-scale engineering will be an indispensable tool in constructing and optimizing yeast cell factories.

Most recently, Bao et al. developed a genome-scale engineering method named CRISPR/Cas9 and homology-directed repair assisted genome-scale engineering (CHAnGE) (Bao et al., 2018). CHAnGE took advantages of CRISPR/Cas9 for precise and efficient genome editing and HI-CRISPR design to physically link the CRISPR guide sequence with the corresponding HDR donor. CHAnGE could readily create a genome-wide knockout library, which was used for directed evolution of yeast genome in an iterative manner. As a proof-of-concept, CHAnGE was used to evolve yeast strains with improved furfural and acetic acid tolerance. After 2 rounds of genome-scale directed evolution, furfural and acetic acid tolerance were increased by ~42- and ~20-fold, respectively. In addition, CHAnGE was used to perform genome editing at a single-nucleotide resolution, which enables direct functional study of large scale genetic variants of an endogenous gene.

5.2.4. Adaptive evolution and reverse metabolic engineering

Although rational pathway design and genome modification has greatly advanced metabolic engineering, random mutagenesis based adaptive evolution also plays an important role in engineering complex phenotypes, particularly for those coupled to cell growth. Adaptive evolution has been demonstrated for improving growth on non-preferred carbon sources and under harsh industrial conditions. For examples, adaptive evolution has been applied to increase the utilization of alternative sugars from renewable sources, such as cellbiose (Oh et al., 2016) and xylose (Kim et al., 2013c) from lignocellulose, galactose from red algae (Hong et al., 2011; Lee et al., 2011), and mannitol and 4-deoxy-L-erythro-5-hexoseulose urinate (DEHU) from brown algae (Enquist-Newman et al., 2014). Similarly, yeast strains have been engineered using adaptive evolution for improved tolerance to high temperature (Caspeta et al., 2014), inhibitors in raw materials (Almario et al., 2013), and high concentration of ethanol (Stanley et al., 2010), alkanes (Ling et al., 2015), and replacement jet fuels (Brennan et al., 2015).

The development of next-generation sequencing and systems biology tools has enabled the identification of causal mutations. Nevertheless, the molecular mechanisms of the evolved phenotypes are still elusive, due to the presence of multiple mutations (Mans et al., 2017). CRISPR/Cas system mediated reintroduction of multiple point mutations to various genomic loci of the non-evolved strain (reverse metabolic engineering) is an invaluable tool for evolutionary engineering (Mans et al., 2015). Besides the elucidation of the genetic basis for complex phenotypes, reverse metabolic engineering can also minimize evolutionary trade-offs by removing undesirable mutations. For example, the evolved thermostolerant yeast strains showed decreased growth and biomass yield when growing at ancestral temperatures. The re-created strain with a point mutation at ERG3 locus not only maintained most of the evolutionary advantages at high temperature, but also restored the growth fitness at lower temperatures (Caspeta et al., 2014).

5.3. Systems biology tools

Mathematical models of yeast metabolism have also been increasing utilized to not only explore shifts in metabolic states under environmental or genetic perturbations, but also design strategies for strain or process engineering (Kerkhoven et al., 2015; Osterlund et al., 2012). With recent advances in data gathering technologies, it is now possible to characterize the performance of mutant strains at various levels of biological processes – namely gene expression levels, protein and metabolite concentrations, and metabolic fluxes. The multi-omics datasets can not only be interpreted but also conveniently integrated with each other in appropriate mathematical frameworks, a few of which are discussed below.

5.3.1. Transcriptomics

Transcriptomics is one of the commonly used analytical techniques with high-throughput omics and involves the measurement of all mRNA transcripts in a cell. Transcriptomics is used to identify genes that are
differentially expressed under various scenarios (McGettigan, 2013; Wang et al., 2009). Transcriptomics studies are often valuable in identifying gene regulatory networks in a host organism, which can help motivate new strategies for rewiring cellular regulation. Transcriptomics analysis has been used extensively to study and engineer *S. cerevisiae* tolerance to various compounds; some examples include l-imonene (Brennan et al., 2015), acetic acid and furfural (Chen et al., 2016a), lignocellulosic derived inhibitors (Thompson et al., 2016), and a mixture of acetic acid, furfural and phenol (Yang et al., 2012). Transcriptomics analyses have also been used to analyze important factors affecting xylose utilization by *S. cerevisiae*. In one such study, three different xylose-utilizing pathways were expressed in two *S. cerevisiae* hosts and RNA-seq analysis employed to identify the role played by host regulation on the heterologous pathways (Feng and Zhao, 2013b). In another study involving a xylose-consuming *S. cerevisiae* strain (Ismail et al., 2013), the authors identified via transcriptomics analyses various genes that have heat-sensitive expression, which could possibly limit the temperature ranges of lignocellulosic fermentation. Transcriptomics analyses have also been used to identify the genes that control increased protein secretion in yeast (Gasser et al., 2007). Comparative analysis of a human trypsin overexpressing *Pichia pastoris* strain with a nonexpressing strain led to the identification of 13 differentially upregulated genes in the overexpression strain; cloning these genes in a separate *P. pastoris* strain resulted in successful protein secretion, showcasing the benefits of this analytical technique in metabolic engineering.

### 5.3.2. Proteomics

Increased development in the multi-omics ideology has also resulted in a significant thrust in the field of proteomics, which measures the protein concentrations of an organism. mRNA levels are often used as a representation of the protein expression levels, but they are only a proxy for actual protein concentrations due to variations in factors such as ribosome binding strength, codon usage frequency, and protein half-lives (Maier et al., 2009). Dynamic variations of cellular protein concentrations, including metabolic enzyme concentrations, can be analyzed for improved understanding of metabolic behavior including changes in regulatory interactions. Thus, proteomics serves as a powerful tool for understanding and modifying an organism’s metabolism.

In the context of yeast engineering, proteomics has been steadily catching on as a powerful tool for analysis and rational design. The first and foremost requirement in proteomics is the development of reproducible protocols for sample preparation and mass spectrometry analysis of yeast proteins. de Godoy et al. performed a highly comprehensive comparison of the protein levels between haploid and diploid *S. cerevisiae* strains by using mass spectrometry-based quantitative proteomic measurements (de Godoy et al., 2008). Stable-isotope labeling of amino acids was used, which resulted in a coverage of 4399 proteins. In other studies, a new Orbitrap hybrid mass spectrometer coupled with liquid chromatography was used to detect 3977 *S. cerevisiae* proteins with a low false discovery rate of 1%, implying nearly a whole-proteome coverage (Hebert et al., 2014; Richards et al., 2015). SWATH-MS is another protocol based on selected reaction monitoring mass spectrometry, which was used to analyze the *S. cerevisiae* proteome and quantified more than 2500 proteins (Selesevk et al., 2015). Paulo et al. analyzed the variations in *S. cerevisiae* proteome due to growth on different carbon sources and discovered that the majority of proteins with significantly altered expression belonged to the class of metabolic enzymes or those localized in the plasma membrane or mitochondria while the unaltered proteins were involved in nuclear functions and localization (Paulo et al., 2015). Matsuda et al. performed targeted proteome analyses of single deletion knockout mutants from the central carbon metabolism to explore the regulatory processes controlling protein abundances in *S. cerevisiae*. The authors discovered that enzyme level changes in glycolysis and trehalose pathways occurred in a coordinated manner under the control of global regulators (Matsuda et al., 2017). In order to achieve high levels of fatty alcohol production in *S. cerevisiae*, D’Espaux et al. used shotgun proteomics as a first level analytical tool along with 13C MFA to identify protein abundances in the reference strain along with potential bottlenecks in the distribution of flux (D’Espaux et al., 2017). A recent study employed quantitative proteomics along with transcript measurements to explore correlations between mRNA and protein abundances in *S. cerevisiae* – mRNA abundances and elongation of translation were highlighted as major factors in influencing protein translation efficiency (Lahtvee et al., 2017). The combined use of transcriptomics and proteomics in this study signals a beneficial shift towards integrative -omics for metabolic engineering. Proteomics is, thus, gradually being incorporated as a tool for systems-level analysis of many biological systems, including *S. cerevisiae*, and subsequent metabolic engineering. Two statistical tools have also been developed that explore correlations among proteomics data and metabolite measurements to identify bottlenecks in metabolic pathways for the purpose of metabolic engineering (Alonso-Gutierrez et al., 2015; George et al., 2014).

### 5.3.3. Genome-scale models

One of the more commonly used frameworks for metabolic modeling, genome-scale models (GEMs) require stoichiometric information for all the reactions specified. Several GEMs have been published for *S. cerevisiae* metabolism (Aung et al., 2013; Duarte et al., 2004; Forster et al., 2003; Heavner et al., 2012; Mo et al., 2009; Nookaew et al., 2008). The GEMs differ in aspects such as coverage of genes and reactions, steady state predictions, and gene essentiality and have been compared in much greater detail elsewhere (Heavner and Price, 2015; Sanchez and Nielsen, 2015). Multiple studies have employed gene expression data for integration with GEMs to improve the accuracy of their prediction. Akesson et al. incorporated a method wherein genes whose mRNA transcripts did not show up in microarray studies were assumed to be inactive, and their corresponding flux values in the yeast GEM were set to zero. The study showed an improvement in the model’s predictions both for product yields as well as some intracellular fluxes (Akesson et al., 2004). Another study developed an algorithm to integrate metabolomics data from various environmental and genetic conditions with GEMs to identify reporter reactions that are more significantly perturbed than the others. The study incorporated metabolomics data from wild-type and engineered *S. cerevisiae* strains grown under aerobic and anaerobic conditions and identified many amino acid biosynthesis pathways as significantly affected reactions across the different growth conditions (Cakir et al., 2006). More recently, Guo et al. presented an algorithm that integrates multi-omics data such as transcriptomics with a GEM to construct a new objective function that can predict cellular phenotypes more accurately. The authors demonstrated its validity by accurately predicting ethanol yields for *S. cerevisiae* in more than 80% of the test cases (Guo and Feng, 2016).

GEM-guided metabolic engineering has been employed in many studies for overproduction of target chemicals. Bro et al. employed the use of a yeast GEM to test various possible strategies in *siliico* for improving ethanol yield. Model predictions, followed by *in vivo* testing showed that the heterologous expression of a non-phosphorylating NADP+-dependent glyceraldehyde-3-phosphate dehydrogenase resulted in a 40% decrease in flux towards glycerol and was accompanied by a 3% increase in flux towards ethanol (Bro et al., 2006). To improve the production of sesquiterpenes in *S. cerevisiae*, Asadollahi et al. studied the effect of knockout of the NADPH-dependent glutamate dehydrogenase (GDH1) gene coupled with overexpression of the NADH-dependent version of the same enzyme (GDH2), predicted by the OptGene algorithm (Patil et al., 2005), in a yeast GEM (Asadollahi et al., 2009). The strategies were validated in batch fermenters, where it was shown that GDH1 knockout resulted in an 85% increase in the final sesquiterpene titer, compared to a 10-fold increase as predicted by *in silico* predictions. GDH2 overexpression was performed to recover the knockout-induced decrease in specific growth rate. Brochado et al.
engineered an *S. cerevisiae* with vanillin production using predictions from the OptGene algorithm (Brochado et al., 2010). Specifically, knockout of the pyruvate decarboxylase gene (*PDC1*) resulted in a titer of 500 mg/L, representing a 5-fold increase from previously reported titers. In another model-guided study for overproduction of fumaric acid in *S. cerevisiae*, Xu et al. discovered that deletion of fumarase (*FUM1*), overexpression of a heterologous pyruvate carboxylase gene from *Rhizopus oryzae* (RoPYC) and overexpression of a succinate-fumarate transporter gene (*SFC1*) resulted in a final titer of 1.6 g/L, compared to a reference strain that produced no detectable fumaric acid (Xu et al., 2012). A great benefit of using genome-scale models to design engineering strategies is the possibility of discovering non-intuitive targets that need not have activity in the local upstream or downstream pathways of the target chemical. This advantage was showcased by a study employing a yeast GEM for overproducing itaconic acid in *S. cerevisiae* (Blazeck et al., 2014). In a sequential fashion, a triple knockout strain of *ade3Δ bna2Δ tes1Δ* was constructed which demonstrated a 7-fold increase in itaconic acid titer to 168 mg/L when compared to the reference strain only carrying the itaconic acid pathway. GEM-aided engineering studies have also been employed to construct *S. cerevisiae* strains overproducing succinic acid (Agren et al., 2013), 3-hydroxypropionic acid (Borodina et al., 2015), and tyrosine (Gold et al., 2015).

Genome-scale models can be used for exploring other biological processes besides enzymatic reactions as well. Feizi et al. developed a GEM that integrated all the known annotated processes involved in the yeast protein secretory machinery (Feizi et al., 2013). The model stores information in the form of a Protein Specific Information Matrix (PSIM), where the rows represent proteins and columns represent seven specific features of post-translational modification, such as signal peptide for ER localization, number of N-linked and O-linked glycosylation sites and number of disulfide bonds among others. To explore the reasons behind phenomena such as the Crabtree effect, which is the preference of some yeast species such as *S. cerevisiae* to switch to fermentation instead of respiration in the presence of high glucose concentrations, Nilsson and Nielsen also employed the use of GEM flux predictions to formulate hypotheses (Nilsson and Nielsen, 2016). The use of genome-scale models for metabolic engineering purposes have also been reviewed elsewhere (Blazeck and Alper, 2010; Borodina and Nielsen, 2014; Patil et al., 2004).

5.3.4. Metabolic flux analysis (MFA)

Metabolic flux analysis, i.e. use of labelled ¹³C substrates coupled with mass spectrometry or nuclear magnetic resonance spectrometry to determine intracellular metabolic fluxes, is a useful analytical tool for metabolic engineering. A better insight into the redirecting of metabolic fluxes can help inform better engineering strategies to repurpose cellular flux towards a target product. Genome-scale ¹³C flux analysis was used to analyze the robustness of cellular flux in *S. cerevisiae* strains containing various gene knockouts, highlighting the network redundancy either via duplicate genes or alternative pathways (Blank et al., 2005a). A study on an arabinofermenting *S. cerevisiae* utilized metabolomics, transcriptomic, and metabolic flux measurements to identify the importance of a galactose transporter for arabinose consumption (Wisselink et al., 2010). The ¹³C MFA technique also employed ¹³C flux states across different strains and conditions, and this knowledge can be used to design better strain engineering strategies. In another study to engineer a free fatty acid-overproducing yeast strain, Ghosh et al. employed a modified version of the ¹³C MFA technique known as the two-scale ¹³C MFA (2S⁻¹³C MFA) (Martin et al., 2015), which employs flux constraints calculated from labeling data instead of evolutionary objective functions to calculate flux values. The study focused on improving production of acetyl-CoA, a precursor for fatty acid pathways, and devised strategies based on 2S⁻¹³C MFA flux predictions (Ghosh et al., 2016). To achieve this goal, the authors introduced a heterologous ACL gene from *Y. lipolytica* for increased acetyl-CoA synthesis, and downregulated malate synthesis along with a deletion of glyceral-3-phosphate dehydrogenase to avoid side reactions that consume acetyl-CoA, resulting in a 70% increase in the fatty acid titer to 780 mg/L. In a study on protein secretion in the yeast *Schizosaccharomyces pombe*, Klein et al. employed ¹³C MFA to identify the key flux changes that accompany the metabolic burden of elevated protein secretion, in this case the protein alpha-glucosidase maltase (Klein et al., 2014). ¹³C MFA has also been used to compare metabolic flux states across different yeast species (Blank et al., 2005b; Jung et al., 2017). The applications of ¹³C MFA in eukaryotic species and its utility in rational engineering have been addressed in other reviews (McAtee et al., 2015; Niklas et al., 2010). The implementation of ¹³C MFA at the genome-scale and the challenges associated with it are addressed elsewhere (Gopalakrishnan and Maranas, 2015).

5.3.5. Kinetic models

Kinetic models of cellular metabolism are based on ordinary or partial differential equations to predict the dynamics of metabolism. The biggest drawback of these models, however, is significant uncertainty in parameter estimates leading to uncertainty in such a model’s predictions as well. For these reasons, kinetic models, while promising in theory, have not been utilized in metabolic engineering efforts as widely as GEMs.

Construction of accurate kinetic models of *S. cerevisiae* is an arduous task due to a variety of reasons – a) compartmentalization is difficult to model as the metabolite fractions as well as transport across compartments are difficult to ascertain, and b) regulatory interactions in yeast and other eukaryotes are not as well understood as in *E. coli*, resulting in models that do not fully capture the dynamic processes. However, there has been a steady increase in work on yeast kinetic models. Smallbone et al. developed a yeast glycolysis model by employing an iterative strategy for *in vitro* measurements of kinetic rate constants such as *k_m* and *K_m* of the enzyme with the greatest flux control within the model at each iteration (Smallbone et al., 2013). A pathway-level model has also been constructed for metabolism and regulation in galactose metabolism (Mitre et al., 2016). Multiple research groups have developed pathway models on yeast lipid metabolism. Alvarez-Vasquez et al. detailed a model on *S. cerevisiae* sphingolipid metabolism that served as a platform for integration of experimental data as well as for predicting the outcome of metabolic perturbations that are difficult to perform experimentally (Alvarez-Vasquez et al., 2005). Savoglidis et al. also developed a curated kinetic model on *S. cerevisiae* sphingolipid metabolism that was used to demonstrate the utility of the Inverse Metabolic Control Analysis (IMCA) method, which uses metabolite changes at the network level to predict changes in protein expression levels (Savoglidis et al., 2016).

Early work was performed in yeast model development where rapid sampling of growth cultures enabled the measurement of metabolite data, which was employed in parameter estimation for a kinetic model that included the glycolysis and TCA pathways (Rizzi et al., 1997; Theobald et al., 1997). A workflow published by Stanford et al. detailed the construction of kinetic models from genome-scale models (GEM) and demonstrated its application on a large-scale *S. cerevisiae* model (Stanford et al., 2013). The workflow integrated the stoichiometric model with flux and metabolite measurements along with geometric and thermodynamic constraints, while employing the use of generalized reversible Michaelis-Menten rate kinetics. Recently, a study utilized large-scale kinetic models of *S. cerevisiae* in a design-build-test cycle to improve xylose uptake (Misiołek et al., 2017). A population of
models was fit to flux and metabolite data generated from a fermentation reaction. One round of iterative experiments, which involved deleting the hexokinase HXK2, enabled the pruning of models that predicted incorrectly, leading to better predictions from the remaining ensemble of models. The study reported a 60% increase in xylose uptake rates as a result of the gene deletion motivated from the population of models, underlining the role that mathematical models can play in the design-build-test cycle that are a part of metabolic engineering. To address the knowledge gaps in regulatory information, Hackett et al. developed the systematic identification of meaningful metabolic enzyme regulation (SIMMER) workflow that utilizes flux, metabolite and protein measurements under different growth conditions and ascertains the best fit from a variety of reaction kinetics models (Hackett et al., 2016). Substrate-level regulations were identified by assessing the data fit in reaction kinetics equations that accounted for different regulators. A Bayesian approach was employed that incorporated prior knowledge of regulation in *S. cerevisiae* along with conditional probabilities calculated using the experimental data, to determine the probability of each putative regulation identified by the workflow. As a result, three previously unknown regulatory interactions were identified (Fig. 4). Efforts towards construction of kinetic models for application in metabolic engineering have resulted in a variety of population modeling-based methods that utilize the predicted ensemble average for strain engineering purposes (Chakrabarti et al., 2013; Khodayari and Maranas, 2016; Tran et al., 2008). While the efficacy of these methods have mainly been displayed in *E. coli* metabolism, their organism-

---

**Fig. 4.** Summary of the SIMMER method which can be used to identify potential (A) inhibitors and (B) activators of a metabolic enzyme by measuring reaction fluxes, protein abundances, and metabolite concentrations across different growth conditions and (C) applying the Bayes theorem to incorporate prior knowledge on metabolite regulation with conditional probability from experimental data (fluxes, proteins, and metabolites) to yield probability of a putative regulation identified by the method. The SIMMER method thus explores different metabolites that potentially influence allosteric regulation over an enzyme, and identifies the candidate that most accurately explains the measured flux data to be a possible regulator of the enzymatic reaction.

**Fig. 5.** Construction of yeast strains for simultaneous utilization of lignocellulosic sugars, such as cellobiose, xylose, and acetate. Intracellular release of glucose alleviates glucose repression to enable xylose and cellobiose co-utilization. Acetate reduction consumes excess NADH generated during xylose assimilation. XR: xylose reductase; XDH: xylitol dehydrogenase; A-ALD: acylating aldehyde dehydrogenase; ADH: alcohol dehydrogenase; CDT: cellobextrin transporter; BGL: β-glucosidase. Overexpressed genes are shown in red and co-factors in blue.
agnostic nature makes them an effective tool for predicting yeast metabolism as well. Population modeling offers a practical way to circumnavigate the issue of uncertainty in parameter estimation and is, thus, a promising direction for model-guided metabolic engineering.

Despite the inherent difficulties in constructing kinetic models and utilizing their predictions, advances in model construction, parameter estimation, and uncertainty analysis can make this genre of models a strong tool in rational metabolic engineering. The benefits of their use and according challenges have been reviewed in significant detail elsewhere (Almqquist et al., 2014; Srinivasan et al., 2015).

6. Metabolic engineering examples

6.1. Co-utilization of lignocellulosic sugars

As the most abundant biomaterials on earth, lignocellulosic biomass is one of the most promising feedstock for renewable production of fuels and chemicals. Nevertheless, efficient and simultaneous utilization of all carbon sources remains the biggest challenge for cellulose biofuel industry (Fig. 5). Although metabolic engineering enabled the construction of efficient xylose utilizing yeast strains (Kim et al., 2013c), xylose will not be consumed until glucose is depleted, a phenomenon known as glucose repression. The sequential sugar fermentation resulted in low biofuel productivity. Cellobiose fermentation via the introduction of a celloextrin transporter and intracellular β-glucosidase was proved to be an effective strategy to alleviate glucose repression, because the gradually released glucose was consumed rapidly. The resultant yeast strain could consume cellobiose and xylose simultaneously (Ha et al., 2011; Li et al., 2010) (Fig. 5). Another challenge is the presence of high concentration of inhibitors in the lignocellulosic hydrolysate, such as acetate. An acetate reduction pathway was firstly introduced to balance cellular redox. By combining an NADH-consuming acetate consumption pathway and an NADH-producing xylose utilization pathway, xylose and acetate were simultaneously converted into ethanol with higher yields (Wei et al., 2013) (Fig. 5). In other words, acetate reduction not only achieved in situ detoxification, but also increased biofuel yield. Recently, these two strategies were combined to enable simultaneous utilization of cellobiose, xylose, and acetate (Wei et al., 2015) (Fig. 5).

6.2. Efficient production of 3-HP in yeast

3-HP is another important platform chemical. Considering the tolerance to relatively low pH, S. cerevisiae is an attractive host for producing organic acids, including 3-HP. Currently, two pathways have been introduced and engineered to enable efficient 3-HP production in yeast (Fig. 6), the malonyl-CoA reductase pathway (Chen et al., 2014b) and the β-alanine pathway (Borodina et al., 2015). The bi-functional malonyl-CoA reductase from Chloroflexus aurantiacus (CaMCR) converted malonyl-CoA to 3-HP directly. Because malonyl-CoA metabolism is highly regulated in yeast, its supply is considered as the bottleneck for efficient 3-HP production. Through overexpression of ADH2, ALD6, AcL2441p, and ACC1 as well as deletion of MLS1, 3-HP production was increased by ~3-fold (~270 mg/L). When coupled with cofactor engineering via overexpression of an NADP+−dependent G3P dehydrogenase gene (GAPN), 3-HP production was further increased to 463 mg/L (Chen et al., 2014b). ACC1, acetyl-CoA carboxylase, is subject to SNF1-mediated inactivation. Shi et al. increased ACC1 activity by mutating the potential phosphorylation sites to abolish the post-translational regulation. Overexpression of the ACC1 mutant (ACCl− er (Sel56ΔA, Ser1157AΔM) produced ~2.2-fold more 3-HP than that of the wild-type ACC1 (Shi et al., 2014a). By combining acetyl-CoA engineering, ACC1 mutant overexpression, and multi-copy genome integration, 3-HP was produced at a titration of ~10 g/L in a carbon-limited fed-batch fermentation (Kildegaard et al., 2016). To better manipulate malonyl-CoA levels, David et al. developed a hierarchical dynamic control system. The upper level of control was to dynamically down-regulate fatty acid biosynthesis using HXT1 promoter. The lower level was based on the malonyl-CoA biosensor, where the expression level of CaMCR was dependent on the intracellular malonyl-CoA concentration. The dual dynamic pathway control increased the production of 3-HP by 10-fold (more than 1 g/L) in yeast (David et al., 2016). Li et al. developed a similar biosensor based on the bacterial TF FapR to monitor intracellular malonyl-CoA concentration. The malonyl-CoA biosensor was combined with the genome-scale cDNA overexpression library to identify two novel gene targets (PMP1 and TP1) that increased intracellular malonyl-CoA levels and the subsequent production of 3-HP (Li et al., 2015b).

Another route for 3-HP production in yeast is via β-alanine, which was identified as the most economically attractive route based on the whole genome-scale metabolic modeling. By testing different enzyme combinations, Bacillus cereus β-alanine-pyruvate aminotransferase YHXA (BcBAPAT) and E. coli 3-hydroxypropanoate dehydrogenase YDFG (EcYDFG) were the most efficient enzymes to convert β-alanine to 3-HP. To enable β-alanine biosynthesis from aspartate, aspartate-1-decarboxylase with different origins were tested in yeast and the enzyme from Tribolium castaneum (TcPAND) was found to be the most active. Native pyruvate carboxylases (PYC1 and PYC2) and cytoplasmic aspartate aminotransferase (AAAT2) were overexpressed to enhance the supply of aspartate from glucose. The optimal strain overexpressing AAAT2, PYC1, PYC2, BcBAPAT, EcYDFG, and multiple copies of TcPAND produced 12.2 g/L and 9.2 g/L of 3-HP from glucose in fed-batch fermentation at pH5 and pH3.5, respectively (Fig. 6) (Borodina et al., 2015).

Besides the rational metabolic engineering efforts to construct efficient 3-HP producers, yeast strains with high 3-HP tolerance has been evolved, with the molecular mechanism being elucidated (Kildegaard et al., 2014). Together with the recently developed malonyl-CoA biosensors (David et al., 2016; Li et al., 2015b) and genome-scale engineering tools (Bao et al., 2018; Si et al., 2017, 2015), large-scale production of 3-HP at low pH using yeast cell factories can be achieved in near future.

6.3. Terpenoid production in yeast: artemisinin precursor and farnesene

Terpenoids are a large family of natural products with broad applications as (precursors of) biofuels, perfume ingredients, and pharmaceuticals. Large-scale production of artemisinin precursor (Paddon et al., 2013; Westfall et al., 2012) and farnesene (Meadows et al., 2016) represent the most successful metabolic engineering and synthetic biology examples in yeast cell factories. The general strategies of metabolic engineering of terpenoid production in yeast are (i) to increase the mevalonate pathway fluxes by overexpressing tHMGR (Ro et al., 2006), UPC2-1 (Ro et al., 2006), and the mevalonate pathway genes (ERG10, ERG13, tHMGR, ERG12, ERG8, ID1, and ERG20) (Paddon et al., 2013); (ii) to downregulate ERG9 expression by replacing the endogenous promoter with MET3p (methylene repressible promoter) (Ro et al., 2006), CTR3p (copper repressible promoter) (Paddon et al., 2013), or HXT1p (glucose inducible promoter) (Xie et al., 2015b); (iii) to knock out some regulators or uncharacterized targets (i.e. ROX1, YIL064W, and YPL062W), which were screened from the YKO Collection for enhanced mevalonate pathway fluxes and carotenoid production (Guzaydin et al., 2012) (Fig. 7).

Ro et al. firstly reported the production of amorpha-4,8-diene in yeast. Through overexpression of tHMGR, UPC2-1, and ERG20, down-regulation of ERG9, and introduction of a heterologous amorpha-4,8-diene synthase (ADS) gene, amorpha-4,8-diene was produced at a titer as high as 800 mg/L in shake flasks (Ro et al., 2006). Westfall et al. followed a similar strategy and integrated additional copies of the mevalonate pathway genes into the S. cerevisiae CEN.PK2 strain, resulting in the highest reported production of amorpha-4,8-diene (~40 g/L) in a well-controlled fed-batch bioreactor (Westfall et al., 2012) (Fig. 7).
Meadows et al. further increased terpenoid production in yeast by precursor supply and redox balance engineering. By replacing the native acetyl-CoA biosynthesis (acs1A acs2A ald63s) with A-ALD and PK/PTA, knocking out glycerol 3-phosphate phosphatase (rth2A) to minimize acetae accumulation, acetyl-CoA supply was significantly enhanced with a reduced ATP requirement. In addition, an NADH dependent HMGFR from *Silicibacter pomeroyi* was introduced to balance cofactor usage for terpenoid biosynthesis from glucose. By combining the strategies mentioned above, commercial production of farnesene (≈150 g/L) has been achieved (Meadows et al., 2016) (Fig. 7).

### 6.4. Production of fatty acid derivatives in yeast

Fatty alcohols such as hexadecanol can be used for detergents, surfactants, and cosmetics. By overexpression of the endogeneous ACCI and fatty acid synthase (FAS) genes and the heterologous mouse fatty acyl-CoA reductase (FAR) gene, Runguphan and Keasling achieved a fatty alcohol titer of nearly 100 mg/L in *S. cerevisiae* (Runguphan and Keasling, 2014). In a related work, Feng et al. expressed an FAR gene from the barn owl, *Tyto alba*, along with ACCI overexpression and a hexadecanol titer of 71 mg/L. Along with the heterologous expression of ACL genes (ACL1 and ACL2), the authors further explored some of the regulatory factors affecting lipid synthesis, and found that the knockout of RPD3, encoding a negative regulator of lipid synthesis, resulted in a titer of 330 mg/L (Feng et al., 2015). Zhou et al. recently engineered a *S. cerevisiae* strain for the overproduction of free fatty acids along with alkanes and fatty alcohols (Zhou et al., 2016b). The genetic modifications included deletion of FAA1 and FAA4 to remove negative feedback by fatty acyl-CoA, POX1 deletion to remove fatty degradation by the β-oxidation pathway, removal of the aldehyde dehydrogenase gene HF1D, heterologous expression of a thioesterase gene from *E. coli*, heterologous expression of various genes for increasing cytosolic acetyl-CoA concentrations, a heterologous fatty acid synthase for increased biosynthesis efficiency, and a TEF1-promoter controlled acetyl-CoA carboxylase overexpression to increase the malonyl-CoA supply, another precursor in the fatty acid pathway. These combined modifications led to a titer of free fatty acids of 10.4 g/L in fed-batch fermentation. Common strategies for synthesis of fatty acids and derived compounds have been reviewed in much greater detail (Fernandez-Moya and Da Silva, 2017).

### 6.5. Biosynthesis of aromatic amino acid derivatives

Recently, *S. cerevisiae* has also been used to synthesize valuable chemicals that serve as precursors for polymers or drug molecules. One such category of chemicals is the class of aromatic amino acids. As described in Section 4.1.3, Suastegui et al. incorporated a mix of model-guided and experimentally-validated strategies to overproduce shikimic acid with a final titer as high as 2.5 g/L (Suastegui et al., 2017). Curran et al. engineered a *S. cerevisiae* strain to produce the valuable bio-plastic monomer muconic acid with a final titer of 141 mg/L (Curran et al., 2013b). The resulting strain involved three heterologous genes that diverted flux from the shikimate pathway towards muconic acid, along with several engineering modifications to enhance the flux towards the diverting HMGR-CoA Reductase; SpHMGR: NADH-dependent HMG-CoA Reductase from *Silicibacter pomeroyi*; FS: farnesene synthase; ADS: amorphadiene synthase; UPC2-1: constitutively active UPC2 mutant; Heterologous genes (including endogenous gene mutants) are shown in red and co-factors in blue.

![Fig. 6. Metabolic engineering of 3-HP production in yeast via malonyl-CoA and β-alanine](image-url)
desired product. Rodriguez et al. constructed a yeast strain for production of p-coumaric acid, another important precursor for the biosynthesis of important secondary metabolites (Li et al., 2015a). The resulting strain involved knockouts of phenylpyruvate decarboxylase and pyruvate decarboxylase to block the synthesis of aromatic alcohols, removal of feedback inhibition of the enzymes DAHP synthase and chorismate mutase, and heterologous overexpression of tyrosine ammonia-lyase and shikimate kinase. Previously, S. cerevisiae strains were engineered to convert a high value precursor such as tyrosine to resveratrol, a nutraceutical (Shin et al., 2012; Wang et al., 2011). A recent study showed resveratrol could also be produced de novo from either glucose or ethanol by a combination of heterologous expression as well as host engineering, and a final titer of 415.65 and 531.41 mg/L on glucose and ethanol respectively was reported (Li et al., 2015a). The metabolic pathways involved in the aromatic amino acid derivatives described above have been illustrated in Fig. 8. In another study, Galanie et al. presented a list of more than 20 enzymes required for overexpression in S. cerevisiae for production of opioids, a highly valuable class of pain management medicines. While the study reported very low titers of the products, it represents a significant step in the direction of engineering S. cerevisiae to produce opioid drugs (Galanie et al., 2015).

7. Conclusions and future perspectives

Although significant progress has been made in the past decade, metabolic engineering efforts are still subject to different degrees of success, probably due to the complex yet robust metabolic and regulatory networks. Genome-scale engineering that can perturb all the ~6000 genes in S. cerevisiae at once provides a new direction of metabolic engineering. Although genome-scale screening based on CRISPR knockout (Bao et al., 2018; Shalem et al., 2014), CRISPR interference (Gilbert et al., 2014), and CRISPR activation (Gilbert et al., 2014; Konermann et al., 2015) have been demonstrated, their metabolic engineering applications have not been fully explored yet. Because HDR is dominant in yeast, co-transformation of gRNA and HDR donor targeting the same gene is nearly impossible at the whole genome scale. The Hi-CRISPR design, where HDR donor and gRNA were coupled in the same vector, provides a solution (Bao et al., 2018, 2015). This design was also used for genome-scale engineering in E. coli (Garst et al., 2017). More importantly, by combining genome-scale engineering (Bao et al., 2018) and the multi-functional CRISPR system (Lian et al., 2017), the most comprehensive library that can control the expression of every gene in the yeast genome to various levels can be created for metabolic engineering applications.

However, the development of genome-scale metabolic engineering is accompanied with the challenge to find the best mutant from millions of possibilities. Currently, genome-scale engineering examples are mainly limited to growth associated phenotypes, such as substrate utilization and tolerance to toxic compounds. Biosensors based on transcription factors (TFs) (Wang et al., 2016a), G-protein-coupled receptors (GPCRs) (Mukherjee et al., 2015), riboswitches (Kim et al., 2015), and responsive promoters (Shi et al., 2017) can be integrated into the genome-scale metabolic engineering pipeline. For example, a PaPR based malonyl-CoA biosensor was developed and used to screen a genome-scale cDNA overexpression library that increased the intracellular malonyl-CoA levels (Li et al., 2015b). In addition, by coupling product formation to an essential cellular process, biosensors can be further developed into synthetic selection circuits for genome-scale metabolic engineering and evolutionary engineering (Mans et al., 2017). For example, an aromatic amino acid responsive promoter was used to drive the expression of a gene in E. coli, providing a platform for integrating multi-dimensional experimental data into a readily accessible format. Mathematical models have often been interpreted as repositories of experimental data – it is a recent trend to utilize this data for inference and prediction. As higher dimensional data becomes easier to generate in the omics era, researchers will require mathematical models to not only interpret...
trends in the data but also use these models to design non-intuitive strategies for better strain engineering. Machine learning, which offers the tools for fitting models to high dimensional complex data, is an attractive avenue for metabolic engineering too. However, there still exist challenges in this regard. Data reproducibility must be strictly ensured whenever the data is used to train a model – this can prove challenging as biological variation is often difficult to control. Despite the possibility of vast amounts of data, fitting model parameters to the training data can still lead to uncertain estimates in parts of the model, which renders the model predictions meaningless. Thus, another challenge in the generation of models besides better parameter estimation algorithms for global optimization is also better uncertainty analysis for model estimates (Heijen and Verheijen, 2013; Vanlier et al., 2013).

Another trend in yeast metabolic engineering is to develop non-Saccharomyces yeast cell factories. Non-model yeasts may be more attractive for producing some certain products, such as oleaginous yeasts (i.e. Y. lipolytica and Rhodospirillum toruloides) for producing fatty acid derived compounds (Shi and Zhao, 2017) and Issatchenka orientalis for producing organic acids at low pH (Xiao et al., 2014). The recently developed synthetic biology tools for an array of non-model yeasts, such as Y. lipolytica (Schwartz et al., 2016) and Pichia stipitis (Cao et al., 2017), make them important alternatives to S. cerevisiae for metabolic engineering applications.

Acknowledgements

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Numbers DE-SC0018420 and DE-SC0018260. J.L. also acknowledges the support of the Shen Postdoc Fellowship from Zhejiang University. Research Funds for the Central Universities (2018QNA4039) and the Startup Fund from Zhejiang University.

References


and chemicals. Front. Microbiol. 8, 2185.


