Engineering plant metabolism into microbes: from systems biology to synthetic biology
Peng Xu1,2, Namita Bhan1,2 and Mattheos AG Koffas1,2

Plant metabolism represents an enormous repository of compounds that are of pharmaceutical and biotechnological importance. Engineering plant metabolism into microbes will provide sustainable solutions to produce pharmaceutical and fuel molecules that could one day replace substantial portions of the current fossil-fuel based economy. Metabolic engineering entails targeted manipulation of biosynthetic pathways to maximize yields of desired products. Recent advances in Systems Biology and the emergence of Synthetic Biology have accelerated our ability to design, construct and optimize cell factories for metabolic engineering applications. Progress in predicting and modeling genome-scale metabolic networks, versatile gene assembly platforms and delicate synthetic pathway optimization strategies has provided us exciting opportunities to exploit the full potential of cell metabolism. In this review, we will discuss how systems and synthetic biology tools can be integrated to create tailor-made cell factories for efficient production of natural products and fuel molecules in microorganisms.

Addresses
1 Department of Chemical and Biological Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, United States
2 Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY 12180, United States

Corresponding author: Koffas, Mattheos AG (mkoffas@eng.buffalo.edu, koffam@rpi.edu)

Current Opinion in Biotechnology 2012, 24:xx–yy
This review comes from a themed issue on Plant biotechnology
Edited by Natalia Dudareva and Dean DellaPenna

0958-1669/$ – see front matter, Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.copbio.2012.08.010

Introduction
Plant metabolism represents an enormous repository of bioactive compounds known as natural products (NPs) that are of pharmaceutical and biotechnological importance [1]. Specialized in skeletal structures and functional groups, plant NPs continue to play a leading role in drug discovery [2]. For example, it has been estimated that more than 75% of FDA-approved antibiotics and anticancer drugs in the last 25 years are NPs or inspired by NPs [3]. As a result, tremendous effort has been dedicated to the development of cost-efficient processes for the production of important NPs. Currently, most of commercialized NPs are still manufactured by extraction from their native plant sources or semi-synthesized from extracted NP intermediates. The major bottleneck in the plant extraction process is the low yield and the complicated downstream purification processes. While organic chemists have been able to elegantly synthesize/modify an array of complex NPs, such chemical synthesis processes are usually overshadowed by inherent disadvantages such as low yield, use of toxic catalysts and extreme reaction conditions that make them not amenable to large scale production [4].

Alternatively, microbial metabolic engineering has emerged as a promising approach to overcome these limitations. With the introduction of a plant heterologous pathway, universal precursors/intermediates (i.e. acetyl-CoA, malonyl-CoA, isopentenyl pyrophosphate) generated from the microbial central carbon metabolism and essential for the cell’s own survival need to be redirected in order to efficiently produce value-added NPs from inexpensive feedstocks. For instance, genetically tractable microorganisms (e.g. Escherichia coli and Saccharomyces cerevisiae) engineered with plant secondary metabolic pathways have demonstrated great success to produce a range of NPs including fatty acids [5], terpenoids [6,7,8**], flavonoids [9,10†], polyketides [11] and alkaloids [12]. Conventional metabolic engineering strategies have been primarily focused on redirecting carbon flux toward a target pathway by overexpression of rate-limiting steps and deletion of precursor competing-pathways [13]. With the advent of the post-genomics era, a staggering volume of genomic, transcriptomic, proteomic and metabolomic information has been deposited; thus it is now possible to manipulate a specific metabolic network by tweaking the entire cell metabolism [14]. Concurrently, as the characterized genetic parts and devices have been accumulated and the cost of DNA synthesis continues to decline, synthetic biology and metabolic engineering has enabled the precise control of cell metabolism to customize the production of plant NPs at a scale and speed never seen before. In this review, we will discuss how systems and synthetic biology tools can be integrated to design and create tailor-made cell factories for efficient production of NPs and fuel molecules in microorganisms.

Synthetic metabolic pathway design
Owing to the highly evolved transcriptional regulatory network and metabolic feedback control, microorganisms tend to maintain a constant level of essential precursor metabolite flux and will not commit to overproduce the
2 Plant biotechnology

Figure 1

Systems biology tools bring together layers of information that will advance our understanding to design and model intertwined metabolic networks. Genomics, transcriptomics, proteomics and constraints-based flux balance analysis (FBA) provide a global landscape of cell metabolism that will benefit the design and control of biological systems for biotechnological applications. Left panel shows the original metabolic network and the right panel shows optimized metabolic network with maximal carbon flux toward a desired product. On the right panel, dark green blocks show overexpression or upregulation targets; red blocks show knockout or downregulation targets. Double arrows show interplay between the neighboring components.

desired recombinant molecules [15]. It is often the case that manipulation of local metabolic networks can only lead to limited performance improvement. In the past decade, the field of systems biology has seen tremendous advances in understanding and predicting the dynamic behavior of complex biological systems [16]. Integrated experimental and computational tools have allowed us to attain a global landscape of molecular interactions among the myriad cellular components. Toward the goal of better engineering cell metabolism for NPs production, a number of high throughput omics tools and constraint-based flux balance models have been employed and our ability to quantitatively dissect cellular phenotypes has been greatly enhanced (Figure 1). These achievements have led to the emergence of the new concept of ‘systems metabolic engineering’ [14] that aims to move strain engineering on a global scale.

The large volume of plant genomic information generated from high throughput DNA sequencing techniques has provided immense possibility for understanding the biosynthetic potential of plant metabolism [17]. Such data can be exploited to identify cryptic pathways or unknown gene clusters that are responsible for valued NP biosynthesis. Sequence comparison between unknown DNAs and canonical genes allows the prediction of conserved catalytic motifs that potentially could lead to new chemical entities [18]. For example, genome mining has been successfully applied to understand the biochemical programming and molecular basis of two types of NP biosynthetic systems: the modular polyketide synthases and nonribosomal polyketide synthetases [19]. Recently, bioinformatics analysis of genomic information and mRNA transcripts has led to the prediction of enzymes and pathways involved in terpenoid biosynthesis in both Catharanthus roseus [20] and Ganoderma lucidum [21]. De novo sequencing, assembly of transcriptomic data has also allowed the prediction of biosynthetic pathways for a range of NPs including prostratins [22], picrosides [23] and hypericins.
Targeted proteomics that directly correlate protein translation efficiency with pathway limitations based on the accurate determination of enzyme abundance have been used to fine-tune/optimise heterologous metabolite [25] and phenylalanine pathways [26] in E. coli, leading to improved production of amorpho-4,11-diene and tyrosine, respectively. These high throughput systems biology tools have largely improved our ability to design and optimize metabolic pathways for efficient production of NPs in microorganisms.

At the same time, in silico computational biology tools based on flux balance analysis (FBA) or kinetics formulation have been developed to identify genetic manipulation targets that would redistribute carbon flux toward a compound of interest. The most popular in silico genome-scale metabolic computational method is constraints-based flux balance analysis [14]. It allows the determination of optimal reaction flux (i.e. maximal cell growth, maximal product formation or minimal precursor dissipation) by searching for a linear space constrained by stoichiometry-based mass balance equations and prescribed physio-biochemical boundaries. More importantly, possible flux distributions solved by linear or quadratic programming can be ranked and chosen to prioritize the implementation of genetic interventions. For example, strain optimization procedures such as OptORF [27], RobustKnock [28] and CiED [29,30] have enabled efficient use of FBA as a tool to predict gene overexpression and/or gene deletion targets. The ever-increasing size of biochemical datasets has motivated the use of algorithms that incorporate layers of information to further constrain flux space and achieve more accurate predictions. For example, transcriptional regulation network [31], chemical reaction thermodynamics [32], RNA-seq data [33], flux measurement data [34] and kinetic constraints [35] have all been incorporated into FBA framework to identify and optimize metabolic states in various organisms. Recently, genetic interventions predicted by OptForce have been successfully applied to improve intracellular malonyl-CoA and the resulting strain demonstrated around a 5-fold increase in flavonoid production in E. coli [10*]. Such achievements demonstrate that constraints-based flux balance models are powerful tools for strain design and improvement.

Synthetic pathway construction: enabling technology to implement the vision of synthetic biology

Synthetic biology is characterized by a constructive approach to understand and manipulate biological systems [36]. Even though the cost of commercial synthesis of genes is declining, our ability to physically construct complex biological devices/pathways from basic DNA parts is still a critical hurdle in implementing the vision of synthetic biology [37*]. These limitations can become prohibitive when constructing multi-gene metabolic pathways and complex regulatory circuits. For example, strain development through metabolic engineering leading to valuable NPs typically involves the manipulation of a dozen precursors or rate-limiting pathways [7,8**]. Conventional pathway construction approaches, which largely rely on smart design to assemble multi-gene fragments in operon form, are limited in terms of automation and context-independent pathway output. Therefore, synthetic metabolic engineering requires efficient tools that allow precise and concerted assembly of multiple gene fragments, leading to programmable or predictable pathway output customized for strain optimization.

Last decade’s advancement in synthetic biology has led to the emergence of several robust gene assembly platforms that are suited for metabolic pathway construction and optimization. These gene assembly tools allow us to efficiently construct directional multi-gene pathways from synthesized or PCR-based gene fragments with 30+ bp overlap regions (Figure 2). For example, Sequence and Ligation-Independent cloning (SLIC) [38,39] and Gibson isothermal assembly [40] rely on in vitro homologous recombination and single-strand annealing. SLIC has been recently modified to allow one-step assembly of multiple medium-size DNA fragments [41] and large genomic DNA fragments (up to 28 kb) [42] using in vitro single-strand overlapping annealing. In vivo assembly tools based on yeast homologous recombination have also been developed for assembling large DNA fragments encoding an entire biochemical pathway [43] or iteratively integrating multi-gene pathways into yeast chromosomes [44]. Coupled with yeast in vivo recombination, Gibson isothermal assembly has been used to assemble a self-replicating synthetic M. genitalium genome (1.08 Mb) [45*] and a high G+C genomic DNA fragment (454 kb) containing Synechococcus elongatus PCC 7942 photosynthetic gene clusters [46]. Circular polymerase extension cloning (CPEC) has been developed for one-step assembly of multi-gene pathways and combinatorial DNA libraries using complementary fragment annealing and PCR overlap extension [47].

Compared with conventional DNA cloning protocols, these advanced DNA assembly tools offer an efficient approach to construct multi-gene pathways in a one-step, scar-less and sequence-independent manner. The downside of these assembly tools is the terminal flanking homology sequence of each DNA fragment must be specifically designed for each assembly junction, which tends to be laborious and error-prone [48]. Assembly of gene fragments with identical end-terminus sequences (such as repeated promoters, repeated ribosome binding site and repeated terminators) can be problematic, as this can lead to constructs with missing, extra or rearranged gene fragments in the wrong order. As such, these
4 Plant biotechnology

Figure 2

DNA assembly tools for synthetic pathway construction. (a) Parallel gene assembly tools provide an efficient approach to construct directional multi-gene pathways from synthesized or PCR-based gene fragments with 30+ bp overlap region. SLIC (sequence and ligation-independent cloning) and Gibson isothermal assembly are based on in vitro homologous recombination and single-strand annealing; specifically, SLIC used a bifunctional DNA polymerase (T4 DNA polymerase) that exhibits 3’ exonuclease activity in the absence of dNTPs, whereas Gibson assembly used enzyme mixtures consisting of 5’ exonuclease, Phusion DNA polymerase and Taq DNA ligase. DNA assembler is based on yeast in vivo homologous recombination. (b) Circular polymerase extension cloning (CPEC) technique for gene assembly. Complementary overhangs attached to insertion gene fragment anneal to linearized vector and can be circularized into an intact vector by PCR extension. (c) ePathBrick vectors provide a versatile platform to combinatorially assemble multi-gene pathways with different configurations. The ePathBricks take advantage of four compatible restriction enzymes (AvrII, XbaI, SpeI and Nhel), which generate compatible cohesive ends and upon ligation result in a scar sequence that cannot be cleaved by either of the original restriction enzymes. Basic molecular components flanked with these isocaudamer pairs can be assembled together by restriction enzyme digestion and ligation.

advanced cloning tools do not offer the capability to refine the regulatory control elements of each of the expression cassettes if identical regulatory sequences (such as promoters, operators, ribosome binding sites and terminators) are used. In addition, due to the pre-determined fragment context specified by the overlap sequence, these gene assembly tools are not scalable for diversifying pathway configurations and shuffling gene
order [49**]. For these reasons, there is a need for new gene assembly tools adept at hierarchical pathway construction and generation of pathway diversities in a combinatorial manner.

Assembly of DNA fragments based on the BioBrick paradigm has become a standard practice in metabolic pathway construction [50,51] and genetic circuit design [52]. A versatile gene assembly platform, ePathBrick, compatible with BioBrick standards has been recently described for metabolic pathway construction and optimization [49**]. The engineered ePathBrick vectors feature four compatible restriction enzyme sites (AvrII, XbaI, SpeI and XhoI) allocated on strategic positions that support the modular assembly of a number of molecular components including regulatory control elements (promoters, operators, ribosome binding sites and terminators) and multi-gene pathways (Figure 2c). In addition, ePathBricks provide a platform for combinatorial generation of pathway diversities with three distinct configurations. Specifically, for a multi-gene pathway, each of the pathway components can be organized either in operon, pseudo-operon, or monocistronic form (Figure 2c) when using different isocaudamer pairs. It was demonstrated that a three-gene flavonoid pathway can be easily diversified to 54 pathway equivalents differing in pathway configuration and gene order; coupled with high throughput screening techniques, we envision that this combinatorial strategy would greatly improve our ability to exploit the full potential of microbial cell factories for recombinant metabolite and NP production. A total pathway size of 36 kb could be assembled using four compatible ePathBrick vectors, which is sufficient for expressing a typical plant secondary metabolic pathway [49**]. The ePathBrick assembly, though limited by one gene at a time, provides a versatile platform to combinatorially construct and optimize multi-gene pathways — a feature that will be proven extremely useful for pathway engineering.

Synthetic pathway optimization through static, dynamic and spatial control

One of the most important goals of metabolic engineering and synthetic biology has been the optimization of metabolic pathways for the production of biofuels [53,54*] and pharmaceuticals [7] in genetically tractable organisms. A major challenge in these efforts has been the accurate tuning and optimization of the expression level of each of the enzymes of the selected pathways. Toward this goal, a number of synthetic biology approaches have been applied including modification of plasmid copy number [26], promoter strength [55] and gene codon usage [56]. In addition, delicately designed molecular control elements or regulatory circuits have been integrated into the cell chassis to enable the host strain to precisely respond to environmental stimuli or cellular intermediates and drive carbon flux toward the target pathway. For example, engineering promoter architecture has resulted in tunable gene expression in both E. coli [57] and yeast [58]; engineered metabolite-responsive riboswitches [59] and synthetic ribosome binding sites [60] can be used to precisely control protein translation; engineering genetic timers based on mutually repressible toggle switches have been used to control the timing of yeast sedimentation [61].

However, engineering microbial overproduction phenotypes remains a daunting task as it usually involves the manipulation of a handful of precursor or rate-limiting pathways that are subjected to tight cellular regulation. For example, precursor flux improvement by overexpression of heterologous pathways may not be accommodated by downstream pathways; as a result, accumulated or depleted intermediates may compromise cell viability and pathway productivity [6]. In an attempt to address these issues, combinatorial transcriptional engineering has been proposed to optimize the expression level of each of the genes involved in a multi-gene pathway.

Central to this endeavor is an approach called ‘multivariate modular metabolic engineering’ (Figure 3a), which aims at removing pathway bottlenecks and optimizing a pathway’s potential through systematic investigation of pathway input variables [62]. For example, strategies including combinatorially tuning promoter strength, plasmid copy numbers and translational initiation rates, have been deployed for optimizing the biosynthesis of important metabolites, such as taxadiene [8**] and free fatty acids [P Xu et al., Modular optimization of multi-gene pathways for fatty acids production in E. coli, unpublished data]. It is important to note here that precisely designed statistical experiments such as response surface methodology [63] can be used to screen main effects and optimize pathway output instead of performing a full factorial search. Combined with synthetic promoter libraries [57] or synthetic ribosome binding sites [60], the multivariate modular metabolic engineering strategies provide a generalized approach to optimize cell factories for valuable metabolite production.

In nature, functionally related enzymes are often organized in close proximity by noncovalent interactions or membrane-tethered anchor proteins. This spatial colocalization strategy enables products formed from one enzyme to be efficiently translocated to another enzyme, thus preventing the accumulation of toxic intermediates and minimizing the loss of reactive intermediates due to diffusion (Figure 3b and c). Recently, a number of scaffold-based approaches that mimic this substrate-channeling effect have been developed to optimize the metabolic flux through a target pathway. For example, heterologous enzymes involved in the mevalonate biosynthetic pathway have been tethered on specifically designed synthetic protein scaffolds at stoichiometric ratios [64**]. This spatial control led to a 77-fold improvement in product titers at low enzyme expression levels. The same strategy has been
applied to improve the production of glucaric acid in *E. coli* [65]. Similarly, yeast cell surface displayed with engineered minihemicellulosome that assembles cellulose-degrading enzymes has enabled ethanol production directly from plant biomass xylan [66]. Potentially, these substrate-channeling strategies could be applied to other systems for the modular control over metabolic flux.

As heterologous pathways become larger and more complicated, it becomes increasingly difficult to optimize...
them with static regulatory control [67]. Generally, optimization through static control is only applicable for a particular environment and any perturbations that deviate from the prescribed condition would likely result in phenotype instability or suboptimal productivity. In contrast to static control, native biological systems typically utilize dynamic regulatory networks to control metabolic flux in response to changing environments (Figure 3d). For example, one of the coherent strategies that occur in most biological systems is gene expression regulation through negative/positive feedback loops. Mediated by a transcriptional regulator, a metabolic intermediate would act as a signaling molecule to induce or repress the expression of enzymes responsible for its synthesis or consumption. Nature has evolved this strategy to allow metabolic pathways to be dynamically modulated so that cellular resources can be more efficiently utilized regardless of the changing environment. In practice, this strategy could be applied to pathway optimization in the case where accumulated toxic intermediates are detrimental to cell growth. For example, a negative feedback loop can be used to control pathways that are responsible for intermediate synthesis (Figure 3d); and a positive feed-forward loop can be used to control pathways that are responsible for intermediate consumption. Successful application of this optimization strategy would require identification of promoters that are both positively and negatively regulated by the same intermediate molecule. Recently, a fatty acyl-CoA responsive promoter has been used to dynamically control the expression of genes coding for enzymes involved in the biodiesel pathway; the resulting dynamic sensor-regulator system has led to a 3-fold increase in FAEEs (fatty acids ethyl esters) production compared with using constitutive promoters [68**] in E. coli.

**Perspective**

The advances in systems and synthetic biology have greatly speeded up our ability to design, construct and optimize cell factories for metabolic engineering applications. Concerns about climate change, rising petroleum prices and emerging health problems will continue to motivate us to explore novel solutions for the production of fuel and pharmaceutical molecules. It is envisioned that synthetic cell factories will partially replace petroleum-based chemical synthesis in the next 5–10 years. Growing interest in modeling cell metabolism will result in the development of new computational tools that incorporate dynamic gene expression patterns and offer a mechanistic view of regulatory networks. Furthermore, gene assembly tools that are adept at high throughput combinatorial pathway construction will physically enable us to exploit the full potential of cell metabolism. Knowledge about hierarchical regulatory architecture [69] and network motif [70] will transform our understanding of how to build modular and orthogonal genetic circuits to control gene expression. Large scale gene-editing tools such as MAGE [71*], CAGE [72] and TALENs [73] will allow the targeted regulation and modification of genomic DNAs. By combining these tools, metabolic engineers will be well positioned to create tailor-made cell factories for efficient, high-yield and economical production of NPs and fuel molecules in the near future.

**Acknowledgements**

This work was supported by NIH 5 R01 GM085323-04 award to Matthew Koffas.

**References and recommended reading**

Papers of particular interest, published within the period of review, have been highlighted as:

* • of special interest
** • of outstanding interest


The authors demonstrated multivariable modular pathway engineering approach to optimize isoprenoid pathway for taxadiene production in E. coli. By combinatorially changing promoter strength and plasmid copy number, the authors successfully identified optimal pathway expression levels that balance the formation and consumption of isopentenyl pyrophosphate and minimized the secretion of toxic metabolite indole.


By using a recently developed OptForce framework, the authors successfully identified minimal set of genetic interventions that cooperatively force carbon flux toward malonyl-CoA. The authors also experimentally validated these predictions and the optimized metabolic network led to a 5.6-fold increase in flavonoid production.

Plant biotechnology


By using an artificial protein scaffold, the authors spatially recruited mevalonate pathway enzymes and achieved modular control over metabolic flux. This substrate channeling strategy has led to a 77-fold increase in mevalonate production in E. coli.


The authors reported the construction of a trifunctional minihemicellulose on the yeast cell surface. Direct ethanol production from birchwood xylan was achieved in this study.


Using a fatty acyl-CoA-responsive promoter, the authors dynamically regulated the genes involved in biodiesel synthesis and achieved 3-fold increase in FAEs production. This strategy provides a generalizable approach for controlling gene expression in almost any pathways if cognate sensors for the pathway intermediate exists.


The authors described multiplex automated genome engineering (MAGE) for large-scale programming and evolution of cells. Using this strategy, the authors have successfully constructed an E. coli strain producing 5-fold more lycoperone compared with parental strain.
