The Future of Metabolic Engineering and Synthetic Biology: Towards a Systematic Practice

Vikramaditya G. Yadav[Δ], **Marjan De Mey**[Δ,†], **Chin Giaw Lim**, **Parayil Kumaran Ajikumar**[‡], and **Gregory Stephanopoulos**[*]

Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Abstract

Industrial biotechnology promises to revolutionize conventional chemical manufacturing in the years ahead, largely owing to the excellent progress in our ability to re-engineer cellular metabolism. However, most successes of metabolic engineering have been confined to over-producing natively synthesized metabolites in *E. coli* and *S. cerevisiae*. A major reason for this development has been the descent of metabolic engineering, particularly secondary metabolic engineering, to a collection of demonstrations rather than a systematic practice with generalizable tools. Synthetic biology, a more recent development, faces similar criticisms. Herein, we attempt to lay down a framework around which bioreaction engineering can systematize itself just like chemical reaction engineering. Central to this undertaking is a new approach to engineering secondary metabolism known as 'multivariate modular metabolic engineering' (MMME), whose novelty lies in its assessment and elimination of regulatory and pathway bottlenecks by re-defining the metabolic network as a collection of distinct modules. After introducing the core principles of MMME, we shall then present a number of recent developments in secondary metabolic engineering that could potentially serve as its facilitators. It is hoped that the ever-declining costs of *de novo* gene synthesis; the improved use of bioinformatic tools to mine, sort and analyze biological data; and the increasing sensitivity and sophistication of investigational tools will make the maturation of microbial metabolic engineering an autocatalytic process. Encouraged by these advances, research groups across the world would take up the challenge of secondary metabolite production in simple hosts with renewed vigor, thereby adding to the range of products synthesized using metabolic engineering.

Keywords

Metabolic Engineering; Synthetic Biology; Multivariate Analysis; Modularization; Natural Products; Secondary Metabolic Pathways; Terpenoids

[*]Corresponding Author. Tel: 617-253-4583; Fax: 617-253-3122; gregstep@mit.edu.
[†]Current Affiliation: Centre of Expertise – Industrial Biotechnology and Biocatalysis, Ghent University, Ghent, Belgium
[‡]Current Affiliation: Manus Biosynthesis, Cambridge, MA, USA
[Δ]Denotes equal authorship

The current state of metabolic engineering

The use of industrial biotechnology for the production of foodstuffs and fuels is centuries-old. However, it was only after the conception of metabolic engineering a little over twenty years ago that employing biological entities for chemical production received a significant boost. This is best evidenced by the scope and impact of the projects presently being pursued (Alper and Stephanopoulos, 2009; Keasling, 2008; Keasling, 2010; Stephanopoulos, 2007). Nevertheless, metabolic engineering, as it is currently practiced, has come a long way from its early days, wherein bottlenecks in the pathway were first identified using metabolic flux analysis, following which, these insights were utilized to improve or redirect fluxes in order to achieve the desired phenotype (Bailey et al., 1996; Stephanopoulos and Sinskey, 1993).

In the two decades following its establishment, including a period that is often referred to as the 'post-genomic era', a staggering volume of gene, protein and metabolite data has been accumulated, the costs of oligonucleotide synthesis have exponentially declined, and more precise techniques for studying cellular metabolism have been developed. Empowered by these developments, the focus of metabolic engineers has gradually shifted away from perturbing individual pathways to manipulating the entire cell itself, giving rise to the concept of 'systems metabolic engineering' (Kholodenko and Westerhoff, 2004; Park and Lee, 2008; Stafford and Stephanopoulos, 2001; Stephanopoulos et al., 2004). This has resulted in the metabolic engineering 'toolbox' greatly expanding from conventional approaches such as rationally deleting and/or over-expressing endogenous genes and introducing heterologous genes to now include tools capable of controlling gene expression and modulating regulatory networks throughout the cell (Blazeck and Alper, 2010; Klein-Marcuschamer et al., 2007; Tyo et al., 2007; Tyo et al., 2010).

Sowing the seeds of systematization of metabolic engineering

In light of the recent advances in the field, one would have expected an explosion in the catalogue of materials produced by industrial biotechnology, and, more importantly, sizable progress towards realizing the chemical industry's vision of integrated 'biorefineries' (Kamm and Kamm, 2004). Yet, much remains to be addressed in order for these goals to become realities. Instead of becoming a systematic study with well-defined principles and tools, metabolic engineering has strangely remained a collection of elegant demonstrations.

A major reason behind this peculiar turn of events is that many of the tools used for manipulating a host's metabolism are not universally applicable and, in some cases, they are specific to only certain pathways or products. This lack of transferability of a majority of tools and techniques between hosts, in turn, can be attributed to an incomplete understanding of the regulatory mechanisms in the cell. The need for generalizable tools and techniques that can characterize and manipulate the regulatory mechanisms that prevail within a host has never been more acute than in the case of demonstrating heterologous over-production of natural products and other secondary metabolites. These molecules are promising targets for metabolic engineers since they either possess desirable pharmacological properties or are potential replacements for chemicals that are currently manufactured from petroleum-derived feedstocks (Ajikumar et al., 2008; Chemler and Koffas, 2008; Gershenzon and Dudareva, 2007). However, the technical challenges associated with achieving the production of these molecules in heterologous hosts are quite grand (Ajikumar et al., 2008; Keasling, 2010; Kirby and Keasling, 2009; Klein-Marcuschamer et al., 2007; Muntendam et al., 2009; Yadav and Stephanopoulos, 2010).

Briefly, interventions to cellular metabolism in order to over-produce secondary metabolites can occur at one or many of the following levels: (1) enhancement in the rate of substrate

uptake, (2) reduction of flux to undesirable by-products and enhancement of precursor and cofactor flux, (3) introduction of the heterologous pathway and optimization of the activity of its constituent enzymes, and (4) export of the product to the extracellular medium in order to shift equilibrium towards product formation (Figure 1). The continually declining costs of oligonucleotide synthesis and a conscious push towards standardization of cloning technologies (Shetty et al., 2008) in recent years, besides germinating the field of synthetic biology, have ensured that pathway assembly is no longer as difficult as it once was. In addition, not only have appreciation and understanding of tightly tuning promoter strength and altering gene copy number to directly modulate gene expression (Ajikumar et al., 2010; Anthony et al., 2009); enhancing the rate of, among others, redox reactions by ensuring optimal cofactor supply (Bond-Watts et al., 2011; Shen et al., 2011); and optimizing the activity of a pathway's enzymes by selectively manipulating its active site residues to enhance turnover (Leonard et al., 2010) now greatly improved, but, more importantly, these tools and techniques are also host-independent.

Once the pathway has been assembled according to the desired specifications, one generally attempts to circumvent the regulatory mechanisms in the cell in order to maximise production. This task, though, is considerably more daunting. Transferring multi-gene pathways into a heterologous production host often leads to flux imbalances since the host typically lacks the complex regulatory mechanisms that are vital for efficient operation of these complex pathways, and, these effects vary from one heterologous host to another. Crucially, tools that alter the regulatory landscape within the cell to favour over-production of a particular molecule are neither as readily available, nor are they as well-understood. Some recent studies attempted to address this difficulty by combining metabolic engineering tools and techniques with those from combinatorial genetics (Ajikumar et al., 2010; Lee et al., 2011; Santos and Stephanopoulos, 2008; Wang et al., 2009). However, combinatorial experimentation necessitates availability of a high-throughput assay to screen and select desirable mutants. Many demonstrations of combinatorial experimentation so far have only focused on improving the production yields, titers or rates of carotenoids and other molecules with similar colorimetric properties (Klein-Marcuschamer et al., 2007; Wang et al., 2009) or demonstrating tolerance improvements. It is apparent that such techniques are not particularly useful for selecting mutants in the absence of high-throughput screens or when, as is usually the case in secondary metabolic engineering, product inhibition is not an impediment

If metabolic engineering is to develop into a more systematic practice such as chemical reaction engineering, the immediate focus of the metabolic engineers ought to be directed at developing techniques capable of evaluating and manipulating the regulatory landscape within the cell in a more rational, generalizable and economical manner. Spanning a larger search space in a fewer number of experiments will be central to this undertaking.

Establishing a blueprint for studying metabolic regulation

A recent study on taxane production in *E. coli* (Ajikumar et al., 2010) offers valuable insights into evaluating the regulatory landscape within the cell. Not only are its conclusions potentially generalizable enough for extension to other hosts and pathways, but the study also lays down a framework for efficiently dealing with regulatory bottlenecks. Called 'multivariate modular metabolic engineering' (MMME), this framework leverages recent developments in the standardization of cloning elements as well as cheaper oligonucleotide synthesis, and, the results of this study effectively debunk the commonly-held notion that *E. coli* is a sub-optimal host for terpenoid production.

Terpenoids are naturally synthesized via one of two possible routes – the mevalonate (MVA) pathway or the non-mevalonate or 2-methyl-(D)-erythritol-4-phosphate (MEP) pathway (Figure 2). The MVA pathway, which is operational in all eukaryotic cells and the cytoplasm and mitochondria of plants, commences with the co-condensation of acetyl-CoA to form acetoacetyl-CoA. The MEP pathway, on the other hand, commences with the condensation of glyceraldehyde-3-phosphate with pyruvate to yield 2-methyl-(D)-erythritol-4-phosphate and is specific to bacteria, other prokaryotes and the plastids in plants. Although both pathways culminate in the production of isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) – the universal, 5-carbon terpenoid precursors, the MEP pathway is energetically balanced and more efficient than the MVA pathway in its ability to convert sugars or glycerol into terpenoids (Ajikumar et al., 2010; Dugar and Stephanopoulos, 2011) (Figure 3).

IPP then successively condenses with DMAPP and the products of its condensation reactions therewith to yield a homologous series of allylic pyrophosphates that, in turn, serve as precursors to specialized terpenoid producing pathways. The precursor-forming condensation reactions are simply head-to-tail heteropolymerization reactions that re-introduce the same functional moiety at spatial intervals dictated by the length of the monomers. The precursors to the individual terpenoid pathways are then de-phosphorylated. The presence of the polymer-forming functional group at repeating intervals in the precursors allows for intra-chain and inter-chain condensation reactions to occur. Intra-chain condensation reactions cyclize the polymer chain whereas inter-chain condensation reactions create highly-branched molecules. The staggering structural diversity that one witnesses across the terpenoid natural product family is simply the product of varying the configuration of a single reaction. The cyclized and branched derivatives of the de-phosphorylated allylic pyrophosphates then undergo a variety of substitution reactions such as hydroxylations, oxidations, esterifications and glycosylations to yield individual terpenoid molecules.

The novelty of MMME (Figure 4) lies in its assessment and elimination of regulatory and pathway bottlenecks by re-defining the metabolic network as a collection of distinct modules. For example, the flux through a linear pathway that is expressed as an operon is the reaction rate of the pathway's slowest step. This implies that, at steady state, the observed turnovers of several enzymes of the pathway are below their maximum values. By grouping enzymes with similar turnovers into a subset, or module, and later equalizing the turnovers of the different subsets by adjusting their concentrations, one can maximize the ratio of pathway turnover to resource expenditure. The practice of re-casting metabolic networks as a collection of interacting modules not only makes the analysis of a very complicated system considerably more tractable, but since the concentrations of a module's enzymes can be adjusted using a variety of transcriptional, post-transcriptional and translational mechanisms, performance of module configurations can be easily assessed using multivariate statistics. Also, modularization is a natural extension of the principle of standardizing cloning elements, and, the ever-declining costs of oligonucleotide synthesis imply that entire modules with specific expression configurations can be synthesized *de novo* and rapidly tested and analyzed.

In their experiments on taxane biosynthesis in *E. coli* (Ajikumar et al., 2010), the authors assembled the MEP pathway enzymes with the lowest turnovers – 1-deoxy-(D)-xylulose-5-phosphate (DOXP) synthase (*dxs*), 2-methyl-(D)-erythritol-4-phosphate cytidyltransferase (*ispD*), 2-methyl-(D)-erythritol-2,4-cyclodiphosphate (MECPP) synthase (*ispF*) and IPP isomerase (*idi*) (Alper et al., 2005a) – into a single module which they termed the 'upstream' module (Figure 5). The enzymes catalyzing the next two steps after IPP and DMAPP formation, namely the synthesis of the allylic pyrophosphate precursor, geranylgeranyl

pyrophosphate (GGPP) and the cyclization of GGPP to the dedicated taxane precursor, taxadien-4(5),11(12)-diene, or simply taxadiene, were grouped into another module called the 'downstream module'. Incidentally, taxol, a blockbuster anti-cancer drug, is one of the taxanes synthesized from taxadiene. Using this approach, Ajikumar *et al.* constructed a total of 32 strains, each differing in the strength of the promoters driving module expression and the copy numbers of the plasmids from which the modules were expressed. The most prolific taxadiene producer amongst these strains achieved an unprecedented average titer of 1 g/L – nearly 15,000 times greater than the control strain harboring the taxadiene pathway (Figure 5).

Besides the improvement in taxadiene production, another striking take-away from this work is the relatively small number of variants that were constructed to arrive at the (presumably) local optimum. Given the low cost of oligonucleotide synthesis these days, even if the number of variants constructed were to rise by an order of magnitude, the rate of convergence to a 15,000-fold greater titer far surpasses that of any combinatorial method.

Another interesting observation that Ajikumar *et al.* made was that strains expressing the aforesaid pathways comprising of enzymes with unbalanced turnovers tended to accumulate high concentrations of indole, often to the detriment of terpenoid production. This observation makes MMME unique in its ability to survey the regulatory landscape that prevails in *E. coli* following expression of heterologous pathways. When the module configurations were adequately tuned, consequently balancing enzyme and pathway turnovers, indole production did decrease. This underscores the utility of the technique to engineering secondary metabolic pathways in simple hosts such as *E. coli*.

While Ajikumar *et al.* 'coarsely' tuned the expression of the two modules by simply varying plasmid copy numbers and altering promoter strengths, a variety of more sophisticated approaches can also be utilized to achieve 'finer' tuning. These include augmenting the activities of native enzymes with homologues from another organism to circumvent regulatory constraints in the cell (Yoon et al., 2009), toggling the order of the genes within the operon (Anthony et al., 2009; Nishizaki et al., 2007), using tunable promoters (Ajikumar et al., 2010; Alper et al., 2005b; De Mey et al., 2007), manipulating the ribosome binding site (RBS) (Salis et al., 2009; Wang et al., 2009), and employing synthetic protein scaffolds for enzyme co-localization in order to mitigate turnover drop-offs due to diffusive losses (Dueber et al., 2009). Some of these facilitating technologies are more effective than others in achieving precise control of module performance and will be discussed in greater detail in the sections to follow.

Homologue augmentation and substitution

Augmenting the activities of rate-limiting enzymes in a native pathway with homologous enzymes from other hosts is an effective strategy to circumvent the native regulatory cascades that maintain tight control of flux in the metabolic network. While such an approach may have been impractical as recently as a decade ago, the vast volumes of genetic and proteomic information collected in the intervening period has made 'homologue augmentation' a realistic and efficient alternative for improving production titers. Similarly, one could also enhance the turnover of heterologous pathways by employing homologues of the non-native enzymes. This approach, known as 'homologue substitution', has also begun to witness increasing use in recent times. Both, homologue augmentation and substitution were effectively utilized in a recent study on carotenoid production in *E. coli* (Yoon et al., 2009) to achieve considerable improvements in β-carotene titers. Interestingly, the study implemented both strategies in the modular framework propounded by MMME.

Briefly, the authors attempted to enhance production of IPP and DMAPP – key intermediates in carotenoid production – by augmenting the native MEP pathway in *E. coli* via expression of a heterologous MVA pathway. The reactions comprising the MVA pathway were grouped into two modules. The first module consisted of reactions catalyzing the conversion of acetyl-CoA to mevalonic acid, whereas the second module comprised of reactions that sequentially transform mevalonic acid to IPP and DMAPP. The enzymes constituting the individual modules were prospected from a variety of microbial sources. Accordingly, the highest β-carotene titers were observed when the first module comprised of enzymes from *Enterococcus faecelis* while the second module consisted of enzymes from *Streptococcus pneumonia*. The average β-carotene production for *E. coli* strains harboring this combination of modules was 465 mg/L, which was significantly higher than previously reported values.

A possible extension of this approach would be to construct a MVA pathway wherein each enzyme has been prospected from a different host. However, the construction of such 'chimeric pathways' greatly encumbered. The activity of an enzyme is a product of both, the evolutionary pressures and the intracellular milieu that prevails within its host, and, unless this context is suitably replicated in the engineered strains, the likelihood of successfully expressing these enzymes to function at their natural propensities is low. Optimizing the throughput of a chimeric pathway that combines enzymes from several sources is not trivial (Yadav and Stephanopoulos, 2010). Besides, the experimental search space can be unfeasibly large. An unanswered question is where does one draw the line with regards to the maximum number of enzyme combinations that must be tested in order to identify an optimal combination. By using modularization, especially of the form based on the 'one module, one source' approach, may enable faster construction, testing and identification of expression configurations that maximize product formation.

Co-localization of pathway enzymes to reduce diffusive losses of intermediates

Secondary metabolic reactions could either be diffusion-limited, reaction-limited or both. Diffusion limitations often arise due to a poor rate of substrate transport to the enzyme. On the other hand, inconsistencies in the reaction mechanism, such as promiscuity to several substrates, poor product selectivity or both, generally cause the enzyme to be reaction-limited (Jones and Firn, 1991). Of the two, manipulating the reactive characteristics of an enzyme is quite protracted – not to mention challenging (Leonard et al., 2010). However, tools to increase substrate fidelity or product selectivity by manipulating the active site of an enzyme are quite mature and more widespread compared to techniques used to eliminate *in vivo* diffusion limitations. In fact, until Dueber *et al.*'s elegant work on enzyme co-localization to eliminate diffusion limitations in a pathway, there were no successful strategies to assess and eliminate this problem (Dueber et al., 2009).

In their study, Dueber *et al.* achieved a 77-fold increase in mevalonic acid production in *E. coli* by localizing engineered isoforms of acetoacetyl-CoA thiolase, hydroxymethylglutaryl-CoA synthase and hydroxymethylglutaryl-CoA reductase of the MVA pathway of *S. cerevisiae* onto a heterologous protein scaffold. The MVA enzymes were suitably modified to incorporate peptide ligands known to interact with specific domains on the scaffold. Significantly, the enzymes and the scaffold were expressed in a programmable manner to optimize their stoichiometric ratios.

Some possible improvements could extend the utility of enzyme scaffolding in the metabolic engineering toolbox. Presently, no more than 3 enzymes can be successfully co-localized onto a synthetic scaffold expressed in a simple host such as *E. coli*. A significant increase in

loading capacity of the scaffolds would allow scaffolds to be employable in engineering secondary metabolic pathways that are longer and more branched than the MVA pathway. More importantly, controlling expression of a high-capacity scaffold accommodating multiple heterologous enzymes will be important to avoid metabolic stress.

Chromosomal integration of heterologous pathways

Plasmid copy number proved to be a key determinant of module performance in taxadiene production in *E. coli* (Ajikumar et al., 2010). More specifically, when the upstream and downstream modules were expressed on plasmids having single-digit copies, taxadiene titers were markedly higher than modules expressed on plasmids replicated to substantially higher numbers. This seems to suggest that issues such as plasmid segregation, imprecise control of plasmid-based gene expression and the metabolic burden incurred on account of plasmid maintenance are possibly amplified with the copy number of the plasmid, and, this observation is consistent with those of previous studies (Anthony et al., 2009; Bentley et al., 1990; Jones et al., 2000; Noack et al., 1981). It is apparent that expressing the modules on low-copy, preferably single-copy plasmids could be a good strategy for minimizing the influence of hitherto poorly understood phenomena such as plasmid segregation and plasmid-associated metabolic burden.

Chromosomal-based expression not only avoids drawbacks such as segregation and plasmid maintenance-associated metabolic stress – not to mention antibiotic supplementation, but such systems can also evolve to eventually accommodate the optimal number of module copies through tandem gene duplication. With these benefits in mind, Tyo et. al developed a methodology for reliably integrating the different modules at specific sites in the chromosome of *E. coli* and controlling copy number (Tyo et al., 2009). Using this technique, which they termed Chemically Inducible Chromosomal Evolution (CIChE), the authors were able to construct and tune copy number of the lycopene biosynthetic pathway. Lycopene production of the CIChE constructs exceeded that of equivalent plasmid-based strains by 60% in batch cultures. More importantly, as opposed to the plasmid-containing strains, the CIChE constructs were propagated without antibiotic selection.

For CIChE to be used in conjunction with MMME, several modifications would be needed for the construction of heterologous pathways, particularly long and branched pathways. As the number of genes in a single module as well as the number of modules being integrated into the chromosome rise, so too would the number of recombination sites required. It is imperative that recombination and transcription efficiencies at each site in the *E. coli* chromosome be carefully catalogued in order to improve the utility of chromosomal integration for metabolic engineering.

Lastly, a gene's position in an operon has also been shown to affect its expression. In *E. coli*, for example, it has been demonstrated that the abundance of mRNA transcripts decreases with increasing distance of the gene from the promoter (Ajikumar et al., 2010; Nishizaki et al., 2007). Pfleger et al. have demonstrated that inequitable transcription and, by extension, inequitable translation could be corrected via use of unique, tunable intergenic regions (TIGRs) (Pfleger et al., 2006). The TIGR approach involves the use of three regions – two of which are variable sequences that form hairpin loops when transcribed. The third region incorporates one of many possible RNase E sites. The hairpin-forming sequences flank the RNase E site. When transcription occurs, the nascent mRNA is cleaved at the RNase E site to yield two new transcripts whose stabilities are individually modulated by the hairpin structures. TIGRs are excellent examples of tools that can be seamlessly assimilated into the modularization concept.

Amorphadiene synthesis in *E. coli*: A modularization framework

Pathway modularization provides practitioners with a robust approach to systematically assess and circumvent a host's complex regulatory mechanisms to improve production titers. More importantly, this approach requires no previous knowledge about the cell's regulatory landscape. It is helpful to frame another successful metabolic engineering effort, maximizing amorpha-4,11-diene (amorphadiene) production in *E. coli* in terms of modularization (Martin et al., 2003a; Newman et al., 2006; Pitera et al., 2007). Amorphadiene is the precursor to the anti-malarial drug, artemisinin – a valuable, plant-derived therapeutic.

One of the plasmids of the 2-plasmid system engineered for amorphadiene synthesis in *E. coli* expressed a chimeric pathway for the production of farnesyl pyrophosphate (FPP) – a 15-carbon allylic pyrophosphate that is subsequently cyclized to yield amorphadiene. The chimeric pathway comprised of two modules, the first of which encoded the synthesis of IPP and DMAPP via the MVA pathway from *S. cerevisiae*. The second module expressed the machinery for the production FPP and consisted of native and *H. pluvialis* genes. The gene encoding synthesis of amorphadiene from FPP was expressed as another module on the second plasmid. Initially, production titers were quite low, arguably due to turnover imbalances in the pathway (Martin et al., 2003b). However, when the turnovers of the rate-limited enzymes were suitably matched using techniques such as promoter tuning, varying the copy number of the plasmids and manipulating the composition of the operons, titers increased to 300 mg/L/OD_{600} (Anthony et al., 2009).

Following replacement of the promoters of the MVA pathway with stronger, tunable promoters, amorphadiene production increased even further. Subsequent experiments by Anthony *et al.* on distributing the biosynthetic machinery over two plasmids as opposed to constructing a single plasmid that harbors the entire apparatus revealed how propagating strongly-expressed gene cassettes on two multi-copy plasmids drastically increases the metabolic burden on the cell, consequently reducing the likelihood of plasmid retention by the hosts(Anthony et al., 2009). Plasmid loss was not an issue for strains harboring a single plasmid. Anthony *et al.* did not attempt to integrate the modules into the chromosome of *E. coli.* It would be interesting to see how chromosomal integration affects amorphadiene titers in this system.

In summary, secondary metabolite production in simple hosts such as *E. coli* has been an attractive proposition for metabolic engineers for quite some time. However, the absence of systematization had hindered the progress towards achieving repeatable, generalizable and efficient production of these molecules. In that regard, modularization and multivariate optimization of metabolic pathways are likely to be useful concepts in new metabolic engineering ventures. Not only were they indispensable for demonstrating taxadiene and amorphadiene production, but they did so in an organism that was discarded as being unfit for production of complex, cyclized secondary metabolites. Going forward, it is likely that modularization and multivariate optimization would witness increasing use and, by extension, refinement in their ability to engineer superior heterologous producers.

Modularization & combinatorial experimentation

Mutating a single or multiple genes in the host's genome could improve its current phenotype to a more desirable state. However, the scarcity of our knowledge on how individual nucleotides influence a host's phenotype prevents us from performing these mutations by rational design. The less one knows about a pathway, its genes and how it is regulated in the cell, the greater is the uncertainty. Under these circumstances, the only way forward is to randomly mutate a gene or several genes, test the influence of these mutations on the phenotype and then iterate this process until a desired phenotype is attained. The

greater the number of genes or the greater the number of nucleotides in a single gene that one can mutate and test, the greater is the likelihood of success, but also the greater the need for efficient high throughput screening. This forms the basis of combinatorial experimentation, and, the past decade has witnessed a significant rise in its use for metabolic engineering.

More often than not, one has to manipulate several genes simultaneously to observe a superior phenotype and this usually involves screening an overwhelming number of strains. For example, identifying 4 genes out of a total of 1000 target genes that when manipulated exhibit superior performance over the native strain necessitates screening 1000C or roughly $>4\times10^{10}$ 4 individual strains. It is obvious that combinatorial experimentation would be infeasible without a reliable, high-throughput screen and more importantly, requires significant investments of time, resources and effort (Figure 6). Nevertheless, combinatorial experimentation warrants a place in the metabolic engineering toolbox, particularly for its applicability to probe distal effects on the performance of a pathway, as evidenced by the improvements in ethanol tolerance of *S. cerevisiae* strains that were randomly mutagenized using a method called 'global transcriptional machinery engineering' (gTME) (Alper et al., 2006; Alper and Stephanopoulos, 2007). But while combinatorial experimentation can aid in identifying superior strains, it is not a panacea that can improve the yields, productivities or titers of poorly-engineered strains. In our experience, significant rational construction is required before combinatorial techniques can be effectively employed.

Given a strain that has been engineered using all available information, an interesting method, Multiplex Automated Genome Engineering (MAGE), allows new capabilities for introducing diversity for the purposes of combinatorial screening. (Wang et al., 2009). MAGE produces combinatorial genomic diversity by simultaneously modifying several locations on the chromosome of a single cell or across a population of cells, thereby scaling-up directed evolution of cells, a notoriously protracted exercise, by several orders of magnitude. Using MAGE, it is now possible to introduce diversity at the nucleotide level simultaneously across the entire genome.

To demonstrate MAGE's capabilities, Wang *et al.* targeted lycopene biosynthesis in *E. coli*. Herein, all the genes in the MEP pathway were targeted for mutagenesis to generate genomic diversity. The 20 genes whose RBS sequences were manipulated to change translational efficiency were previously known to influence the production of lycopene and other carotenoids(Alper et al., 2005c). Four genes that were targeted for inactivation by introducing nonsense mutations in their sequences were also previously known to negatively impact lycopene production. While the genes were known to affect lycopene synthesis, the exact expression level of each gene was not known. Previous studies had coarsely balanced the MEP pathway, however, MAGE allows for a much finer precision in adjusting expression (Yuan et al., 2006). Although production levels were lower compared to previous attempts to rationally engineer the *E. coli* MEP pathway for the production of lycopene (Alper et al., 2005c; Farmer and Liao, 2000), MAGE was able to make significant improvements in a short amount of time. One limitation of MAGE and other combinatorial approaches, is the need to screen 10^5+ mutants. MAGE was dependent on the colorimetric properties of lycopene that make it amenable for rapid assaying. Implementing MAGE, using a multivariate, modular approach, may have significant advantages if a smaller search space (<100 mutants) may be screened by low-throughput approaches.

Moving forward

Recent breakthroughs in the field have already demonstrated the ability to assess and circumvent regulatory blockades to secondary metabolite overproduction in *E. coli*.

However, it remains to be seen if the regulatory responses observed during taxadiene overproduction are also elicited during over-production of other secondary metabolites. Extending the methodology to other hosts would also be desirable.

One hopes that the ever-declining costs of *de novo* gene synthesis; the improved use of bioinformatic tools to mine, sort and analyze biological data; and the increasing sensitivity and sophistication of investigational tools makes the development of techniques such as MMME and MAGE autocatalytic processes. The improvements in experimental and analytical techniques should encourage other groups to also apply modularization and multivariate optimization to overproduce other value-added chemicals. These demonstrations, in turn, would add to the information base that would then enable metabolic engineers and synthetic biologists to design modules in an objective-oriented manner, making judicious use of technologies such as enzyme scaffolding and chromosomal integration in the process.

The next decade holds great excitement and challenges for metabolic engineers and synthetic biologists. The repertoire of bio-derived products will grow and many new technologies that modulate and control cellular processes will emerge. However, analytical techniques will continue to remain the rate-limiting step in strain engineering, and, developing better, faster and broader analytical platforms must assume top priority. In the absence of high-throughput screening, combinatorial techniques cannot be applied to further improve the performance of rational design and engineering. Better analytical capabilities are central to systematizing metabolic engineering along the framework of modularization presented herein.

Acknowledgments

References

Ajikumar PK, et al. Terpenoids: Opportunities for biosynthesis of natural product drugs using engineered microorganisms. Molecular Pharmaceutics. 2008; 5:167–190. [PubMed: 18355030]

Ajikumar PK, et al. Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. Science. 2010; 330:70–4. [PubMed: 20929806]

Alper H, et al. Tuning genetic control through promoter engineering. PNAS. 2005a; 102:12678–12683. [PubMed: 16123130]

Alper H, et al. Tuning genetic control through promoter engineering. Proceedings of the National Academy of Sciences of the United States of America. 2005b; 102:12678. [PubMed: 16123130]

Alper H, et al. Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets. Nat Biotechnol. 2005c; 23:612–6. [PubMed: 15821729]

Alper H, et al. Engineering yeast transcription machinery for improved ethanol tolerance and production. Science. 2006; 314:1565–8. [PubMed: 17158319]

Alper H, Stephanopoulos G. Global transcription machinery engineering: a new approach for improving cellular phenotype. Metab Eng. 2007; 9:258–67. [PubMed: 17292651]

Alper H, Stephanopoulos G. Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? Nature Reviews Microbiology. 2009; 7:715–723.

Anthony JR, et al. Optimization of the mevalonate-based isoprenoid biosynthetic pathway in Escherichia coli for production of the anti-malarial drug precursor amorpha-4,11-diene. Metab Eng. 2009; 11:13–9. [PubMed: 18775787]

Bailey JE, et al. Inverse metabolic engineering: A strategy for directed genetic engineering of useful phenotypes. Biotechnol Bioeng. 1996; 52:109–21. [PubMed: 18629857]

Bentley WE, et al. Plasmid-encoded protein: the principal factor in the "metabolic burden" associated with recombinant bacteria. Biotechnol Bioeng. 1990; 35:668–81. [PubMed: 18592563]

Blazeck J, Alper H. Systems metabolic engineering: Genome-scale models and beyond. Biotechnology Journal. 2010; 5:647–569. [PubMed: 20151446]

Bond-Watts BB, et al. Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways. Nat Chem Biol. 2011; 7:222–7. [PubMed: 21358636]

Chemler JA, Koffas MAG. Metabolic engineering for plant natural product biosynthesis in microbes. Current opinion in biotechnology. 2008; 19:597–605. [PubMed: 18992815]

De Mey M, et al. Construction and model-based analysis of a promoter library for E. coli: an indispensable tool for metabolic engineering. BMC Biotechnol. 2007; 7:34. [PubMed: 17572914]

Dueber JE, et al. Synthetic protein scaffolds provide modular control over metabolic flux. Nat Biotechnol. 2009; 27:753–9. [PubMed: 19648908]

Dugar D, Stephanopoulos G. Relative potential of biosynthetic pathways for biofuels and bio-based products. Nature Biotechnology. 2011; 29:1074–1078.

Farmer WR, Liao JC. Improving lycopene production in Escherichia coli by engineering metabolic control. Nat Biotechnol. 2000; 18:533–7. [PubMed: 10802621]

Gershenzon J, Dudareva N. The function of terpene natural products in the natural world. Nature chemical biology. 2007; 3:408–414.

Jones CG, Firn RD. On the evolution of plant secondary chemical diversity. Phil. Trans. R. Soc. Lond. B (Biological Sciences). 1991; 333:273–280.

Jones KL, et al. Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. Metabolic Engineering. 2000; 2:328–338. [PubMed: 11120644]

Kamm B, Kamm M. Principles of biorefineries. Applied Microbiology and Biotechnology. 2004; 64:137–145. [PubMed: 14749903]

Keasling JD. Synthetic biology for synthetic chemistry. ACS Chem Biol. 2008; 3:64–76. [PubMed: 18205292]

Keasling JD. Manufacturing molecules through metabolic engineering. Science. 2010; 330:1355–8. [PubMed: 21127247]

Kholodenko BN, Westerhoff HV. Metabolic engineering in the post genomic era. 2004

Kirby J, Keasling JD. Biosynthesis of plant isoprenoids: perspectives for microbial engineering. Annu Rev Plant Biol. 2009; 60:335–55. [PubMed: 19575586]

Klein-Marcuschamer D, et al. Engineering microbial cell factories for biosynthesis of isoprenoid molecules: beyond lycopene. Trends Biotechnol. 2007; 25:417–24. [PubMed: 17681626]

Lee JY, et al. Engineering butanol-tolerance in escherichia coli with artificial transcription factor libraries. Biotechnology and Bioengineering. 2011; 108:742–749. [PubMed: 21404248]

Leonard E, et al. Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. Proc Natl Acad Sci U S A. 2010; 107:13654–9. [PubMed: 20643967]

Martin VJ, et al. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. Nat Biotechnol. 2003a; 21:796–802. [PubMed: 12778056]

Martin VJJ, et al. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. Nature Biotechnology. 2003b:796–802.

Muntendam R, et al. Perspectives and limits of engineering the isoprenoid metabolism in heterologous hosts. Appl Microbiol Biotechnol. 2009; 84:1003–19. [PubMed: 19669755]

Newman JD, et al. High-level production of amorpha-4,11-diene in a two-phase partitioning bioreactor of metabolically engineered Escherichia coli. Biotechnol Bioeng. 2006; 95:684–91. [PubMed: 16878333]

Nishizaki T, et al. Metabolic engineering of carotenoid biosynthesis in Escherichia coli by ordered gene assembly in Bacillus subtilis. Appl Environ Microbiol. 2007; 73:1355–61. [PubMed: 17194842]

Noack D, et al. Maintenance and genetic stability of vector plasmids pBR322 and pBR325 in Escherichia coli K12 strains grown in a chemostat. Mol Gen Genet. 1981; 184:121–4. [PubMed: 6278246]

Park JH, Lee SY. Towards systems metabolic engineering of microorganisms for amino acid production. Current opinion in biotechnology. 2008; 19:454–460. [PubMed: 18760356]

Pfleger BF, et al. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nature Biotechnology. 2006; 24:1027–1032.

Pitera DJ, et al. Balancing a heterologous mevalonate pathway for improved isoprenoid production in Escherichia coli. Metab Eng. 2007; 9:193–207. [PubMed: 17239639]

Salis HM, et al. Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol. 2009; 27:946–50. [PubMed: 19801975]

Santos CNS, Stephanopoulos G. Combinatorial engineering of microbes for optimizing cellular phenotype. Current opinion in chemical biology. 2008; 12:168–176. [PubMed: 18275860]

Shen CR, et al. Driving forces enable high-titer anaerobic 1-butanol synthesis in Escherichia coli. Appl Environ Microbiol. 2011; 77:2905–15. [PubMed: 21398484]

Shetty RP, et al. Engineering BioBrick vectors from BioBrick parts. Journal of Biological Engineering. 2008:2. [PubMed: 18304325]

Stafford DE, Stephanopoulos G. Metabolic engineering as an integrating platform for strain development. Current opinion in microbiology. 2001; 4:336–340. [PubMed: 11378489]

Stephanopoulos G. Challenges in engineering microbes for biofuels production. Science. 2007; 315:801. [PubMed: 17289987]

Stephanopoulos G, et al. Exploiting biological complexity for strain improvement through systems biology. Nature Biotechnology. 2004; 22:1261–1267.

Stephanopoulos G, Sinskey AJ. Metabolic engineering--methodologies and future prospects. Trends in biotechnology. 1993; 11:392–396. [PubMed: 7764086]

Tyo KE, et al. Stabilized gene duplication enables long-term selection-free heterologous pathway expression. Nat Biotechnol. 2009; 27:760–5. [PubMed: 19633654]

Tyo KE, et al. Expanding the metabolic engineering toolbox: more options to engineer cells. Trends Biotechnol. 2007; 25:132–7. [PubMed: 17254656]

Tyo KEJ, et al. Toward design-based engineering of industrial microbes. Current opinion in microbiology. 2010; 13:255–262. [PubMed: 20226723]

Wang HH, et al. Programming cells by multiplex genome engineering and accelerated evolution. Nature. 2009; 460:894–8. [PubMed: 19633652]

Yadav VG, Stephanopoulos G. Reevaluating synthesis by biology. Curr Opin Microbiol. 2010; 13:371–6. [PubMed: 20447859]

Yoon SH, et al. Combinatorial expression of bacterial whole mevalonate pathway for the production of beta-carotene in E. coli. J Biotechnol. 2009; 140:218–26. [PubMed: 19428716]

Yuan LZ, et al. Chromosomal promoter replacement of the isoprenoid pathway for enhancing carotenoid production in E. coli. Metabolic Engineering. 2006; 8:79–90. [PubMed: 16257556]
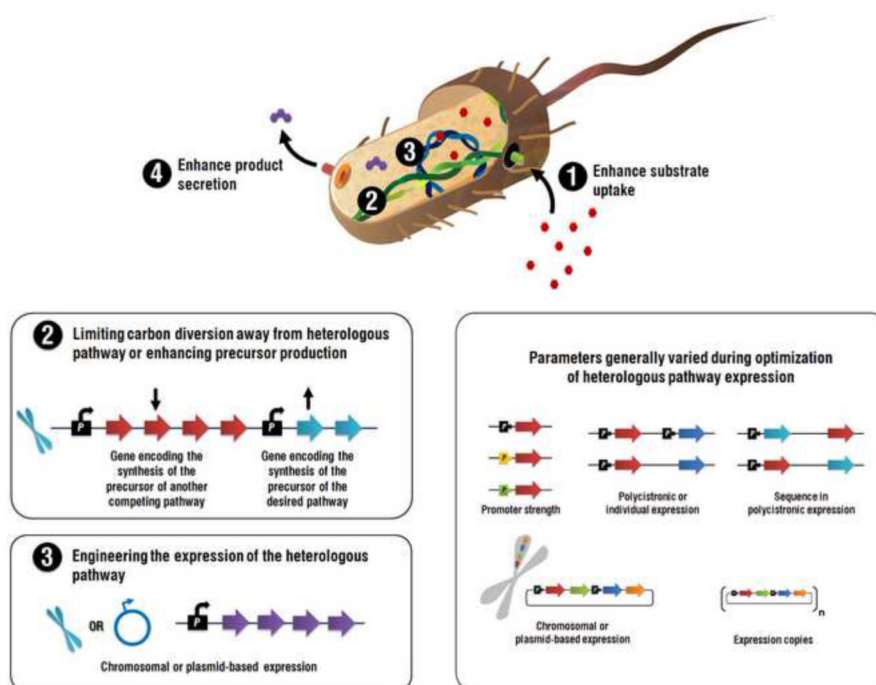
Figure 1. A holistic view of metabolic and cellular engineering
Methods for manipulating the flux from a substrate towards the product can be grouped into four categories: (1) enhancement in the rate of substrate uptake, (2) reduction of flux to undesirable by-products and enhancement of precursor and cofactor flux, (3) introduction of the heterologous pathway and optimization of the activity of its constituent enzymes, and (4) export of the product to the extracellular medium in order to shift equilibrium towards product formation.
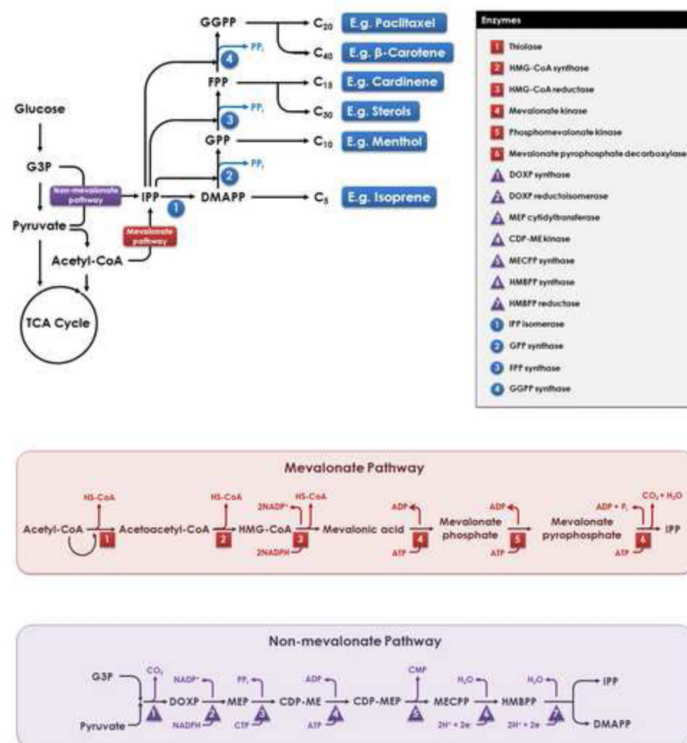
Figure 2. Terpenoid biosynthetic pathways
Terpenoids are naturally synthesized via one of two possible routes – the mevalonate (MVA) pathway or the non-mevalonate or 2-methyl-(*D*)-erythritol-4-phosphate (MEP) pathway. The MVA pathway commences with the co-condensation of acetyl-CoA to form acetoacetyl-CoA. The MEP pathway, on the other hand, commences with the condensation of glyceraldehyde-3-phosphate with pyruvate to yield 2-methyl-(D)-erythritol-4-phosphate. Both pathways culminate in the production of isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) – the universal, 5-carbon terpenoid precursors. IPP then successively condenses with DMAPP and the products of its condensation reactions therewith to yield a homologous series of allylic pyrophosphates that, in turn, serve as precursors to specialized terpenoid producing pathways. The cyclized and branched derivatives of the de-phosphorylated allylic pyrophosphates then undergo a variety of substitution reactions such as hydroxylations and esterifications to yield individual terpenoid molecules.
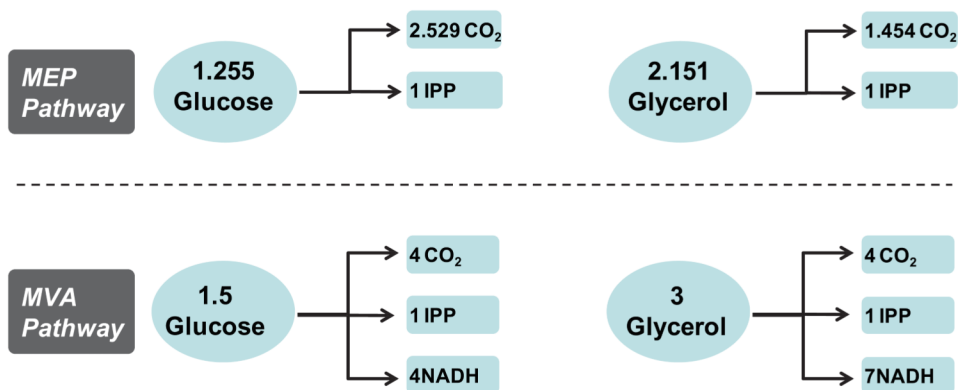
Figure 3. Stoichiometry and redox balance in isoprenoid pathways
Although both, the MVA and MEP pathways culminate in the production of IPP and DMAPP, the MEP pathway is energetically balanced and more efficient than the MVA pathway in its ability to convert sugars or glycerol into terpenoids. The mevalonate MVA consumes 1.5 moles of glucose or 3 moles of glycerol for every mole of IPP produced but generates excess reducing equivalents (NADH) that must be expended in cell growth or secreted as reduced metabolites. This diverts carbon away from the product and decreases the overall yield of IPP. Moreover, excessive NADH can further heighten co-factor imbalances. On the other hand, the MEP pathway utilizes 1.255 moles of glucose or 2.151 moles of glycerol for every mole of IPP produced, but, in contrast with the MVA pathway, is redox balanced.
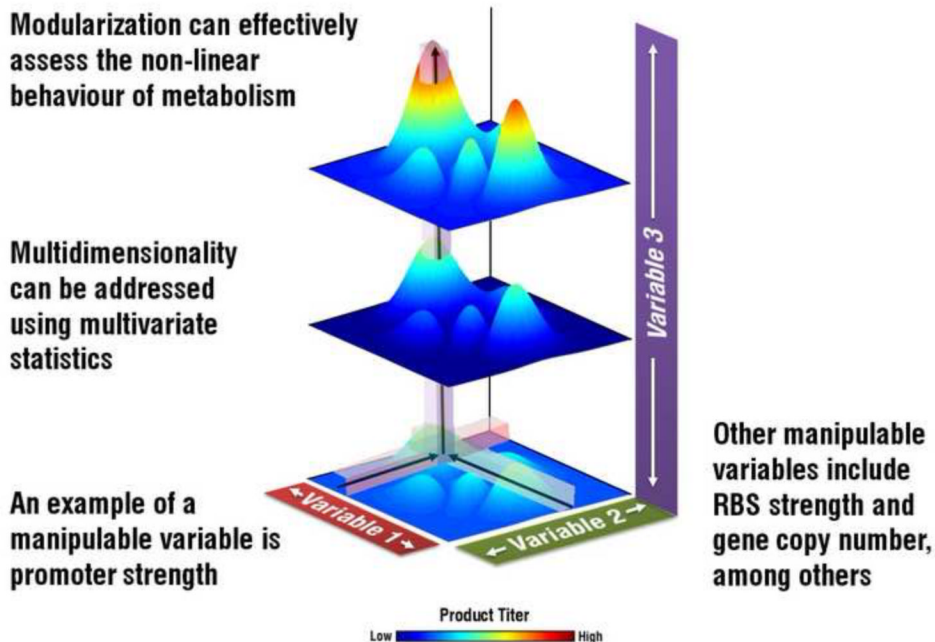
Figure 4. Multivariate modular metabolic engineering (MMME)
The technique assesses and eliminates regulatory and pathway bottlenecks by re-defining the metabolic network as a collection of distinct modules. Each module generally consists of enzymes with similar turnovers. Next, the turnovers of the different modules are equalized by adjusting their concentrations. This maximizes the ratio of pathway turnover to resource expenditure. The practice of re-casting metabolic networks as a collection of interacting modules not only makes the analysis of a very complicated system considerably more tractable, but since the concentrations of a module's enzymes can be adjusted using a variety of transcriptional, post-transcriptional and translational mechanisms, performance of module configurations can be easily assessed using multivariate statistics.
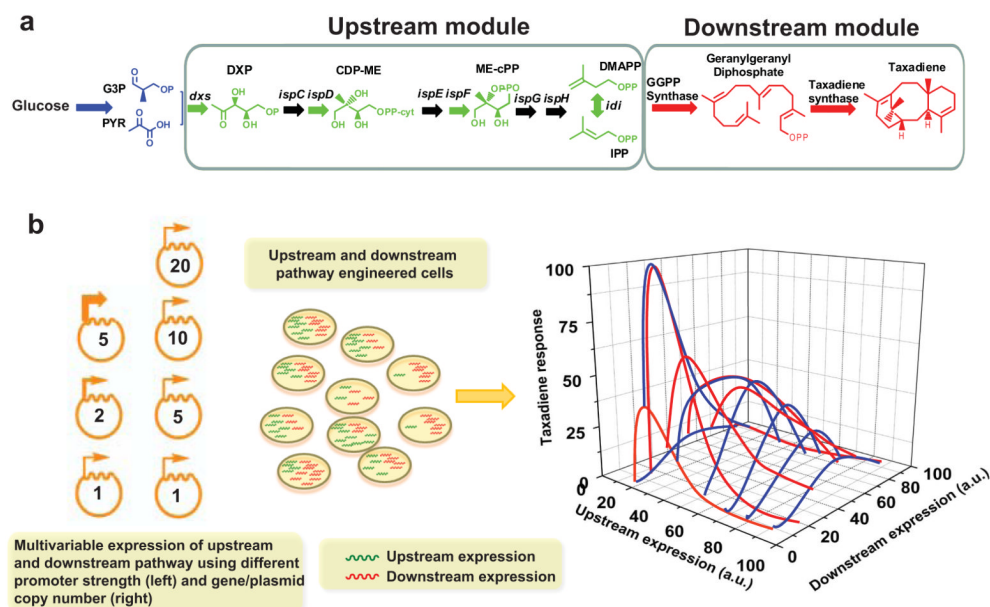
Figure 5. Optimization of taxadiene production by applying multivariate-modular engineering
(a) Schematic of the two modules, the native upstream MEP isoprenoid pathway (green) and synthetic taxadiene pathway (red). **(b)** Schematic of the multivariate-modular isoprenoid pathway engineering approach for probing the non-linear response in terpenoid accumulation from upstream and downstream pathway engineered cells. Expression of upstream and downstream pathways is modulated by varying the promoter strength [**1** (Trc), **2** (T_5) and **5** (T_7)] or gene/plasmid copy number (right). Variation of upstream and downstream pathway expression gives different maxima in taxadiene accumulation.

Multivariate Modular Metabolic Engineering

Combinatorial Methodologies

Modularization of the pathway makes it amenable to identify the most influential variable through multivariate statistical analysis

The success of combinatorial methods is contingent on screening throughput & the starting point

Legend
- Landscape traversed by strains from experiment #1
- Landscape traversed by strains from experiment #2
- Landscape traversed by strains from experiment #3
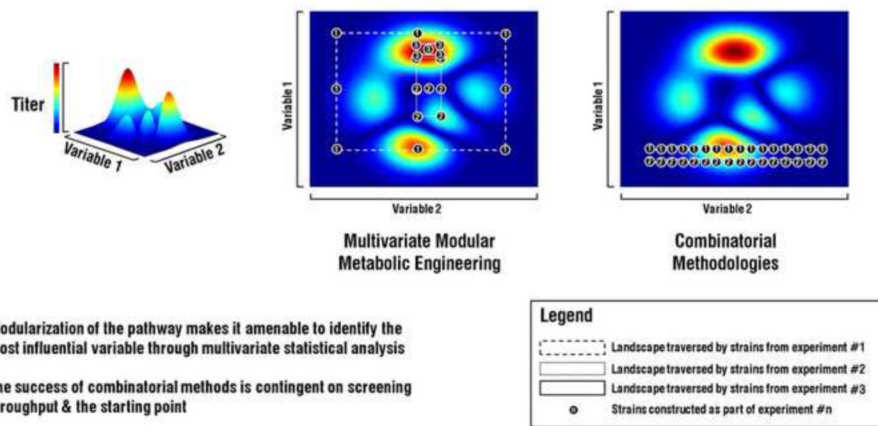- Strains constructed as part of experiment #n

Figure 6. Comparison between MMME and combinatorial approaches
In combinatorial strain optimization, one must manipulate several genes in order to observe a superior phenotype and this usually involves screening an overwhelming number of strains. Combinatorial experimentation may be infeasible without a reliable, high-throughput screen and more importantly, may require significant investments of time, resources and effort. Moreover, the success of a combinatorial screen is often heavily reliant on the starting point. A bad starting mutation could make the process laborious and unsuccessful.