

Metabolic design and control for production in prokaryotes

Swapnil R. Chhabra^{1,3} & Jay D. Keasling^{1,2,3}

*Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California¹;
Departments of Chemical Engineering and Bioengineering, University of California, Berkeley, California²;
Joint BioEnergy Institute, Emeryville, California³.*

This work conducted by the Joint BioEnergy Institute was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

Synopsis

Prokaryotic life on earth is manifested by its diversity and omnipresence. These microbes serve as natural sources of a large variety of compounds with the potential to serve the ever growing, medicinal, chemical and transportation needs of the human population. However, commercially viable production of these compounds can be realized only through significant improvement of the native production capacity of natural isolates. The most favorable way to achieve this goal is through the genetic manipulation of metabolic pathways that direct the production of these molecules. While random mutagenesis and screening have dominated the industrial production of such compounds in the past our increased understanding of microbial physiology over the last five decades has shifted this trend towards rational approaches for metabolic design. Major drivers of this trend include recombinant DNA technology, high throughput characterization of macromolecular cellular components, quantitative modeling for metabolic engineering, targeted combinatorial engineering and synthetic biology. In this chapter we track the evolution of microbial engineering technologies from the black box era of random mutagenesis to the science and engineering-driven era of metabolic design.

Introduction

Prokaryotic life forms have been discovered in almost all conceivable environmental niches on earth – from sub-zero temperatures of the arctic sea-ice to hydrothermal vents of the mid-Atlantic ridge where temperatures exceed the normal boiling point of water, from the safe confines of the human gut to some of the most toxic nuclear and chemical waste sites on this planet. The remarkable ability of microbes to thrive in such diverse environments is enabled through the deliberate evolution of their intrinsic metabolic pathways, which has allowed these organisms to adapt to local milieus and sequester carbon and energy sources therein. Not surprisingly these metabolic pathways are capable of executing a mind-numbingly large number of chemical transformations, to produce an equally diverse array of compounds that, coincidentally, satisfies many of the medicinal, chemical and transportation needs of our planet. Compounds of economic importance produced by microorganisms include recombinant proteins, primary metabolites such as amino acids, nucleotides, vitamins, organic acids, first-generation biofuels, and secondary metabolites such as antibiotics, anti-cancer agents, immunosuppressants, cholesterol-lowering agents, anti-parasitic drugs, monomeric constituents of synthetic polymers, bioplastics and most recently advanced biofuels. Natural isolates of microbial strains usually produce only tiny amounts of these products, however. Therefore from a commercial standpoint, strategies have to be developed to improve the production of these compounds.

Metabolic design strategies may be broadly grouped into two types (Fig. 1):
(i) those that can be implemented independently of genome sequence information

or advanced genetic tools, and (ii) those that employ mathematical and synthetic biology tools for the direct genetic manipulation of metabolic pathways.

In the pre-genomics era, the simplest form of metabolic control intervention in the fermentation development laboratory was manipulation of the growth medium or growth conditions to support product biosynthesis. Typically, the best conditions for growth are different from those for product formation. Strategies for optimizing growth for metabolite production included addition of limiting precursors and variation of carbon, nitrogen and inorganic compounds. For instance, use of glucose as the sole carbon source was shown to repress the production of antibiotics in several organisms, e.g. actinomycin (*Streptomyces antibioticus*), puromycin (*Streptomyces alboniger*), cephalosporin (*Cephalosporium acremonium* and *Streptomyces clavuligerus*) and penicillin (*Penicillium chrysogenum*) [34]. Industrial production of these compounds was therefore carried out using lactose or other sugars, including starch, which was only slowly broken down to glucose, as the carbon source. Manipulation of the culture conditions, for example adjusting the level of dissolved oxygen in the fermenter, or regulating the feeding of a nutrient to prevent the accumulation of acetate, is also widely used to improve product titers. The single-most successfully employed approach, used by all pharmaceutical companies to improve the titers of their desired fermentation products (drugs), termed *mutagenesis and screening* does not provide understanding of how the desired compound is produced (biochemical pathway) or how it is regulated. Thus, it is not possible to apply a rational design method to this approach.

Technologies for microbial engineering have come a long way from simple mutagenesis techniques of the 1960s to present-day *de novo* construction of customized genetic circuits and complete genomes. Contemporaneous developments in high throughput characterization of macromolecular cellular components and computational tools for subsequent data management and analyses have vastly increased our understanding of microbial physiology. There has been a near exponential growth in the pace of microbial genome sequencing over the past fifteen years since the first bacterial genome (*Haemophilus influenzae*) was reported. Today, annotated genomes of well over a thousand bacterial species are available in online databases. Several bacterial genomes have been functionally characterized via gene expression profiling facilitated by DNA microarray technology and next-generation sequencing tools, as well as protein abundance profiling facilitated by mass spectrometry. Concurrently, targeted manipulations of genomes and components therein have been facilitated by developments in DNA assembly methods, *de novo* syntheses, mathematical modeling and computer aided-design. In the following sections we track the evolution of microbial engineering technologies from the black box era of random mutagenesis to the science and engineering-driven era of metabolic design.

Classical Mutagenesis

Major developments in industrial strain improvement in the second half of the twentieth century relied on “brute force” genetic manipulations using chemical or physical agents called mutagens. Examples of mutagens include radiation (X rays, ultraviolet rays), chemicals (5-chlorouracil, nitrous acid), alkylating agents (N-

methoxy-*N'*-nitro *N*-nitrosoguanidine (NTG), ethyl methanesulfonate) and intercalating agents (ethidium bromide, acridine dyes) [1].

Exposure of microbes to mutagens results in permanent alteration of nucleotides in their respective genomes. The process of selecting improved strains follows three basic steps: (i) mutagenesis of the population to create genetic variants, (ii) random selection or screening from the surviving population to find an improved strain, and (iii) identification of improved strains through assay of fermentation broth for products [1]. Each time an improved strain is obtained, it is reused as a parent strain and the process is repeated under the constraints of resource availability and time. The efficiency of random selection depends on several factors such as the mutagen dose and exposure time, culture type, type of DNA damage, conditions of treatment and post-treatment and the extent of yield increase.

Despite being laborious and time consuming, this approach continues to be used in industry due to its long history of success. Classical strain improvement has been the main factor behind improved production titers of several antibiotics such as penicillin (70 gL⁻¹), cephalosporin C (30 gL⁻¹), tylosin (15 gL⁻¹), salinomycin (60 gL⁻¹), chlortetracycline (30 gL⁻¹) and tetracycline (35 gL⁻¹) in the past few decades [1]. Continued development in high throughput screening and analytical technologies such as liquid chromatography-mass spectrometry (LCMS), fluorescence activated cell sorting (FACS), and robotic miniaturization of assays, which enable screening and evaluating large mutant libraries, has kept the classical approach alive [29]. Also from a regulatory perspective, this approach has always had more acceptance than some of the newer recombinant DNA-based technologies

[29]. One of the drawbacks of the classical approach is that it introduces secondary mutations into the genome, which may accumulate in one background over the course of many cycles of mutagenesis. The final production strains may therefore display poor growth compared to the corresponding wild type strains [18]. Protoplast fusion of the former with a low-titer, rapidly growing strain may be used to address this issue. An alternative, recently proposed strategy, called “genome-based strain reconstruction,” merges the classical approach with rapid genome sequencing and genetic manipulations. In this approach, the genome of the producer strain is analyzed post-mutagenesis, and a minimal set of beneficial mutations for high-level production is identified and then assembled in a single wild-type background. This approach was applied to L-lysine production in *Corynebacterium glutamicum* [18]. The resulting mutant strain displayed higher growth rate and sugar consumption as well as improved stress-tolerance as compared to the classically-derived producer strain. In principle, this approach may be applied for secondary metabolite producers provided the pathways for these molecules are known and methods for genetic manipulation have been established.

Protoplast Fusion and Genome Shuffling

In addition to induced mutations, new genotypes may be generated in microorganisms by recombination between different bacterial genera or mutant strains of a single species. Protoplast fusions between strains for combining distinct phenotypic traits have been applied extensively in yeasts. In prokaryotes, protoplast fusion methods have been developed for many species of *Streptomyces*, *Lactobacillus* and *Escherichia coli*. Verma and coworkers reported intergeneric protoplast fusion between *Gluconobacter oxydans* and *Corynebacterium* spp. (ATCC

9937) to generate a hybrid strain capable of converting glucose to 2-keto-L-gulonic acid, a precursor for the production of vitamin C [39]. The hybrid strain combined traits of both parents albeit with low conversion efficiency: *G. oxydans* can oxidize glucose to gluconate and its 2,5-diketo derivative while *Cornebacterium* spp. (ATCC 9937) encodes for a ketoreductase enzyme system, which can reduce 2,5-diketogluconate to 2-keto-L-gluconic acid. Reports of intergeneric fusion in prokaryotes are relatively rare, perhaps due to low recombination/regeneration efficiencies.

In 2002, Zhang and coworkers used recursive protoplast fusion for fusing multiple parent strains [44]. This approach termed as 'whole-genome shuffling' was used for rapid evolution of *Streptomyces fradiae* for the production of the antibiotic tylosin from a population of classically improved strains of this organism. The authors were able to demonstrate that only two rounds of genome-shuffling, corresponding to 24,000 assays performed in a period of one year, resulted in the same final titers (8 gL⁻¹) as roughly 1,000,000 assays performed in a period of 20 years through the classical approach. This approach has been extended to other bacteria such as *Lactobacillus rhamnosus* for enhancing acid tolerance and improving lactic acid production, *Sphingobium chlorophenicum* for improved pesticide (pentachlorophenol) degradation, *Streptomyces* sp. U121 for improved hydroxycitric acid production and *Sorangium cellulosum* for improved production of epothilones.

Recombinant DNA Technology and First Generation Metabolic Engineering

Classical mutagenesis and protoplast fusions for strain improvement are essentially 'black box' techniques in that they can be applied without prior

knowledge of the genomic properties of the host microbe. The advent of recombinant DNA technology in the early 1970s enabled modifications of microorganisms by direct genetic manipulation and revolutionized the application of design strategies for controlling metabolic pathways. Recombinant DNA technology enabled not only the export of genes and pathways from one organism to another but also precise control of their expression through the manipulation of genetic features such as the plasmid copy number, promoter strength and mRNA stability. In addition site-directed and random mutagenesis could be used to improve properties of individual genes in the native or non-native pathways. The term 'Metabolic Engineering' is defined as the directed improvement of cellular activities through the modification of specific biochemical processes or the introduction of heterologous ones, with the use of recombinant DNA technology [38]. In essence it involves combining analytical methods to quantify fluxes and their control with molecular biological techniques to construct novel metabolic configurations or to manipulate the flux through natural pathways [38]. First generation metabolic engineering efforts demonstrated successful incorporation of new enzymatic functions and synthetic operons in genetically amenable hosts as well as amplification of existing metabolic fluxes towards desired compounds by enhanced expression of pathway components, removal of competing pathways or relief of feedback inhibition in these organisms. Among prokaryotes, *E. coli* and *C. glutamicum* have been engineered for the synthesis of a wide variety of metabolites. This can be attributed to the extensive physiological knowledge of these microbes, well-known features of central carbon metabolism, availability of genome sequence information and well-established genetic tools (Fig. 1). Expanding these knowledge

areas to 'non-model' microbes with superior native-metabolite-production capabilities may be a worthwhile exercise. For instance, *Saccharopolyspora erythraea* and many *Streptomyces* species may be better suited for polyketide assembly [42]. Similarly, *Actinobacillus succinogenes*, *Mannheimia succiniciproducens* and *Anaerobiospirillum succiniciproducens* may be better suited for succinic acid production [37], while *Klebsiella pneumoniae*, *Citrobacter freundii* and *Clostridium butyricum* hold great promise for the production of 1,3-propanediol [7]. In this section we highlight some of the major developments in metabolic design and engineering of *E. coli* and *C. glutamicum*.

In the early 1980s, Ensley *et al.* demonstrated biosynthesis of the textile dye indigo in *E. coli* by expressing a naphthalene dioxygenase gene contained on the self-transmissible plasmid NAH7 from *Pseudomonas putida* PpG7 [11]. This biochemical route offered an environmentally safer alternative to the chemical process employed for generating this dye molecule. However, this first generation strain was able to produce indigo only when supplemented with the high-cost pathway-intermediates indole or tryptophan. A decade later Murdock and coworkers developed a second-generation strain capable of direct production of indigo from glucose [25]. In their metabolic engineering effort the authors employed a mixture of several strategies: (i) amplification of the ferredoxin component of the naphthalene dioxygenase complex to improve its half life in *E. coli*; (ii) stabilization of the iron-sulfur center of ferredoxin by amino acid substitution of His47 with a Cys residue; (iii) engineering of the native *E. coli* tryptophan synthetase by two amino acid substitutions to increase indole production; and (iv) incorporation of a tryptophan overproducing plasmid in the same system. The resulting strain was

able to produce up to 135 mgL⁻¹ of indigo from glucose in shake flasks, which was further improved by manipulation of growth conditions.

Most metabolic engineering efforts in *C. glutamicum* to date have focused on the production of amino acids. Aromatic amino acids are widely used in the food and drug industry: L-tryptophan is used as a feed additive and mood enhancer, L-phenylalanine is a major component of the low-calorie sweetener aspartame while L-tyrosine is used in the production of the anti-Parkinson's disease drug L-dopa. In the early 1990s, Katsumata and Ikeda reported the first successful attempt of tryptophan overproduction in *C. glutamicum* using rational design [17]. The parent strain used for this effort (a phenylalanine and tyrosine double auxotroph) was generated through multiple rounds of classical mutagenesis producing ~8 gL⁻¹ tryptophan. Subsequent improvements in this strain involved a combination of rational design and combinatorial screening of feedback regulated genes from the aromatic acid biosynthesis pathway. The following strategies were implemented: (i) amplification of feedback-deregulated, 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase (DS) that was obtained from a regulatory mutant of *C. glutamicum* (from another study), and (ii) amplification of tryptophan insensitive versions of the genes encoding anthranilate synthase and anthranilate phosphoribosyltransferase. Feedback insensitive versions of the aforementioned enzymes were obtained by screening a mutant library of the corresponding plasmid-encoded genes. These regulatory modifications resulted in an engineered strain capable of producing ~43 gL⁻¹ tryptophan, a 54% yield increase compared to its parent. Further improvement was achieved by redirecting carbon flux from central metabolism to the aromatic pathway by amplifying the transketolase gene.

This resulted in an increase in the supply of the pathway precursor, erythrose 4-phosphate. The genes for feedback insensitive enzymes described above were also incorporated into the same low copy plasmid bearing the transketolase gene. The final strain was reported to produce 58 gL⁻¹ of tryptophan using sucrose as the carbon source. Since aromatic amino acids have a common biosynthetic pathway, tryptophan-producing strains of *C. glutamicum* could potentially be re-engineered to overproduce tyrosine or phenylalanine as well. This was demonstrated by Ikeda and Katsumata [16]. Three biosynthetic genes encoding the first enzyme in the common pathway, DS, and end-product-desensitized versions of the branch-point enzymes chorismate mutase and prephenate dehydratase were individually cloned from regulatory mutants of *C. glutamicum*. Overexpression of different combinations of these genes in a tryptophan-producing (18 gL⁻¹) *C. glutamicum* parent strain resulted in redirection of the carbon flow for the overproduction of tyrosine (26 gL⁻¹) or phenylalanine (28 gL⁻¹).

E. coli has also been engineered for the production of the aromatic amino acids L-phenylalanine and L-tryptophan using similar approaches. Tyrosine overproduction in *E. coli*, however, had not been described until recently. Lütke-Eversloh and Stephanopoulos used strategies similar to those described for *C. glutamicum* [23]: (i) employment of feedback resistant versions of DS (AroG) and chorismate mutase (TyrA), (ii) over-expression of phosphoenolpyruvate synthase (PpsA) and transketolase (TktA) responsible for the supply of pathway precursors, phosphoenolpyruvate and erythrose-4-phosphate; and (iii) elimination of a global regulator (TyrR) of aromatic amino acid biosynthesis. The final *E. coli* strain bearing all three modifications was capable of producing ~9.7 gL⁻¹ L-tyrosine with a yield of

L-tyrosine per glucose of 0.102 g/g. In another study, a L-phenylalanine-producing *E. coli* strain (generated through the classical approach) was converted to a L-tyrosine-producing strain using the following approach: (i) deletion of the chromosomal region encoding for *pheA* (encoding chorismate mutase/prephenate dehydratase), its leader peptide (*pheL*) and its associated promoter, and (ii) over-expression of *tyrA*, (chorismate mutase/prephenate dehydrogenase) driven by a strong non-native *trc* promoter [30]. Optimization of fermentation conditions for the engineered strain resulted in final titers of 55 gL⁻¹ tyrosine.

Unlike *C. glutamicum*, *E. coli* can readily metabolize a variety of hexose and pentose sugars derived from renewable feedstocks using only mineral salts as nutrients. During sugar fermentation, however, *E. coli* produces a mixture of compounds to balance reducing equivalents generated during glycolysis. The distribution of fermentation products can be altered by blocking branches of the existing fermentative pathways or by introducing critical branch point enzymes. These manipulations have been applied in *E. coli* for the production of several commodity chemicals such as lactic acid, succinic acid, pyruvate, ethanol and 1,3-propanediol, some of which are discussed below.

Among biologically-derived commodity chemicals, ethanol has been touted as a biofuel additive with great commercial impact. It has the potential to be a fully sustainable energy resource when coupled with efficient utilization of renewable lignocellulosic biomass. Glucose fermentation in wild-type *E. coli* can result in ethanol formation from pyruvate in two NADH-dependent steps. Considering that the conversion of sugars to pyruvate results in only one NADH, which may also be used in other fermentation reactions such as succinic acid formation, it is not

surprising that ethanol production in wild-type *E. coli* is highly inefficient. Organisms such as *Zyomonas mobilis* encode for a pyruvate decarboxylase (PDC) that depends on thiamine diphosphate and Mg²⁺ as cofactors for non-oxidative decarboxylation of pyruvate to acetaldehyde and CO₂. Based on this property of the *Z. mobilis* enzyme, Ingram and coworkers demonstrated that *lac* promoter-driven expression of PDC and alcohol dehydrogenase from *Z. mobilis* in *E. coli* resulted in significant conversion of glucose to ethanol. In the first generation strain, the succinate pathway was disrupted by elimination of the terminal pathway gene, fumarate reductase (*frd*). Second generation strains had the aforementioned plasmid-encoded genes integrated into the *E. coli* chromosome for enhanced strain stability [26]. Chromosomal incorporation, however, resulted in lower titers of ethanol than the plasmid-bearing strains, most likely due to the reduction in the number of copies of the heterologous genes resulting in the lowered effective concentrations of the required enzymes. Additional strain engineering efforts involved selection of mutants with increased expression of the chromosomally-integrated genes and those with higher tolerance to ethanol than the wild-type strain. The final engineered strain, KO11, was reported to produce ethanol close to theoretical maximum of 0.51 g ethanol/g sugar under optimal growth conditions.

Other important metabolites derived from the central metabolic pathway include succinic and pyruvic acids. Succinic acid is used as an intermediate in the chemical synthesis of other commodity chemicals such as 1,4-butanediol, tetrahydrofuran, and adipic acid, while pyruvic acid is used as a food additive, nutraceutical, and weight-control supplement. Wild-type succinic acid overproducers such as *A. succinogenes*, *M. succiniciproducens* and *A.*

succiniciproducens and their variants have been evaluated for commercial production solely by manipulation of growth conditions [37]. However, the lack of well developed genetic tools and genomic sequence information until recently has hampered improvement and elimination of by products such as acetic, formic, propionic and pyruvic acids in these microbes. Wild-type *E. coli* ferments glucose producing minor amounts of succinic acid (0.13 g/g glucose) under anaerobic conditions. Metabolic engineering strategies developed to enhance succinic acid production by *E. coli* have typically included (i) inactivation of routes that compete with succinic acid production (lactate dehydrogenase, pyruvate formate lyase, glucose-specific phosphotransferase system permease), (ii) amplification of the corresponding genes for the enzymes involved in succinic acid pathways (phosphoenolpyruvate carboxylase, malic enzyme, pyruvate carboxylase) and (iii) the introduction of heterologous enzymes to improve succinic acid formation [37]. In a recent study Sanchez *et al.* constructed an *E. coli* strain capable of producing succinate using a combination of two pathways such that the combined pathways resulted in efficient conversion of glucose to succinate using 1.25 moles of NADH per mole of succinate in contrast to the sole fermentative pathway, which uses 2 moles of NADH per mole of succinate [35]. The metabolic design strategies involved in this effort included (i) elimination of competing pathways (deletion of lactate dehydrogenase, aldehyde dehydrogenase, acetyl kinase and phosphate acetyl transferase genes) to conserve carbon atoms in the form of acetyl-CoA and help channel the acetyl-CoA to the glyoxylate pathway leading to succinate formation; (ii) repression-relief of the enzymes of the glyoxylate pathway (deletion of the repressor IclR); and (iii) heterologous overproduction of pyruvate carboxylase from

Lactococcus lactis to increase the flux from pyruvate to oxaloacetate. The final recombinant strain was shown to produce ~0.98 g succinic acid/g glucose with minute quantities of other metabolites, such as formate and acetate. Although metabolically engineered strains of *E. coli* are remarkable in their performance compared with the wild-type strain, specific and volumetric succinic acid productivities of these strains are much lower than those obtained from natural overproducers such as *A. succiniciproducens* and other rumen bacteria. In the latter, the problem of undesirable fermentation products needs to be solved [37].

Another example of purposeful redirection of *E. coli*'s central metabolism is that of pyruvate overproduction [5]. The goal of this work was to demonstrate development of a strain that could thrive in minimal medium supplemented with an inexpensive carbon source, a step towards improving process economics over current microbial production routes for this metabolite that employ complex fermentation broths. The design strategies in this endeavor included (i) elimination of pathways that consume pyruvate – those involved in the formation of acetate, lactate, formate, ethanol and succinate – and (ii) introduction of mutations that reduce the utilization of pyruvate for cell growth – minimizing ATP yield and CO₂ production. The mutation to minimize ATP yield ($\Delta atpFH$) resulted in an increased rate of glycolysis by providing a cytoplasmic F₁-ATPase. This increased the availability of ADP for glycolysis and minimized allosteric inhibition attributed to ATP accumulation. In mineral salts medium containing glucose as the sole carbon source, the engineered strain TC44 ($\Delta focA$ -*pflB*, $\Delta frdBC$, $\Delta ldhA$, $\Delta atpFH$, $\Delta adhE$, $\Delta sucA$, *poxB::FRT* $\Delta ackA$) converted glucose to pyruvate with a yield of 0.75 g pyruvate/g glucose (77.9% of theoretical yield).

One of the most celebrated examples in recent years demonstrating the power of novel metabolic design, is the production of 1,3-propanediol in *E. coli* K12 [7]. 1,3-Propanediol serves as the monomeric component in the synthesis of polyesters and polyurethanes. Natural producers of this metabolite that have been well studied include *K. pneumoniae*, *C. freundii* and *C. butyricum*, all of which use glycerol for their growth [7]. A more desirable industrial process would convert a low-cost carbon source such as glucose to the desired 1,3-propanediol end product. Efforts to introduce the glycerol pathway into a natural 1,3-propanediol producer or the 1,3-propanediol pathway into a natural glycerol producer were met with limited success associated with natural regulation of the native pathways in the base organisms. A third approach attempted by Genencor International and DuPont involved introduction of both the glycerol and the 1,3-propanediol pathways into *E. coli* K12. Since *E. coli* has only a weak capacity to produce glycerol and no capacity to produce 1,3-propanediol, it was expected that there would be no natural regulation to overcome for the production 1,3-propanediol. Expanding the glycerol-based natural process to a more efficient glucose-based process involved several steps: (i) switching from an anaerobic process to an aerobic one; (ii) replacing the glucose uptake mechanism of *E. coli*; (iii) intergeneric transfer of complex metabolic pathways; and (iv) design and implementation of an optimum solution to the balance of carbon, redox and energy with respect to microbial growth and product formation. The engineered *E. coli* strain relied on a heterologous carbon pathway that diverted carbon from dihydroxyacetone phosphate to 1,3-propanediol through the following modifications: (i) conversion of glucose to glycerol by incorporation of glycerol 3-phosphate dehydrogenase and glycerol 3-phosphate phosphatase genes,

from *Saccharomyces cerevisiae*; (ii) conversion of glycerol to 3-hydroxypropionaldehyde by incorporation of the glycerol dehydratase complex and its reactivating factors from *K. pneumoniae*; (iii) conversion of 3-hydroxypropionaldehyde to 1,3-propanediol using an oxidoreductase (yqhD) native to *E. coli* (the advantage of using this native enzyme was in its utilization of NADPH as a cofactor in contrast to NADH used by dehydrogenases from natural 1,3-propanediol producers); (iv) deletion of pathways that re-direct glycerol to central carbon metabolism (glycerol kinase and glycerol dehydrogenase). Additional reactions, such as modulation of rationally targeted enzyme expression levels within and outside of the direct carbon pathway to 1,3-propanediol, were examined to maximize efficiency resulting in a metabolically engineered strain capable of producing 1,3-propanediol at a titer of 135 gL⁻¹, and a weight yield of 51% in D-glucose in fed-batch fermentations, significantly higher than the corresponding anaerobic process based on natural producers.

Recombinant DNA technologies for the production of therapeutic compounds date back to the late 1970s when production of human insulin was first demonstrated in recombinant *E. coli*. Microbial production of drugs and drug precursors has several advantages over total chemical synthesis or extraction from natural resources or engineered mammalian/plant cells. Microbial processes are environmentally benign, relatively easy to engineer, and can be scaled-up for commercial production. Several groundbreaking studies within the past decade have demonstrated microbial synthesis of major classes of plant- and non-plant based therapeutic products in engineered strains of *E. coli* (discussed below) and

other bacteria. These include macrolides, cyclic peptides, terpenes, alkaloids and aromatic polyketides.

Streptomyces species, natural producers of a large variety of antibiotics and other secondary metabolites such as antihelminthic, antitumor and antifungal agents have been further engineered for enhanced production of these compounds [42]. An elegant example of macrolide synthesis in a heterologous host was demonstrated by Pfeifer *et al.*, who engineered *E. coli* for the production of 6-deoxyerythronolide B (6dEB), the macrocyclic core of the antibiotic erythromycin, using propionate as the starting substrate [31]. This substantial metabolic engineering effort involved heterologous expression and assembly of two large protein complexes, post translational modifications of complex components, disruption of a pathway that consumed the starting substrate, enhanced production of a native precursor from the starting substrate and balanced generation of an exogenous precursor from the native precursor. This was accomplished through the following steps. First, the genes encoding the deoxyerythronolide B synthase (DEBS) complex that synthesizes the macrocyclic core of erythromycin from propionyl-CoA and (2*S*)-methylmalonyl-CoA were optimally expressed. The 2-megadalton DEBS complex is composed of three distinct dimeric-subunits – $\alpha_2\beta_2\gamma_2$ which provide 28 distinct active sites for catalysis. The three genes required for complex assembly (DEBS1~3) were ported from the natural producer of 6dEB, *Saccharopolyspora erythraea*. The heterologous complex was found to properly assemble in *E. coli* only when the growth temperature was lowered from 30°C to 22°C. Second, post-translational modification (pantetheinylation) of a subset of the 28 active site residues was accomplished through chromosomal insertion and T7-promoter

driven expression of the phosphopantetheinyl transferase gene (*sfp*) from *Bacillus subtilis*. Third, the integration site of *sfp* was designed to disrupt the native propionate catabolism pathway. Fourth, the supply of propionyl-CoA from propionate was increased by placing the propionyl-CoA synthetase gene (*prpE*) from *E. coli* under a T7 promoter. Fifth, the exogenous metabolite, (2S)-methylmalonyl-CoA, was produced from propionyl-CoA through heterologous expression of the genes encoding the propionyl CoA carboxylase complex (*pccA-pccB*) from *Streptomyces coelicolor*. Sixth, the activity of the biotinylated subunit of the aforementioned complex, PccA, was improved through co-expression of the biotin ligase gene (*birA*) from *E. coli*. The resulting *E. coli* strain was capable of converting exogenous propionate into 6dEB with a specific productivity (0.1 mmol 6dEB/gm total protein) comparable to a high-producing mutant of *S. erythraea* that had been incrementally enhanced through the classical approach for the industrial production of erythromycin.

The first report on *de novo* microbial synthesis of a non-ribosomal peptide (echinomycin) was published by Watanabe and coworkers in 2006 [41]. In this work, the 13-gene echinomycin biosynthetic cluster from *Streptomyces lasaliensis* (a natural producer of this antibiotic) was distributed in three plasmids for expression in *E. coli*. In addition, *fabC*, which encodes the fatty acid biosynthesis acyl carrier protein from *E. coli*, necessary to complete the biosynthetic pathway, and the *sfp* gene (described above) from *B. subtilis*, which is necessary for posttranslational modification of the foreign proteins, were incorporated in the same plasmid system. The multiple plasmids were characterized by compatible origins of replication and antibiotic resistance markers to ensure stability. The authors chose a monocistronic

arrangement in the plasmid assembly process such that each gene was flanked by its own T7 promoter, ribosome-binding site and T7 transcriptional terminator. This was done to minimize both potential premature terminations as well as mRNA degradation associated with large transcripts. The final yield of echinomycin after an 8-day long fed-batch fermentation run using a minimal medium was 0.3 mgL⁻¹.

In a recent study, Zhang and Tang demonstrated the synthesis of type II aromatic polyketides in a heterologous bacterial host [43]. Known bacterial polyketide synthases (PKSs) differ from their fungal counterparts in their final protein form. The minimal bacterial PKS consists of a type II 'ketosynthase (KS)-chain length factor' heterodimer and a dissociated acyl carrier peptide (ACP) that generates the polyketide backbone, which is further modified by dissociated cyclases/aromatases into different aromatic structures. In contrast, the fungal PKS is a megasynthase that contains all of the above catalytic domains in a single polypeptide. Apart from the work done by Pfeifer *et al.*, most attempts at expressing an active soluble form of the bacterial PKS complex in *E. coli* have met with limited success. In order to bypass the problems associated with heterologous assembly of large bacterial protein complexes, Zhang and Tang attempted the 'dissection' and 'reassembly' of a fungal (*Gibberella fujikuroi*) megasynthase into a protein structure that would resemble the bacterial minimal PKS. This notable protein engineering effort was carried out in two steps: (i) the N-terminal SAT (starter-unit:ACP acyltransferase) and the C-terminal TE/CLC (thioesterase/claisen-cyclase) domains of the fungal megasynthase were removed, and (ii) the middle PT (product-template) domain, which controls the C2-C7 cyclization reaction, was replaced with a linker peptide from the DEBS3 (described above) subunit from *S. erythraea*. The

resulting synthetic megasynthase retained the minimal PKS components (KS, malonyl CoA:ACP transferase and ACP) on a single polypeptide with the polyketide synthase activity still intact but the native cyclization capability completely abolished. This engineered PKS was tested for the biosynthesis of anthraquinone, which is derived from a nonaketide backbone after ketoreduction at C9 by a ketoreductase (KR), C7-C12 cyclization/dehydration by an aromatase/cyclase (CYC1) and C5-C14 cyclization by a second-ring cyclase (CYC2), followed by spontaneous third-ring cyclization and second-ring oxidation. Two plasmids bearing the engineered fungal megasynthase and heterologous components described above were introduced into an *E. coli* strain harboring a chromosomally integrated copy of the *sfp* gene. The resulting strain was shown to produce a maximum titer of 3 mgL⁻¹ of anthraquinone, 60 h post induction in a fed batch process using minimal medium.

In 2003, Martin *et al.* successfully engineered *E. coli* for the production of amorpha-4,11-diene the sesquiterpene olefin precursor to artemisinin, a powerful antimalarial drug [24]. All terpenoids originate from the same universal precursors (isopentenyl pyrophosphate (IPP) and its isomer dimethylallyl pyrophosphate (DMAPP)) that are generated through two known biosynthetic pathways – the mevalonate-dependent (MEV) isoprenoid pathway found in eukaryotes and the deoxyxylulose 5-phosphate (DXP) pathway found in most prokaryotes. The precursor molecules, IPP and DMAPP essential to *E. coli* for the prenylation of tRNAs and the synthesis of farnesyl pyrophosphate (FPP), which is used for quinone and cell wall biosynthesis. To bypass the regulatory effects of the native DXP pathway, which synthesizes these essential metabolites in *E. coli*, Martin *et al.* engineered the expression of the mevalonate-dependent pathway from *S. cerevisiae* in this

organism. This eight-gene biosynthetic pathway from *S. cerevisiae* was split into two operons expressed independently across two different plasmids - the 'top' operon, MevT, which transformed acetyl-CoA to (*R*)-mevalonate in three enzymatic steps, and the 'bottom' operon, which converted (*R*)-mevalonate to FPP for subsequent terpenoid synthesis. A third plasmid bearing the amorphadiene synthase gene from *Artemisia annua* (a natural producer of artemisinin) that converts FPP to amorphadiene was incorporated into the same system. To optimize expression of the plant-based synthase, the corresponding gene was synthesized chemically such that its codon usage was optimized for expression in *E. coli*. The final strain was estimated to produce ~23 and 112 mgL⁻¹ of amorphadiene when grown on LB and LB + 0.8% glycerol cultures, respectively. While further optimization efforts were not attempted in this study, this first generation strain was used as a parent for further improvement using combinatorial and synthetic biology tools as described later in this chapter (Fig. 2). In a follow-up study to this work, Chang *et al.*, engineered two derivatives of this strain, capable of producing 8-hydroxycadinene and artemisinic acid, precursors to their medically relevant counterparts, gossypol and artemisinin, respectively [6] (Fig. 2). This study demonstrated the first examples of in vivo production of functionalized terpenoids in *E. coli* at high titer using native plant P450s. P450s carry out selective oxidations of their substrates, usually essential for further chemical transformation of the resulting biosynthetic intermediates to their final bioactive structures. Several strategies were explored to address the challenge of functional expression (such as folding and translation, membrane insertion, co-factor incorporation, post-translational modification, and protein-protein interactions) of these plant-derived enzymes in *E. coli*. These

included (i) codon optimization achieved through *in vitro* gene synthesis; (ii) N-terminal transmembrane engineering, involving deletion of the predicted transmembrane sequence followed by replacement with various N-terminal sequences from heterologous P450s, secretion/solubilization sequences, and a self-assembling membrane protein; (iii) switching the native interacting redox partner to one from a different plant species; and (iv) use of different expression vectors and *E. coli* strain types. The maximum reported titers for *in vivo* production were 60 mgL⁻¹ of 8-hydroxycadinene (*E. coli* DH10B) and 105 mgL⁻¹ of artemisinic acid (*E. coli* DH1) in terrific broth medium supplemented with 2% glycerol.

The examples discussed thus far highlight the myriad of design strategies employed to incorporate foreign genes or even entire pathways from organisms with novel and germane metabolic capabilities into an organism for which there is a well-established genetic toolbox. These strategies made use of the expansive knowledge of biochemical reactions, properties of individual proteins and their interacting partners, as well as intuitive approaches for redirecting existing pathways towards the desired output. In the following sections we discuss design strategies that enable further improvement of these first generation strains through quantitative analyses of metabolic fluxes, sophisticated combinatorial engineering of both pathway and non-pathway components and the application of synthetic biology tools.

Quantitative Approaches for Metabolic Design

Metabolic fluxes are the time-dependent movement of metabolites through a network that may be inferred from measurable quantities using mathematical models [36]. Measurable quantities include rates of substrate uptake,

product/byproduct secretion that can be garnered using stable isotope tracers. Metabolic fluxes are useful for the calculation of theoretical yields, the determination of non-measured metabolite fluxes and the observation of the function of metabolic pathways *in vivo* [38]. Most importantly, flux modeling can establish a quantitative framework for metabolic engineering by examination and simulation of metabolism as a whole, an appreciably better alternative to manual assessment of a limited number of interactions that may fail to detect non-intuitive causal interactions [12]. Two approaches that have been widely proposed for the analysis of metabolic fluxes under the pseudo-steady state assumption are ^{13}C -based flux analysis [36] and constraints-based flux analysis [19]. The former utilizes an isotope labeled carbon source to follow ^{13}C enrichment patterns of intracellular metabolites, and the resulting data are used to estimate *in vivo* fluxes using mathematical models. Difficulties in experimental design and subsequent calculations for large metabolic networks have been cited as the reasons for the low popularity of this method [19]. The constraint-based approach is currently the dominant method used for flux modeling and the basis for a number of algorithms that predict cellular conditions such as intracellular flux distribution, gene expression, gene knockout candidates and others. These include FBA (Flux Balance Analysis), ROOM (Regulatory On/Off Minimization), OMNI (Optimal Metabolic Network Identification), MOMA (Minimization of Metabolic Adjustment), OptKnock, OptStrain and others [19]. Each algorithm is distinguished from the other by its final objective, the unique constraints applied to the respective mathematical model as well as the final output. However, the general principle behind model development is the same. A stoichiometric model is reconstructed using genomic and biochemical

information from the existing literature. This involves setting mass balance equations for the cellular metabolites and accounting for their stoichiometry in the form of matrices and vectors. The model is then simulated by a linear optimization technique using an appropriate objective function (such as maximization of biomass, metabolite production, etc) and constraints that restrict the solution space within the cell's capacity. Common constraints include substrate uptake rate, metabolite excretion rate and maintenance energy. The *in silico* model now has the potential to allow one to evaluate the effects of genetic and environmental perturbations on the cell at a global level and predict what parts of the system need to be modified for a given objective [19]. Model development efforts for host organisms in the pre-genomic era were restricted to their respective central metabolic pathways. The explosion of annotated genomic sequence information in the past two decades led to a rapid increase in the understanding of gene-metabolite-pathway relationships and a corresponding rise in the incorporation of these entities in successive iterations of reconstructed models. Besides *E. coli*, genome-scale stoichiometric models have been developed for several other prokaryotes [19]: *H. influenza*, *Helicobacter pylori*, *M. succiniciproducens*, *B. subtilis*, *Geobacter sulfurreducens*, *Lactobacillus plantarum*, *Lactococcus lactis*, *Mycobacterium tuberculosis*, *Neisseria meningitides*, *Staphylococcus aureus*, and *S. coelicolor*. Models developed for *E. coli* have been applied for improving the production of carotenoids, organic acids and amino acids [12]. As pointed out by Feist and Palsson, while there have been multiple *E. coli* genome-scale models, there is only one unique network for *E. coli* and successive iterations strive to best represent this content [12]. The latest iteration of the reconstructed metabolic

network of *E. coli* generated in 2007, incorporated 2077 reactions, 1260 open reading frames (of which 1159 had been experimentally verified) and 1039 metabolites [12]. In addition, this reconstruction included characterization and quantification of the biomass components and maintenance requirements associated with growth of *E. coli* plus thermodynamic information for the included chemical reactions. The applications of genome-scale models span beyond metabolic engineering into the realms of bacterial evolution, network analysis, phenotypic behavior, biological discovery and synthetic biology. Here we highlight two examples of model based metabolic engineering for improving the production of succinic acid and L-threonine in *E. coli*.

In the previous section we discussed the work of Sanchez *et al.* who constructed an *E. coli* strain capable of efficient production of succinate from glucose by clever manipulation of its central metabolic pathways [35]. In 2005, Lee *et al.* used a model-based approach to identify the best combination of knockout candidates that would enhance succinate production in *E. coli* [22]. The genome-scale model (GSM) used for this study – *E. coli* iJR904 GSM/GPR was the most comprehensive reconstruction of its time. The iJR904 GSM/GPR reconstruction improved upon its parent iJE660a GSM by incorporation of 904 genes and 931 unique biochemical reactions in addition to gene-protein reaction (GPR) associations and other features. The objective function in the optimization run was maximum biomass formation rate with metabolic fluxes corresponding to knockout candidates being set to zero. Under the pseudo-steady-state assumption, the unknown internal fluxes within the underdetermined metabolic reaction network were evaluated, subject to the constraints provided by mass conservation,

thermodynamics, and reaction stoichiometry. Because of limited computational power, knockout combinations of only select genes within the central metabolic pathway of *E. coli* were evaluated. These candidate genes were identified through comparative analysis of the genomes of *E. coli* and succinic acid-overproducing *M. succiniciproducens*. Computational effort notwithstanding, the *in silico* analyses relieved the monumental task of experimentally testing all possible gene-knockout combinations identified through the comparative genomics analysis. The *in silico* simulation results suggested that reducing metabolic flux to pyruvate would be crucial for improving succinic acid production in *E. coli*. Predicted knockout candidates *ptsG* (glucose-specific permease) and *pykFA* (pyruvate kinase I/II) were experimentally validated for the enhanced production of succinic acid in *E. coli*. This triple mutation increased the succinic acid yield (0.2 g/g glucose) over the wild type strain by more than seven fold and the ratio of succinic acid to undesired fermentation products by nine fold.

In another study, Lee *et al.* engineered *E. coli* for L-threonine production using a combination of traditional metabolic design, transcriptome profiling and *in silico* flux response analyses [21]. First, a parent strain was constructed using metabolic and regulatory information from the literature in the following steps: (i) feedback deregulation of enzymes from the native pathway – aspartokinase I (inhibited by threonine) and aspartokinase III (inhibited by lysine) – by active site modifications; (ii) removal of transcriptional attenuation of the mutated threonine biosynthesis operon (*thrA*^{C1034T}*BC*) by replacing the native promoter with a *tac* promoter, which enables constitutive expression in the base strain (Δ *lacI* mutant of WL3110) (The mutated *thrA*^{C1034T}*BC* operon was further amplified by incorporating

a medium copy plasmid bearing these genes.); and (iii) disabling pathways that degrade threonine or compete with its formation: deletion of diaminopimelate decarboxylase, homoserine succinyl transferase and threonine dehydrogenase. Also, the activity of threonine dehydratase was decreased by its active site modification. The final recombinant strain (TH07 (pBRThrABC)) yielded 0.2 g threonine/g glucose with a titer of 10 gL⁻¹ when grown in TPM1 medium containing 50 gL⁻¹ of glucose for 48h. This first generation strain was further improved using transcriptome analyses and *in silico* metabolic simulation.

The transcriptome profiles of TH07 (pBRThrABC) were compared to those of its unmodified parent (WL3110) under identical conditions, and the following genes were targeted for further manipulation using this information: (i) phosphoenolpyruvate carboxylase (*ppc*); (ii) malate synthase (*aceB*), isocitrate lyase (*aceA*) and their transcriptional regulator (*iclR*); (iii) threonine transporter (*tdcC*); and (iv) threonine efflux transporter (*rhtC*).

The *in silico* flux response analyses were performed using the genome-scale metabolic model *E. coli* MBEL979 reconstructed from 979 metabolic reactions and 814 metabolites, a slightly modified network of iJR904. For examining the effect of changes in PPC or ICL fluxes on threonine production, the corresponding fluxes were perturbed from the minimum to the maximum values, and maximization of threonine production rate was set as the objective function. Based on the simulation results, an optimal expression level of *ppc* was achieved by chromosomal replacement of its native promoter with the *trc* promoter, which resulted in a 27% increase in threonine production in the engineered strain as compared to its parent, TH07 (pBRThrABC). Modifications identified through transcriptome analyses, such

deletion of *iclR* and *tdcC* as well as overexpression of *rhtC*, led to further improvements in threonine production. Batch culture of the final engineered strain TH27C (pBRThrABCR3) produced 11.8 gL⁻¹ threonine with a yield of 0.393 g threonine/g glucose. A fed-batch fermentation run of this strain produced 77 gL⁻¹ threonine in 56.1 h, and a volumetric productivity of 1.37 gL⁻¹h⁻¹. However, accumulation of acetic acid resulted in growth retardation and lowered the rates of glucose uptake and threonine production in the later stages of fermentation. To investigate the effect of the central metabolic flux on the acetic acid formation rate, each flux was perturbed, *in silico*, from the minimum value to the maximum value, with maximization of acetic acid production rate as an objective function. Among several gene knockout/overexpression solutions, amplification of acetyl-CoA synthetase (*acs*) was chosen to relieve toxicity. The recombinant strain incorporating this final modification gave a threonine titer of 82 gL⁻¹ in 50 h, (volumetric productivity of 1.65 gL⁻¹h⁻¹) comparable to numbers generated by industrial strains engineered through the laborious classical approach.

The same group extended the approach described above (combining traditional metabolic design with transcriptome profiling and *in silico* flux analyses) to engineer L-valine production in *E. coli* [28]. Using the same genome-scale metabolic model of *E. coli* described above (MBEL979), gene knockout targets were identified by applying the MOMA algorithm. Batch culture of the final engineered strain resulted in 7.55 gL⁻¹ L-valine from 20 gL⁻¹ glucose with a yield of 0.378 g L-valine/g glucose.

Metabolic engineering applications employing *in silico* analyses such as the ones described above have been reported only recently. Current research efforts are

being directed towards the expansion of existing models to incorporate (i) additional equations and constraints pertaining to regulatory and signaling circuits; (ii) transcript/protein/metabolite abundance profiles; (iii) predictions in non-exponential growth phases, which are used for real-world metabolite production, and (iv) measured fluxes generated from isotopomer-based studies to improve the accuracy of constraint based algorithms. The power of global scale models will only continue to grow as cell-wide data being collected at an unprecedented pace across multiple laboratories around the world are incorporated in future iterations of current reconstructions. While successive iterations of metabolic models continue to improve, next generation combinatorial strategies that do not rely on predictive models for improving cellular phenotype have shown promising results. These are highlighted in next section.

Targeted Combinatorial Engineering

As discussed previously, the classical approach for strain improvement is based on creating random mutations throughout the entire genome of a host organism. Mutations can be beneficial, silent or deleterious to growth, thus precluding facile recognition of genotype-phenotype relationships to guide further rational engineering efforts after mutations have been identified. In contrast, targeted combinatorial approaches aim to accelerate the process of rewiring existing genetic circuits through screening of random variants of *select* cellular components. While still random by construction, these approaches have been found to reduce the time scale for identifying improved variants and offer the potential of uncovering regulatory, kinetic, or poorly understood gene targets not accounted for in existing genome scale models. Strategies that have been examined so far include

(i) engineered promoter libraries; (ii) transposon-mediated random chromosomal insertions; (iii) random overexpression libraries; (iv) tunable intergenic regions; (v) transcription/translation machinery engineering, and (vi) multiplex genome engineering. In this section we discuss the last three strategies in more detail.

Heterologous pathways engineered on a plasmid backbone may be assembled in a monocistronic (as seen in the case of echinomycin biosynthesis, discussed above) or a polycistronic arrangement (as seen in the case of amorpho-4,11-diene biosynthesis, discussed above). The former resembles the eukaryotic mode of gene expression, while the latter is akin to its prokaryotic counterpart. Either arrangement requires balanced and coordinated expression of multiple genes for optimal production of the desired metabolite. Expression levels of plasmid-borne genes from heterologous pathways may be fine-tuned at the pre- or post-transcriptional stage by ‘tweaking’ sequences (or ‘parts’) that flank the respective genes – promoters, ribosomal binding sites, and terminators among others. Comprehensive mathematical models that can predict not only the right sequences, but also the correct combinations of assembled ‘parts’ for optimal production have yet to be developed. Thus far, ‘brute-force’ combinatorial techniques have been attempted to generate libraries of these ‘parts’, which are then screened for the best producers. Along these lines, Pflieger *et al.* targeted intergenic regions in a plasmid-borne, synthetic operon designed for the production of the amorpho-4,11-diene precursor – mevalonate (discussed above) [32] (Fig. 2). Intergenic regions can affect a number of post-transcriptional features such as mRNA stability, transcription termination and translation initiation. In the aforementioned study, a library of tunable intergenic regions (TIGRs) was generated such that the variable sequences

would incorporate mRNA secondary structures of various lengths, GC contents, asymmetries and mismatched bulges. TIGRs were implemented for improving flux through the ‘top’ portion of the mevalonate pathway (described above) because simple overexpression of all three genes of this operon led to reduced growth and product formation, a phenotype attributed to toxicity of imbalanced gene expression. Combinatorial screening of TIGR variants was aided by coupling the mevalonate-producing strains with a GFP-producing mevalonate-auxotroph. Mevalonate production from the best library strains, which displayed improved growth and increased flux through the pathway, was seven-fold greater than those carrying the unmodified control plasmid. Surprisingly, the improvement in production was attributed to decreased expression of two genes from the pathway – HMG-CoA synthase and truncated HMG-CoA reductase. This counterintuitive result highlights an interesting facet of metabolic control – ‘more is not always better!’

The approach discussed above may be considered as a local optimization effort, in that only components directly related to the pathway of interest were targeted for combinatorial screening. However variations in non-pathway genes, especially ones that display pleiotropic features, by virtue of their inherent activity, can affect the output of heterologous pathways. Engineering complexes that control the transcriptional or translational processes in the cell would allow complete reprogramming of the cellular machinery, and therefore serve as valuable targets for generating strain diversity. The approach dubbed global transcription machinery engineering (gTME) has been shown recently to be an effective tool for generating strains optimized for several phenotypes [2]. It involves targeted mutagenesis of select components of the transcriptional machinery, such as sigma

factors to perturb cell-wide transcriptional levels. Sigma factors are involved in DNA sequence recognition prior to transcription initiation and therefore slight variations in sequences of these proteins (generated using error prone PCR), have the potential to greatly affect the subset of genes that would eventually be expressed by the RNA polymerase complex. In terms of generating diversity, this approach was directly compared to classical mutagenesis, and on the basis of colony growth and size, gTME was found to be superior to the classical approach [20]. The increased diversity generated by the gTME approach would be expected to increase the chances of finding a desirable mutant during phenotypic screening or selection. On an additional note, it was found that phenotypic variation generated through this approach could be effectively tweaked by changing the frequency of error generation during the mutagenic PCR reaction with higher mutation frequencies leading to a broader range of phenotypes [20]. gTME has been demonstrated to work in *L. plantarum* and *E. coli* for improved resistance as well as production phenotypes [2, 20]. For the latter, lycopene (a powerful antioxidant) was chosen as a representative metabolic phenotype. Several studies have demonstrated recombinant production of lycopene in *E. coli* through the DXP pathway (described above) to form the precursor monomer geranyl geranyl pyrophosphate, which undergoes polymerization to form this bright red colored 40-carbon carotenoid. Improvement of lycopene production in *E. coli* has been attempted through several approaches including but not limited to transposon insertion mutagenesis, genomic library overexpressions as well as global stoichiometric model based predictions to identify single- and multiple- gene knockout/overexpression candidates. The transcriptional reprogramming approach was compared to strains engineered

through the aforementioned traditional approaches and the following conclusions were drawn [2]: (i) a single round of selection using gTME was more effective than successive rounds of single-gene knockouts or overexpression modifications based on predictive models; (ii) strains engineered through traditional routes could be further improved through gTME; and (iii) transcriptional reprogramming was genotype specific, i.e., each parental strain required its own optimized sigma factor to elicit an improved phenotype.

Along the lines of the approach discussed above, translational reprogramming enabled through modulation of ribosomal proteins or ribosomal RNA, in theory, could also be used to generate diverse phenotypes. While there have been no reports of generating random variants of ribosomal protein-complex-components, site-directed mutagenesis of specific amino acid residues of the ribosomal protein S12 (RpsL) from *Streptomyces lividans* resulted in significant overproduction of the antibiotic undecylprodigiosin [27]. Similarly antibiotic production in some species of *Bacillus* and *Pseudomonas* was shown to improve through targeted mutation of the respective RpsL proteins [27]. These studies suggest the potential utility of ribosomal engineering in generating diverse phenotypes.

An extension of pathway-localized approaches, such as TIGRs, to multiple targets within the host chromosome may be envisioned as a third alternative to global cellular reprogramming. Wang and coworkers, in a recent report, coined this approach as MAGE – multiplex automated genome engineering, a highly efficient, inexpensive and automated solution to concurrently modify many genomic locations (genes, regulatory regions, etc) across different length scales, from the

nucleotide to the genome level [40]. The technique, mediated via bacteriophage λ -Red protein, involves simultaneous allelic replacement, of multiple target regions on the genome using divergent oligonucleotides (oligos), within a single cell, or across a population, to generate combinatorial genomic diversity. MAGE was tested for the optimization of the DXP pathway in *E. coli* for the overproduction of lycopene. As noted above, a bevy of traditional techniques have been employed to improve production of this carotenoid in *E. coli*. A total of twenty-four endogenous gene targets (twenty overexpressed and four knocked out) have been documented by these studies to affect lycopene production. All twenty-four candidates were simultaneously targeted through MAGE for further improvement of lycopene production in an *E. coli* strain harboring the production plasmid. Specifically for each of the 20 overexpression candidates, the native ribosomal binding sites were targeted for enhanced translation efficiency, using 90-mer oligos. For the knockout candidates, oligos were designed to introduce non-sense mutations in the open reading frames. MAGE produced as many as 15 billion genetic variants, which were screened on the basis of pigmentation intensity of the lycopene produced. Under experimental conditions similar to those used in previous studies, the MAGE-generated variants produced the highest lycopene yields ($\sim 9 \text{ mg g}^{-1}$ dry cell weight) documented to date. Prototype devices constructed to automate the MAGE procedure enabled isolation of improved variants in as little as 3 days, a significant improvement over all of the combinatorial techniques described in this chapter. In summary, the MAGE multiplex approach offers an unparalleled scale of improvement in the design and evolution of improved phenotypes.

While state-of-the-art combinatorial strategies such as the ones described above offer great hope in the rapid design and generation of optimized strains, the inherent 'brute-force' nature of these methods, overdependence on facile screening tools, inability to generate multiple phenotypic traits in a single strain and difficulty in establishing causal effects, underscore the importance of applying rigorous engineering principles for biological design. This brings us to the world of synthetic biology.

Synthetic Biology: Parts, Devices and Circuits

Synthetic biology is based on the principle of developing reusable components so that they can be assembled into networks of increasing complexity [4]. Traditional recombinant DNA technology relied on simple yet powerful tools of molecular biology to 'cut-copy-paste' naturally occurring DNA fragments from one organism to another. In contrast, computer-aided network design, standardized parts' libraries, custom DNA syntheses and automated assembly are the hallmarks of synthetic biology. Contemporary synthetic biologists are already accustomed to an ever-growing and constantly updated repertoire of web-based tools and resources that greatly facilitate metabolic design. These include comprehensive enzyme knowledge bases (like [BRENDA](#) and [ENZYME](#)), pathway databases and solvers (such as [MetaCyc](#) and [FMM](#)), genome-aggregator sites with sophisticated search algorithms (such as [MicrobesOnline](#) and [KEGG](#)), CAD programs (such as [Tinkercell](#)), sequence analyzers (such as the [RBS Calculator](#) and [ESSA](#)), and a dedicated encyclopedia of *E. coli* K12 genes and metabolism ([EcoCyc](#)), and others. These have been complemented well by next generation experimental tools for DNA assembly such as MAGIC, SLIC, Golden Gate Shuffling, DNA Assembler and others.

Adding to these lists are the recently developed repositories of standardized parts – the MIT and JBEI Registries of Standard Biological Parts (<http://partsregistry.org> and <https://www.jbeir.org>). The concept of a standard part, one inspired from the electronics industry, in biological terms refers to a genetically encoded object that performs a biological function and that has been engineered to meet specific design or performance requirements [10]. Precise construction of standardized DNA entities with properties more reliable than their natural counterparts is best accomplished through the chemical syntheses of such fragments. However, standardization of genes, in which many allelic variants may be known, may be problematic. Significant technological improvements over the past decade have brought down costs of DNA synthesis from ~\$20/bp in the late 90s' to less than \$0.5/bp by 2009, thus paving the way for economical construction of increasingly complex circuits.

The conceptual analogy between synthetic gene circuits and their electronic counterparts lends itself naturally to the application of several concepts from control theory (such as stability analysis, robustness, response kinetics and oscillatory behavior) for dynamic gene regulation. Indeed, seminal work at the turn of the last century by Gardner *et al.* on the development of a synthetic 'toggle switch' [13] and Elowitz *et al.* on the first generation ring oscillator [9], sparked the synthetic biology 'revolution'. The first wave of this 'revolution' was marked by the development of several such devices and modules with specified behaviors [33] – cascades, pulse generators, time delayed circuits, spatial patterning and logic formulas. Considering that most of these were early examples of prototypical devices, standardized datasheets of these devices, summarizing rigorous

performance characteristics of the type reported recently [4] have yet to be developed. Here we point out salient features of two prototypical devices with future potential for metabolic engineering applications – an inversion recombination switch and a synthetic protein scaffold.

In the previous section we talked about pathway-specific or cell-wide transcriptional control strategies based on engineered promoters or gTME respectively. While engineered promoters can fine-tune flux through a heterologous pathway, additional levels of control may be desired in certain situations such as minimizing leaky expression of a toxic protein or minimizing usage of an expensive inducer. In 2006, Ham *et al.* developed a FimE invertase- based switch that could be used to address such cases [14]. This device was composed of the following parts: a *fimE* gene, an *araBAD* promoter driving *fimE*, a constitutive *trc* promoter (P_{trc}) and a Fim invertible DNA segment bearing P_{trc} . In terms of device operation, a given target gene is driven by P_{trc} oriented in its OFF position initially, resulting in no transcribed product. When arabinose is added to the system, expression of *fimE* switches P_{trc} from its initial OFF to a permanently ON position. This means that subsequent removal or addition of arabinose does not alter the state of the switch. This device therefore offers an inducible, unidirectional mode of operation for binary control of gene expression. In a follow-up study the same group extended this work to construct a heritable sequential memory switch, using the FimB (from *E. coli*) and Hin (from *Salmonella* spp.) recombination systems that could output multiple states, for potential applications in complex regulatory circuits [15].

The TIGR and MAGE approaches discussed above improve desired metabolic fluxes by the manipulation of translational levels of pathway or non-pathway

components. In a recent study, Dueber *et al.* used post-translational control, mediated by synthetic protein scaffolds, as an approach to improve substrate channeling and therefore metabolic fluxes through heterologous pathways [8]. These artificial protein complexes (or devices) were based on the signal processing machinery from eukaryotic cells and could be modularly programmed to accommodate different component stoichiometries for fine-tuning flux through pathways of interest (Fig. 2). Pathways for the production of mevalonate and glucaric acid in *E. coli* were tested by engineering the respective proteins to 'dock' on pre-designed scaffolds. For the three-gene 'top' portion of the mevalonate pathway a 77-fold improvement in product titer was observed with low enzyme expression and reduced metabolic load. Production of glucaric acid (a building block for nylons and hyper-branched polyesters) in *E. coli* was first demonstrated by Moon *et al.* [8] through the expression of *myo*-inositol-1-phosphate synthase (Ino1) from *S. cerevisiae*, *myo*-inositol oxygenase (MIOX) from mice and uronate dehydrogenase from *Pseudomonas syringae*, resulting in a final titer of 1 gL⁻¹. Scaffolding of the first two pathway components, Ino1 and MIOX, resulted in a small but appreciable improvement over a non-scaffolded control and a final titer of 1.7 gL⁻¹, thus suggesting a potential generalizability of this approach.

As mentioned above, the examples discussed offer only a glimpse of early prototypical devices and simple modules constructed from these devices. Current research in this area is aimed towards adapting existing frameworks for describing engineered devices, such as development of standardized datasheets describing quantitative and dynamic behavior of biological objects, providing open source access to such information through 'Parts Registries' and continuous refinement of

these parts, devices and modules by the research community [4]. This was exemplified in a recent report describing a device called BBa_F2620 – a cell-cell communication receiver developed by Canton *et al* [4]. As pointed out by Purnick and Weiss [33], to view cells as true ‘programmable’ entities, it is essential to develop effective strategies for assembling devices and modules into intricate customizable large-scale systems. This concept of a true bottom-up approach for metabolic design marks the second wave of the synthetic biology ‘revolution’. The development of minimalist organisms with the intent to reduce system complexity for improving performance and the synthesis of entire genomes from scratch may be considered as harbingers of this second wave [3, 33].

References

1. Adrio JL and Demain AL, (2006) *Genetic improvement of processes yielding microbial products*. Fems Microbiology Reviews. 30(2): p. 187-214.
2. Alper H and Stephanopoulos G, (2007) *Global transcription machinery engineering: A new approach for improving cellular phenotype*. Metabolic Engineering. 9(3): p. 258-67.
3. Bhalerao KD, (2009) *Synthetic gene networks: The next wave in biotechnology?* Trends in Biotechnology. 27(6): p. 368-374.
4. Canton B, Labno A, and Endy D, (2008) *Refinement and standardization of synthetic biological parts and devices*. Nature Biotechnology. 26(7): p. 787-93.
5. Causey TB, Shanmugam KT, Yomano LP, and Ingram LO, (2004) *Engineering escherichia coli for efficient conversion of glucose to pyruvate*. Proceedings of the National Academy of Sciences U S A. 101(8): p. 2235-40.
6. Chang MC and Keasling JD, (2006) *Production of isoprenoid pharmaceuticals by engineered microbes*. Nature chemical biology. 2(12): p. 674-81.
7. Chotani G, Dodge T, Hsu A, et al., (2000) *The commercial production of chemicals using pathway engineering*. Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology. 1543(2): p. 434-455.
8. Dueber JE, Wu GC, Malmirchegini GR, et al., (2009) *Synthetic protein scaffolds provide modular control over metabolic flux*. Nature Biotechnology. 27(8): p. 753-9.
9. Elowitz MB and Leibler S, (2000) *A synthetic oscillatory network of transcriptional regulators*. Nature. 403(6767): p. 335-8.
10. Endy D, (2005) *Foundations for engineering biology*. Nature. 438(7067): p. 449-53.
11. Ensley BD, Ratzkin BJ, Osslund TD, et al., (1983) *Expression of naphthalene oxidation genes in escherichia coli results in the biosynthesis of indigo*. Science. 222(4620): p. 167-9.
12. Feist AM and Palsson BO, (2008) *The growing scope of applications of genome-scale metabolic reconstructions using escherichia coli*. Nature Biotechnology. 26(6): p. 659-667.
13. Gardner TS, Cantor CR, and Collins JJ, (2000) *Construction of a genetic toggle switch in escherichia coli*. Nature. 403(6767): p. 339-42.
14. Ham TS, Lee SK, Keasling JD, and Arkin AP, (2006) *A tightly regulated inducible expression system utilizing the fim inversion recombination switch*. Biotechnology and bioengineering. 94(1): p. 1-4.
15. Ham TS, Lee SK, Keasling JD, and Arkin AP, (2008) *Design and construction of a double inversion recombination switch for heritable sequential genetic memory*. PLoS ONE. 3(7): p. e2815.
16. Ikeda M and Katsumata R, (1992) *Metabolic engineering to produce tyrosine or phenylalanine in a tryptophan-producing corynebacterium glutamicum strain*. Applied and Environmental Microbiology. 58(3): p. 781-785.
17. Ikeda M and Katsumata R, (1999) *Hyperproduction of tryptophan by corynebacterium glutamicum with the modified pentose phosphate pathway*. Applied and Environmental Microbiology. 65(6): p. 2497-502.

18. Ikeda M, Ohnishi J, Hayashi M, and Mitsuhashi S, (2006) *A genome-based approach to create a minimally mutated corynebacterium glutamicum strain for efficient l-lysine production*. Journal of Industrial Microbiology and Biotechnology. 33(7): p. 610-5.
19. Kim HU, Kim TY, and Lee SY, (2008) *Metabolic flux analysis and metabolic engineering of microorganisms*. Molecular BioSystems. 4(2): p. 113-20.
20. Klein-Marcuschamer D and Stephanopoulos G, (2008) *Assessing the potential of mutational strategies to elicit new phenotypes in industrial strains*. Proceedings of the National Academy of Sciences U S A. 105(7): p. 2319-24.
21. Lee KH, Park JH, Kim TY, Kim HU, and Lee SY, (2007) *Systems metabolic engineering of escherichia coli for l-threonine production*. Molecular Systems Biology. 3(p. 149).
22. Lee SJ, Lee DY, Kim TY, *et al.*, (2005) *Metabolic engineering of escherichia coli for enhanced production of succinic acid, based on genome comparison and in silico gene knockout simulation*. Applied and Environmental Microbiology. 71(12): p. 7880-7.
23. Lutke-Eversloh T and Stephanopoulos G, (2007) *L-tyrosine production by deregulated strains of escherichia coli*. Applied Microbiology and Biotechnology. 75(1): p. 103-10.
24. Martin VJ, Pitera DJ, Withers ST, Newman JD, and Keasling JD, (2003) *Engineering a mevalonate pathway in escherichia coli for production of terpenoids*. Nature Biotechnology. 21(7): p. 796-802.
25. Murdock D, Ensley BD, Serdar C, and Thalen M, (1993) *Construction of metabolic operons catalyzing the de novo biosynthesis of indigo in escherichia coli*. Biotechnology (N Y). 11(3): p. 381-6.
26. Ohta K, Beall DS, Mejia JP, Shanmugam KT, and Ingram LO, (1991) *Genetic improvement of escherichia coli for ethanol production: Chromosomal integration of zymomonas mobilis genes encoding pyruvate decarboxylase and alcohol dehydrogenase ii*. Applied and Environmental Microbiology. 57(4): p. 893-900.
27. Okamoto-Hosoya Y, Okamoto S, and Ochi K, (2003) *Development of antibiotic-overproducing strains by site-directed mutagenesis of the rpsL gene in streptomyces lividans*. Applied and Environmental Microbiology. 69(7): p. 4256-9.
28. Park JH, Lee KH, Kim TY, and Lee SY, (2007) *Metabolic engineering of escherichia coli for the production of l-valine based on transcriptome analysis and in silico gene knockout simulation*. Proceedings of the National Academy of Sciences U S A. 104(19): p. 7797-802.
29. Patnaik R, (2008) *Engineering complex phenotypes in industrial strains*. Biotechnology Progress. 24(1): p. 38-47.
30. Patnaik R, Zolandz RR, Green DA, and Kraynie DF, (2008) *L-tyrosine production by recombinant escherichia coli: Fermentation optimization and recovery*. Biotechnology and bioengineering. 99(4): p. 741-52.
31. Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, and Khosla C, (2001) *Biosynthesis of complex polyketides in a metabolically engineered strain of e-coli*. Science. 291(5509): p. 1790-1792.

32. Pflieger BF, Pitera DJ, D Smolke C, and Keasling JD, (2006) *Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes*. Nature Biotechnology. 24(8): p. 1027-1032.
33. Purnick PEM and Weiss R, (2009) *The second wave of synthetic biology: From modules to systems*. Nature Reviews Molecular Cell Biology. 10(6): p. 410-422.
34. Revilla G, Lopez-Nieto MJ, Luengo JM, and Martin JF, (1984) *Carbon catabolite repression of penicillin biosynthesis by penicillium chrysogenum*. Journal of Antibiotics (Tokyo). 37(7): p. 781-9.
35. Sanchez AM, Bennett GN, and San KY, (2005) *Novel pathway engineering design of the anaerobic central metabolic pathway in escherichia coli to increase succinate yield and productivity*. Metabolic Engineering. 7(3): p. 229-239.
36. Sauer U, (2006) *Metabolic networks in motion: 13c-based flux analysis*. Molecular Systems Biology. 2(p. 62.
37. Song H and Lee SY, (2006) *Production of succinic acid by bacterial fermentation*. Enzyme and Microbial Technology. 39(3): p. 352-361.
38. Stephanopoulos G, (1999) *Metabolic fluxes and metabolic engineering*. Metabolic Engineering. 1(1): p. 1-11.
39. Verma V, Qazi GN, and Parshad R, (1992) *Intergeneric protoplast fusion between gluconobacter-oxydans and corynebacterium species*. Journal of biotechnology. 26(2-3): p. 327-330.
40. Wang HH, Isaacs FJ, Carr PA, *et al.*, (2009) *Programming cells by multiplex genome engineering and accelerated evolution*. Nature. 460(7257): p. 894-8.
41. Watanabe K, Hotta K, Praseuth AP, *et al.*, (2006) *Total biosynthesis of antitumor nonribosomal peptides in escherichia coli*. Nature chemical biology. 2(8): p. 423-8.
42. Weissman KJ and Leadlay PF, (2005) *Combinatorial biosynthesis of reduced polyketides*. Nature Reviews Microbiology. 3(12): p. 925-936.
43. Zhang W, Li Y, and Tang Y, (2008) *Engineered biosynthesis of bacterial aromatic polyketides in escherichia coli*. Proceedings of the National Academy of Sciences U S A. 105(52): p. 20683-8.
44. Zhang YX, Perry K, Vinci VA, *et al.*, (2002) *Genome shuffling leads to rapid phenotypic improvement in bacteria*. Nature. 415(6872): p. 644-646.

Websites:

1. [BRENDA](http://www.brenda-enzymes.info/): http://www.brenda-enzymes.info/
2. [ENZYME](http://www.expasy.ch/enzyme/): http://www.expasy.ch/enzyme/
3. [MetaCyc](http://metacyc.org/): http://metacyc.org/
4. [FMM](http://fmm.mbc.nctu.edu.tw/): http://fmm.mbc.nctu.edu.tw/
5. [MicrobesOnline](http://www.microbesonline.org/): http://www.microbesonline.org/
6. [KEGG](http://www.genome.jp/kegg/): http://www.genome.jp/kegg/
7. [TinkerCell](http://www.tinkercell.com/): http://www.tinkercell.com/
8. [RBS Calculator](http://www.voigtlab.ucsf.edu/software/instructions.html): http://www.voigtlab.ucsf.edu/software/instructions.html
9. [ESSA](http://www.inra.fr/internet/Departements/MIA/T/essa/Doc/essa_home.html): http://www.inra.fr/internet/Departements/MIA/T/essa/Doc/essa_home.html
10. [EcoCyc](http://www.ecocyc.org): http://www.ecocyc.org

Figure 1

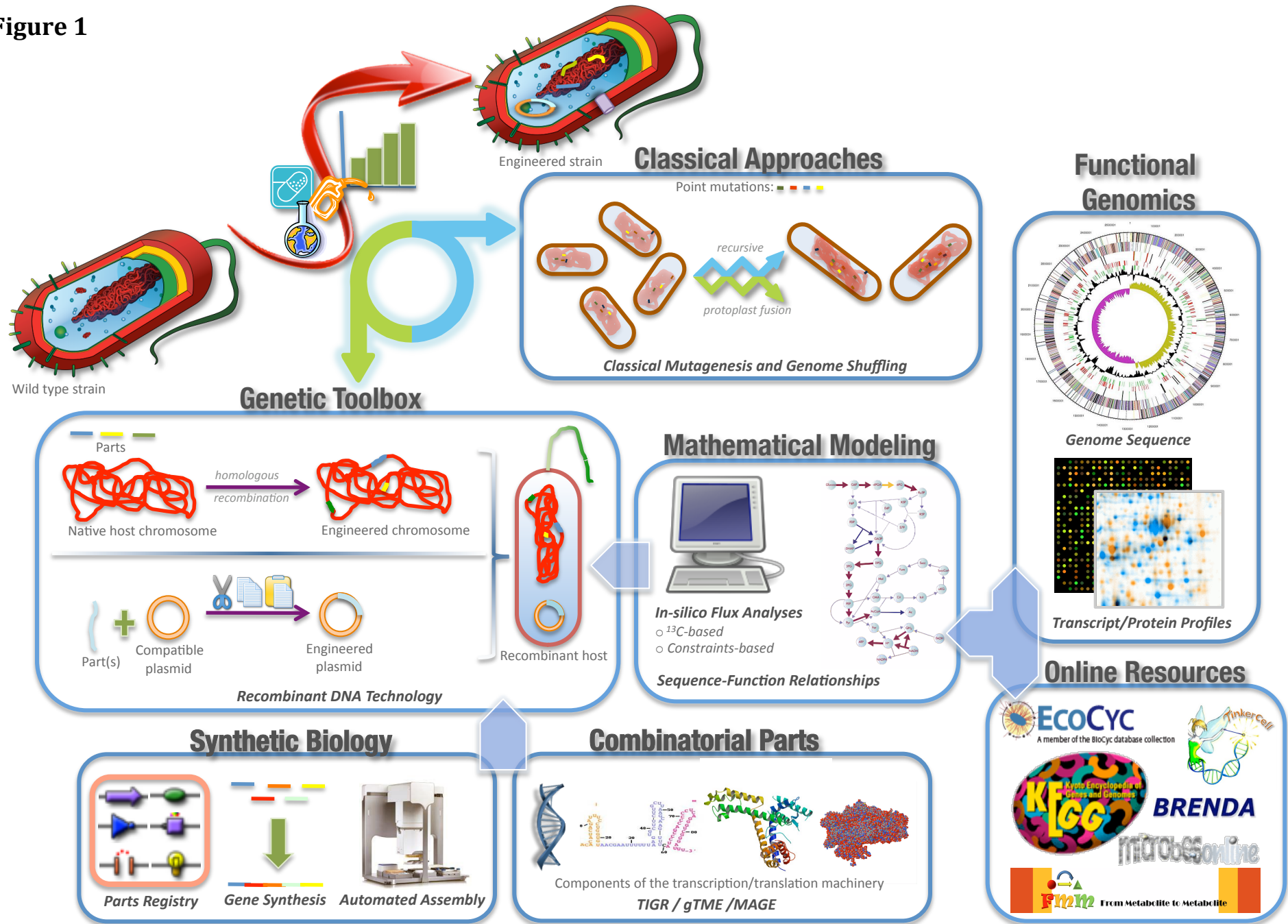
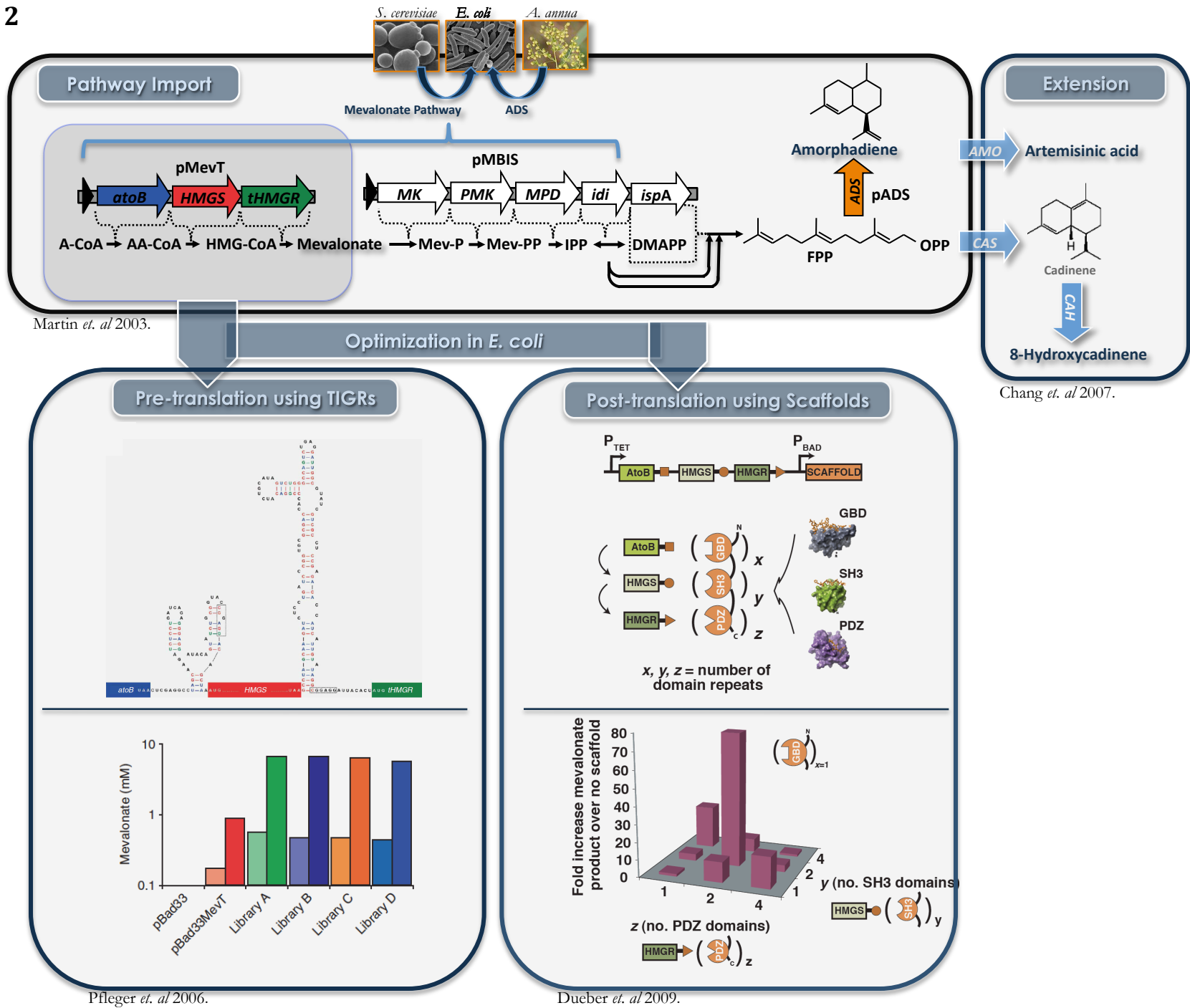


Figure 1: Approaches for Metabolic Design in Prokaryotes

Strategies for metabolic design in prokaryotes as outlined in this chapter are summarized here. Broadly speaking, metabolic design strategies may be grouped into two types: (i) strategies that can be implemented independently of genome sequence information or advanced genetic tools. Classical mutagenesis and genome shuffling are two examples of this group; (ii) those that employ mathematical and synthetic biology tools for the direct genetic manipulation of metabolic pathways. Predictive model based approaches combine information from organism-specific genomic, fluxomic, transcriptomic, proteomic data in addition to information from a variety of online databases. Model-independent approaches rely on existing biochemical knowhow, parts registries and combinatorial libraries of cellular components coupled with high throughput screening of improved phenotypes. Details on individual approaches are provided in the text.

Figure 2



Martin *et. al* 2003.

Chang *et. al* 2007.

Pfleger *et. al* 2006.

Dueber *et. al* 2009.

Figure 2: Engineering the mevalonate pathway in *E. coli*, optimization of the top portion and routes for terpenoid synthesis through the engineered pathway.

The mevalonate pathway from *S. cerevisiae* was engineered in *E. coli*. The pathway was split between two plasmids, pMevT and pMBIS carrying three and five genes respectively. Subsequently, mevalonate production was optimized at the pre- and post-translation levels. Representative data from each strategy are shown. The bottom portion of the engineered pathway was directed for the production of terpenoids, artemisinic acid and 8-hydroxycadenine, precursors for important drug molecules - artemisinin and gossypol respectively. Gene names are from the respective publications. Details on individual strategies are provided in the text.