Review in Advance first posted online on March 27, 2013. (Changes may still occur before final publication online and in print.)

Metabolic Engineering: Past and Future

Benjamin M. Woolston,* Steven Edgar,* and Gregory Stephanopoulos

Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; email: woolston@mit.edu, sedgar@mit.edu, gregstep@mit.edu

Annu. Rev. Chem. Biomol. Eng. 2013. 4:259–88

The *Annual Review of Chemical and Biomolecular Engineering* is online at chembioeng.annualreviews.org

This article's doi:
10.1146/annurev-chembioeng-061312-103312

*These authors contributed equally to the work.

Keywords

metabolic flux analysis, MFA, metabolic control analysis, MCA, inverse metabolic engineering, synthetic biology, systems biology

Abstract

We present here a broad overview of the field of metabolic engineering, describing in the first section the key fundamental principles that define and distinguish it, as well as the technological and intellectual developments over the past approximately 20 years that have led to the current state of the art. Discussion of concepts such as metabolic flux analysis, metabolic control analysis, and rational and combinatorial methods is facilitated by illustrative examples of their application drawn from the extensive metabolic engineering literature. In the second section, we present some of the rapidly emerging technologies that we think will play pivotal roles in the continued growth of the field, from improving production metrics to expanding the range of attainable compounds.

INTRODUCTION

Soon after Cohen & Boyer (1) first successfully introduced foreign genes into a bacterial cell, it became apparent that bacterial and other cells could be engineered to convert these genes into little chemical factories for the overproduction of numerous chemical and pharmaceutical products. However, it was soon realized that, whereas the production of a very complex molecule such as insulin could be achieved through the overexpression of a single gene in the bacterium *Escherichia coli*, the overproduction of a simple compound such as ethanol required the coordinated modulation of the expression of many genes comprising the ethanol production pathway. As a result, attention turned quickly to medical applications of biotechnology, and other production opportunities of recombinant DNA technology fell lower on the priority list. During the same period (the 1980s), researchers in academia investigated fundamental questions of design and functioning of systems (or networks) of reactions. This led to the publication of the first papers enumerating all possible routes connecting a substrate with a target product, the thermodynamic analysis of such pathways, investigations on the distribution of kinetic control, and the design of genetic circuits to bring about a desired pattern of gene expression and product synthesis, among others.

These activities culminated in two seminal papers (2, 3) that essentially initiated the field of metabolic engineering. A conference followed soon thereafter (in 1996), along with a journal (*Metabolic Engineering*) and the first book in the field, *Metabolic Engineering: Principles and Methodologies* (4). The new field was defined as the "directed improvement of product formation or cellular properties through the modification of specific biochemical reactions or the introduction of new genes with the use of recombinant DNA technology" (4, p. 2). Metabolic engineering was differentiated from genetic engineering by its distinct focus. Namely, the field was occupied with investigating the properties of integrated metabolic pathways and genetic regulatory networks as opposed to individual genes and enzymes, which were the subject of most molecular biological research at that (pre–systems biology) time (5). In this sense, metabolic engineering preceded systems biology by championing the need for a systemic view of metabolic pathways and approaches for their optimal functioning.

Metabolic engineering has a distinct industrial dimension, because it aims to construct microbes that can be used as biocatalysts for the cost-effective production of fuels, chemicals, and pharmaceuticals. As such, it encompasses a lot more than simply stitching genes together to build a basic functioning pathway, and this is another characteristic that differentiates metabolic engineering from other overlapping fields, such as synthetic biology. One can successfully express the totality of pathway genes to produce a few milligrams of product, but a cost-effective process cannot be realized until all three factors of titer, rate (or productivity), and yield have been optimized. So, although a functional pathway can be built in a few months, it can take much longer to improve it to the point that it can support a commercial process. This underlines the basic elements of metabolic engineering as the field aiming at pathway design, construction, and optimization. These elements encompass a lot more than genetic engineering and molecular biology and include components from graph theory, chemical reaction engineering, biochemistry, and optimization.

This review comprises two major parts. The first (Past and Present) deals with advances in metabolic engineering during the past 20 years. Owing to the great diversity of topics that comprise metabolic engineering and that have also been covered in other reviews, this paper is designed not to exhaustively cover all these advances in great depth but rather to define them and highlight their importance in the development of the intellectual foundation and identity of the field. For more in-depth analyses, the reader is directed to recent excellent reviews where applicable (see **Table 1**). This allows us to present a very broad review encompassing issues such as pathway design, thermodynamic feasibility of metabolic pathways, metabolic fluxes, and flux determination

Table 1 List of topical review papers relevant to metabolic engineering

Subject	Reference(s)
Metabolic flux analysis	14, 16–18, 31
Metabolic control analysis	27, 147–149
Flux balance analysis	30, 31, 150
Genome scale models	42–44, 151–153
Inverse metabolic engineering	81
High-throughput screening	88
Global transcription machinery engineering	92, 154
Spatial organization of enzymes	126, 155
Probabilistic pathway construction	156
Computational tools for design of pathways	157
Gene circuits	158
Production of nonnatural chemicals	137

along with the use of ^{13}C isotopic tracers, distribution of kinetic control, and metabolic control analysis (MCA). These analytical tools generate information leading to the identification of specific enzymes that carry the majority of the product flux and thus, as part of the rational design and analysis of metabolic pathways, become targets for modulation. Additionally, one can apply combinatorial methods in recognition of the inadequacy of available kinetic and regulatory models to support optimal pathway design on a global scale. Application of combinatorial methods to a rationally designed pathway usually improves performance by a further two- to fourfold. Of particular importance to combinatorial methods, and to the associated approach of inverse metabolic engineering (IME), are high-throughput screening techniques that allow selection of improved mutants and identification of the particular genetic element(s) responsible for the enhanced performance. Interwoven into this section are illustrative applications showing that, after a period of small hesitant steps with pathways characterized by well-understood kinetics and regulation, researchers are now engineering pathways for the production of high-volume commodity chemicals in addition to the initially targeted high-priced therapeutics and chemicals. Examples include biopolymers, fuels (ethanol, isobutanol, *n*-butanol, hydrocarbons, oils, and lipids), chemicals (succinic acid, butanediol, acrylic acid, lactic acid, and isoprene), pharmaceuticals, and numerous specialty chemicals.

Looking into the future, we can expect metabolic engineering applications to increase dramatically. The ability to chemically synthesize DNA has made possible heterologous expression of genes of any kind in the preferred microbial host cell. At the same time, the explosion of -omics based tools, in no small part due to the lowering cost of genome sequencing, has facilitated the development of novel metabolic engineering tools for both standard lab organisms and noncanonical organisms, opening up a new era in which the choice of host is no longer limited to a small subset of organisms and one can take advantage of the diverse capabilities of microbes from all branches of life. Together, these two developments create an essentially infinite number of possibilities in terms of pathway construction and microbial production of virtually any product. As well as enabling an expanded set of products, the rapidly increasing capabilities of synthetic DNA production present the possibility of synthesizing and expressing cellular compartments (e.g., endoplasmic reticulum, lipid bodies) in host cells lacking such compartments. Ultimately, one can envision the era of completely synthetic cells. Although this goal is as yet many years away, the synthetic biology tools being developed for its realization will have an unignorable and profound impact on metabolic engineering. Therefore, the second part of the paper (Visions of the Future)

reviews these technologies and their impact on the field of metabolic engineering and concludes with an assessment of the future prospects of metabolic engineering and biotechnology.

PART I: PAST AND PRESENT

Enumerating and Assessing Pathways for Converting a Specific Substrate to a Target Product

With the goal of overproduction of specific products in mind, the first question that must be answered when embarking on a metabolic engineering endeavor is, What pathways can be used to produce the compound of interest? This is fundamentally the same question that preoccupies a synthetic organic chemist, with the additional caveats that the reactions taking place must be thermodynamically favorable under biological conditions, the intermediate metabolites must not be toxic, and the enzymes required for the conversions must be expressible in the host organism. Often, the interconnectedness of metabolic reactions is such that the number of potential pathways linking substrate to product is huge; thus, these pathways must be not only enumerated but also assessed against these other criteria. The earliest complete method for pathway enumeration from a database of biochemical reactions emerged in the early 1990s (6) and was used to identify several possible routes in the biosynthesis of lysine. Analysis of these pathways provided information on the maximal yields that could be expected, suggested approaches for bypassing critical bottlenecks, and identified key intermediates. A more generalized algorithm, the Biochemical Network Integrated Computational Explorer (BNICE), was developed in the early 2000s (7). This method considers the chemical structure of the substrates and products, in addition to a set of enzymatic rules based on the elasticity coefficient (EC) classification, to describe pathways between them, thus facilitating the design of pathways de novo. Its efficacy was demonstrated by analyzing the aromatic amino acid pathways and showing a large array of compounds that could be synthesized from these starting points, in addition to multiple novel pathways for the production of phenylalanine, tyrosine, and tryptophan. These pathways were further evaluated through consideration of their thermodynamic feasibility, facilitated by application of group contribution theory to establish the free energy changes of reaction (8, 9). This type of methodology is under continual development and was employed recently for the automated design and evaluation of pathways for production of 3-hydroxypropanoate, an important industrial precursor (10), and for the production of 1,4-butanediol from a variety of sugar streams (11).

Understanding Metabolic Fluxes and Metabolic Control

Once the possible pathways have been established, the next step in engineering a production phenotype is to analyze both the rates at which these pathways operate and their control architecture, to then propose rational targets for modification (12). Metabolic engineering thus differs from applied molecular biology because it considers integrated pathways as opposed to individual reactions. The primary method for analyzing these networks is the platform of metabolic flux analysis (MFA), whereby fluxes are determined under different conditions and their deviations from the control conditions are used to identify kinetic limiting steps (13). MFA has been the subject of several recent reviews and a book chapter (14–18). In MFA, intracellular fluxes are calculated by using a stoichiometric model for the major intracellular reactions and applying mass balances around intracellular metabolites. A set of measured extracellular fluxes, typically uptake rates of substrates and secretion rates of products, is used as input to the calculations. The final outcome of flux calculation is a metabolic flux map showing a diagram of the biochemical reactions included in

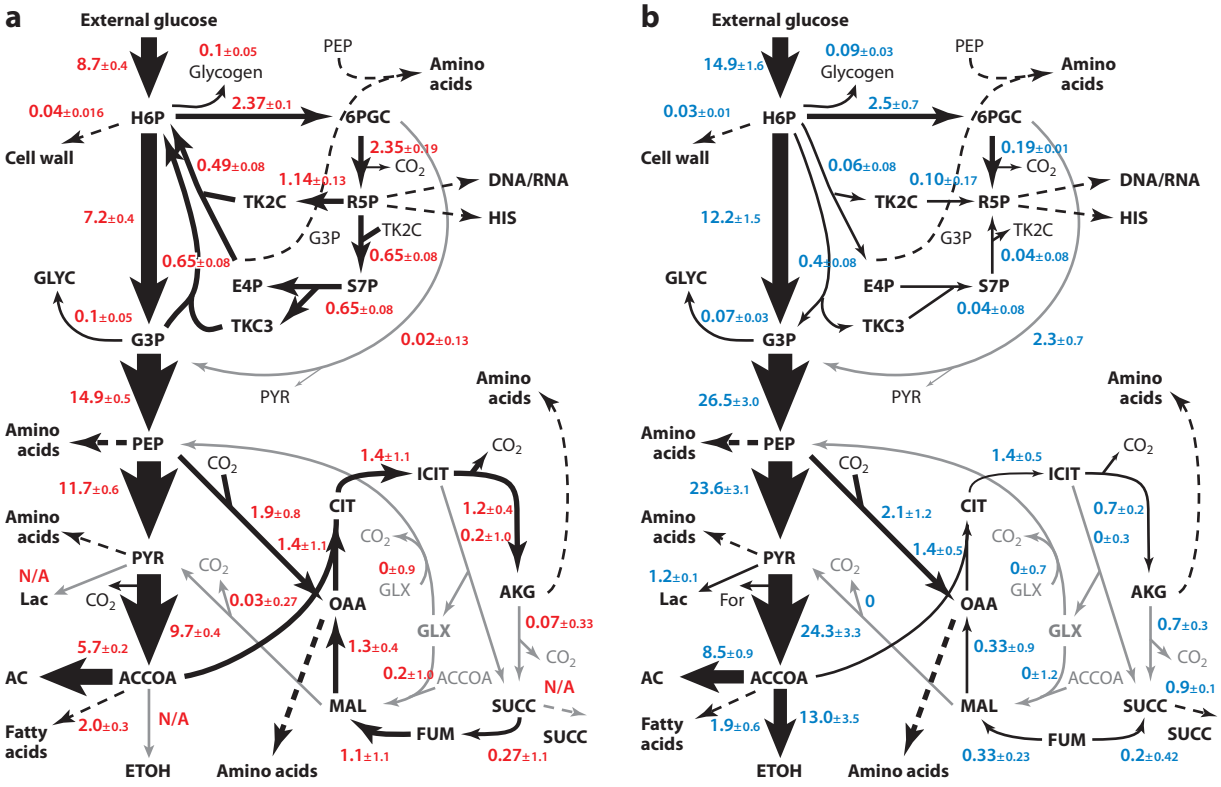

Figure 1

^{13}C flux maps comparing aerobic and anaerobic growth of *Escherichia coli*. ^{13}C net fluxes of central metabolism determined from best fitting with (*a*) 153 (aerobic) and (*b*) 157 (anaerobic) measurements of uptake, secretion, biomass accumulation rates, and ^{13}C labeling. Size of black arrows is proportional to flux values relative to glucose uptake rates. Reproduced with permission from Reference 178.

the calculations along with an estimate of the steady state rate (i.e., the flux) at which each reaction in the diagram occurs. **Figure 1** shows two such flux map examples, which analyze central carbon metabolism of *E. coli* grown aerobically and anaerobically.

The information thus obtained can be enormously helpful in metabolic engineering, helping to identify critical branch points in pathways (19, 20), discover unusual pathways in less-characterized species (21), and define the maximum theoretical yield for the synthesis of products from complex integrated pathways producing and consuming multiple cofactors and intermediates (22). Knowledge of these factors can be used to select targets for genetic manipulation. Of critical importance are not the fluxes per se, but their deviations from a control (base) state and their flexibility to alteration by genetic modulation and environmental manipulation. This gives rise to the concept of flexibility and rigidity of enzymes and branch points of metabolic networks, whereby enzymes and branch points that are rigid (i.e., resist change when the expression of neighboring genes or activity of other enzymes is altered) harbor most of the kinetic control, whereas flexible enzymes and branch points are amenable to easy modulation.

A rigorous methodology for the analysis of these concepts is found in MCA, a theory developed during the 1970s as a logical extension of traditional control theory (23, 24). Using experimentally determined fluxes and their changes in response to a perturbation of the system, one can calculate for each enzyme in a pathway the flux control coefficient (FCC), or the degree of control that

enzyme exerts on the flux of a given metabolite; the concentration control coefficient (CCC), or the degree of control the enzyme has over the concentration of a given metabolite; and the EC, which defines an enzyme's capacity to change its rate in response to a perturbation (e.g., substrate concentration, inhibitor concentration). The FCC and CCC are properties of the network as a whole, whereas the EC is a function of the specific enzyme in question. Taken together, this information can help direct the rational application of genetic engineering to divert more flux to a specified product. For example, in a linear pathway, if one particular enzyme exhibits a higher FCC than the rest, then this enzyme can be considered rate limiting. Thus, its overexpression leads to increased flux through the pathway. However, if the control coefficients do not vary significantly between the enzymes, then control of the network is said to be distributed. In most biological systems, the latter behavior is found, which explains why early genetic engineering approaches, which centered around the idea of overcoming a rate-limiting step, resulted in little success. In contrast, in a landmark study in the early 1990s, Niederberger et al. (25) were able to increase flux to tryptophan in an overproducing strain of yeast by simultaneously overexpressing five of the genes from a plasmid. Other conceptual strategies for metabolic engineering based on MCA have been well documented and employed for the overproduction of a wide range of products (see 26 and 27 and references therein).

Computational Methods of Metabolic Engineering

To provide targets for metabolic engineering as described above, MCA relies on accurate determination of fluxes throughout a metabolic pathway. However, in all but the simplest pathways, the basic implementation of MFA does not provide a high enough resolution. This can result from factors such as parallel internal pathways, futile cycles, and bidirectional reaction steps (18). To provide more insight, techniques using stable ^{13}C isotopes were developed. In this approach, cells are cultivated by using labeled carbon substrates, and, once metabolic and isotopic steady state is reached (usually in a chemostat), the distribution of isotopes in the metabolites is measured by GC-MS or ^{13}C-NMR. An iterative algorithm is then used to give a much more accurate assessment of the intracellular fluxes, based on the so-called isotopomer distribution. Sophisticated computational algorithms to accomplish this were developed in the late 1990s and refined in the 2000s (28, 29). An inherent problem with this approach, however, is the cost and expertise associated with the experimental procedures needed. This has led more recently to the development of purely computational methods to estimate fluxes under different conditions. This approach, termed flux balance analysis (FBA), uses the same genome-scale metabolic stoichiometric model (GSM) and material balances as those used in MFA. At this point, the system is underdetermined (a large number of possible flux solutions are possible), but rather than incorporating isotopomer concentrations to satisfy the remaining degrees of freedom, the equations are formulated as a constrained optimization problem, with a particular objective function to be maximized or minimized. Common objective functions are the maximization of cell growth and the minimization of metabolic adjustment (30). In the first case, the fluxes in the system are determined such that they fall into a reasonable range while providing for the maximum possible biomass accumulation. There are three key advantages to this methodology over the experimental approach: (*a*) minimal experimental data are needed; (*b*) perturbations (e.g., knockouts) can be made by changing a 1 to a 0 in silico, whereas to accomplish the same modifications in vivo requires much more elaborate experimental procedures; (*c*) as new -omics data become available and genome-annotation methodologies improve, the models can be continually refined (in the form of modified constraints), for example, with the introduction of regulatory information or thermodynamic constraints about the constituent reactions (31). Of course, the major disadvantage of this approach and FBA-derived fluxes is that

the latter are not the real fluxes of the system but fluxes yielding maximum biomass, an important distinction often missed in FBA-based computations of metabolic engineering gene-modulation strategies. Comparisons of maximum-biomass fluxes with real ones derived from isotope labeling and MFA computations have yielded mixed results: In some cases, the agreement is satisfactory, whereas in other cases, the two types of fluxes differ significantly (32). Generally, the two cannot be assumed to be the same. Despite these shortcomings, the FBA framework and GSM models continue to advance. Recently, for example, a very comprehensive GSM for *E. coli* was developed, including 2,077 reactions and 1,039 metabolites (33). Such models have found considerable use in metabolic engineering. For example, using FBA, researchers identified and then experimentally validated knockout targets for the overproduction of lycopene, with an increase of 37% yield over the parent strain (34). Some of these methods have been formalized and made widely available, such as the algorithm OptKnock, which allows a user to determine a set of targeted knockouts that will increase flux to a desired product (35). These models are not limited to *E. coli*, and GSMs have been developed for numerous organisms, including *Saccharomyces cerevisiae* (36), *Bacillus subtilis* (37), *Clostridium acetobutylicum* (38), *Corynebacterium glutamicum* (39), *Arabidopsis thaliana* (40), and even humans (41). For a more complete review on the implementation and applications of GSM, the recent reviews by Blazeck & Alper (42) and Palsson et al. (43, 44) can be consulted.

The Rational Approach to Metabolic Engineering

The rational approach to metabolic engineering encompasses methods relying on knowledge of network topography, as well as kinetics and regulation of a metabolic pathway, identification of kinetic bottlenecks through labeling methods and flux determination, elimination of side reactions through pruning the corresponding genes, and assessment of dysfunction in the seamless operation of a pathway through measurements of accumulating intermediates. Most initial metabolic engineering efforts relied on such rational approaches to pathway analysis and flux optimization to rationally manipulate metabolism and shift flux toward a desired product. The field is now approximately 20 years old, and a vast array of success stories of this type exist. Instead of an exhaustive survey of all these applications, we provide a few choice examples that exemplify the general methodologies used.

A classic example that illustrates the importance of integrated pathway analysis is the overproduction of lysine in *C. glutamicum*, shown in **Figure 2**. A key objective here was to increase lysine yield on glucose, and numerous genes had been targeted to this end. For example, assuming that NADPH supply was limiting lysine biosynthesis, the reactions of the pentose phosphate pathway had been modulated. Or, assuming a kinetic bottleneck in the aspartate pathway, other enzymes had been similarly targeted for overexpression. Other efforts to modulate the central glycolytic pathway had similarly yielded minimal results. Using labeled glucose and flux analysis, researchers first determined that up to 90% of the lysine flux was channeled through pyruvate carboxylase (*pyc*) (45–48). *pyc* overexpression thus became an obvious first target; however, the resultant *pyc*-overexpressing strain exhibited reduced specific lysine productivity with a concomitant increase in cell growth. Although unexpected, this result identified aspartokinase (*ask*) as a next possible bottleneck, because overexpression of pyruvate kinase without simultaneous amplification of the fluxes downstream of oxaloacetate in the lysine pathway would result in accumulation of tricarboxylic acid (TCA)-cycle intermediates and thus fuel cell growth instead of lysine product formation. This suggested that *ask* should be overexpressed to divert the additional flux out of the TCA cycle toward lysine synthesis. Expressed alone, this enzyme increased lysine productivity but drastically reduced growth rate, presumably owing to depletion of TCA-cycle intermediates. However, the simultaneous expression of both enzymes (*pyc* and *ask*) increased flux both into and

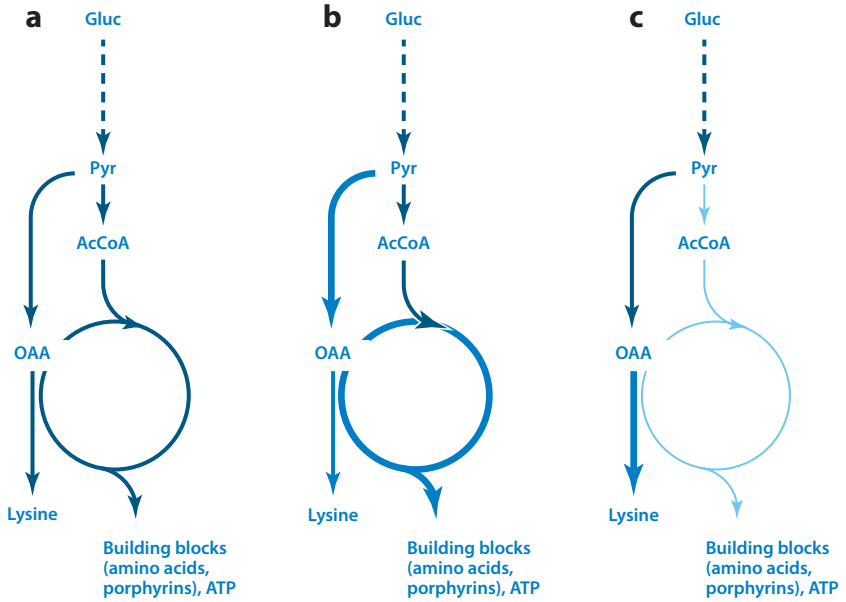

Figure 2

Metabolic engineering of lysine production in *Corynebacterium glutamicum*. Schematic model of the metabolic effects of pyruvate carboxylase and aspartate kinase overexpression: (*a*) wild type, (*b*) *pyc*-overexpressing strain, and (*c*) *ask*-overexpressing strain. The overexpression of *pyc* gene increases the carbon flux entering the tricarboxylic acid (TCA) cycle. The presence of bottlenecks in the lysine pathway minimizes the effect of *pyc* overexpression on lysine. Instead, increased carbon flux to the TCA cycle supplies precursors to other pathways fueling cell growth. However, the overexpression of aspartate kinase increases carbon flux to the lysine pathway. Without a commensurate increase of the anaplerotic fluxes, the increase in lysine flux drains TCA cycle intermediates, thus negatively impacting cell growth. As a result, lower growth rate and biomass formation are observed for the recombinant strain. Reproduced with permission from Reference 49. Abbreviations: AcCoA, acetyl-CoA; Gluc, glucose; OAA, oxaloacetate; Pyr, pyruvate.

out of the TCA cycle, balancing the intermediates while increasing lysine productivity 150% and maintaining the same growth rate as the control strain (49).

Another example is the engineering of *E. coli* for the production of 1,3-propanediol from glucose, reviewed by researchers at DuPont (50) and shown in **Figure 3**. This compound is naturally produced from glycerol, by conversion first to 3-HPA via glycerol dehydratase and then to 1,3-propanediol via an oxidoreductase. To facilitate production from glucose, a combination of rational approaches was taken: First, knockouts of nonproductive reactions were introduced to prevent reincorporation of glycerol (produced from glucose) into central metabolism. Second, the glucose phosphotransferase system was eliminated. This system is PEP-dependent and less energetically efficient than the ATP-based alternative. Finally, several other enzymes were targeted for expression-level variation. These modifications together allowed the production of 1,3-propanediol at a rate of 3.5 g/liter/h, titers of 135 g/liter, and a yield of 51%.

Glycerol is a by-product of ethanol fermentation that reduces the yield of converting glucose to ethanol. The ethanol pathway is cofactor balanced, with no net production or consumption of reducing equivalents (NADH or NADPH). This allows a perfect conversion of glucose to ethanol with 100% energetic yield. This balance, however, is disturbed if some glucose is consumed for biomass synthesis, yielding a surplus of NADH production. Glycerol synthesis provides a sink

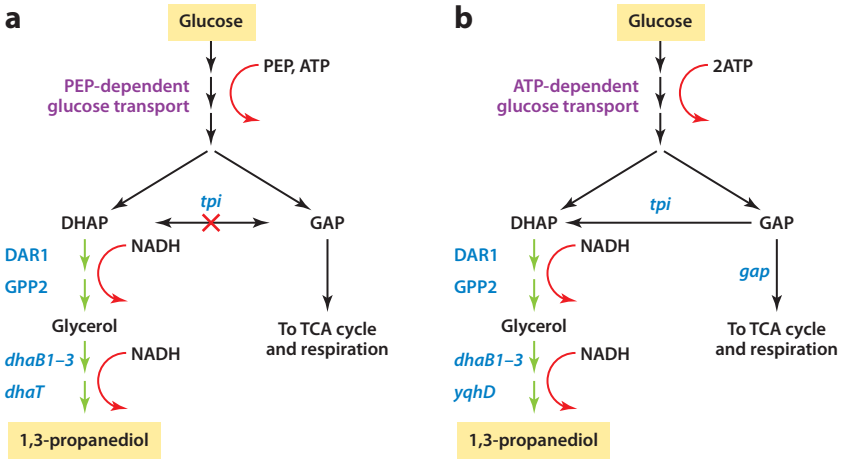

Figure 3

Strain development for 1,3-propanediol production. Engineering metabolic pathways from D-glucose to 1,3-propanediol: development from (*a*) an early strain to (*b*) a late strain. Gene names are indicated in blue; intermediate metabolites are indicated in black. Reactions encoded by genes native to the host organism are indicated by black arrows (deleted native genes indicated by an X); reactions encoded by genes obtained from donor organisms are indicated by green arrows. Energy (ATP or PEP) and redox demand (NADH or NADPH) are shown by red arrows. The further metabolism of glyceraldehyde-3-phosphate (GAP) via the tricarboxylic acid (TCA) cycle and respiration (oxygen consumption) contribute to cell mass, energy, NADH, and NADPH generation. Abbreviation: DHAP, dihydroxyacetone phosphate. Reproduced with permission from Reference 50.

for the extra NADH molecules at the cost of negatively impacting the economics of ethanol production from corn and sugarcane. Glycerol production was reduced by introducing *gapN*, which encodes nonphosphorylating, $NADP^+$-dependent glyceraldehyde-3-phosphate dehydrogenase (GAPN) from *Streptococcus mutans*. This bypasses GAPDH and PGK reactions in glycolysis with the simultaneous synthesis of NADPH, which is easily used in biosynthesis. The net result was reduction of glycerol production by 40% (51).

Another example from the area of biofuels is the engineering of *S. cerevisiae* for the use of xylose, a substrate that is not naturally assimilated by yeast. This challenge has occupied researchers for the past 30 years and has resulted in quite efficient strains converting xylose to ethanol through the expression and balancing of the xylose reductase and xylitol dehydrogenase enzymes (52), or xylose isomerase (53, 54). To enhance the simultaneous use of glucose and xylose, researches sought to avert catabolite repression, which causes assimilation of xylose to occur only after glucose has been depleted, with an intermediate lag phase. To this end, they expressed a cellodextrin transporter and an intracellular β-glucosidase that allow direct assimilation of cellobiose, which led to the concomitant consumption of both glucose and xylose, and an overall faster process (55). In the same area, the engineering of *E. coli* to allow it to convert glucose (and xylose) to ethanol through the introduction of pyruvate decarboxylase and alcohol dehydrogenase II from *Zymomonas mobilis* should be mentioned (56). Deletion of additional genes for the synthesis of lactate and acetate improved the overall yield of ethanol production from glucose.

Increases in product titers have been traditionally achieved through the elimination of so-called bottlenecks. Isoprenoids represent a large class of important pharmaceuticals, fragrances, and fuels, and as such, their production in microbial hosts has occupied researchers for many years. [Isoprene (57), artemisinin (58), taxol (59), and geranylgeraniol (60) are a few examples

in this large field.] The initial step in isoprenoid biosynthesis involves the condensation of two central carbon metabolites, glyceraldehyde-3-phosphate (G3P) and pyruvate. It was hypothesized that an imbalance in the availability of these two compounds acted as a bottleneck in the pathway, so that expression of several genes in central carbon metabolism was altered to eliminate this bottleneck. It was found that limited supply of G3P was the cause of the bottleneck and that expression of a pathway step that converts pyruvate to phosphoenolpyruvate (PEP) (which is then in equilibrium with G3P) can yield a fivefold titer improvement for the isoprenoid pathway (61). Further down in the isoprenoid pathway, it was found that *dxs*, *ispG*, *ispH*, and *idi* were limiting the flux of the MEP pathway. As such, they were collectively overexpressed in a recent study aimed at overproduction of taxadiene, the first committed taxane precursor in the taxol pathway (62). This study also applied a multivariate modular approach to balance the upstream pathway, leading to the synthesis of isopentenyl pyrophosphate (IPP), a central, key isoprenoid intermediate from which a large number of terpenoids can be synthesized, with the downstream pathway leading to GGPP and taxadiene synthesis from IPP. The correct balancing of the two pathway modules eliminated inhibitory intermediates and allowed accumulation of gram-level quantities of taxadiene, more than one-thousandfold improvement over the state of the art.

As mentioned, the above are only a few examples in a fast-growing field that is expanding to encompass a great diversity of applications spanning several industries. We do not provide comprehensive coverage here. However, we would be remiss to not mention (*a*) recent efforts to engineer higher and branched alcohol pathways in various organisms, notably isobutanol from glucose in *E. coli* (63) and yeast (64–67), through the engineering of the branch chain decarboxylase Erlich (valine) pathway; (*b*) synthesis of hydrocarbons in *E. coli*, facilitated by the discovery of a gene encoding a decarbonylase enzyme (68), numerous accomplishments in engineering flavonoid synthesis in *E. coli* (69, 70) and alkaloids in yeast (71), and various terpenoids, such as levopimaradiene, in *E. coli* (72); (*c*) ongoing efforts for the synthesis of commodity chemicals such as isoprene (73), acrylic acid (74), succinic acid (75), butane diol (76), and glucaric acid (77); and (*d*) burgeoning activity in the synthesis of natural products, such as tetracycline derivatives (78).

The common theme in the examples presented above is the construction of a pathway through the combined expression of the constituent genes followed by selective targeting of a few key enzymes for modification, be it overexpression, knockout, or inhibition removal, aiming at the optimal performance of the pathway. Over the years, these efforts have been enhanced by advances in molecular biology tools that have facilitated incredibly the pace of research and accomplished progress to the point where work that was the foundation of a PhD thesis in the 1990s can now be accomplished in a few days. For example, whole libraries of promoters, ribosomal binding sites, and terminators have been characterized and are available in a registry (BioBricks). Methods for the rapid generation of knockout strains have been developed (79), and the ability to integrate specific DNA into the host chromosome is now commonplace. This burgeoning repertoire in no small part explains the rapid increase in the number of publications in the field that has been seen over the past few years (**Figure 4**)

Combinatorial Methods for Metabolic Engineering

As shown above, application of rational metabolic engineering and proper modulation of a few well-selected pathway genes are often sufficient to allow production of target molecules at titers on the order of g/liter. These methods are centered around a detailed understanding of the pathways responsible for converting a reactant to a product as well as their associated fluxes and regulatory systems. However, this road map alone is insufficient for optimization of a production phenotype.

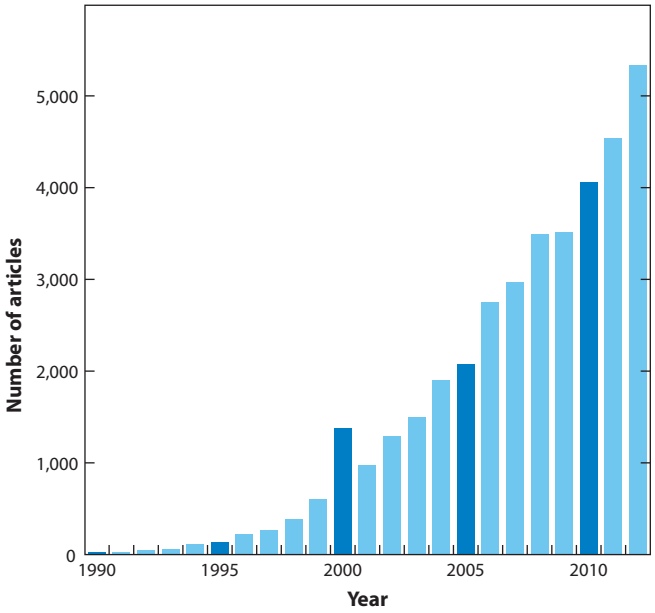

Figure 4

Growth of the metabolic engineering literature. The number of scientific papers published each year from 1990–2012 containing the term "Metabolic Engineering," obtained through GoogleScholar search.

Other factors, including global regulatory networks, metabolic burden, and product toxicity, must also be considered to achieve high levels of production. Often the connection between these factors and the genotype is complex, involving multiple genes in a poorly understood manner, and thus their manipulation can be difficult to target through a rational approach. Combinatorial techniques have been very useful here, because, unlike rational techniques, they do not necessarily rely on a priori knowledge of biosynthetic bottlenecks. Instead, random or semirandom changes are made to the genotype of a strain, then the resulting mutants are screened for an improved phenotype. Several of these methods are described below, along with representative examples of their successes in metabolic engineering. Though they differ in implementation, they all share a common theme of stochastically introducing diversity into a population then selecting for an improved strain in a high-throughput manner, in certain cases identifying the nature of the genetic element that brought about the desirable phenotypic change.

The earliest examples of combinatorial metabolic engineering date back to the use of random mutagenesis in the development of industrial microorganisms. Mutagenesis was carried out by toxic reagents or UV light to increase mutation frequency, followed by screening of the resultant strains for an improved production phenotype. This approach was very popular and was applied extensively to industrial strains, which yielded a long list of production organisms that heralded the dawn of industrial microbiology, including the flagship example of the engineering of penicillin production levels in *Penicillium chrysogenum* to 50 g/liter, an improvement of 4,000-fold over the original strain (80). In the early 2000s, a new approach, termed IME, extended the scope of combinatorial metabolic engineering to include isolation of the specific mutation that conferred the productive phenotype, to facilitate the rational reconstruction of the strain via molecular biology (81). This approach has been used extensively since then and is fundamentally similar to directed evolution of proteins, except that in this case genetic diversity is targeted not only to the open

reading frame but to any number of genetic elements, and the ultimate goal is directly associated not with the mutated protein but with the improved phenotype of the whole cell. An illustrative example of IME was the improvement of xylose uptake and ethanol production in *S. cerevisiae* (82). A library of genomic fragments from the xylose-consuming yeast *Pichia stipitis* was transformed into *Saccharomyces*, and the resulting strains were screened based on growth rate in xylose-containing medium. DNA sequencing of the highest-performing strains identified two genes present in the majority of the plasmids, *xyl3* and *psTal1*, which were responsible for the improved phenotype. These findings were verified by expressing these genes, either from an expression vector or from the chromosome, and observing a 100% increase in growth rate and a 70% increase in ethanol production rate over the parent strain. Recently, IME has been used to generate an *E. coli* strain for improved heterologous protein production (83). By expressing *E. coli* genome fragments in an antisense orientation, such that mRNA would bind to the complementary native mRNA and downregulate translation, knockdown targets that retarded growth and improved GFP expression were identified.

Three important factors affect the success of IME: (*a*) A high-throughput screen must be developed for the desired phenotype; (*b*) the library generated must be of high quality and must be able to generate sufficiently large diversity, thus increasing the probability of harboring a genotype with the phenotype of interest; and (*c*) one must rapidly determine which genetic changes relate to the desired phenotype. Advances in DNA sequencing and microarray technologies have facilitated the latter of these endeavors (84, 85). The first is arguably more complicated and case dependent, because in many cases the phenotype of interest is production of a small molecule for which conventional assay procedures (e.g., HPLC, GC-MS) are not conducive to screening a library of thousands of mutants. Thus, development of facile, high-throughput screens is an active area of research. For example, L-tyrosine is a valuable target for microbial production, but there are no convenient direct assays for its production in vivo. Recently, researchers developed an assay whereby a heterologous tyrosinase enzyme catalyzed the production of the black pigment melanin from tyrosine, which could be measured both visually and spectroscopically (86). This assay was then used to identify highly producing strains in a combinatorial library. In general, intracellular products are easier to assay because they are retained within the cell, where they can be detected by a variety of staining methods using instruments specifically designed for the measurement of intracellular products, such as flow cytometry. Secreted products, however, represent a special challenge because such products do not retain their connection with the (overproducing) cell that secreted them. One recently developed method makes use of microfluidic technology whereby cells are encapsulated within aqueous microdroplets suspended in an oil solution. In this environment, secreted products are trapped within the microdroplet, where they can be assayed to identify and sort microdroplets with high product concentration harboring overproducing cells. This idea was implemented to assay yeast cells capable of consuming xylose at high rates (87). Analysis of the genetic inserts responsible for high xylose assimilation rate identified multiple copies (as many as 50 or more) of the gene encoding xylose isomerase as the key determinant of the high xylose assimilation phenotype. The above are but two examples of successful generation of a novel assay. For a detailed review on current and ongoing work in this field, the reader is directed to the recent review by Dietrich et al. (88).

Finally, library quality is essential to the successful implementation of a combinatorial method. A recent study aimed at improving solvent tolerance in *C. acetobutylicum* via the generation and transformation of a genomic library demonstrates this point. The first aspect of library quality is completeness. In this study, only 73% of the *C. acetobutylicum* genome was incorporated into the library, possibly precluding the identification of important genes. An additional common flaw in library design is that cells can be preferentially transformed with plasmids with smaller insert sizes,

which in this case prevented large operons responsible for a desired trait from being identified (89). Thus, consideration must be given to how to best generate an unbiased library for best results.

With the rapid development of recombinant DNA technologies, methods for introducing test mutations into a strain have improved drastically in both their specificity (i.e., changes can be localized to a particular region of the genome that is more likely to be effective) and their efficiency in generating large strain libraries rapidly (90). Thus, global transcription machinery engineering (gTME) introduces mutations to the DNA-binding region of the transcription factor via error-prone PCR. These mutations change the affinity of the transcription factor for specific promoters, thus altering the expression profile of the cell on a global scale. This approach was first used for improving the tolerance of *Saccharomyces* to ethanol (91), and its applicability soon broadened to include other phenotypes and microbial systems (92). Similarly, optimal balancing of expression levels of a specific set of pathway genes can be accomplished in a high-throughput manner by using the newly developed COMPACTER method (93). Each gene in the pathway is placed under control of an individual promoter and terminator. Nucleotide analog mutagenesis is then used to introduce mutations into the promoters, which thus generates a library of strains that can be tested with a relevant high-throughput screen.

Combinatorial methods for strain improvement need not necessarily be targeted to a global scale. Often, improvements to production can be evidenced by manipulating a single enzyme. One common method involves mutagenesis and directed evolution of an existing enzyme to create new functionality or altered specificity. The power of this tool is evidenced in the recent creation of a novel pathway to polylactic acid (PLA), a biodegradable polymer from lactate. The pathway conceived by the authors required enzymes for the conversion of D-lactate to lactyl-CoA and the subsequent transformation to PLA. Neither of these enzymes exists in nature, but enzymes performing similar reactions but on different substrates were successfully engineered by a combination of random mutagenesis and site-directed and saturation mutagenesis to perform the desired transformations (94). Another example displays the ability to engineer enzyme activity by random mutation and screening. In this example, a high-throughput screen was constructed by generating a temporary pathway that included the enzyme of interest to a colorimetric reporter, lycopene. The enzyme was then mutated, and high-producing colonies were selected based on color, which yielded a significantly more active version of the enzyme. The more active form of the enzyme was then applied to a levopimaradiene-producing pathway, which significantly improved the titer (95).

PART II: A VISION OF THE FUTURE OF METABOLIC ENGINEERING

In the previous section, we described the development of the key principles and technologies that define metabolic engineering as a field, illustrating them with a few examples of their successful application. Many of these technologies are now reaching intellectual maturity, and one might ask what the future of metabolic engineering will be. We believe the answer lies in the integration of the highly developed theoretical framework of metabolic engineering with the rapidly expanding tools and understanding stemming from other areas of fundamental biological research. Thus, as molecular biology, synthetic biology, and systems biology become more advanced, their application in the context of metabolic engineering will rapidly expand the array of attainable products, both natural and nonnatural, as well as the titers and rates at which they can be produced and even the specific organisms used for the task. A review of some of the most applicable of these technologies is the subject of the next section.

Alternative Hosts for Metabolic Engineering

To date, the majority of applications of metabolic engineering have used a small number of model systems, with *E. coli* as the foremost prokaryotic organism and *S. cerevisiae* as the predominant

eukaryote. This is primarily because of the tractability of these organisms: They are easily cultured in the laboratory; a host of molecular biology techniques has been developed for their genetic manipulation; and their sequences were some of the first publically available, facilitating the development of bioinformatics tools to aid in research. In addition, the use of these organisms in the investigation of fundamental biological phenomena during the early days of molecular biological research led to a wealth of knowledge about their physiology, metabolism, and genetics, which further aided their development as useful production organisms. The ability to have DNA synthesized allowed researchers to optimize the codon usage of nonnative genes for improved expression in the host. As shown in the first section, these attributes and developments have led to a multitude of success stories in producing compounds in these organisms from all branches of life.

However, several other organisms have been explored, albeit on a smaller scale, for metabolic engineering and have found niches owing to their ability to grow well in specific environments (e.g., thermophiles, anaerobes), use a variety of alternative feedstocks (e.g., pentoses derived from cellulosic material, gases), or naturally produce and tolerate large amounts of a desired chemical product (e.g., solventogenic bacteria). The ability of an organism to grow at high temperatures, for example, can be a very attractive industrial phenotype: Intrinsic enzymatic rates are often higher with thermostable enzymes, leading to higher productivities. High temperatures also exclude the growth of common mesophilic contaminants and can facilitate simplified product recovery in the case of volatile targets through bioreactor operation strategies analogous to reactive distillation. To this end, *Thermoanaerobacter saccharolyticum* and *Geobacillus thermoglucosidasius* have recently been engineered to produce ethanol at high yields at elevated temperatures by using cellulose-derived sugars (96, 97). Production of several other important biofuel molecules has been targeted to alternative organisms; for example, members of the genus *Clostridium* have recently been engineered for the production of isobutanol (from cellulose) (98). These organisms, which naturally produce solvents through what is known as the acetone-butanol-ethanol (ABE) pathway, are relatively resilient to many solvents toxic to *E. coli*, which makes them a preferable host for their production (99). Recently, the cyanobacterium *Synechococcus* was engineered to produce 1-butanol from carbon dioxide, the first example of a direct gas-to-liquid fuel production platform (100). These efforts serve to highlight only a very few examples of metabolic engineering of nonstandard organisms. **Table 2** lists a small (and by no means comprehensive) collection of other molecules whose production has been successfully engineered into a diverse group of organisms, from thermophiles to acidogens, methanotrophs, cyanobacteria, and even plants.

In many cases, attempts to transfer useful pathway characteristics of these organisms into hosts such as *E. coli* have met with limited success (99). Why, then, given that these capabilities exist naturally in other organisms, are there so few examples of their use for metabolic engineering in the literature? The answer is that tools for their genetic manipulation have lagged significantly behind those for typical hosts. After the genomes of many source organisms had been sequenced, it was far easier, in general, to import genes from them into more tractable hosts (such as *E. coli*) than to attempt to develop tools to engineer them directly, a process that could take years. In *Clostridium*, for example, although replicating plasmids were discovered approximately 20 years ago, efficient systems for chromosomal manipulation were developed only recently (101). Holding back the development of genetic tools is the fact that most transformation methods that have been developed for these other organisms are very species specific; there is no convenient systematic approach for attempting the transformation of a novel organism.

Now, however, owing to accessibility of sequence data and associated bioinformatics tools, as well as general progress in molecular biology, developing tools for such organisms presents a much less daunting barrier. Promoters and terminators likely to function well for heterologous expression can be identified through a variety of publicly available server-based tools. Similarly, targets

Table 2 Examples of metabolic engineering accomplishments in nontraditional hosts

Organism	Product(s)	Reference
Bacillus subtilis	Carotenoids	(159)
	Riboflavin	(160)
	Ethanol	(161)
Bacillus licheniformis	2,3-Butanediol	(162)
Synechococcus elongatus	1-Butanol	100
Clostridium acetobutylicum	1-Butanol	(163) (Review)
Clostridium tyrobutyricum	1-Butanol	(164)
Geobacillus thermoglucosidasius	Ethanol	(165) (Review)
Thermoanaerobacterium saccharolyticum	Ethanol	(165) (Review)
Thermoanaerobacter mathranii	Ethanol	(165) (Review)
Thermophiles (general)	Various	(166) (Review)
Methylomonas sp. strain 16a	Carotenoids	(167)
Bacillus methanolicus	L-Lysine	(168)
Algae (general)	Various	(169) (Review)
Plants (general)	Various	(170) (Review)

for gene knockouts to serve as selectable markers can be inferred from homology. For example, the *pyrF* gene, a component of uracil biosynthesis that can be used for either positive or negative selection, has been identified in numerous genomes and used to develop genetic engineering strategies (102). Potential barriers to transformation, such as active restriction-modification systems that destroy foreign DNA, can be identified in silico, and strategies can be developed to mitigate them (103, 104). Genomics data can also inform the development of the general transformation strategy. A recent comparative genomics study showed that the components of homologous recombination, which is used natively both for repairing DNA breaks and for allowing the horizontal transfer of DNA between species, are present in almost all forms of life (105). The use of this process in biotechnology for the integration of heterologous genes into the chromosome (termed recombineering) therefore represents a general approach for transformation that can theoretically be applied to a multitude of species, especially those for which stably replicating plasmids have yet to be found or are inefficiently maintained. For example, a suicide vector (a plasmid that cannot replicate in the target organism and whose genetic material is delivered by homologous recombination with the chromosome) was recently developed for *Thermoanaerobacterium* spp. and used to integrate the cellobiohydrolase gene from *Clostridium thermocellum* without the need to maintain the recombinant strain on selective media (106). In the past few years, many other reports have been published on the first genetic tools for a variety of organisms. Once the initial barrier of successful transformation is crossed, molecular tools can be rapidly developed and refined based on the long legacy of lessons learned through their development for the organisms in common usage today. There is good cause to be optimistic, therefore, that metabolic engineering will soon be efficiently applied to a less restricted group of organisms, which will allow us to capitalize on the huge diversity of biology.

This raises the interesting question of how to choose a host for a particular application. Is it better to transfer a production phenotype into a traditional host (despite its potential shortcomings in terms of, for example, tolerance to product toxicity or adverse growth environment) or to attempt engineering of an organism better suitable to the conditions of application, despite a less-refined genetic system? We think that this will depend primarily on the complexity of the

desirable traits in the nonstandard organism. For example, researchers have long aspired to generate an autotrophic strain of *E. coli* to enable the production of high value compounds from CO_2. However, the complexity of the pathways required makes it more efficient to develop the tools needed for producing the compound of interest in an organism already capable of autotrophic growth, such as the case described where butanol production was engineered into cyanobacterium (100). However, when the desired trait can be distilled to a single gene or simple set of genes, then the more facile approach may be to port this capability into an easily manipulable organism. Further, there are often advantages to expressing a pathway in a heterologous host, because the native regulatory systems are not present (107). The decision of what host to employ will thus vary on a case-by-case basis; here, we simply highlight this paradigm shift, which has been generated by the unprecedented advances in genetic manipulation of noncanonical organisms, coupled with the wealth of sequencing information now available from genome sequencing data and bioinformatics tools.

Emerging Methodologies for Improving Production Titers

As shown in the first section, the traditional rational methods for strain optimization have focused on overexpressing or knocking out specific enzymes to modulate their activity. However, as molecular biology advances, more finely tuned methods for regulating pathway flux are being developed. One promising method for improving pathway flux is to optimize the spatial organization of enzymes within a multistep pathway, which can reduce buildup of inhibitory intermediates, reduce siphoning of intermediates toward alternative metabolic routes, and aid in overcoming diffusional limitations (107). To date, three methods with varying degrees of complexity have been used to colocalize pathway enzymes: generation of chimeric fusion proteins, binding of proteins to scaffolds, and localization to a subcellular compartment or organelle, as outlined in **Figure 5**. These approaches represent a significant departure from the traditional paradigm, where heterologous enzymes have been targeted either to the cytosol or to the cell membrane, with little consideration for pathway sequestration that occurs natively.

Fusion proteins. A simple, but limited, approach to colocalization involves the generation of chimeric fusion proteins connected by a linker sequence. This approach has been the most widely used historically and is implemented by fusing two or more genes with a linker region, which generates a single large polypeptide with two or more functionalities. The linker region reduces folding interference, enabling the individual proteins to retain their native activity, and either separates domains spatially or allows them to interact if necessary (108, 109). This methodology has been applied to proteins that are sequential steps in a pathway to increase local concentration of intermediates, and it has also been applied to proteins in which protein-protein interactions are required for functionality, such as in passing activated intermediates (110). A recent study on a fusion protein of 4-coumaroyl-CoA ligase::stilbene synthase, linked with a three–amino acid sequence, improved production of resveratrol 15-fold relative to expression of the individual enzymes (111). A follow-up kinetic and structural study confirmed that the improvement in yield resulted from proximity of the enzyme active sites rather than from improvement in catalytic activity of either enzyme (112). These studies illustrate the ability of colocalized enzymes to overcome diffusional limitations to improve concentrations and yields. Another recent example includes improvements in α-farnesene production by creating a protein fusion of FPP-synthase and α-farnesene synthase. An interesting result from this study is that fusion of the two proteins eliminated production of farnesol, produced via an alternative metabolic route, which illustrates that colocalization of enzymes can also reduce or eliminate unwanted side reactions (113).

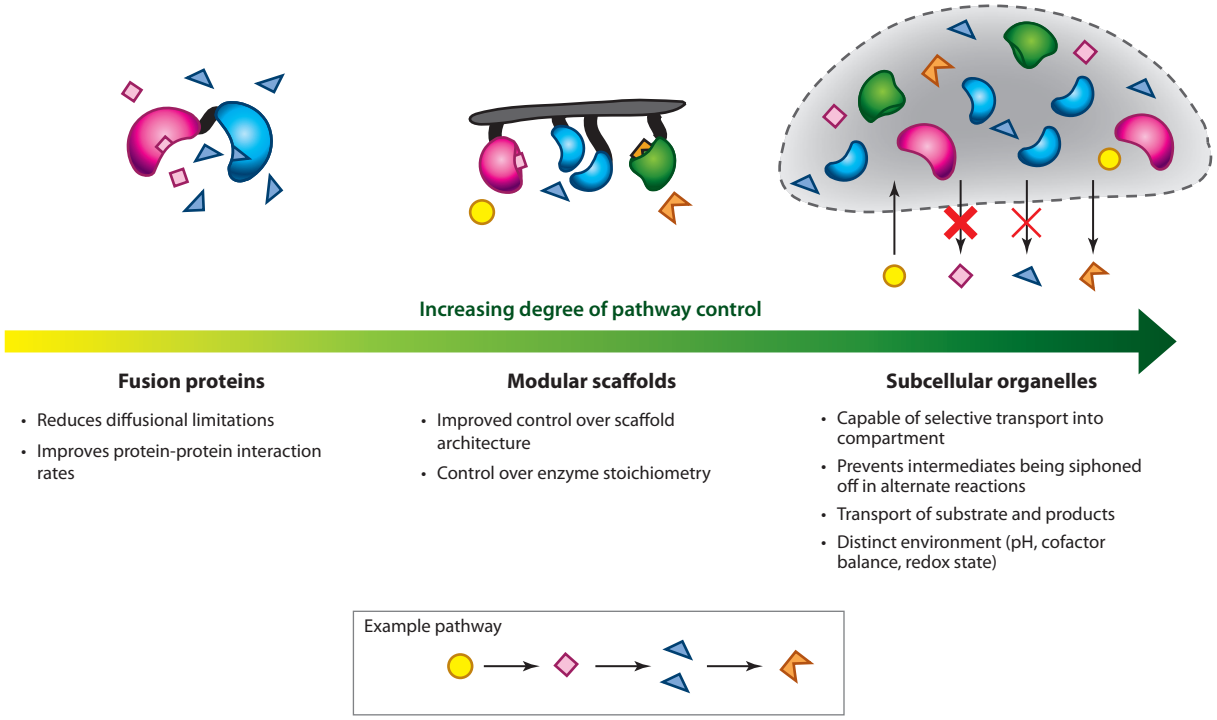

Figure 5

Colocalization of metabolic pathway enzymes. Example methods for colocalization of metabolic pathways with complexity and degree of control increasing from left to right. Fusion proteins exhibit the most basic benefits of colocalization by reducing diffusional limitations and improving protein-protein interaction rates. Modular scaffolds bear further improvements, such as control over pathway architecture and enzyme stoichiometry. Localization to subcellular organelles further improves upon this because it can directly limit or allow transport of substrates, intermediates, and products, preventing side-reactions, loss of intermediates, or product inhibition, among other factors. Furthermore, subcellular organelles offer the benefit of a distinct environment with pH, cofactor balance, and redox state differing from cytosolic values.

Scaffolds. Attaching pathway enzymes to a scaffold has more recently emerged as a versatile method for colocalization. Although more complex, this method has several advantages over fusion proteins, such as the ability to modulate scaffold architecture and the number of bound enzymes. The scaffolds used have included RNA molecules (114); DNA molecules, to which proteins bind via attached zinc-finger domains (115); and coexpressed scaffold proteins with protein-protein interaction domains for attachment of pathway enzymes (116). Architecture is key to pathway optimization when using scaffolds, because the ratio and spatial orientation of individual steps is essential to obtain high throughput. One recent example, which made use of a protein scaffold, yielded a remarkable 77-fold improvement in mevalonate pathway flux with the optimum scaffold architecture (116). Further application of this strategy, again involving colocalization of three-pathway enzymes, recently improved production of D-glucaric acid by fivefold (117). As with fusion proteins, scaffolds are not limited to only catalyzing successive pathway steps but are also capable of increasing the rate of protein-protein interaction (118). Additionally, scaffold proteins may also be engineered to generate new or hybrid signaling pathways to generate new phenotypes, as illustrated in a recent study generating an engineered MAP kinase pathway with nonnatural inputs and outputs (119). For a more

extensive discussion on engineered signaling pathways, see the recent review by Good et al. (120).

Compartmentalization. The final, and most complex, method for colocalization involves the localization of pathway enzymes to cellular compartments or organelles. These organelles can be orders of magnitude larger than scaffold proteins, capable of holding a large number of proteins, and better able to limit mass transfer across their boundaries. These organelles can thus sequester intermediates to prevent toxic effects to the cells and prevent metabolic side-routes as well as increase local concentration for higher turnover.

Although prokaryotes lack true organelles, proteinaceous bacterial microcompartments (BMCs) (121, 122), along with other, smaller protein shells, do occur naturally. Because the study of BMCs is still in its infancy, they have not yet been applied for the production of target compounds. However, preliminary work has been performed on heterologous expression and targeting. For example, CO_2-fixing BMCs from *Halothiobacillus neapolitanus* were heterologously expressed in *E. coli* through the expression of 10 genes and were shown to possess their CO_2 fixation activity in vivo (123). Similar work has resulted in expression of empty propanediol-utilization BMC shells in *E. coli* (124), and localization of proteins to the protein shell has been achieved by using an N-terminal peptide tag in *Salmonella* (125), which suggests the possibility of targeting new proteins and entire synthetic pathways to BMCs in the near future. For a more in-depth review of recent scaffold- and protein-based compartments research, see the recent review by Lee et al. (126).

In eukaryotic hosts, targeting pathway enzymes to organelles can potentially improve pathway productivity through the same means imparted by other localization methods. Additionally, these organelles can possess a pH, redox potential, cofactor balance, and metabolite composition different from the cytosol, which allows enzymatic reactions to occur that may otherwise be thermodynamically unfavorable in the cell (127). Currently, the main body of work involved in organelle engineering focuses on protein targeting using N-terminal targeting sequences for preexisting organelles, such as the peroxisome (128), vacuole (129), mitochondria, and chloroplast. Recent work demonstrating the utility of organelle targeting has included production of plant terpenoids in yeast mitochondria, which improves yield by 8- to 20-fold, likely owing to increased precursor availability (130).

The true potential of organelle-based localization lies not in just hijacking existing cellular organelles but in generating entirely new organelles with designed functionalities. As of yet, no true synthetic organelles (excluding BMCs) have been generated; however, a recent study introduced the cyanobacterium *Synechococcus* strain PCC 7942 into eukaryotic animal cell lines through direct injection, invasion, and phagocytosis and showed these new organelles could be maintained in the cell lines (131), which suggests that synthetic organelles are not out of the realm of possibility.

Design of Novel Genetic Circuits for System-Level Control

In most applications of metabolic engineering to date, the activity of heterologous pathways is typically modulated with static promoters, which fix expression at a constant level regardless of the state of the cellular environment or substrate availability. However, for optimal overproduction of a compound, it may become necessary to generate dynamic (or flexible) regulatory action in the host metabolism to respond to varying cellular conditions.

A significant advance in this direction was made recently with the construction of a dynamic sensor-regulator system (DSRS), used for the production of the biodiesel fatty acid ethyl ester

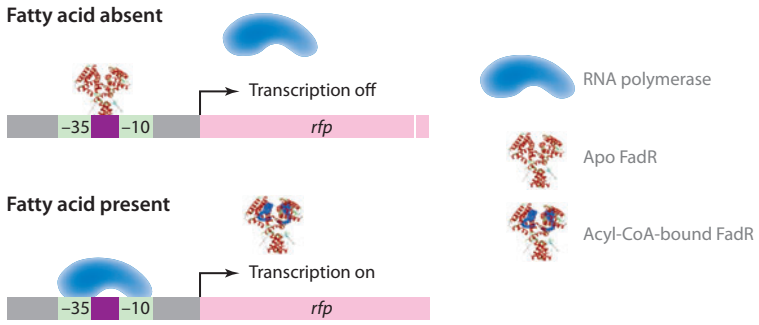

Figure 6

Design of biosensors for dynamic sensor-regulator system. Transcription of pathway genes is regulated by pathway precursors (in this case, acyl-CoA), which allows the transcription of pathway enzymes to be proportional to substrate availability. In this example, in the absence of acyl-CoA, FadR binds within the RNA-polymerase-binding site and blocks transcription. Alternatively, an excess of acyl-CoA results in binding to FadR, which reduces its ability to bind DNA and allows transcription of pathway genes. Used with permission from (132).

(FAEE). This system uses a transcription factor that senses a key pathway intermediate (fatty acids) and dynamically regulates gene expression of the downstream pathway genes, such that ethanol and the enzymes required for esterification to form FAEE are produced only if there is a high enough concentration of free fatty acids (**Figure 6**). This keeps the two components (the fatty acid and the ethanol) in balance to reduce the metabolic burden on the cell. Thus, the DSRS system resulted in a threefold increase in product yield with a 1.5 g/liter FAEE titer (132). This system, and others like it, may foreshadow more complex regulation in pathways and yield the first applications of synthetic biological circuits in metabolic engineering.

Borrowing from extensive research in synthetic biology, synthetic circuits may also play an important role in enabling system-level control in metabolic engineering. Few practical applications have emerged to date, because most research focuses on replicating functions performed by electronic circuits within biological systems. Examples illustrating the types of complex behavior possible with synthetic circuits include a bistable toggle switch that requires only transient rather than continuous induction (133), a robust biological oscillator (134, 135), and complex circuits capable of performing logical functions (combinations of AND, OR, and NOR gates) (136). One can envision that these types of circuits could be integrated with cellular metabolites, media components, or heterologous pathway intermediates, yielding complex cellular signaling and expression phenotypes. These genetic circuits and multiple sensing transcription factors have the potential to create a complex platform capable of balancing the needs of the cell with the desire for high productivity, which would generate a strain that is both high performing and robust.

Approaches to Nonnatural Product Synthesis

An important early step in metabolic engineering involves extending the paradigm of overproduction of a native compound in a host (such as lysine in *C. glutamicum*) to overproduction of a natural compound from another organism using a heterologous pathway (ethanol in *E. coli*). The next major leap in the spectrum of compounds that can be produced biologically involves creation of nonnatural compounds that are not known to be produced naturally in any organism.

Natural products will likely remain a staple of metabolic engineering owing to the far-reaching uses of natural compounds, ranging from potential fuels to treatments for disease [artemisinin

Table 3 List of notable nonnatural compounds generated through metabolic engineering

Compound	Reference
P-Hydroxystyrene	171
Phenol	172
P-Hydroxybenzoate	173
Styrene	174
Polylactic acid	175
1,4-Butanediol	11
Long-chain alcohols	176
L-Homoalanine	177
1,3-Propanediol	50

(58) and taxol (59)]. However, biological systems also have several advantages for producing nonnatural chemicals, including the ability to produce molecules with complex stereochemistry and regiospecificity, reach molecules that would otherwise require complex and expensive synthesis, and reduce the use of potentially harmful or toxic chemicals. These attributes make biologically based production a promising platform for the manufacture of industrial and specialty chemicals. Recently, metabolic engineering has made strides toward the production of these nonnative chemicals, where in many cases no natural enzymes have been found for use. [For more detail, see the recent review by Lee et al. (137)]. By improving the toolset used to reach these new compounds, it will be possible to produce nearly any compound imaginable through metabolic engineering. **Table 3** lists several notable nonnatural chemicals recently produced through metabolic engineering.

As discussed above, modification of existing enzymes through rational design or directed mutagenesis has been used to create new enzyme functionalities or improve catalytic activity. These mutational studies, generating new functionalities and pathways, have proven to be essential to create nonnatural compounds, such as PLA. However, these minor mutational changes to existing proteins may not be enough in themselves to reach a broad range of desired compounds. Although a more daunting task, in some circumstances it may be necessary to generate new pathways or enzymes de novo to reach a desired chemical.

De Novo Construction of Pathways

Production of the common industrial monomer styrene represents one example of a simple de novo pathway. The ability to biologically produce new compounds, such as styrene, allows for green production of many products that are typically derived from nonrenewable fossil feedstocks. Although small amounts of styrene are known to accumulate naturally, no natural pathway or enzyme has yet been been identified, and an entirely novel pathway had to be generated for production. This nonnatural pathway was created by the introduction of two genes encoding phenylalanine ammonia lyase from *A. thaliana*, which converts the natural amino acid L-phenylalanine to trans-cinnamate (tCA), and phenylacrylate decarboxylase from *S. cerevisiae*, which converts tCA to styrene. Although it was unclear at the time whether the latter enzyme would successfully catalyze the second reaction in the pathway, it turned out that the enzyme functioned properly, and construction of this novel pathway resulted in a titer of 260 mg/liter in *E. coli* (138). This is a simple example because the pathway could be hypothesized without computational methods, and the enzymes for the pathways existed but were used for reactions not believed to be their natural substrate.

In other cases, no structural analogs are available for mutation (such as in the case of PLA), and a pathway to a given compound is not immediately obvious (such as in the case of styrene). Under these circumstances, the computational methods for pathway enumeration described in the section on Enumerating and Assessing Pathways for Converting a Specific Substrate to a Target Product, above, become increasingly important in facilitating metabolic engineers to reach novel compounds. The de novo pathway generation approach is exemplified by the engineering of an *E. coli* strain capable of producing 1,4-butanediol. In this case, the pathway was predicted by examining transformation of functional groups rather than known enzymatic reactions, and notably, the algorithm permitted reactions that were not performed by any known enzyme to be included in the search space, which allowed for enzyme creation if necessary. The initial search yielded more than 10,000 pathways, which were then ranked and eliminated on criteria such as thermodynamic feasibility, maximum theoretical yield, pathway length, number of nonnative steps, and number of steps for which no enzyme is known. An optimal pathway was then selected and expressed, leading to low-level production of the novel compound by using five heterologous enzymatic steps from α-ketoglutarate. From here, traditional metabolic engineering approaches were applied to improve yield, including host organism metabolism optimization using OptKnock, codon optimization, use of nonnative enzymes known to function anaerobically in place of native enzymes, and reduction of NADH sensitivity of enzymes. Overall, this multipronged approach led to production increases of more than three orders of magnitude to approximately 20 g/liter of 1,4-butanediol (11). This approach shows how a nonnatural chemical, even one that requires a large number of enzymatic steps, can be produced in high yield by using several tools from the metabolic engineering toolbox.

De Novo Generation of Enzymes

As metabolic engineering continues to reach into new chemical spaces to generate novel compounds, it may be found that simply no enzyme exists to perform some of the necessary reactions. In the past, this has been overcome by performing mutagenesis or directed evolution on existing enzymes to impart new or improve activity or selectivity (discussed above). However, there may be no precedent for a given reaction, and it may be necessary to generate an entirely new enzyme. Though this area is in its infancy, it is driven forward by the desire for a more diverse range of products for chemical and pharmaceutical use, and it becomes increasingly feasible as computational power and speed continue to increase. This area likely will grow in importance and relevance to the main body of metabolic engineering, further broadening the types of fundamental chemical transformations possible.

To date, the most common method for de novo enzyme generation derives from the Rosetta algorithm. This algorithm involves first generating a set of idealized active site residues, dubbed a theozyme, and conformations by using saddle points in transition-state energy maps of the intermediates (139). Next, compatible scaffolds from a library of existing protein structures are identified and select residues are rechosen to better enable the reaction, and the model is reoptimized. This is exactly the process used in the Rosetta3 software to generate enzymes (140), and the methodology has remarkably been applied to generate several entirely novel proteins, albeit with low activity, capable of catalyzing Diels-Alder reactions (141), retro-aldol reactions (142), and Kemp elimination reactions (143).

Although these enzymes have the benefit of performing novel catalytic functions, they also exhibit low activity. The activity of these enzymes may be improved significantly by using traditional tools of metabolic engineering, such as directed evolution and mutagenesis. As an example of this, a kemp-elimination enzyme, designed de novo computationally, was subjected

to a directed evolution experiment, which yielded an enzyme with a greater than 200-fold improvement in catalytic efficiency (K_{cat}/K_m) with a final value of 2,600 liters/mol/s (144). This improvement illustrates that traditional metabolic engineering approaches, such as directed evolution, can improve enzyme functionality; however, improvements in the computationally designed enzyme are necessary before industrial applications become viable.

Although the proteins described above have a de novo designed active site, the protein backbones are chosen from existing templates and are highly modified to perform the appropriate catalytic function. A suitable backbone template is essential to obtaining functional enzymes, and the ability to correctly predict protein structure and folds has a strong influence on the ability to predict and design catalytic activity. A backbone may not be found, or found backbones may fit poorly. Although more computationally intense, engineering of protein tertiary structure in addition to the active site may improve the ability to achieve new catalytic functions. A recent study, which illustrated the emerging capability to design protein backbones computationally, generated a small, 93-residue protein, named Top7. Upon expression in vivo, determination of the X-ray crystal structure revealed that the predicted structure was very similar to that observed (145). Even more remarkably, di-iron protein engineered from first principles was shown to exhibit correct cofactor incorporation and phenol oxidase activity (146). This further demonstrates that engineering proteins without referencing known structures is feasible, suggesting the possibility of creating protein backbones and folds not found in nature, which improves the likelihood that effective catalysts could be found for any desired reaction. Whereas in the past metabolic engineering has been limited to overproducing natural and nonnatural compounds with close structural analogs, in the future, synthesis of any conceivable molecule, with a high degree of activity, likely will be possible.

CLOSING REMARKS

Clearly, much has changed since the field of metabolic engineering was defined approximately 20 years ago. What started out as the manipulation of a few enzymes in well-defined pathways has morphed into a field where almost any target organic compound can be produced in a microbial host. This transformation has been facilitated by a combination of well-developed theory on metabolic fluxes and network control and novel tools from molecular biology, bioinformatics, systems biology, and synthetic biology that can be applied in a rational or combinatorial manner. The driving force behind all these developments has been concern about sustainability and the associated increasing interest in the production of products from renewable resources, namely, sugars derived directly from sugarcane or corn but also from cellulosic biomass, and even gases at some point in the near future. Combined with these market forces, technology advances are providing the enabling platform required to realize the vision of a biobased economy. Still, when one looks at the metabolic engineering literature, it is hard not to see a series of elegant but disparate examples, with little in the way of a guiding instruction list. What limits us in this regard is still our understanding of fundamental biology. As this understanding grows, through both applied and fundamental research, we think metabolic engineering will evolve to eliminate the requirement for some of the guesswork that is currently necessary, to further transform the field into the truly robust analytical framework it was envisioned to be at its conception.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

1. Cohen SN, Chang ACY, Boyer HW, Helling RB. 1973. Construction of biologically functional bacterial plasmids *in vitro*. *Proc. Natl. Acad. Sci. USA* 70(11):3240–44
2. Stephanopoulos G, Vallino J. 1991. Network rigidity and metabolic engineering in metabolite overproduction. *Science* 252(5013):1675–81
3. Bailey JE. 1991. Toward a science of metabolic engineering. *Science* 252(5013):1668–75
4. Stephanopoulos GN, Aristidou AA, Nielsen J. 1998. *Metabolic Engineering: Principles and Methodologies*. San Diego, CA: Acad. Press
5. Stephanopoulos G, Sinskey AJ. 1993. Metabolic engineering—methodologies and future prospects. *Trends Biotechnol.* 11(9):392–96
6. Mavrovouniotis ML, Stephanopoulos G. 1990. Computer-aided synthesis of biochemical pathways. *Biotechnol. Bioeng.* 36(11):1119–32
7. Hatzimanikatis V, Li C, Ionita Ja, Henry CS, Jankowski MD, Broadbelt LJ. 2005. Exploring the diversity of complex metabolic networks. *Bioinformatics* 21(8):1603–9
8. Mavrovouniotis ML. 1990. Group contributions for estimating standard Gibbs energies of formation of biochemical compounds in aqueous solution. *Biotechnol. Bioeng.* 36(10):1070–82
9. Mavrovouniotis ML. 1991. Estimation of standard Gibbs energy changes of biotransformations. *J. Biol. Chem.* 266(22):14440–45
10. Henry CS, Broadbelt LJ, Hatzimanikatis V. 2010. Discovery and analysis of novel metabolic pathways for the biosynthesis of industrial chemicals: 3-hydroxypropanoate. *Biotechnol. Bioeng.* 106(3):462–73
11. Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, et al. 2011. Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat. Chem. Biol.* 7(7):445–52
12. Stephanopoulos G, Aristidou A, Nielsen J. 1998. *Metabolic Engineering: Principles and Practices*. San Diego, CA: Acad. Press
13. Antoniewicz MR, Kraynie DF, Laffend La, González-Lergier J, Kelleher JK, Stephanopoulos G. 2007. Metabolic flux analysis in a nonstationary system: fed-batch fermentation of a high yielding strain of *E. coli* producing 1,3-propanediol. *Metab. Eng.* 9(3):277–92
14. Stephanopoulos G. 1999. Metabolic fluxes and metabolic engineering. *Metab. Eng.* 1(1):1–11
15. Kim HU, Kim TY, Lee SY. 2008. Metabolic flux analysis and metabolic engineering of microorganisms. *Mol. BioSyst.* 4(2):113–20
16. Sauer U. 2006. Metabolic networks in motion: ^{13}C-based flux analysis. *Mol. Syst. Biol.* 2:62
17. Nielsen J. 2003. Guest commentary: It is all about metabolic fluxes. *J. Bacteriol.* 185(24):7031–35
18. Wiechert W. 2001. ^{13}C metabolic flux analysis. *Metab. Eng.* 3(3):195–206
19. Vallino J, Stephanopoulos G. 1994. Carbon flux distributions at the pyruvate branch point in *Corynebacterium glutamicum* during lysine overproduction. *Biotechnol. Prog.* 10(3):320–26
20. Stephanopoulos G. 1994. Metabolic engineering. *Curr. Opin. Biotechnol.* 5:196–200
21. Nissen TL, Schulze U, Nielsen J, Villadsen J. 1997. Flux distributions in anaerobic, glucose-limited continuous cultures of *Saccharomyces cerevisiae*. *Microbiology* 143(Pt. 1):203–18
22. Vallino JJ, Stephanopoulos G. 1993. Metabolic flux distributions in *Corynebacterium glutamicum* during growth and lysine overproduction. *Biotechnol. Bioeng.* 41(6):633–46
23. Kacser H, Burns JA. 1973. The control of flux. In *Rate Control of Biological Processes*, ed. Soc. Exp. Biol., pp. 65–104. London: Cambridge Univ. Press
24. Properties G, Strength E, Rapoport TA. 1974. A linear steady-state treatment of enzymatic chains. *FEBS J.* 95:89–95
25. Niederberger P, Prasad R, Miozzari G, Kacser H. 1992. A strategy for increasing an in vivo flux by genetic manipulations. The tryptophan system of yeast. *Biochem. J.* 287:473–79
26. Fell D. 1998. Increasing the flux in metabolic pathways: a metabolic control analysis perspective. *Biotechnol. Bioeng.* 58(2–3):121–24
27. Moreno-Sánchez R, Saavedra E, Rodríguez-Enríquez S, Olín-Sandoval V. 2008. Metabolic control analysis: a tool for designing strategies to manipulate metabolic pathways. *J. Biomed. Biotechnol.* 2008:597913
28. Schmidt K, Nielsen J, Villadsen J. 1999. Quantitative analysis of metabolic fluxes in *Escherichia coli*, using two-dimensional NMR spectroscopy and complete isotopomer models. *J. Biotechnol.* 71(1–3):175–89

29. Antoniewicz MR, Kelleher JK, Stephanopoulos G. 2007. Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metab. Eng.* 9(1):68–86
30. Raman K, Chandra N. 2009. Flux balance analysis of biological systems: applications and challenges. *Brief. Bioinforma.* 10(4):435–49
31. Lee SY, Park JM, Kim TY. 2011. Application of metabolic flux analysis in metabolic engineering. *Methods Enzymol.* 498:67–93
32. Schuetz R, Kuepfer L, Sauer U. 2007. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Mol. Syst. Biol.* 3:119
33. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, et al. 2007. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* 3(121):121
34. Alper H, Jin Y-S, Moxley JF, Stephanopoulos G. 2005. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metab. Eng.* 7(3):155–64
35. Burgard AP, Pharkya P, Maranas CD. 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* 84(6):647–57
36. Förster J, Famili I, Fu P, Palsson BØ, Nielsen J. 2003. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* 13(2):244–53
37. Oh Y-K, Palsson BO, Park SM, Schilling CH, Mahadevan R. 2007. Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *J. Biol. Chem.* 282(39):28791–99
38. Senger RS, Papoutsakis ET. 2008. Genome-scale model for *Clostridium acetobutylicum*: part I. Metabolic network resolution and analysis. *Biotechnol. Bioeng.* 101(5):1036–52
39. Kjeldsen KR, Nielsen J. 2009. In silico genome-scale reconstruction and validation of the *Corynebacterium glutamicum* metabolic network. *Biotechnol. Bioeng.* 102(2):583–97
40. de Oliveira Dal'Molin CG, Quek L-E, Palfreyman RW, Brumbley SM, Nielsen LK. 2010. AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis. *Plant Physiol.* 152(2):579–89
41. Duarte NC, Becker Sa, Jamshidi N, Thiele I, Mo ML, et al. 2007. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. USA* 104(6):1777–82
42. Blazeck J, Alper H. 2010. Systems metabolic engineering: genome-scale models and beyond. *Biotechnol. J.* 5(7):647–59
43. Oberhardt Ma, Palsson BØ, Papin Ja. 2009. Applications of genome-scale metabolic reconstructions. *Mol. Syst. Biol.* 5(320):320
44. Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ. 2009. Reconstruction of biochemical networks in microorganisms. *Nat. Rev. Microbiol.* 7(2):129–43
45. Park SM, Sinskey AJ, Stephanopoulos G. 1997. Metabolic and physiological studies of *Corynebacterium glutamicum* mutants. *Biotechnol. Bioeng.* 55(6):864–79
46. Park SM, Shaw-Reid C, Sinskey AJ, Stephanopoulos G. 1997. Elucidation of anaplerotic pathways in *Corynebacterium glutamicum* via ^{13}C-NMR spectroscopy and GC-MS. *Appl. Microbiol. Biotechnol.* 47(4):430–40
47. Petersen S, de Graaf AA, Eggeling L, Möllney M, Wiechert W, Sahm H. 2000. *In vivo* quantification of parallel and bidirectional fluxes in the anaplerosis of *Corynebacterium glutamicum*. *J. Biol. Chem.* 275(46):35932–41
48. Peters-Wendisch PG, Kreutzer C, Kalinowski J, Pátek M, Sahm H, Eikmanns BJ. 1998. Pyruvate carboxylase from *Corynebacterium glutamicum*: characterization, expression and inactivation of the *pyc* gene. *Microbiology* 144:915–27
49. Koffas Ma, Jung GY, Stephanopoulos G. 2003. Engineering metabolism and product formation in *Corynebacterium glutamicum* by coordinated gene overexpression. *Metab. Eng.* 5(1):32–41
50. Nakamura CE, Whited GM. 2003. Metabolic engineering for the microbial production of 1,3-propanediol. *Curr. Opin. Biotechnol.* 14(5):454–59
51. Bro C, Regenberg B, Förster J, Nielsen J. 2006. In silico aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production. *Metab. Eng.* 8(2):102–11

52. Kötter P, Ciriacy M. 1993. Xylose fermentation by *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* 38(6):776–83
53. Kuyper M, Harhangi HR, Stave AK, Winkler AA, Jetten MSM, et al. 2003. High-level functional expression of a fungal xylose isomerase: the key to efficient ethanolic fermentation of xylose by *Saccharomyces cerevisiae*? *FEMS Yeast Res.* 4(1):69–78
54. van Maris AJA, Winkler AA, Kuyper M, de Laat WTAM, van Dijken JP, et al. 2007. Development of efficient xylose fermentation in *Saccharomyces cerevisiae*: xylose isomerase as a key component. *Adv. Bioeng. Biotechnol.* 108:179–204
55. Ha S, Galazka JM, Kim SR, Choi J, Yang X, et al. 2011. Engineered *Saccharomyces cerevisiae* capable of simultaneous cellobiose and xylose fermentation. *Proc. Natl. Acad. Sci. USA* 108(2):504–9
56. Ohta K, Beall DS, Mejia JP, Shanmugam KT, Ingram LO. 1991. Genetic improvement of *Escherichia coli* for ethanol production: chromosomal integration of *Zymomonas mobilis* genes encoding pyruvate decarboxylase and alcohol dehydrogenase II. *Appl. Environ. Microbiol.* 57:893–900
57. Lindberg P, Park S, Melis A. 2010. Engineering a platform for photosynthetic isoprene production in cyanobacteria, using *Synechocystis* as the model organism. *Metab. Eng.* 12(1):70–79
58. Ro D-K, Paradise EM, Ouellet M, Fisher KJ, Newman KL, et al. 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 440(7086):940–43
59. Ajikumar PK, Xiao W-H, Tyo KEJ, Wang Y, Simeon F, et al. 2010. Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli*. *Science* 330(6000):70–74
60. Tokuhiro K, Muramatsu M, Ohto C, Kawaguchi T, Obata S, et al. 2009. Overproduction of geranylgeraniol by metabolically engineered *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 75(17):5536–43
61. Farmer WR, Liao JC. 2001. Precursor balancing for metabolic engineering of lycopene production in *Escherichia coli*.*Biotechnol. Prog.* 17(1):57–61
62. Ajikumar PK, Xiao W-H, Tyo KEJ, Wang Y, Simeon F, et al. 2010. Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* 330:70–74
63. Atsumi S, Wu T-Y, Eckl E-M, Hawkins SD, Buelter T, Liao JC. 2010. Engineering the isobutanol biosynthetic pathway in *Escherichia coli* by comparison of three aldehyde reductase/alcohol dehydrogenase genes. *Appl. Microbiol. Biotechnol.* 85(3):651–57
64. Anthony LC, Huang LL, Ye RW. 2010. Production of isobutanol in yeast mitochondria. *US Patent No. 2010/0129886 A1*
65. Feldman RMR, Gunawardena U, Urano J, Meinhold P, Aristidou A, et al. 2012. Yeast organisms producing isobutanol and a high yield. *US Patent No. 2011/0020889*
66. Festel G, Boles E, Weber C, Brat D. 2011. Fermentative production of isobutanol with yeast. *US Patent No. 2011/0053235 A1*
67. Chen X, Nielsen KF, Borodina I, Kielland-Brandt MC, Karhumaa K. 2011. Increased isobutanol production in *Saccharomyces cerevisiae* by overexpression of genes in valine metabolism. *Biotechnol. Biofuels* 4(1):21
68. Schirmer A, Rude Ma, Li X, Popova E, Del Cardayre SB. 2010. Microbial biosynthesis of alkanes. *Science* 329(5991):559–62
69. Leonard E, Yan Y, Fowler ZL, Li Z, Lim C-G, et al. 2008. Strain improvement of recombinant *Escherichia coli* for efficient production of plant flavonoids. *Mol. Pharm.* 5(2):257–65
70. Santos CNS, Koffas M, Stephanopoulos G. 2011. Optimization of a heterologous pathway for the production of flavonoids from glucose. *Metab. Eng.* 13(4):392–400
71. Hawkins KM, Smolke CD. 2008. Production of benzylisoquinoline alkaloids in *Saccharomyces cerevisiae*. *Nat. Chem. Biol.* 4(9):564–73
72. Leonard E, Ajikumar PK, Thayer K, Xiao W-H, Mo JD, et al. 2010. Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. *Proc. Natl. Acad. Sci. USA* 107(31):13654–59
73. Whited GM, Feher FJ, Benko DA, Cervin MA, Chotani GK, et al. 2010. Development of a gas-phase bioprocess for isoprene-monomer production using metabolic pathway engineering. *Ind. Biotechnol.* 6(3):152–63
74. de Jong E, Higson A, Walsh P, Wellisch M. 2012. Product developments in the bio-based chemicals arena. *Biofuels Bioprod. Biorefining* 6(6):606–24

75. Zhang X, Jantama K, Moore JC, Jarboe LR, Shanmugam KT, Ingram LO. 2009. Metabolic evolution of energy-conserving pathways for succinate production in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 106(48):20180–85
76. Burgard AP, Van Dien SJ, Burk M. 2011. Methods and organisms for the growth-coupled production of 1-4-butanediol. *US Patent No. 2011/0201071 A1*
77. Moon TS, Dueber JE, Shiue E, Prather KLJ. 2010. Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered *E. coli*. *Metab. Eng.* 12(3):298–305
78. Wang P, Kim W, Pickens LB, Gao X, Tang Y. 2012. Heterologous expression and manipulation of three tetracycline biosynthetic pathways. *Angew. Chem.* 51(44):11136–40
79. Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* 97(12):6640–45
80. Gonzalez JB, Fernandez FJ, Tomasini A. 2003. Microbial secondary metabolites production and strain improvement. *Indian J. Biotechnol.* 2:322–33
81. Bailey JE, Sburlati A, Hatzimanikatis V, Lee K, Renner WA, Tsai PS. 2002. Inverse metabolic engineering: a strategy for directed genetic engineering of useful phenotypes. *Biotechnol. Bioeng.* 79(5):568–79
82. Jin Y-S, Alper H, Yang Y-T, Stephanopoulos G. 2005. Improvement of xylose uptake and ethanol production in recombinant *Saccharomyces cerevisiae* through an inverse metabolic engineering approach *Appl. Environ. Microbiol.* 71(12):8249–56
83. Ghosh C, Gupta R, Mukherjee KJ. 2012. An inverse metabolic engineering approach for the design of an improved host platform for over-expression of recombinant proteins in *Escherichia coli*. *Microb. Cell Fact.* 11(1):93
84. Gill RT. 2003. Enabling inverse metabolic engineering through genomics. *Curr. Opin. Biotechnol.* 14(5):484–90
85. Gill RT, Wildt S, Yang YT, Ziesman S, Stephanopoulos G. 2002. Genome-wide screening for trait conferring genes using DNA microarrays. *Proc. Natl. Acad. Sci. USA* 99(10):7033–38
86. Santos CNS, Stephanopoulos G. 2008. Melanin-based high-throughput screen for L-tyrosine production in *Escherichia coli*. *Appl. Environ. Microbiol.* 74(4):1190–97
87. Wang BL. 2009. *High throughput screen for cells with high extracellular metabolite consumption–secretion rates using microfluidic droplets*. PhD thesis. Mass. Inst. Technol.
88. Dietrich JA, McKee AE, Keasling JD. 2010. High-throughput metabolic engineering: advances in small-molecule screening and selection. *Annu. Rev. Biochem.* 79:563–90
89. Borden JR, Papoutsakis ET. 2007. Dynamics of genomic-library enrichment and identification of solvent tolerance genes for *Clostridium acetobutylicum*. *Appl. Environ. Microbiol.* 73(9):3061–68
90. Rubin-Pitel S, Cho CM-H, Chen W, Zhao H. 2006. Directed evolution tools in bioproduct and bioprocess development. In *Bioprocessing for Value-Added Products from Renewable Resources: New Technologies and Applications*, ed. S-T Yang, pp. 49–72. Amsterdam: Elsevier B.V.
91. Alper H, Moxley J, Nevoigt E, Fink GR, Stephanopoulos G. 2006. Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* 314:1565–68
92. Alper H, Stephanopoulos G. 2007. Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metab. Eng.* 9(3):258–67
93. Du J, Yuan Y, Si T, Lian J, Zhao H. 2012. Customized optimization of metabolic pathways by combinatorial transcriptional engineering. *Nucleic Acids Res.* 40(18):e142
94. Yang TH, Kim TW, Kang HO, Lee S-H, Lee EJ, et al. 2010. Biosynthesis of polylactic acid and its copolymers using evolved propionate CoA transferase and PHA synthase. *Biotechnol. Bioeng.* 105(1):150–60
95. Leonard E, Kumaran P, Thayer K, Xiao W, Mo JD. 2010. Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. *Proc. Natl. Acad. Sci. USA* 107(31):13654–59
96. Shaw AJ, Podkaminer KK, Desai SG, Bardsley JS, Rogers SR, et al. 2008. Metabolic engineering of a thermophilic bacterium to produce ethanol at high yield. *Proc. Natl. Acad. Sci. USA* 105(37):13769–74
97. Cripps RE, Eley K, Leak DJ, Rudd B, Taylor M, et al. 2009. Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production. *Metab. Eng.* 11(6):398–408

98. Higashide W, Li Y, Yang Y, Liao JC. 2011. Metabolic engineering of *Clostridium cellulolyticum* for production of isobutanol from cellulose. *Appl. Environ. Microbiol.* 77(8):2727–33
99. Tracy BP, Jones SW, Fast AG, Indurthi DC, Papoutsakis ET. 2012. Clostridia: the importance of their exceptional substrate and metabolite diversity for biofuel and biorefinery applications. *Curr. Opin. Biotechnol.* 23(3):364–81
100. Lan EI, Liao JC. 2011. Metabolic engineering of cyanobacteria for 1-butanol production from carbon dioxide. *Metab. Eng.* 13(4):353–63
101. Heap JT, Pennington OJ, Cartman ST, Carter GP, Minton NP. 2007. The ClosTron: a universal gene knock-out system for the genus *Clostridium*. *J. Microbiol. Methods* 70(3):452–64
102. Heap JT, Ehsaan M, Cooksley CM, Ng Y-K, Cartman ST, et al. 2012. Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic Acids Res.* 40(8):e59
103. Suzuki T, Yasui K. 2011. Plasmid artificial modification: a novel method for efficient DNA transfer into bacteria. *Methods Mol. Biol.* 765:309–26
104. Yasui K, Kano Y, Tanaka K. 2009. Improvement of bacterial transformation efficiency using plasmid artificial modification. *Nucleic Acids Res.* 37(1):e3
105. Rocha EPC, Cornet E, Michel B. 2005. Comparative and evolutionary analysis of the bacterial homologous recombination systems. *PLoS Genet.* 1(2):e15
106. Mai V, Lorenz WW, Wiegel J. 2006. Transformation of *Thermoanaerobacterium* sp. strain JW/SL-YS485 with plasmid pIKM1 conferring kanamycin resistance. *FEMS Microbiol. Lett.* 148(2):163–67
107. Martin VJJ, Pitera DJ, Withers ST, Newman JD, Keasling JD. 2003. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* 21(7):796–802
108. Li J, Nayak S, Mrksich M. 2010. Rate enhancement of an interfacial biochemical reaction through localization of substrate and enzyme by an adaptor domain. *J. Phys. Chem. B* 114(46):15113–18
109. Wriggers W, Chakravarty S, Jennings PA. 2005. Control of protein functional dynamics by peptide linkers. *Biopolymers* 80(6):736–46
110. Sakaki T, Shibata M. 1990. Expression of bovine cytochrome P450c21 and its fused enzymes with yeast NADPH-cytochrome P450 reductase in *Saccharomyces cerevisiae*. *DNA Cell Biol.* 9(8):603–14
111. Zhang Y, Li S-Z, Li J, Pan X, Cahoon RE, et al. 2006. Using unnatural protein fusions to engineer resveratrol biosynthesis in yeast and mammalian cells. *J. Am. Chem. Soc.* 128(40):13030–31
112. Wang Y, Yi H, Wang M, Yu O, Jez JM. 2011. Structural and kinetic analysis of the unnatural fusion protein 4-coumaroyl-coa ligase::stilbene synthase. *J. Am. Chem. Soc.* 133(51):20684–87
113. Wang C, Yoon S-H, Jang H-J, Chung Y-R, Kim J-Y, et al. 2011. Metabolic engineering of *Escherichia coli* for α-farnesene production. *Metab. Eng.* 13(6):648–55
114. Delebecque CJ, Lindner AB, Silver Pa, Aldaye FA. 2011. Organization of intracellular reactions with rationally designed RNA assemblies. *Science* 333(6041):470–74
115. Collins CH, Yokobayashi Y, Umeno D, Arnold FH. 2003. Engineering proteins that bind, move, make and break DNA. *Curr. Opin. Biotechnol.* 14(4):371–78
116. Dueber JE, Wu GC, Malmirchegini GR, Moon TS, Petzold CJ, et al. 2009. Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* 27(8):753–59
117. Moon TS, Dueber JE, Shiue E, Prather KLJ. 2010. Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered *E. coli*. *Metab. Eng.* 12(3):298–305
118. Hirakawa H, Nagamune T. 2010. Molecular assembly of P450 with ferredoxin and ferredoxin reductase by fusion to PCNA. *Chembiochem* 11(11):1517–20
119. Park S-H, Zarrinpar A, Lim WA. 2003. Rewiring MAP kinase pathways using alternative scaffold assembly mechanisms. *Science* 299(5609):1061–64
120. Good MC, Zalatan JG, Lim WA. 2011. Scaffold proteins: hubs for controlling the flow of cellular information. *Science* 332(6030):680–86
121. Cheng S, Liu Y, Crowley CS, Yeates TO, Bobik TA. 2008. Bacterial microcompartments: their properties and paradoxes. *BioEssays* 30(11–12):1084–95
122. Yeates TO, Crowley CS, Tanaka S. 2010. Bacterial microcompartment organelles: protein shell structure and evolution. *Annu. Rev. Biophys.* 39:185–205
123. Bonacci W, Teng P. 2012. Modularity of a carbon-fixing protein organelle. *Proc. Natl. Acad. Sci. USA* 109(2):478–83

124. Parsons JB, Frank S, Bhella D, Liang M, Prentice MB, et al. 2010. Synthesis of empty bacterial microcompartments, directed organelle protein incorporation, and evidence of filament-associated organelle movement. *Mol. Cell* 38(2):305–15
125. Fan C, Cheng S, Liu Y, Escobar CM, Crowley CS, et al. 2010. Short N-terminal sequences package proteins into bacterial microcompartments. *Proc. Natl. Acad. Sci. USA* 107(16):7509–14
126. Lee H, DeLoache WC, Dueber JE. 2012. Spatial organization of enzymes for metabolic engineering. *Metab. Eng.* 14(3):242–51
127. Feldman R, Sigman D. 1983. The synthesis of ATP by the membrane-bound ATP synthase complex from medium ^{32}Pi under completely uncoupled conditions. *J. Biol. Chem.* 258(20):12178–83
128. Roggenkamp R. 1992. Targeting signals for protein import into peroxisomes. *Cell Biochem. Funct.* 10(3):193–99
129. Bayer TS, Widmaier DM, Temme K, Mirsky EA, Santi DV, Voigt CA. 2009. Synthesis of methyl halides from biomass using engineered microbes. *J. Am. Chem. Soc.* 131(18):6508–15
130. Farhi M, Marhevka E, Masci T, Marcos E, Eyal Y, et al. 2011. Harnessing yeast subcellular compartments for the production of plant terpenoids. *Metab. Eng.* 13(5):474–81
131. Agapakis CM, Niederholtmeyer H, Noche RR, Lieberman TD, Megason SG, et al. 2011. Towards a synthetic chloroplast. *PLoS ONE* 6(4):e18877
132. Zhang F, Carothers JM, Keasling JD. 2012. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat. Biotechnol.* 30(4):354–59
133. Gardner TS, Cantor CR, Collins JJ. 2000. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403(6767):339–42
134. Elowitz MB, Leibler S. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature* 403(6767):335–38
135. Stricker J, Cookson S, Bennett MR, Mather WH, Tsimring LS, Hasty J. 2008. A fast, robust and tunable synthetic gene oscillator. *Nature* 456(7221):516–19
136. Tamsir A, Tabor JJ, Voigt CA. 2011. Robust multicellular computing using genetically encoded NOR gates and chemical "wires". *Nature* 469(7329):212–15
137. Lee JW, Na D, Park JM, Lee J, Choi S, Lee SY. 2012. Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nat. Chem. Biol.* 8(6):536–46
138. McKenna R, Nielsen DR. 2011. Styrene biosynthesis from glucose by engineered *E. coli*. *Metab. Eng.* 13(5):544–54
139. Zhang X, DeChancie J, Gunaydin H, Chowdry AB, Clemente FR, et al. 2008. Quantum mechanical design of enzyme active sites. *J. Org. Chem.* 73(3):889–99
140. Richter F, Leaver-Fay A, Khare SD, Bjelic S, Baker D. 2011. De novo enzyme design using Rosetta3. *PLoS One* 6(5):e19230
141. Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, et al. 2010. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* 329(5989):309–13
142. Jiang L, Althoff EA, Clemente FR, Doyle L, Röthlisberger D, et al. 2008. De novo computational design of retro-aldol enzymes. *Science* 319(5868):1387–91
143. Röthlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, et al. 2008. Kemp elimination catalysts by computational enzyme design. *Nature* 453(7192):190–95
144. Khersonsky O, Röthlisberger D, Dym O, Albeck S, Jackson CJ, et al. 2010. Evolutionary optimization of computationally designed enzymes: kemp eliminases of the KE07 series. *J. Mol. Biol.* 396(4):1025–42
145. Schaeffer RD, Daggett V. 2011. Protein folds and protein folding. *Protein Eng. Des. Sel.* 24(1–2):11–19
146. Faiella M, Andreozzi C, De Rosales RTM, Pavone V, Maglio O, et al. 2009. An artificial di-iron oxoprotein with phenol oxidase activity. *Nat. Chem. Biol.* 5(12):882–84
147. Nielsen J. 1998. Metabolic engineering: techniques for analysis of targets for genetic manipulations. *Biotechnol. Bioeng.* 58(2–3):125–32
148. Fell D. 1992. Metabolic control analysis: a survey of its theoretical and experimental development. *Biochem. J.* 330:313–30
149. Liao JC, Delgado J. 2008. Advances in metabolic control analysis. *Biotechnol. Prog.* 9(3):221–33
150. Kauffman KJ, Prakash P, Edwards JS. 2003. Advances in flux balance analysis. *Curr. Opin. Biotechnol.* 14(5):491–96

151. Price N, Reed J, Palsson BØ. 2004. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.* 2(11):886–97
152. Feist AM, Palsson BØ. 2008. The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat. Biotechnol.* 26(6):659–67
153. Patil KR, Åkesson M, Nielsen J. 2004. Use of genome-scale microbial models for metabolic engineering. *Curr. Opin. Biotechnol.* 15(1):64–69
154. Lanza A, Alper H. 2012. Using transcription machinery engineering to elicit complex cellular phenotypes. *Methods Mol. Biol.* 813:229–48
155. Agapakis CM, Boyle PM, Silver PA. 2012. Natural strategies for the spatial optimization of metabolism in synthetic biology. *Nat. Chem. Biol.* 8(6):527–35
156. Yousofshahi M, Lee K, Hassoun S. 2011. Probabilistic pathway construction. *Metab. Eng.* 13(4):435–44
157. Medema MH, van Raaphorst R, Takano E, Breitling R. 2012. Computational tools for the synthetic design of biochemical pathways. *Nat. Rev. Microbiol.* 10(3):191–202
158. Nandagopal N, Elowitz MB. 2011. Synthetic biology: integrated gene circuits. *Science* 333(6047):1244–48
159. Yoshida K, Ueda S, Maeda I. 2009. Carotenoid production in *Bacillus subtilis* achieved by metabolic engineering. *Biotechnol. Lett.* 31(11):1789–93
160. Shi S, Chen T, Zhang Z, Chen X, Zhao X. 2009. Transcriptome analysis guided metabolic engineering of *Bacillus subtilis* for riboflavin production. *Metab. Eng.* 11(4–5):243–52
161. Romero S, Merino E, Bolívar F, Gosset G, Martinez A. 2007. Metabolic engineering of Bacillus subtilis for ethanol production: lactate dehydrogenase plays a key role in fermentative metabolism. *Appl. Environ. Microbiol.* 73(16):5190–98
162. Wang Q, Chen T, Zhao X, Chamu J. 2012. Metabolic engineering of thermophilic *Bacillus licheniformis* for chiral pure D-2,3-butanediol production. *Biotechnol. Bioeng.* 109(7):1610–21
163. Lütke-Eversloh T, Bahl H. 2011. Metabolic engineering of *Clostridium acetobutylicum*: recent advances to improve butanol production. *Curr. Opin. Biotechnol.* 22(5):634–47
164. Yu M, Zhang Y, Tang I-C, Yang S-T. 2011. Metabolic engineering of *Clostridium tyrobutyricum* for *n*-butanol production. *Metab. Eng.* 13(4):373–82
165. Taylor MP, Eley KL, Martin S, Tuffin MI, Burton SG, Cowan DA. 2009. Thermophilic ethanologenesis: future prospects for second-generation bioethanol production. *Trends Biotechnol.* 27(7):398–405
166. Frock AD, Kelly RM. 2012. Extreme thermophiles: moving beyond single-enzyme biocatalysis. *Curr. Opin. Chem. Eng.* 1(4):363–72
167. Sharpe PL, DiCosimo D, Bosak MD, Knoke K, Tao L, et al. 2007. Use of transposon promoter-probe vectors in the metabolic engineering of the obligate methanotroph *Methylomonas* sp. strain 16a for enhanced C_{40} carotenoid synthesis. *Appl. Environ. Microbiol.* 73(6):1721–28
168. Jakobsen ØM, Brautaset T, Degnes KF, Heggeset TMB, Balzer S, et al. 2009. Overexpression of wild-type aspartokinase increases L-lysine production in the thermotolerant methylotrophic bacterium *Bacillus methanolicus*. *Appl. Environ. Microbiol.* 75(3):652–61
169. Lü J, Sheahan C, Fu P. 2011. Metabolic engineering of algae for fourth generation biofuels production. *Energy Environ. Sci.* 4(7):2451
170. Wu S, Chappell J. 2008. Metabolic engineering of natural products in plants; tools of the trade and challenges for the future. *Curr. Opin. Biotechnol.* 19(2):145–52
171. Verhoef S, Wierckx N, Westerhof RGM, De Winde JH, Ruijssenaars HJ. 2009. Bioproduction of *p*-hydroxystyrene from glucose by the solvent-tolerant bacterium *Pseudomonas putida* S12 in a two-phase water-decanol fermentation. *Appl. Environ. Microbiol.* 75(4):931–36
172. Wierckx NJP, Ballerstedt H, de Bont JAM, Wery J. 2005. Engineering of solvent-tolerant *Pseudomonas putida* S12 for bioproduction of phenol from glucose. *Appl. Environ. Microbiol.* 71(12):8221–27
173. Verhoef S, Ruijssenaars HJ, De Bont Ja M, Wery J. 2007. Bioproduction of *p*-hydroxybenzoate from renewable feedstock by solvent-tolerant *Pseudomonas putida* S12. *J. Biotechnol.* 132(1):49–56
174. McKenna R, Nielsen DR. 2011. Styrene biosynthesis from glucose by engineered *E. coli*. *Metab. Eng.* 13(5):544–54
175. Jung YK, Kim TY, Park SJ, Lee SY. 2010. Metabolic engineering of *Escherichia coli* for the production of polylactic acid and its copolymers. *Biotechnol. Bioeng.* 105(1):161–71

176. Zhang K, Sawaya MR, Eisenberg DS, Liao JC. 2008. Expanding metabolism for biosynthesis of nonnatural alcohols. *Proc. Natl. Acad. Sci. USA* 105(52):20653–58
177. Zhang K, Li H, Cho KM, Liao JC. 2010. Expanding metabolism for total biosynthesis of the nonnatural amino acid L-homoalanine. *Proc. Natl. Acad. Sci. USA* 107(14):6234–39
178. Chen X, Alonso AP, Allen DK, Reed JL, Shachar-Hill Y. 2011. Synergy between ^{13}C-metabolic flux analysis and flux balance analysis for understanding metabolic adaptation to anaerobiosis in *E. coli*. *Metab. Eng.* 13(1):38–48