nature biotechnology

PERSPECTIVE

Systems strategies for developing industrial microbial strains

Sang Yup Lee[1–4] & Hyun Uk Kim[1,2,4]

Industrial strain development requires system-wide engineering and optimization of cellular metabolism while considering industrially relevant fermentation and recovery processes. It can be conceptualized as several strategies, which may be implemented in an iterative fashion and in different orders. The key challenges have been the time-, cost- and labor-intensive processes of strain development owing to the difficulties in understanding complex interactions among the metabolic, gene regulatory and signaling networks at the cell level, which are collectively represented as overall system performance under industrial fermentation conditions. These challenges can be overcome by taking systems approaches through the use of state-of-the-art tools of systems biology, synthetic biology and evolutionary engineering in the context of industrial bioprocess. Major systems metabolic engineering achievements in recent years include microbial production of amino acids (L-valine, L-threonine, L-lysine and L-arginine), bulk chemicals (1,4-butanediol, 1,4-diaminobutane, 1,5-diaminopentane, 1,3-propanediol, butanol, isobutanol and succinic acid) and drugs (artemisinin).

The metabolic engineering literature is replete with studies describing microbial strains engineered to produce a chemical, fuel or material of economic interest, and yet relatively few such strains actually are ever scaled-up in an industrial context. Although conventional metabolic engineering studies have succeeded in developing microorganisms capable of overproducing bioproducts[1], the current approach for developing an industrial product still requires a great deal of time and effort (e.g., 50–300 person-years of work) and investment (up to several hundred million US dollars)[1]. One reason that the process of translation of academic projects is so challenging is that researchers often fail to consider a fully integrated industrial bioprocess when developing microbial strains with new activities; on the other hand, many companies have in the past lagged behind academia in adopting state-of-the-art metabolic engineering techniques. Thus, there is a need for academia and industry to enter into more active collaboration (as recently discussed[2]) and for a more rapid transfer of knowledge between the public and private sectors.

Since its formal recognition more than two decades ago[3], metabolic engineering has evolved into a more systematic and high-throughput discipline, sometimes termed 'systems metabolic engineering'[4,5]. This emerging discipline integrates traditional metabolic engineering approaches with other fields, such as systems biology (e.g., omics analysis and genome-scale computational simulation), synthetic biology (e.g., various molecular approaches, tools and pathway modules allowing fine control and regulation of gene expression levels and precise genome engineering) and evolutionary engineering (e.g., laboratory evolution of cells for enhanced product tolerance), while continuing to consider upstream (strain development) to midstream (fermentation) to downstream (separation and purification) processes as a whole (**Fig. 1**). Implementation of such tools in a systems metabolic engineering framework is already reaping dividends and has resulted in the development of *Escherichia coli* strains overproducing L-valine[6–8] and L-threonine[9,10] in 10 person-years—considerably fewer years than previously required to create high-performance industrial producer strains. Similar approaches have also enabled the successful development of an increasing number of industrial strains capable of efficiently producing various bioproducts, such as 1,4-butanediol[11], the antimalarial drug artemisinin[12] and succinic acid[13,14].

Here, we discuss ten general strategies behind the successful development of industrial microbial strains through systems metabolic engineering. We illustrate their implementation using several case studies of microbial overproduction at large-scale: L-arginine[15] (**Box 1** and **Fig. 2**); 1,4-butanediol[11] (**Box 2** and **Fig. 3**); L-lysine[16] and 1,5-diaminopentane[17] (**Box 3** and **Fig. 4**); L-valine[6–8] (**Supplementary Fig. 1**); and L-threonine[9,10] (**Supplementary Fig. 2**). Depending on the target strain, product, substrate, fermentation and downstream processes, such strategies are not always applied sequentially in the same order; often they may be implemented in an iterative manner[18]. By describing each strategy in turn, together with examples, we hope to inform readers as to the considerations important for the commercialization of strains developed in laboratories, which academic researchers may sometimes overlook. Details of specific techniques available for systems metabolic engineering have been thoroughly reviewed elsewhere, so are not discussed here (for reviews, see refs. 4, 19–26).

Project design (strategy 1)

In any metabolic engineering application, project objective and strain metrics should be designed based on various technical, economical, legal and regulatory factors. This is a process that is considered in detail in the private sector, but is often neglected in the academic

[1]Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical and Biomolecular Engineering (BK21 Plus Program), Center for Systems and Synthetic Biotechnology, Institute for the BioCentury, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea. [2]BioInformatics Research Center, KAIST, Daejeon, Republic of Korea. [3]BioProcess Engineering Research Center, KAIST, Daejeon, Republic of Korea. [4]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Hørsholm, Denmark. Correspondence should be addressed to S.Y.L. (leesy@kaist.ac.kr).

Received 4 August 2014; accepted 23 August 2015; published online 8 October 2015; doi:10.1038/nbt.3365

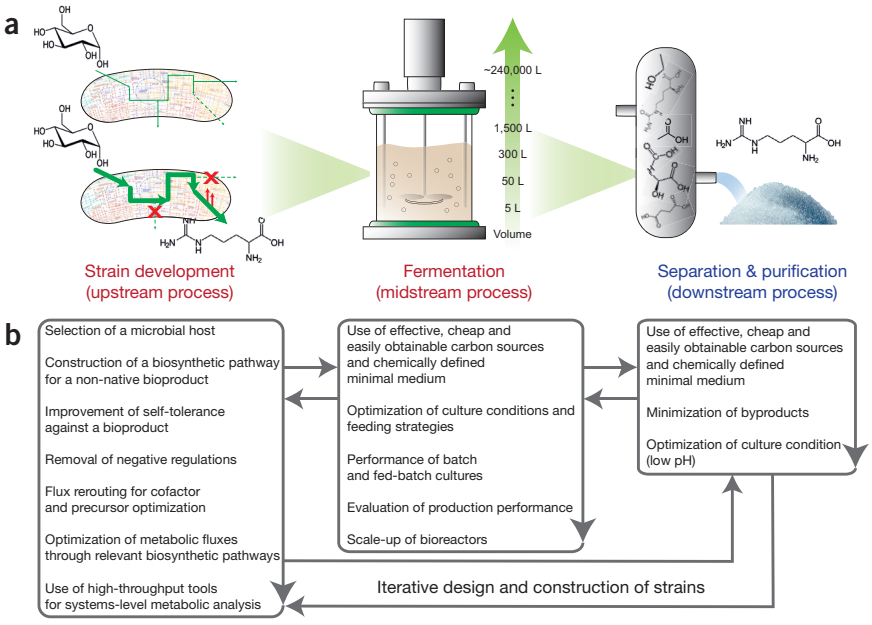


Figure 1 Concept of the systems metabolic engineering framework. (**a**) Three major bioprocess stages. (**b**) Considerations in systems metabolic engineering to optimize the whole bioprocess. List of considerations for the strain development and fermentation contribute to improving microbial strain's production performance (red), whereas those for the separation and purification help in reducing overall operation costs by facilitating the downstream process (blue). Some of the considerations can be repeated in the course of systems metabolic engineering.

milieu. Once the bioproduct to be produced has been chosen, a preliminary techno-economic analysis should be performed using several candidate host strains under the most plausible scenarios (e.g., cost-effective carbon sources, aerobic versus anaerobic fermentation, batch, fed-batch or continuous culture, fermenters and downstream processing equipment to be employed)[27]. Depending on the natural or synthetic nature of the product, different strategies can be employed for strain design[4]. In all cases, three key performance indices, namely, product concentration (titer), yield, and productivity, need to be carefully estimated to assess whether the bioprocess can be industrially competitive. For the industrial fermentation mode, fed-batch culture is most often preferred over batch or continuous culture[28], mainly owing to the higher product concentration, yield and productivity that can be achieved, flexibility of fermentation operation and reduced chance of contamination (compared with continuous culture).

Because at this stage, the best-performing microbial strain is not yet known, several candidate strains should be modeled and compared. Some of the most widely examined production microorganisms include *E. coli*, *Corynebacterium* sp., *Bacillus* sp., *Pseudomonas* sp., *Clostridium* sp. and *Saccharomyces cerevisiae*. Because abundant information on these microorganisms is available in the literature—from whole genome sequence to detailed metabolic characteristics—it is possible to estimate performance metrics through genome-scale metabolic simulation. However, it should be noted that one can employ less studied or newly isolated microorganisms if much higher performance is expected; for example, a high C4 flux bacterium *Mannheimia succiniciproducens* was isolated from the rumen of Korean cows and employed for succinic acid production at very high efficiencies[13,14].

Beyond these technical aspects of project design, one must also consider economic aspects. Typically, products sold at greater market premiums can tolerate lower performance metrics, whereas competitive products with lower premiums, including bulk chemicals, require the achievement of near-theoretical maximum metrics for all performance indices (i.e., highest possible concentration, yield and productivity). It should be noted that titer, yield and productivity are tightly linked to downstream process cost, substrate cost and bioreactor size, respectively[29]. Also, the high cost of expression systems requiring chemical inducers like isopropyl β-D-1-thiogalactopyranoside (IPTG) usually precludes their use in industrial-scale fermentations (see strategy 10, below).

Depending on the final product, another economic aspect that should be considered is the cost of any chemical modification required following biological production of precursors to efficiently yield final products. In many cases, one does not have to rely 100% on biological conversion. For example, bio-based nylon (polyamide 5/10), artemisinin, γ-butyrolactone and hydrocarbons are produced by chemical conversion of microbially overproduced precursors: 1,5-diaminopentane from *Corynebacterium glutamicum*[17], artemisinic acid from *S. cerevisiae*[12], 4-hydroxybutyric acid from *M. succiniciproducens*[30] and acetone-butanol-ethanol fermentation products of *Clostridium acetobutylicum*[31], respectively.

A last economic aspect is where to locate a manufacturing facility. As early as possible, project design should consider the optimal geographical location of a fermentation site because the cost of raw material transportation can be very high for countries where feedstocks are not abundantly available; in the case of South Korea, for example, most of the fermentation substrates are imported from outside the country.

Apart from the economic aspects, legal and regulatory considerations must also be incorporated into project design. If the intention is to take a project forward in an industrial context, an essential step is to protect intellectual property in the form of patents or trade secrets. Any project coming out of an academic laboratory that has aspirations to develop a commercial product should ensure that patents are filed before any public disclosure of the invention in the literature.

Finally, researchers should bear in mind the 'use case' for their end product and how this relates to international regulations. Microbial strains used in metabolic engineering are genetically modified organisms (GMOs), and, therefore they, and the products derived from them, fall under different GMO regulations in different jurisdictions. These regulations are particularly stringent when the final intention is for the products to be used for human and/or animal consumption (**Box 1** and **Fig. 2**).

Selection of host strain (strategy 2)

Although *E. coli* and *S. cerevisiae* are often considered as workhorses for metabolic engineering due to the availability of a wide range of established tools, in certain cases other microorganisms may be more suitable host strains. For example, in the case of L-arginine, our group selected *C. glutamicum* ATCC 21831 (AR0) as a host strain because it was already capable of producing 17 g/L of product by fed-batch culture[15] (**Box 1** and **Fig. 2**). If *C. glutamicum* had been too difficult to metabolically engineer, it would have been necessary to explore other host strains. Another consideration is whether the organism can utilize the desired carbon feedstock. Evidently, all strains of a microbe are not equal; thus, performance of, and yield from, different strains of

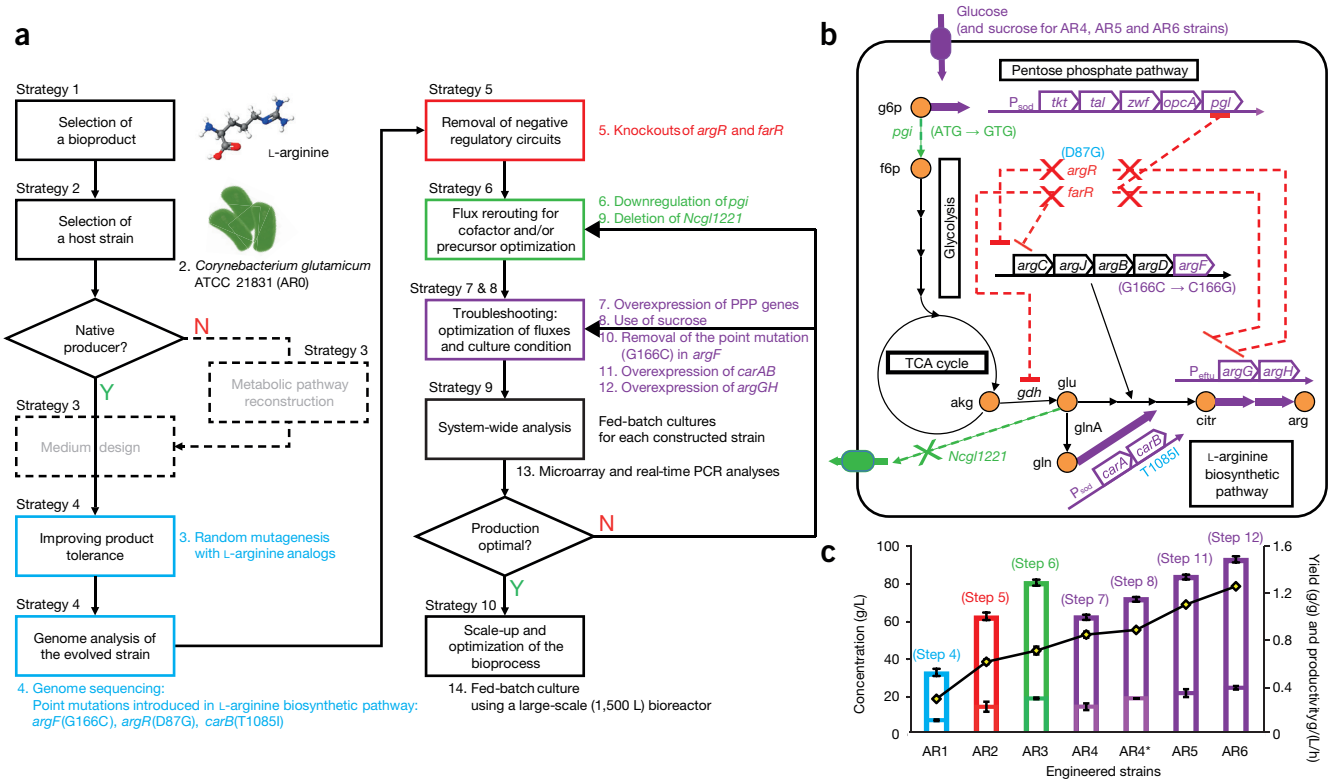


Figure 2 General scheme of systems metabolic engineering and its case study on the overproduction of L-arginine using *C. glutamicum*[15]. (**a**) Flowchart illustrates the use of each of the ten strategies, except strategy 3, which was not required as *C. glutamicum* was a native producer that did not require medium design. Metabolic engineering experiments (steps) conducted are shown next to their corresponding strategy (box). (**b**) A metabolic map of *C. glutamicum* corresponding to the metabolic engineering changes made in **a**. Dotted lines with 'X' indicate inactivated reactions, and those without 'X' indicate downregulated reactions. (**c**) L-arginine concentrations, yields and productivities of the engineered *C. glutamicum* strains obtained from their fed-batch cultures using 5-liter bioreactors. Each step of systems metabolic engineering shown in **a** and its corresponding experiments (**b**) and resulting strains (**c**) share the same colors. Only glucose was used as a feedstock for the AR1, AR2, AR3 and AR4 strains, whereas glucose and sucrose were simultaneously used for the AR4*, AR5 and AR6 strains (note that AR4 and AR4* are the same strain, the latter with a different substrate). Upper and lower bars in **c** for each strain represent L-arginine concentrations and yields, respectively; line graph with yellow diamonds represents L-arginine productivities of different engineered strains. PPP, pentose phosphate pathway; akg, α-ketoglutarate; arg, L-arginine; citr, L-citrulline; f6p, D-fructose 6-phosphate; g6p, L-glucose 6-phosphate; gln, L-glutamine; glu, L-glutamate.

a producer organism can vary substantially as can the ability of those strains to utilize different feedstocks.

Increasingly, the availability of tools to make a microorganism tractable to genetic manipulation is becoming less of an issue. With our ability to achieve genome-scale characterization in a short period of time[32], rapid advances in systems biology now allow the use of relatively unknown species as expression systems. Also, efficient synthetic biology tools, such as CRISPR (clustered, regularly interspaced, short palindromic repeats)-Cas9 (CRISPR-associated protein)-based systems are making microorganisms that were previously relatively recalcitrant amenable to gene manipulation[33,34]. For example, our group has selected the native succinic acid producer bacterium *M. succiniciproducens* as a host strain for metabolic engineering, with the goal of overproducing succinic acid[13]. In this case, by developing genetic manipulation tools from scratch, followed by systems metabolic engineering, we were able to succeed in fully exploiting its high succinic acid production capability; an engineered *M. succiniciproducens* strain was capable of producing succinic acid concentrations of ~100 g/L, close to the theoretical maximum yield, and a productivity >10 g/L/h in fed-batch culture, demonstrating higher performance indices than other microorganisms.

Thus, as new microbial species continue to emerge as new potential production hosts, one has a choice to make: developing tools for a more exotic organism previously intractable to genetic manipulation or changing the host strain to another that can be more easily engineered. Other factors, such as carbon substrate utilization range, ease of fermentation in an inexpensive medium, ease of scale-up, requirement for anaerobic versus aerobic conditions and the suitability of cost-effective downstream processes, need to be considered when selecting a host strain. For example, for the production of acids (e.g., lactic acid or succinic acid), microbial strains tolerant of low pH may be preferable because the purification of undissociated-acid product can be costly if fermentation is performed at around neutral pH.

Metabolic pathway reconstruction (strategy 3)

If metabolic pathways leading to the formation of a desired bioproduct are absent in the candidate microbial host, the potential metabolic pathways need to be established in the strain by carefully examining the candidate enzymes and/or genes through mining genomes and metagenomes[11,35,36]. For obvious reasons, this strategy is not needed for producer organisms that already generate the desired product naturally (**Box 1** and **Fig. 2**). In general, however, this strategy is extremely important in systems metabolic engineering due to our increasing interest in the biobased production of a diverse range of chemicals[37] and fuels[38,39] that are either non-natural or inefficiently produced in natural hosts.

Heterologous reactions and pathways can be introduced to a candidate host strain based not only on intuitive biological knowledge but also on the results of genome mining and cheminformatic analyses[35]. Several chemo-bioinformatic tools for this purpose have been developed (**Box 4** and **Table 1**). Construction of such heterologous pathways extends our ability to produce other diverse bulk chemicals and drugs, including 1,5-diaminopentane[17] (**Fig. 4**), adipic acid[40], 3-hydroxypropionic acid[41], artemisinic acid (a precursor of antimalarial artemisinin)[12] and antibiotic erythromycin A and its derivatives[42] (**Supplementary Fig. 3**). Once constructed, heterologous reactions need to be optimized by several well-known techniques, such as gene expression optimization and codon optimization[11], according to the fermentation results. It should also be noted that native enzymes can be evolved or even redesigned to modify their substrate specificities, thereby endowing more diverse metabolic reactions, even those not found in nature, to be performed[35,43]. Furthermore, if carbon substrate utilization was not considered in choosing the host strain in performing strategy 2 (see above), selecting the least expensive substrate available at this stage can reduce the overall operation cost of the bioprocess.

Thanks to the relatively low cost of DNA synthesis and rapid advances in synthetic biology research, efficient techniques for the combinatorial assembly of genes, operons and pathways are becoming readily available[44]. Even so, it is still difficult to predict how to construct such pathways for the most efficient production of a desired chemical. For example, a multigene pathway can be designed in many ways by assessing different genes encoding enzymes of the same function from different sources, by varying the order of the genes, by using separate promoters versus operons and by varying expression levels for each

Box 1 Metabolic engineering of a *C. glutamicum* strain overproducing L-arginine[15]

A challenge in the development of L-arginine-overproducing *C. glutamicum* was to achieve a production performance better than previous studies at laboratory scale and reproducibly at pilot-scale (**Fig. 2**). The final, high-performance AR6 strain was developed in about 9 person-years by the nine strategies described below (strategy 3 was skipped).

- Strategy 1. The end market for L-arginine was identified as for non-food, non-animal feed applications, such as for use as a plant growth stimulant[68], thus the use of a genetically modified *C. glutamicum* strain was considered acceptable. The preferred carbon substrates were derived from glucose and/or sucrose from industrially relevant feedstocks, corn starch hydrolysate and/or decomposed raw sugar from sugarcane. The initial target concentration, yield and productivity in the available 1,500-liter demo-plant scale fermenter were set at >80 g/L, 0.3~0.4 g/g and ~1.0 g/L/h, respectively.
- Strategy 2. *C. glutamicum* ATCC 21831 (AR0) was selected for its ability to produce a substantial amount of L-arginine and to use the preferred carbon substrates, corn starch hydrolysate and/or decomposed raw sugar; the organism can also be metabolically engineered using reasonably well-established gene manipulation tools.
- Strategy 3 was skipped because L-arginine is the natural product of the strain.
- Strategy 4. Random mutagenesis was performed to select an L-arginine-tolerant strain using L-arginine analogs, isolating the AR1 strain capable of producing 34.2 g/L of L-arginine in fed-batch culture (**Fig. 2c**), a more than twofold higher concentration than that obtained with the AR0 strain. Subsequent genome sequencing and analysis of the AR1 strain revealed mutations in genes associated with the L-arginine biosynthetic pathway.
- Strategy 5. Negative feedback regulation in the AR1 strain was removed by inactivating two regulatory genes, *argR* and *farR*. The resulting AR2 strain was able to produce 61.9 g/L of L-arginine by fed-batch culture (**Fig. 2c**).
- Strategy 6. Downregulation of the *pgi* gene was achieved by replacing its start codon ATG with GTG to reroute fluxes toward the pentose phosphate pathway (PPP) for more NADPH, while still permitting flux through glycolysis for good enough cell growth. As a result, the AR3 strain capable of producing 80.2 g/L of L-arginine by fed-batch culture was developed.
- Strategy 7. After fed-batch cultures of AR3 revealed decreased consumption of glucose substrate and consequently longer fermentation times than AR2, it was found that the flux through the PPP was not streamlined after rerouting the flux from the glycolytic pathway. Thus, an operon associated with PPP flux was amplified using a stronger *sod* promoter. The resulting AR4 strain gave the highest L-arginine productivity yet of 0.85 g/L/h (**Fig. 2c**); however, L-arginine titer and yield were lower than those obtained with the AR3 strain, suggesting the need for further optimization.
- Strategy 8. Fed-batch culture of the AR4 strain using sucrose and glucose as carbon sources resulted in the production of 71.7 g/L of L-arginine with productivity and yield of 0.89 g/L/h and 0.31 g/g, respectively; these values are all higher than those obtained by the fed-batch culture using glucose as a single carbon source.
- Strategy 9. The AR4 strain was subjected to final rounds of engineering by analyzing its fed-batch fermentation profiles in connection with metabolic characteristics. Because fed-batch culture of the AR4 strain consistently showed the excretion of L-glutamate, which can be further converted to L-arginine for an improved L-arginine yield and titer, the L-glutamate exporter was deleted. Analysis of genes associated with L-arginine biosynthesis using the sequenced genome data of the AR0 and AR1 strains also identified two additional gene manipulation targets: removal of a negative mutation in the *argF* gene and overexpression of the *carAB* genes. The AR5 strain constructed this way produced 82 g/L of L-arginine, with a yield of 0.35 g/g by fed-batch culture. Finally, to solve the problem of the excretion of about 5 g/L of citrulline, two rate-controlling enzymes encoded by the *argGH* genes were overexpressed using a strong promoter, resulting in the AR6 strain. Fed-batch culture of the AR6 strain allowed production of 92.5 g/L of L-arginine with a high yield of 0.40 g/g and productivity of 1.26 g/L/h. Intended changes in the AR6 strain were confirmed with transcriptome analysis, enzyme assay and real-time PCR.
- Strategy 10. When the AR6 strain was cultured in a fed-batch mode in a 1,500-liter bioreactor, 81.2 g/L of L-arginine was produced with a yield and productivity of 0.353 g/g and 0.91 g/L/h, respectively. Even though these values are somewhat lower than those obtained by the lab-scale fed-batch culture, they met the project objectives defined in strategy 1 (high enough for initial industrial-scale fermentation). The AR6 strain's production performance was found to be highly reproducible even under scale-up fermentation.

Table 1 Key features and performance metrics of pathway modeling tools used in systems metabolic engineering.

	Chemo-bioinformatics tools for pathway construction	Constraint-based reconstruction and analysis of genome-scale metabolic models	^{13}C-metabolic flux analysis	Elementary mode analysis
Key features	Use of reaction rules that describe rearrangements of atoms observed in biological system Enumeration of pathways from one metabolite to another using reactions rules Selection of enumerated candidate pathways using criteria often involving total pathway length and thermodynamic feasibility of the predicted reactions	Simulation using numerical optimization Description of comprehensive gene-protein-reaction associations in metabolic network model Easy implementation without kinetic parameters except for uptake/secretion data of several metabolites Overall good accuracy for predicting mid-exponential growth and primary metabolite production Application as a platform for omics data integration Experimentally well validated for metabolic studies of various prokaryotes and eukaryotes, and increasingly of human cells/tissues	Implementation of isotopic-labeling experiments and analytical measurements—direct biochemical evidence of metabolic status Implementation of model fitting with ^{13}C-labeling data and statistical analysis to estimate *in vivo* fluxes Greatest accuracy of estimating *in vivo* flux distribution Increasingly widely used in metabolic studies of prokaryotes and eukaryotes including mammalian cells	Analysis of stoichiometric matrix using linear algebra Easy implementation without kinetic parameters and any modeling assumptions except for reaction reversibility—by far simplest approach among these four tools Comprehensive analysis of metabolic network topology through calculation of all the unique flux vectors Easy applicability to organisms not characterized well
Limitations	Expensive computation Lack of consideration of biological feasibility (e.g., prediction of completely new proteins that cannot be created *in vivo* or *in vitro*; incompatibility of host metabolism with heterologous reactions) Very few experimental validations conducted so far	Inaccurate prediction of regulation-associated phenotypes; downregulation and amplification of genes; growth and production phenotypes outside mid-exponential phase (e.g., stationary phase growth and secondary metabolites) Necessity of manual curation to improve the model accuracy, a time-consuming procedure	Difficult implementation—proficiency in both analytical and computational processes required Use of a small-scale metabolic network Ambiguous criteria for model structure whose slight changes can yield very different *in vivo* flux estimations	Use of a small-scale metabolic network—expensive computation for large metabolic networks Difficulty in analysis of the generated flux vectors in biological context
Motivation for selecting the tool in systems metabolic engineering	Pathway reconstruction with unknown reactions Detailed reaction mechanism needed	Prediction of genome-wide counterintuitive gene manipulation targets Metabolic analysis needed, but isotopic labeling experiments unavailable Pathway design largely with homologous reactions and without detailed reaction mechanism	Accurate metabolic analysis needed Isotopic labeling experiments available	Genome information and/or genome-scale metabolic model unavailable Prediction of gene manipulation targets largely from central carbon metabolism

gene in the pathway. As there is currently no easy way of reducing the time to develop optimal synthetic pathways other than massively parallel experiments, this step is a bottleneck in pathway design and is an active subject of research.

Increasing tolerance to product (strategy 4)

It is important to increase tolerance of the strain to the desired product, which can be performed at any stage of strain development[45]. When this is performed early in strain development, it can open up metabolic flux to the formation of a desired product. However, increasing product tolerance is typically only recommended after the expression system under study has been optimized to generate product at a requisite level—say arbitrarily 50–80% of the level that starts to inhibit cell growth. The usual approach for developing a product-tolerant strain is serial subculturing of cells with increasing concentrations of product or analogs with or without mutagen treatment, followed by identifying those cells that outgrow (i.e., grow faster)[46]. By repeating this process, the tolerance level can gradually be increased.

Improved product tolerance has also been demonstrated using rational engineering approaches, such as the use of an efflux pump for biofuel in *E. coli*[47] or manipulation of ionic membrane gradients of *S. cerevisiae* for ethanol[48]. Another example is L-valine production, where overexpression of genes encoding the L-valine exporter led to an increase in L-valine production titer by more than 40% in *E. coli*[6]. Thus, engineering efflux pumps to improve the product tolerance and/or titer is a strategy that deserves investigation.

An important caveat to either of these approaches is that producer cells selected from serial subculturing or rational engineering do not necessarily overproduce the product. This is because enhanced product (or analog) tolerance at high concentrations does not necessarily correlate with an ability to synthesize product at increased specific productivity and yield. One can, of course, design the bioprocess by coupling in situ product removal with fermentation if no better ways of increasing the product tolerance are identified.

Removing negative regulatory circuits limiting overproduction (strategy 5)

Negative feedback loops are ubiquitous features of biological networks, influencing gene expression and signal transduction, generating oscillations and modulating noise. Cells have evolved their metabolic, gene regulatory and signaling networks over millennia to optimally allocate resources so as to maximize growth and prevent overproduction of metabolites that penalize fitness.

Typical negative regulatory circuits relevant to the metabolic engineer involve transcriptional attenuation control and feedback inhibition during the synthesis of amino acids. Transcriptional attenuation control is often relieved by removing genes encoding regulators that repress amino acid biosynthesis or replacing the promoters of metabolic enzymes with constitutive ones. Feedback inhibition can be handled in several different ways. If residues responsible for the feedback inhibition are already known, point mutations can be introduced by means of site-directed mutagenesis. If they are unknown, two other approaches can be taken. First, homology-based rational mutation screening, or even random mutagenesis, followed by screening at the enzyme level can be performed. Second, the production host can be mutagenized to increase product tolerance (see strategy 4), and then the relevant regulated genes or even the whole-genome can be sequenced to identify responsible mutations for increased product tolerance, which often results from the removal of feedback inhibition. In some

Box 2 Metabolic engineering of an *E. coli* strain overproducing 1,4-butanediol[11]

There were two major challenges in developing a strain for 1,4-butanediol production. First, a new synthetic pathway needed to be designed and constructed to enable production of a non-natural chemical. Second, high demand for redox power needed to be addressed at the whole metabolism level because 1,4-butanediol is a highly reduced product. The development of an efficient 1,4-butanediol-producing microorganism (**Fig. 3**) can thus be considered a landmark of systems metabolic engineering that addressed both of these challenges. The exact person-years spent on this project is not known.

- Strategies 1 and 2 were not described in the literature. For the establishment of a new synthetic pathway, the recombinant workhorse *E. coli* is commonly employed as the first choice, at least if it meets the criteria needed for the desired task.
- Strategy 3. The SimPheny Biopathway Predictor (http://www.genomatica.com/) was used to facilitate reconstructing the 1,4-butanediol biosynthetic pathway, which is absent in the host *E. coli*. Based on the predicted reactions, a set of heterologous genes were screened and their expression levels were optimized (pZA33S-sucCD-sucD-4hbd/sucA and pZE23S-cat2-002C).
- Strategy 4. The original study[11] did not discuss in detail the issue of product tolerance probably because the titer of 1,4-butanediol at that time was not high enough. However, it is known that Genomatica developed a strain highly tolerant to 1,4-butanediol.
- Strategy 5. The native *E. coli lpdA* gene was replaced with the mutated *Klebsiella pneumonia lpdA* gene to provide cells with pyruvate dehydrogenase activity under oxygen-limited conditions.
- Strategy 6. Flux rerouting toward 1,4-butanediol through the oxidative branch of the tricarboxylic acid (TCA) cycle was achieved by knocking out metabolic genes, all predicted from *in silico* genome-scale metabolic simulation (strategy 9).
- Strategy 7. The fluxes toward the oxidative branch of the TCA cycle were reinforced by deleting the *arcA* gene to avoid aerobic stress and mutating the *gltA* gene to inhibition by NADH. Furthermore, a reaction converting 4-hydroxybutyryl-CoA into 4-hydroxybutyraldehyde, which was presumed to be a bottleneck, was optimized by rescreening responsible enzymes, subsequently identifying and employing a codon-optimized *Clostridium beijerinckii* aldehyde dehydrogenase (025B). The use of this enzyme led to a slightly increased 1,4-butanediol yield, but much less ethanol byproduct, compared with the codon-optimized *C. acetobutylicum adhE2* gene initially used.
- Strategy 8. Several different carbon sources other than glucose were explored, including sucrose, xylose and biomass hydrolysate (i.e., a crude extract containing glucose, xylose, glycerol, acetic acid, 5-hydroxymethylfurfural, furfural, oligosaccharides and acetyls).
- Strategy 9. The gene knockout prediction algorithm OptKnock[69] was applied to an *E. coli* genome-scale metabolic model to identify metabolic genes that were candidates for deletion. Four genes were selected and removed, including genes encoding alcohol dehydrogenase (*adhE*), pyruvate formate lyase (*pfl*), lactate dehydrogenase (*ldh*) and malate dehydrogenase (*mdh*). Microaerobic fed-batch culture of the final strain produced over 18 g/L of 1,4-butanediol from glucose in a 2-liter fermenter (**Fig. 3c**).
- Strategy 10. In 2013, Genomatica (San Diego) announced its first successful microbial production of 1,4-butanediol on a commercial scale (greater than five-million pounds; http://www.genomatica.com/news/press-releases/successful-commercial-production-of-5-million-pounds-of-bdo/).

cases, a feedback-resistant heterologous gene, if available, can replace the native feedback-controlled gene in the production host. After removal of these negative regulatory circuits, amino acid biosynthetic genes can be further upregulated as discussed in strategy 9 (see below).

Thus, removal or attenuation of negative regulatory circuits is essential for overproduction of amino acids and polyamide precursors, as our group and others have demonstrated for the production of L-arginine in *C. glutamicum* (**Fig. 2**), L-lysine in *C. glutamicum* (**Fig. 4**), L-valine in *E. coli*[6] (**Supplementary Fig. 1**) and L-threonine in *E. coli* (**Supplementary Fig. 2**). Negative regulatory feedback mechanisms are also common in other metabolic pathways of industrial interest; thus, they have been engineered in strains overproducing aliphatic alcohols (**Box 2** and **Fig. 3**) and dicarboxylic acids (e.g., adipic acid[40]; **Supplementary Fig. 3**). Such regulatory engineering will be increasingly performed as more regulatory mechanisms related to the formation of other products are uncovered.

Rerouting fluxes to optimize cofactor and/or precursor availability (strategy 6)

Cofactors play an essential role in many biochemical reactions, and their availability are thus key consideration for a metabolic engineer. For example, NADH, NADPH, ATP and coenzyme A (CoA) are cofactors involved in hundreds of reactions. Also, many intracellular metabolites serve as precursors to the desired product. It is often necessary to reroute the metabolic fluxes in a pathway to optimize the availability of cofactors and precursor metabolites, even though the desired product can be generated through the original, designed pathways.

Several approaches can be taken to manipulate cofactor and/or precursor availability. Gene knockout is the most obvious choice to reroute fluxes to increase the level of cofactors or precursors by removing competing pathways (strategy 6 in **Boxes 1**–**3** and **Figs. 2**–**4**). Gene knockdown, however, using synthetic small regulatory RNAs[49,50], becomes useful if the genes to be removed are essential (strategy 6 in **Boxes 1** and **3** and **Figs. 2** and **4**). Also, use of synthetic small regulatory RNAs and other high-throughput genome-scale engineering tools discussed in strategy 9 (see below) enables high-throughput screening of strains having various combinations of downregulated genes. These parallel high-throughput methods are extremely powerful as they overcome the disadvantages of gene knockout experiments, which are often time-consuming, laborious, sequential, irreversible, nontransferable to another strain and only applicable to nonessential genes[49,50]. Once combinatorial knockdown targets are identified using plasmid-based experiments, industrial strains can be developed by transferring that genotype into the chromosome for more stable operation.

Using an alternative enzyme or pathway of heterologous origin can boost the product formation rate[11]. Also, cofactor (e.g., NADH or NADPH) specificity of the enzyme can be delicately manipulated, for example, by swapping in a cofactor that optimizes performance or developing an enzyme capable of using both NADH and NADPH as a cofactor, to maximize redox power available for the enzyme while improving production performance[51,52]. However, introduction of heterologous

enzymes into a production host needs careful scrutiny because it can trigger unexpected inhibitory metabolic reactions caused by the promiscuous activities of the introduced enzymes[53]. In general, cofactor and precursor optimization usually requires more system-wide approaches than removal of pathway-specific negative regulatory circuits, because in the former case, global mass, energy and redox balances need to be considered. In the case of engineering enhanced L-arginine production, even though the central carbon flux was rerouted toward the pentose phosphate pathway to increase the intracellular NADPH pool necessary for boosting product formation, glycolysis was not completely blocked so as to prevent substantial growth retardation (even after the flux rerouting; **Box 1** and **Fig. 2**).

Another striking example is the reversal of the β-oxidation cycle in *E. coli*, which led to efficient production of alcohols and carboxylic acids with varied carbon-chain lengths directly from acetyl-CoA[39]. In this work, genes responsible for the β-oxidation cycle were constitutively expressed by mutating or inactivating repressive regulatory genes, and different termination enzymes were overexpressed for the production of specific target products.

Diagnosing and optimizing metabolic fluxes toward product formation (strategy 7)

The metabolic status of the strain needs to be systematically diagnosed and optimized not only to remove any flux bottlenecks but also to send as much flux as possible toward the formation of the end product. This strategy should be performed under conditions that are as similar as possible to the final industrial fermentation conditions. Because standard industrial fermentation operations are often performed in a fed-batch mode, it is essential to perform fed-batch cultures, at least under laboratory-scale conditions, of intermediate strains during strain development[24]. By doing so, problems, such as potential metabolic bottlenecks and byproduct formation in fed-batch conditions, can be prioritized for further metabolic engineering of the production strain toward enhanced product formation.

A key point here is that analysis of an intermediate strain grown under fed-batch culture conditions at each stage facilitates evaluation and/or diagnosis of production performance measured by performance indices, such as productivity, yield and titer; this then leads to new objectives for the next round of metabolic engineering. For example, in our engineering of a high-yield L-arginine producer (**Box 1** and **Fig. 2**), when fed-batch culture revealed that the substrate glucose consumption rate was low in one strain, use of a stronger promoter enabled greater flux through the pentose phosphate pathway, which then resulted in a strain with much higher productivity; however, this strain had lower yield and product titer, again prompting diagnosis and further optimization. **Box 3** describes a similar approach, including changes to direct flux toward end product and boost flux through the pentose phosphate pathway for L-lysine production in *C. glutamicum*. Thus, flux to end product can be optimized by removing bottlenecks, diverting flux from branch pathways or even stopping secretion of byproducts (**Box 2** and **Fig. 3**), which can reduce the operation costs for product separation and purification in downstream processes[54,55] (**Fig. 1**).

Diagnosis and optimization of microbial culture conditions (strategy 8)

Although undesirable characteristics that emerge during the fed-batch cultivation of an intermediate strain can be corrected by gene manipulation (strategy 7), another approach is to optimize culture conditions

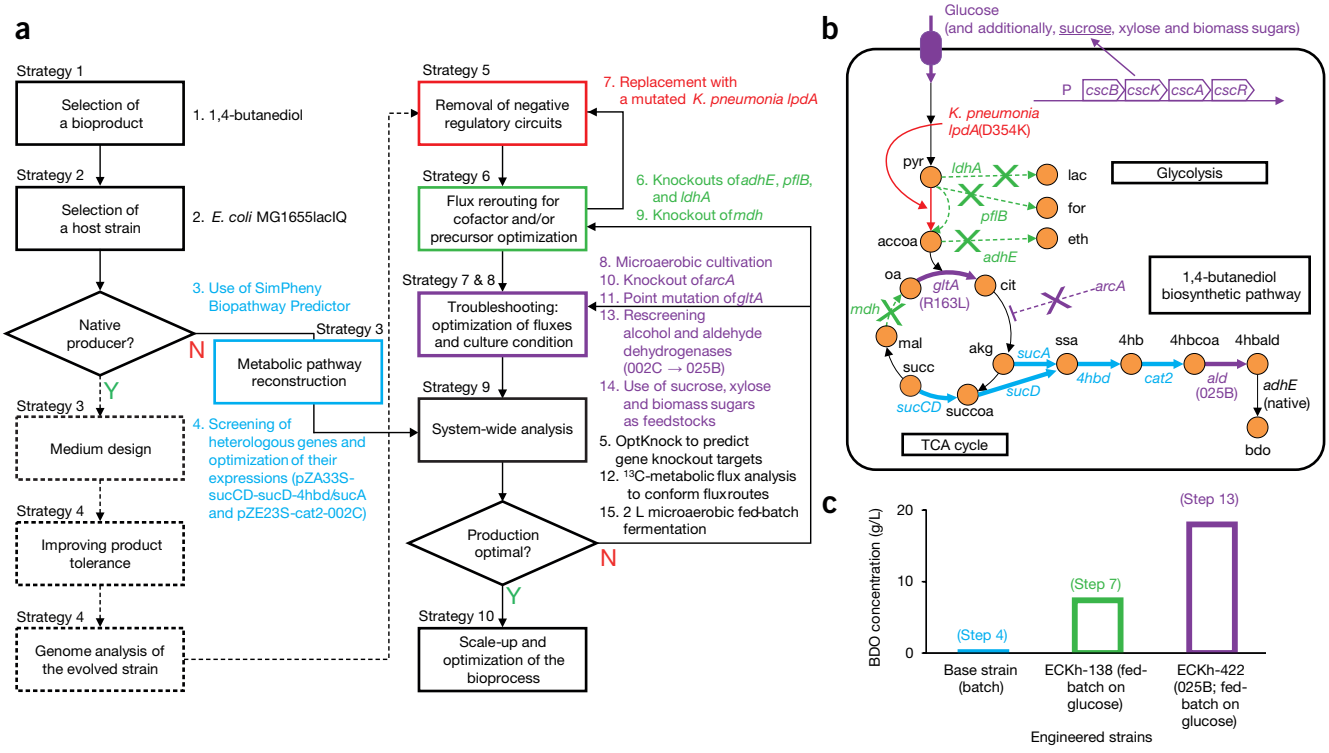


Figure 3 General scheme of systems metabolic engineering and its case study on the overproduction of 1,4-butanediol using *E. coli*[11]. (**a**) Flowchart illustrates how strategies 1–3, 5–10 were applied. Metabolic engineering experiments (steps) conducted are shown next to their corresponding strategy (box). (**b**) A metabolic map of *E. coli* corresponding to the metabolic engineering changes made in **a**. Dotted lines with 'X' indicate inactivated reactions and those without 'X' indicate downregulated reactions. (**c**) Titers of 1,4-butanediol produced by the middle and final engineered *E. coli* strains from their batch and fed-batch cultures. 4hb, 4-hydroxybutyrate; 4hbald, 4-hydroxybutyraldehyde; 4hbcoa, 4-hydroxybutyryl-CoA; accoa, acetyl-CoA; akg, α-ketoglutarate; bdo, 1,4-butanediol; cit, citrate; eth, ethanol; for, formate; lac, lactate; mal, malate; oa, oxaloacetate; pyr, pyruvate; ssa, succinyl semialdehyde; succ, succinate; succoa, succinyl-CoA.

Box 3 Metabolic engineering of a *C. glutamicum* strain overproducing L-lysine[16] and bio-nylon[17]

Our aim in this work was to develop the first genetically defined strain that overproduces L-lysine, rather than employing a random mutagenesis and selection approach. Furthermore, this L-lysine-overproducing strain was further engineered to produce the bio-nylon monomer 1,5-diaminopentane. Along with our previous studies on bio-based production of 1,4-diaminobutane and 1,5-diaminopentane by engineered *E. coli*[4], this research demonstrated the potential of microbial production of nylon in a biosustainable manner (**Fig. 4**).

- Strategy 1 and 2. The wild-type *C. glutamicum* strain was selected for L-lysine production, and the *C. glutamicum* strain LYS-12 engineered to overproduce L-lysine was used as the starting strain to produce 1,5-diaminopentane because L-lysine is a direct precursor of 1,5-diaminopentane.
- Strategy 3. A codon-optimized heterologous *ldcC* gene encoding lysine decarboxylase was introduced into the *C. glutamicum* LYS-12 strain to produce 1,5-diaminopentane.
- Strategy 4 was not reported in the literature.
- Strategy 5. The *lysC* gene encoding aspartokinase was mutated by nucleotide substitution (C932T) to release feedback inhibition by L-lysine and L-threonine in *C. glutamicum* for enhanced L-lysine production.
- Strategy 6. Fluxes to L-lysine were reinforced by overexpressing *ddh* and removing competing pathways (knockout of *pck* and downregulation of *hom*). Flux rerouting to the pentose phosphate pathway (PPP) was conducted for NADPH generation by overexpressing the gluconeogenic gene *fbp*. For 1,5-diaminopentane production, *NCgl1469* encoding *N*-acetyltransferase and *lysE* encoding L-lysine exporter were both removed as they divert L-lysine away from 1,5-diaminopentane.
- Strategy 7. For L-lysine production, the engineered *C. glutamicum* went through metabolic flux optimization by overexpressing genes involved in the L-lysine biosynthetic pathway (*dapB*, *lysA* and the mutated *lysC*), mutating a *pycA* gene and downregulating the *icd* gene, which increased fluxes toward L-lysine biosynthesis. As in the case of L-arginine production, the PPP operon was amplified in the engineered *C. glutamicum* strain to enhance L-lysine production. For 1,5-diaminopentane production, amplification of its exporter (*cg2893*) led to further improvement in the production titer.
- Strategy 8. A complex medium based on molasses was used for the fed-batch culture of *C. glutamicum* LYS-12 in order to evaluate its L-lysine production performance in an industrial setting. Similarly, industrial glucose medium was used for cultivating 1,5-diaminopentane-overproducing *C. glutamicum* DAP-16 strain. As a result, *C. glutamicum* LYS-12 and *C. glutamicum* DAP-16 strains produced 120 g/L of L-lysine and 88 g/L of 1,5-diaminopentane, respectively, by fed-batch cultures using industrial-like media (**Fig. 4c**).
- Strategy 9. The ideal flux distributions obtained through elementary model analysis and the *in vivo* flux distributions obtained by ^{13}C-metabolic flux analysis of the strain were compared to identify gene manipulation targets for L-lysine production in *C. glutamicum*. The comparison suggested changes in the following pathways: first, upregulation of the lysine biosynthetic pathway; second, upregulation of anaplerotic carboxylation; third, downregulation of the TCA cycle; fourth, upregulation of the PPP; and fifth, downregulation of anabolic pathways.
- Strategy 10 was not reported in these studies.

in an industrially relevant manner. As with other aspects of the systems metabolic engineering scheme, choice of alternative substrates should take into account such considerations as the availability of feedstock and its economics. In general, chemically defined media are desirable for both laboratory- and industrial-scale fermentations because they allow precise metabolic analysis of a production host and greater experimental reproducibility compared with complex media. However, for industrial-scale fermentations, medium cost is often considered more important than the nature of culture media being either defined, semi-defined or complex. Thus, it is important to develop an industrial strain capable of efficiently producing a desired product using rather impure low-price substrates, in particular, for the production of rather inexpensive bulk products.

For the microbial production of L-arginine, the decreased product titer and yield of the AR4 strain (strategy 7, **Box 1**) appeared to be caused by too much increased biomass formation, although the strain showed the desired productivity due to the higher glucose uptake and specific L-arginine production rates. Therefore, we reexamined substrate utilization to assess whether this could be used to influence the growth rate of the AR4 strain. Because the *Corynebacterium* strain employed for L-arginine production was found to grow more slowly on sucrose than on glucose, we investigated the effects of employing a mixture of sucrose and glucose (**Box 1** and **Fig. 2**). An additional rationale behind our selection of sucrose and glucose was that these carbohydrates are present in two of the most popular industrial carbon feedstocks, raw sugar and corn starch hydrolysate, respectively. The use of different carbon sources other than glucose was also studied for culturing the engineered, 1,4-butanediol-producing *E. coli* to use industrially preferred fermentation substrates (**Box 2** and **Fig. 3**).

System-wide gene manipulation to further increase production (strategy 9)

Once the above interventions have been carried out, a system-wide examination of metabolism can be carried out to identify potential gene manipulation targets that may further enhance production capability. This strategy encompasses the final rounds of engineering required to construct the industrial strain. They are performed by taking systems and synthetic biology approaches, such as cultivation profile-based system-wide analyses[7], high-throughput genome-scale engineering[26], omics-based approaches[6] and/or *in silico* metabolic simulations[6,56] (**Box 4**). In this context, the use of an increasing array of high-throughput genome-scale engineering tools, such as (co-selection) multiplex automated genome engineering[57,58], trackable multiplex recombineering[59] and synthetic small regulatory RNAs[49,50], enables the isolation of mutants with desired phenotypes by identifying optimal combinations of genetic targets within a short period of time.

The current challenge for wider application of these methods is to overcome low transformation efficiencies of host strains other than *E. coli* and to develop screening methods for mutants overproducing a desired bioproduct. Among these, genome-scale metabolic simulations have been successfully used for developing industrial strains, whereas high-throughput genome-scale engineering tools are coming into more widespread use.

It is important to stress that although strategy 9 comprises all the final rounds of engineering, the computational tools may be applied at a relatively early stage of strain development. Thus, in the case of L-arginine production, system-wide gene analysis and/or manipulation was applied

to intermediate strains before the final industrial strain of microbe, AR6, was obtained (**Box 1** and **Fig. 2**).

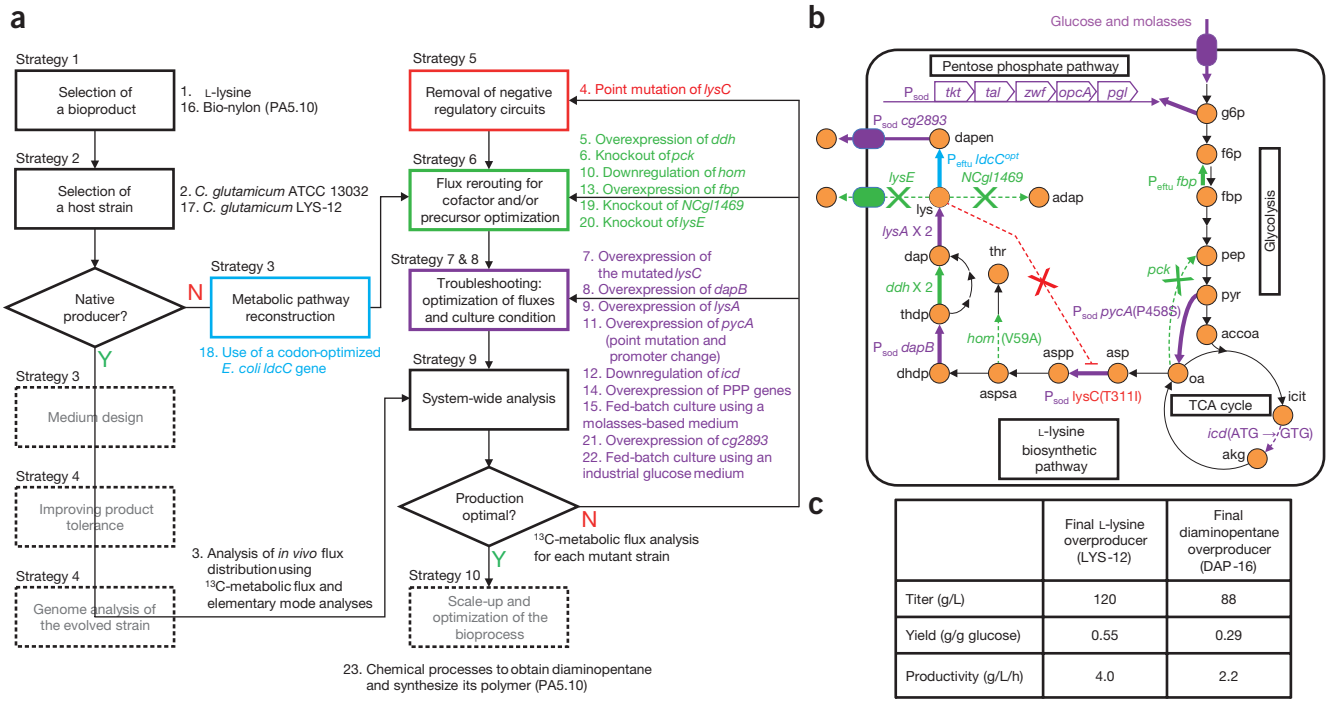
Scale-up fermentation of the developed strain and diagnosis (strategy 10)

The developed microbial strain should be validated in a pilot plant–scale or demo plant–scale fermenter because additional problems can be found during the scale-up cultivation of the developed strain. This process can be laborious if the pilot-scale fermentation results vary widely from the laboratory-scale fermentation results. The extent of such discrepancies will be different from one strain to another, and thus they cannot be easily predicted until actual pilot-scale fermentation is performed. Discrepancies in strain performance are found mainly due to different mixing and aeration in laboratory-scale and pilot-scale fermenters, and thus aerobic fermentation, which is sensitive to the oxygen transfer rate, is much more often affected by scale-up issues.

Another critical problem is reversion of a high-performance strain back to the low-performance strain through the loss of production capacity and phenotype alteration during the course of industrial fermentation, which can be caused by genetic instability and/or fermentation conditions having different mass transfer rates of nutrients and oxygen, and substrate and product concentration profiles in the industrial-sized fermenter[60]. At the moment, no tools exist that guarantee 100% strain stability, but permanent chromosomal manipulation (instead of plasmid-mediated engineering) and gene-level engineering (e.g., knockout of an entire gene) rather than changes of a few base pairs (e.g., single-point mutations) can mitigate the loss of the desired phenotype.

In industrial fermentation, companies regularly isolate pure colonies of high performance at the end of the fermentation to maintain the desired strain phenotype in subsequent fermentations. As mentioned earlier, inducers, such as IPTG, should not be used in industrial fermentation; constitutive promoters are standard practice for gene expression in industrial contexts. Other synthetic biology–based regulatory circuits, such as an auto-inducer, which dynamically detects the presence of specific target molecules produced *in vivo*[61], are also being actively developed. These new tools will be useful in fine-controlling the fluxes during fermentation on a large scale.

Because of the time and effort required to perform fed-batch cultures, even on a laboratory scale, many researchers end up carrying out only flask or batch cultures during the strain development. This increases the probability of failing in scale-up fermentation. Of course, even if fed-batch cultures have been explored in development in an academic setting, it is the job of industrial fermentation engineers to fine-tune the fermentation conditions (e.g., using industrial-grade carbon substrates and other medium components and manipulating pH, temperature and oxygen transfer) to match or even exceed the high performance of laboratory-scale fed-batch culture.



	Final L-lysine overproducer (LYS-12)	Final diaminopentane overproducer (DAP-16)
Titer (g/L)	120	88
Yield (g/g glucose)	0.55	0.29
Productivity (g/L/h)	4.0	2.2

Figure 4 General scheme of systems metabolic engineering and its case study on the overproduction of L-lysine[16] and bio-nylon[17] using *C. glutamicum*. (**a**) Flowchart illustrates the use of each of the ten strategies, except strategies 4 and 10, which were not reported. Although strategy 10 was not discussed, chemical polymerization studies were conducted to synthesize bio-nylon from 1,5-diaminopentane and diacid. The whole procedure was repeated twice, first for the production of L-lysine and second for the production of 1,5-diaminopentane, a monomer of bio-nylon. The wild-type *C. glutamicum* ATCC 13032 strain was used to develop the L-lysine overproducer and the L-lysine overproducer LYS-12 was used as the starting host for 1,5-diaminopentane production. Metabolic engineering experiments (steps) conducted are shown next to their corresponding strategy (box). (**b**) A metabolic map of *C. glutamicum* ATCC 13032 and LYS-12 strains corresponding to the metabolic engineering changes made in **a**. Dotted lines with 'X' indicate inactivated reactions and those without 'X' indicate downregulated reactions. Note that the point mutant in the *lysC* gene (i.e., T311I) is shown in red to indicate removal of the feedback inhibition on this gene, whereas its new promoter (e.g., *sod* promoter) is in purple to present overexpression of this gene in the later steps. adap, *N*-acetyldiaminopentane; akg, α-ketoglutarate; asp, L-aspartate; aspp, L-4-aspartyl phosphate; aspsa, L-aspartate semialdehyde; dap, D,L-diaminopimelate; dapen, 1,5-diaminopentane; dhdp, 2,3-dihydrodipicolinate; f6p, D-fructose 6-phosphate; fbp, D-fructose 1,6-bisphosphate; g6p, D-glucose 6-phosphate; icit, isocitrate; lys, L-lysine; oa, oxaloacetate; pep, phosphoenolpyruvate; pyr, pyruvate; thdp, tetrahydrodipicolinate; thr, L-threonine. (**c**) Production performances of the final *C. glutamicum* strains, each overproducing L-lysine and 1,5-diaminopentane, from their fed-batch cultures.

Box 4 Pathway modeling tools in systems metabolic engineering

Several pathway modeling tools have contributed to systems metabolic engineering, including but not limited to the following: chemo-bioinformatic tools for pathway construction[35], constraint-based reconstruction, and analysis (COBRA) of genome-scale metabolic models[56], ^{13}C-metabolic flux analysis[70] and elementary mode analysis[71]. The chemo-bioinformatic tools for pathway construction use predefined reaction rules to enumerate candidate biosynthetic pathways from a target chemical product to a known precursor metabolite[35]. Use of reaction rules makes this approach chemically logical, and most distinct from the other three approaches. Genome-scale metabolic models, ^{13}C-metabolic flux analysis and elementary mode analysis are rooted in the same mathematical logic that develops a stoichiometric matrix (i.e., a metabolic network) describing a set of metabolic reactions with stoichiometric coefficients[71,72]. However, the stoichiometric matrix is processed in different ways in these tools, which are selected based on the study objective and the availability of experimental support. All these tools can help identify optimal metabolic pathways to drive fluxes from one metabolite (e.g., precursor) to another (e.g., target product) in common. However, because each tool is based on different logic, each has different strengths and thus works best under different circumstances. Key features and performance metrics of these tools are summarized in **Table 1**.

Another major challenge in industrial-scale fermentation is contamination control. Phage infection is a well-known problem for many microbial fermentations. In the case of L-arginine production, *C. glutamicum* has a longer doubling time than *E. coli* (~1.5–2.5 h versus 20–30 min), and thus is subject to a higher contamination risk. Also, excessive use of certain preferred medium constituents, such as corn steep liquor, complicates downstream processing because of their complex components and their tendency to increase medium viscosity, which hinders complete sterilization of the fermenter and affects subsequent rounds of fermentation. Although these issues have been successfully addressed in industry through experience, researchers should consider these factors carefully during strain development and establishment and operation of the fermentation plant.

Conclusions and perspectives

In this Perspective, we describe ten strategies that comprise systems metabolic engineering for industrial strain development. These strategies encompass project design, host selection, metabolic pathway reconstruction, regulatory engineering, byproducts removal, energy and redox requirements, flux distribution at the branch points, carbon source and nutrient utilization, fermentation profile and characteristics, product (and metabolic intermediate) export and import, the toxicity of the product and its intermediate metabolites, and recovery and purification of the target product. It is obvious that microbial production of different types of bioproduct (both complex and simple molecules) will require adjustment of the aforementioned strategies of systems metabolic engineering. Nonetheless, we anticipate that the increasing adoption of the strategies described here will allow development of strains capable of efficiently producing various types of bioproducts on an industrial scale with reduced effort, time and cost.

Nonetheless, systems metabolic engineering is not without its challenges. First, metabolic engineers must work within the confines of the catalytic efficiency of the enzymes they are working with, the range of chemical reactions carried out, and the compatibility of the desired product and its associated new pathway with a microbial host's metabolism. For the production of non-natural products for which biosynthetic enzymes and pathways are unknown, chemo-informatic and enzyme engineering approaches need to be employed to better understand the substrate promiscuity of native and engineered enzymes. A recent work on the production of thebaine and hydrocodone (both opioid compounds) using an engineered yeast strain expressing 21 and 23 genes, respectively, encoding inherent yeast enzymes and heterologous enzymes from plants, mammals and bacteria, nicely demonstrates such approaches[62]. Also, interactions between a novel pathway and the host metabolism need to be systematically studied[53].

Second, optimization of metabolic fluxes is frequently complicated by a trade-off between the formation of biomass and target product, redox balances and complex inhibitory regulation. Although metabolic problems have been systematically handled using genome-scale metabolic modeling and simulation, more often than not, regulatory effects have not. Whole-cell modeling[63] that also captures other cellular processes, such as regulatory and signaling circuits, in addition to metabolism can be useful in tackling these metabolic engineering problems.

Third, enhancement of product tolerance still relies on random mutagenesis and/or evolutionary engineering rather than rational design because the molecular mechanisms involved are mostly unknown to us. Combined use of transport protein engineering and rationally designed strain evolution will continue to play important roles in developing product-tolerant strain. Some intracellular enzymes might also become inhibited when product concentration becomes higher, which also needs to be overcome by evolution of enzymes or the use of more tolerant enzymes of heterologous origin.

Fourth, most industrial strains engineered at process are not necessarily capable of fully using newly emerging, more sustainable non-food or waste substrates. As the impetus grows to use sustainable feedstocks, further effort will need to be placed into developing strains capable of using alternative carbon substrates, including natural gas, food wastes, non-food lignocellulosic hydrolysates of various origins and even carbon dioxide. Methods for preparing fermentable substrates from these alternative carbon substrates need to be much improved. Research effort needs to be exerted to develop optimally combined physicochemical and biological methods to derive fermentable carbon sources from them. In the case of food waste, various stakeholders need to participate in its collection and processing, following accepted policy to make such use economically and socially acceptable.

Fifth, bioprocesses use large volumes of water, which makes it difficult to homogeneously mix nutrients and oxygen within the fermenter and to separate the product cost effectively. This will become an increasing challenge due to decreasing worldwide water resources. One important solution is 'use and reuse' of the fermentation broth; that is, recycling fermentation broth after fermentation[64]. Other interesting solutions include using seawater and halophilic host strains for fermentation[65] and thermotolerant strains to reduce the use of cooling water[66].

Sixth, overall limitations in strain development come from incomplete knowledge of the link between microbial genotype and phenotype. Application of carefully designed genetic perturbations to the cell and observation of the consequent phenotypes[67] as well as integrative analysis of various omics data[32] should help better link microbial genotype and phenotype. All of these lead to a grand

seventh challenge of developing industrial chassis strains that can be easily used to swap in different pathways and as the starting point for developing industrial strains capable of producing novel chemicals and biomaterials. This final challenge may well be insurmountable, at least in terms of there being one, unique chassis strain. An alternative might be the creation of chassis strains that work for different product groups (e.g., aromatic chemicals, C4 chemicals or fatty acid–derived products).

As systems metabolic engineering is applied more widely, we expect rapid progress in addressing the above challenges. Together with an increasing body of knowledge on metabolic systems in a diverse array of microbial species and strains, a greater appreciation and application of the strategies described here will galvanize industrial strain development. Greater adoption of these strategies will also become imperative for accelerating the development of innovative products in a biosustainable manner to meet the needs of society in the coming decades.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank Seok Hyun Park, Sol Choi, Joungmin Lee, Chan Woo Song, Jae Ho Shin and Won Jun Kim for thoughtful discussions. This work was supported by the Technology Development Program to Solve Climate Changes on Systems Metabolic Engineering for Biorefineries (NRF-2012M1A2A2026556) and by Intelligent Synthetic Biology Center through the Global Frontier Project (2011-0031963) from the Ministry of Science, ICT and Future Planning (MSIP) through the National Research Foundation (NRF) of Korea. This work was also supported by the Novo Nordisk Foundation.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Hong, K.K. & Nielsen, J. Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries. *Cell. Mol. Life Sci.* **69**, 2671–2690 (2012).
2. Pronk, J.T. *et al.* How to set up collaborations between academia and industrial biotech companies. *Nat. Biotechnol.* **33**, 237–240 (2015).
3. Bailey, J.E. Toward a science of metabolic engineering. *Science* **252**, 1668–1675 (1991).
4. Lee, J.W. *et al.* Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nat. Chem. Biol.* **8**, 536–546 (2012).
5. Sagt, C.M. Systems metabolic engineering in an industrial setting. A*ppl. Microbiol. Biotechnol.* **97**, 2319–2326 (2013).
6. Park, J.H., Lee, K.H., Kim, T.Y. & Lee, S.Y. Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and *in silico* gene knockout simulation. *Proc. Natl. Acad. Sci. USA* **104**, 7797–7802 (2007).
7. Park, J.H., Kim, T.Y., Lee, K.H. & Lee, S.Y. Fed-batch culture of *Escherichia coli* for L-valine production based on *in silico* flux response analysis. *Biotechnol. Bioeng.* **108**, 934–946 (2011).
8. Park, J.H., Jang, Y.S., Lee, J.W. & Lee, S.Y. *Escherichia coli* W as a new platform strain for the enhanced production of L-valine by systems metabolic engineering. *Biotechnol. Bioeng.* **108**, 1140–1147 (2011).
9. Lee, K.H., Park, J.H., Kim, T.Y., Kim, H.U. & Lee, S.Y. Systems metabolic engineering of *Escherichia coli* for L-threonine production. *Mol. Syst. Biol.* **3**, 149 (2007).
10. Lee, J.W. *et al.* Development of sucrose-utilizing *Escherichia coli* K-12 strain by cloning β-fructofuranosidases and its application for L-threonine production. *Appl. Microbiol. Biotechnol.* **88**, 905–913 (2010).
11. Yim, H. *et al.* Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat. Chem. Biol.* **7**, 445–452 (2011).
12. Paddon, C.J. *et al.* High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496**, 528–532 (2013).
13. Lee, S.J., Song, H. & Lee, S.Y. Genome-based metabolic engineering of *Mannheimia succiniciproducens* for succinic acid production. *Appl. Environ. Microbiol.* **72**, 1939–1948 (2006).
14. Becker, J. *et al.* Systems-wide analysis and engineering of metabolic pathway fluxes in bio-succinate producing *Basfia succiniciproducens*. *Biotechnol. Bioeng.* **110**, 3013–3023 (2013).
15. Park, S.H. *et al.* Metabolic engineering of *Corynebacterium glutamicum* for L-arginine production. *Nat. Commun.* **5**, 4618 (2014).
16. Becker, J., Zelder, O., Häfner, S., Schröder, H. & Wittmann, C. From zero to hero—design-based systems metabolic engineering of *Corynebacterium glutamicum* for L-lysine production. *Metab. Eng.* **13**, 159–168 (2011).
17. Kind, S. *et al.* From zero to hero—production of bio-based nylon from renewable resources using engineered *Corynebacterium glutamicum*. *Metab. Eng.* **25**, 113–123 (2014).
18. Lee, S.Y., Lee, D.Y. & Kim, T.Y. Systems biotechnology for strain improvement. *Trends Biotechnol.* **23**, 349–358 (2005).
19. Medema, M.H., van Raaphorst, R., Takano, E. & Breitling, R. Computational tools for the synthetic design of biochemical pathways. *Nat. Rev. Microbiol.* **10**, 191–202 (2012).
20. Paddon, C.J. & Keasling, J.D. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nat. Rev. Microbiol.* **12**, 355–367 (2014).
21. Esvelt, K.M. & Wang, H.H. Genome-scale engineering for systems and synthetic biology. *Mol. Syst. Biol.* **9**, 641 (2013).
22. Yadav, V.G., De Mey, M., Lim, C.G., Ajikumar, P.K. & Stephanopoulos, G. The future of metabolic engineering and synthetic biology: towards a systematic practice. *Metab. Eng.* **14**, 233–241 (2012).
23. Cobb, R.E., Si, T. & Zhao, H. Directed evolution: an evolving and enabling synthetic biology tool. *Curr. Opin. Chem. Biol.* **16**, 285–291 (2012).
24. Lee, S.Y. High cell-density culture of *Escherichia coli*. *Trends Biotechnol.* **14**, 98–105 (1996).
25. Lee, B.R., Cho, S., Song, Y., Kim, S.C. & Cho, B.K. Emerging tools for synthetic genome design. *Mol. Cells* **35**, 359–370 (2013).
26. Song, C.W., Lee, J. & Lee, S.Y. Genome engineering and gene expression control for bacterial strain development. *Biotechnol. J.* **10**, 56–68 (2015).
27. Zhuang, K.H. & Herrgård, M.J. Multi-scale exploration of the technical, economic, and environmental dimensions of bio-based chemical production. *Metab. Eng.* **31**, 1–12 (2015).
28. Croughan, M.S., Konstantinov, K.B. & Cooney, C. The future of industrial bioprocessing: batch or continuous? *Biotechnol. Bioeng.* **112**, 648–651 (2015).
29. Van Dien, S. From the first drop to the first truckload: commercialization of microbial processes for renewable chemicals. *Curr. Opin. Biotechnol.* **24**, 1061–1068 (2013).
30. Choi, S. *et al.* Production of 4-hydroxybutyric acid by metabolically engineered Mannheimia succiniciproducens and its conversion to γ-butyrolactone by acid treatment. *Metab. Eng.* **20**, 73–83 (2013).
31. Anbarasan, P. *et al.* Integration of chemical catalysis with extractive fermentation to produce fuels. *Nature* **491**, 235–239 (2012).
32. Guan, Y., Dunham, M., Caudy, A. & Troyanskaya, O. Systematic planning of genome-scale experiments in poorly studied species. *PLOS Comput. Biol.* **6**, e1000698 (2010).
33. Cobb, R.E., Wang, Y. & Zhao, H. High-efficiency multiplex genome editing of Streptomyces species using an engineered CRISPR/Cas system. *ACS Synth. Biol.* **4**, 723–728 (2015).
34. Tong, Y., Charusanti, P., Zhang, L., Weber, T. & Lee, S.Y. CRISPR-Cas9 based engineering of actinomycetal genomes. *ACS Synth. Biol.* doi:10.1021/acssynbio.5b00038 (25 March 2015).
35. Shin, J.H., Kim, H.U., Kim, D.I. & Lee, S.Y. Production of bulk chemicals via novel metabolic pathways in microorganisms. *Biotechnol. Adv.* **31**, 925–935 (2013).
36. Lee, J.W., Kim, T.Y., Jang, Y.S., Choi, S. & Lee, S.Y. Systems metabolic engineering for chemicals and materials. *Trends Biotechnol.* **29**, 370–378 (2011).
37. Lee, J.W., Kim, H.U., Choi, S., Yi, J. & Lee, S.Y. Microbial production of building block chemicals and polymers. *Curr. Opin. Biotechnol.* **22**, 758–767 (2011).
38. Choi, Y.J. & Lee, S.Y. Microbial production of short-chain alkanes. *Nature* **502**, 571–574 (2013).
39. Dellomonaco, C., Clomburg, J.M., Miller, E.N. & Gonzalez, R. Engineered reversal of the β-oxidation cycle for the synthesis of fuels and chemicals. *Nature* **476**, 355–359 (2011).
40. Yu, J.L., Xia, X.X., Zhong, J.J. & Qian, Z.G. Direct biosynthesis of adipic acid from a synthetic pathway in recombinant *Escherichia coli*. *Biotechnol. Bioeng.* **111**, 2580–2586 (2014).
41. Borodina, I. *et al.* Establishing a synthetic pathway for high-level production of 3-hydroxypropionic acid in *Saccharomyces cerevisiae* via β-alanine. *Metab. Eng.* **27**, 57–64 (2015).
42. Jiang, M. & Pfeifer, B.A. Metabolic and pathway engineering to influence native and altered erythromycin production through *E. coli*. *Metab. Eng.* **19**, 42–49 (2013).
43. Zhang, K., Li, H., Cho, K.M. & Liao, J.C. Expanding metabolism for total biosynthesis of the nonnatural amino acid L-homoalanine. *Proc. Natl. Acad. Sci. USA* **107**, 6234–6239 (2010).
44. Smanski, M.J. *et al.* Functional optimization of gene clusters by combinatorial design and assembly. *Nat. Biotechnol.* **32**, 1241–1249 (2014).
45. Ling, H., Teo, W., Chen, B., Leong, S.S. & Chang, M.W. Microbial tolerance engineering toward biochemical production: from lignocellulose to products. *Curr. Opin. Biotechnol.* **29**, 99–106 (2014).
46. Utrilla, J. *et al.* Engineering and adaptive evolution of *Escherichia coli* for D-lactate fermentation reveals GatC as a xylose transporter. *Metab. Eng.* **14**, 469–476 (2012).
47. Dunlop, M.J. *et al.* Engineering microbial biofuel tolerance and export using efflux pumps. *Mol. Syst. Biol.* **7**, 487 (2011).
48. Lam, F.H., Ghaderi, A., Fink, G.R. & Stephanopoulos, G. Biofuels. Engineering alcohol tolerance in yeast. *Science* **346**, 71–75 (2014).
49. Na, D. *et al.* Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. *Nat. Biotechnol.* **31**, 170–174 (2013).
50. Yoo, S.M., Na, D. & Lee, S.Y. Design and use of synthetic regulatory small RNAs to control gene expression in *Escherichia coli*. *Nat. Protoc.* **8**, 1694–1707 (2013).
51. King, Z.A. & Feist, A.M. Optimal cofactor swapping can increase the theoretical yield for chemical production in *Escherichia coli* and *Saccharomyces cerevisiae*. *Metab. Eng.* **24**, 117–128 (2014).

52. Rellos, P., Ma, J. & Scopes, R.K. Alteration of substrate specificity of *Zymomonas mobilis* alcohol dehydrogenase-2 using *in vitro* random mutagenesis. *Protein Expr. Purif.* **9**, 83–90 (1997).
53. Kim, J. & Copley, S.D. Inhibitory cross-talk upon introduction of a new metabolic pathway into an existing metabolic network. *Proc. Natl. Acad. Sci. USA* **109**, E2856–E2864 (2012).
54. Jantama, K. *et al.* Eliminating side products and increasing succinate yields in engineered strains of *Escherichia coli* C. *Biotechnol. Bioeng.* **101**, 881–893 (2008).
55. Kim, T.Y., Park, J.M., Kim, H.U., Cho, K.M. & Lee, S.Y. Design of homo-organic acid producing strains using multi-objective optimization. *Metab. Eng.* **28**, 63–73 (2015).
56. Lewis, N.E., Nagarajan, H. & Palsson, B.O. Constraining the metabolic genotype-phenotype relationship using a phylogeny of *in silico* methods. *Nat. Rev. Microbiol.* **10**, 291–305 (2012).
57. Wang, H.H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).
58. Wang, H.H. *et al.* Genome-scale promoter engineering by coselection MAGE. *Nat. Methods* **9**, 591–593 (2012).
59. Warner, J.R., Reeder, P.J., Karimpour-Fard, A., Woodruff, L.B. & Gill, R.T. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat. Biotechnol.* **28**, 856–862 (2010).
60. Hollinshead, W., He, L. & Tang, Y.J. Biofuel production: an odyssey from metabolic engineering to fermentation scale-up. *Front. Microbiol.* **5**, 344 (2014).
61. Dahl, R.H. *et al.* Engineering dynamic pathway regulation using stress-response promoters. *Nat. Biotechnol.* **31**, 1039–1046 (2013).
62. Galanie, S., Thodey, K., Trenchard, I.J., Filsinger, I.M. & Smolke, C.D. Complete biosynthesis of opioids in yeast. *Science* **349**, 1095–1100 (2015).
63. Karr, J.R. *et al.* A whole-cell computational model predicts phenotype from genotype. *Cell* **150**, 389–401 (2012).
64. Hsiao, T.Y., Glatz, C.E. & Glatz, B.A. Broth recycle in a yeast fermentation. *Biotechnol. Bioeng.* **44**, 1228–1234 (1994).
65. Yin, J., Fu, X.Z., Wu, Q., Chen, J.C. & Chen, G.Q. Development of an enhanced chromosomal expression system based on porin synthesis operon for halophile Halomonas sp. *Appl. Microbiol. Biotechnol.* **98**, 8987–8997 (2014).
66. Caspeta, L. *et al.* Biofuels. Altered sterol composition renders yeast thermotolerant. *Science* **346**, 75–78 (2014).
67. Bonneau, R. *et al.* A predictive model for transcriptional control of physiology in a free living cell. *Cell* **131**, 1354–1365 (2007).
68. Xia, J., Yamaji, N. & Ma, J.F. An appropriate concentration of arginine is required for normal root growth in rice. *Plant Signal. Behav.* **9**, e28717 (2014).
69. Burgard, A.P., Pharkya, P. & Maranas, C.D. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* **84**, 647–657 (2003).
70. Antoniewicz, M.R. ^{13}C metabolic flux analysis: optimal design of isotopic labeling experiments. *Curr. Opin. Biotechnol.* **24**, 1116–1121 (2013).
71. Trinh, C.T., Wlaschin, A. & Srienc, F. Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Appl. Microbiol. Biotechnol.* **81**, 813–826 (2009).
72. Kim, H.U., Kim, T.Y. & Lee, S.Y. Metabolic flux analysis and metabolic engineering of microorganisms. *Mol. Biosyst.* **4**, 113–120 (2008).