REVIEW ARTICLE
PUBLISHED ONLINE: 17 MAY 2012 | DOI: 10.1038/NCHEMBIO.970
nature chemical biology

Systems metabolic engineering of microorganisms for natural and non-natural chemicals

Jeong Wook Lee[1,2], Dokyun Na[1,3], Jong Myoung Park[1,3], Joungmin Lee[1], Sol Choi[1] & Sang Yup Lee[1–3]*

Growing concerns over limited fossil resources and associated environmental problems are motivating the development of sustainable processes for the production of chemicals, fuels and materials from renewable resources. Metabolic engineering is a key enabling technology for transforming microorganisms into efficient cell factories for these compounds. Systems metabolic engineering, which incorporates the concepts and techniques of systems biology, synthetic biology and evolutionary engineering at the systems level, offers a conceptual and technological framework to speed the creation of new metabolic enzymes and pathways or the modification of existing pathways for the optimal production of desired products. Here we discuss the general strategies of systems metabolic engineering and examples of its application and offer insights as to when and how each of the different strategies should be used. Finally, we highlight the limitations and challenges to be overcome for the systems metabolic engineering of microorganisms at more advanced levels.

Microorganisms can be used for the production of diverse chemicals and materials from renewable resources. However, microbial metabolism has unfortunately not evolved to suit the practical outcomes desired by humankind; thus, when microbes are isolated from nature, their efficiency in producing any given molecule is rather low. Metabolic engineering[1,2] is used to convert these reluctant biofactories into highly efficient, focused machines capable of generating huge quantities of a molecule of interest. The recent development of high-throughput techniques for deciphering genomes, transcriptomes, proteomes, metabolomes and fluxomes, together with computational tools, have changed the landscape of what is possible in the field, with systems metabolic engineering emerging as a conceptual framework to encompass the spectrum of strategies used in metabolic engineering. Systems metabolic engineering can thus be defined as systems-level metabolic engineering integrating the 'omic' and computational techniques of systems biology, the fine design capabilities of synthetic biology and the rational and random mutagenesis methods of evolutionary engineering. Using systems metabolic engineering, scientists synthesize both information and biological activities to create new metabolic products and pathways, cellular regulatory circuits and functions[3]. Through systems metabolic engineering, a variety of microbial cell factories are being developed to efficiently manufacture both natural and non-natural chemicals and materials[1,2].

The overall trajectory of any particular metabolic engineering project varies according to the identity of the target molecule. Chemicals can be broadly classified into four categories on the basis of whether they have thus far been found or reported to exist in nature (natural versus non-natural) and whether they can be produced by inherent pathways of microorganisms (inherent, noninherent or created; **Fig. 1**). Natural-inherent chemicals are endogenous metabolites in naturally isolated microorganisms and thus can be produced inherently through a native pathway. Natural-noninherent chemicals are those that are found in nature but are best produced in a heterologous host strain using noninherent pathways introduced from other hosts or metagenomes. Non-natural–noninherent chemicals are those that have not yet been found in nature but can be produced in a noninherent host strain by the establishment of heterologous pathways and enzymes, often using genes found from various sources in combination. Non-natural–created chemicals are those that have not yet been found in nature and, owing to the lack of any known metabolic enzymes and pathways leading to their formation, can only be produced by creating synthetic enzymes and pathways with new functions. These definitions are thus somewhat fluid depending on available information; for example, chemicals currently classified as non-natural might be reclassified as natural if they are discovered in nature; similarly, created pathways might be reclassified as inherent or noninherent if corresponding biosynthetic enzymes and metabolic pathways are discovered.

In their efforts to obtain these different categories of molecules, metabolic engineers consider not only the efficiency of a potential metabolic pathway but also the most efficient means to construct it. For example, natural-inherent chemicals can often be overproduced by directly modifying the host strain to optimize the fluxes of native pathways at the systems level. As a result, their engineering may rely on more intuitive approaches that use traditional metabolic and bioprocess engineering strategies to cope with well-defined and widely known problems. However, cells are finely tuned to avoid wasteful accumulation of chemicals, and so efforts to produce some molecules may require application of systems biology, synthetic biology and evolutionary engineering—a strategy for tuning the expression of multiple genes and adapting cellular physiology simultaneously and autonomously by mimicking the natural selection process—to tackle more difficult problems that result from the complex nature of metabolic and cellular regulatory networks. In contrast, the production of natural-noninherent, non-natural–noninherent and non-natural–created chemicals that are not synthesized via native pathways begins with the design of appropriate metabolic pathways, constructed via heterologous and/or combinatorial expression of known genes or creation of new genes. Once the synthetic pathways are successfully established, additional approaches can be taken to

[1]Department of Chemical and Biomolecular Engineering–Brain Korea 21 program, Center for Systems and Synthetic Biotechnology, Institute for the BioCentury, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea. [2]BioProcess Engineering Research Center, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea. [3]Bioinformatics Research Center, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea. *e-mail: leesy@kaist.ac.kr

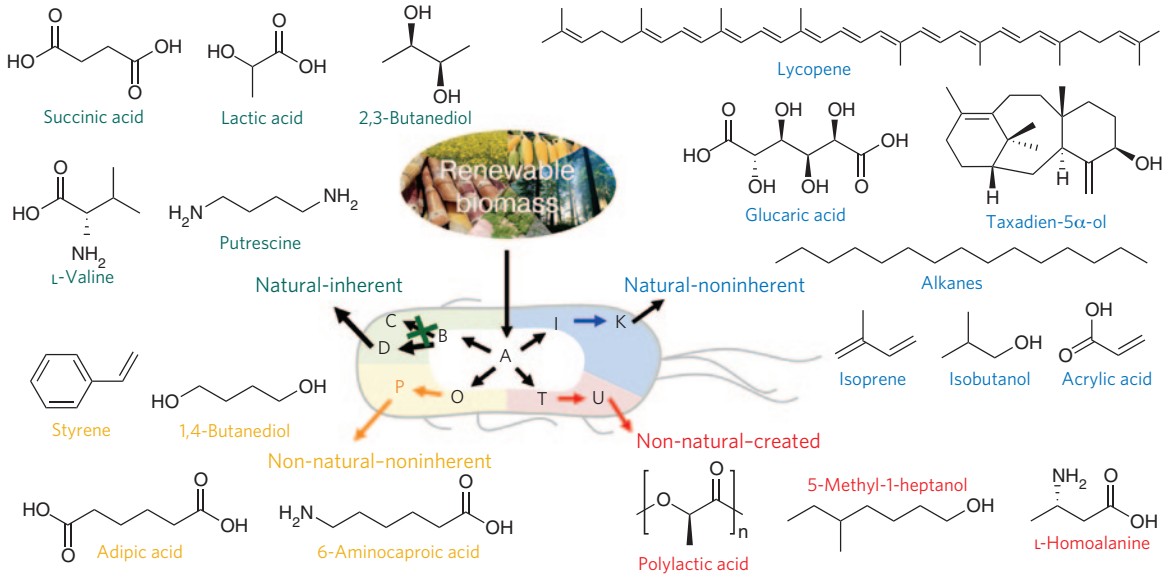

Figure 1 | Categories of chemicals produced by microbial cell factories. Inherent metabolites and pathways are indicated in black, and noninherent metabolites and pathways are represented by different colors according to the category: natural-inherent, green; natural-noninherent, blue; non-natural–noninherent, orange; non-natural–created, red. A, carbon source; B, I, O and T, metabolic intermediates; C, byproduct; D, K, P and U, target products.

further engineer the host strain for the enhanced production of desired chemicals.

In this review, we focus our attention on the techniques and strategies that, in combination, constitute systems metabolic engineering. We use recent examples of engineering successes to exemplify these ideas and provide guidance as to when each of the methods should be used and what limitations should be considered. Finally, we offer a perspective on the challenges facing the field and the potential of systems metabolic engineering as an essential enabling strategy for the successful establishment of biorefineries.

Construction of synthetic metabolic pathways

As discussed above, many target chemicals cannot be synthesized via native metabolic pathways; as a result, innovative strategies and tools are required to build synthetic pathways leading to the efficient formation of targeted chemicals. Some engineering projects use known enzymes to catalyze their canonical metabolic reactions but with improved performance in their respective host strains. In other cases, the numerous enzymes and pathways in nature are used as a diverse trove of genes for reconstructing a synthetic pathway, a resource that is made even more powerful by the accumulation of genomic and metagenomic sequencing data and advances in inexpensive gene-synthesis technology. Computational algorithms based on the accumulated genetic, genomic and enzymatic information can help design the most efficient metabolic routes using enzymes originating from diverse organisms and metagenomes. Directed evolution and rational protein design based on protein structure information can also contribute to the creation of new enzymes with new catalytic functions. Thus, the diversity of chemicals that can be manufactured by metabolic engineering is virtually limitless and is indeed increasing rapidly. The principles and tools for pathway prediction and design have recently been comprehensively reviewed[4]. Instead of iterating these, the pathways constructed using rational and computational strategies are reviewed here.

***De novo* pathway design.** The first step in reconstructing synthetic metabolic pathways for natural-noninherent and non-natural–noninherent chemicals is the design of optimal pathways leading to their formation. Next, the best candidate enzymes originating from diverse organisms or metagenomes can be heterologously and/or combinatorially introduced to establish a new metabolic pathway. Sometimes, filling in the gaps between disconnected metabolic reactions in the production host is all that is needed to establish a continuous pathway leading to the formation of the desired product. Such simple 'gap filling' with heterologous enzymes has been successfully applied to the development of strains for the production of various natural and non-natural chemicals.

A good example of a natural-noninherent chemical produced through metabolic engineering is fatty acid ethyl ester (FAEE). FAEE, an alternative diesel fuel, was produced from hemicellulosic biomass in an engineered *Escherichia coli* strain by combining genes from various organisms, including both plants and bacteria[5]. The heterologous enzymes, wax ester synthase from *Acinetobacter baylyi* and thioesterases from plants, as well as pyruvate decarboxylase and alcohol dehydrogenase from an ethanol producer, were introduced for free fatty acid production and for ethanol production, respectively. The combined introduction of these two systems yielded an *E. coli* strain that produced 674 mg l^{-1} of FAEE from a renewable resource. Another example is isoprene, which was intended for use as a jet fuel as well as a platform chemical. Isoprene production was obtained by introducing heterologous *ispS* genes encoding isoprene synthase from *Populus nigra* and *Pueraria montana* into *E. coli* and *Synechocystis* sp. PCC6803, respectively[6,7]. Similarly, various natural alcohols, including isobutanol, 2-methyl-1-butanol, 3-methyl-1-butanol and 2-phenylethanol, could be produced in *E. coli* by introducing metabolic reactions catalyzed by 2-ketoacid decarboxylase from *Lactococcus lactis* and alcohol dehydrogenase from *Saccharomyces cerevisiae*, which convert 2-ketoacids to alcohols (**Fig. 2a**)[8].

There has recently been a report describing the biological production of styrene, a non-natural–noninherent chemical that is a monomer of industrially important polymers, by establishing a synthetic pathway composed of heterologous genes of eukaryotic origin in *E. coli* (**Fig. 2a**)[9]. In this study, a pathway composed of phenylalanine:ammonia lyase from *Arabidopsis thaliana* and cinnamate decarboxylase from *S. cerevisiae* was introduced into *E. coli*, generating an engineered strain that was able to sequentially convert phenylalanine, endogenously produced from glucose, to styrene, yielding a styrene titer of 260 mg l^{-1}.

***In silico* pathway prediction.** What can scientists do when obvious gap-filling strategies are not available? Several pathway prediction

tools that inspect all of the possible routes to a target chemical have been developed for the accurate prediction of synthetic pathways. These tools use network-based pathway analysis, considering pathway distance from substrate to product, and take into account substrate specificities and binding sites, reaction mechanisms, structural changes in substrate-product pairs and/or thermodynamic favorability[10,11]. Using these tools, it is possible to design multistep synthetic pathways for the biosynthesis of non-natural chemicals. Prediction methods can be broadly classified into two types: chemical structure–based and knowledge-based ones.

Chemical structure–based methods are applicable to reconstructing multistep metabolic pathways on the basis of changes in chemical structures from the substrate to the product[10,12–15]. Various plausible pathways that can be predicted by the chemical structure–based method are not restricted by the set of known enzymes, providing several possibilities for the production of the target compound. For instance, an algorithm based on transformation of functional groups by known specific chemistry was used to design *E. coli* capable of producing 1,4-butanediol (1,4-BDO), a non-natural chemical of industrial importance for the manufacture of polyesters and spandex fibers. It predicted more than 10,000 possible routes of four to six steps starting from central metabolites such as acetyl-CoA, α-ketoglutarate, glutamate and succinyl-CoA[10]. Two synthetic pathways for 1,4-BDO production were selected by evaluating the maximum theoretical 1,4-BDO yield, pathway distance, thermodynamic feasibility and number of non-native and new steps. The construction of these synthetic pathways in *E. coli* led to the first biological production of 18 g l^{-1} of 1,4-BDO from glucose.

In contrast to chemical structure–based methods, knowledge-based methods predict metabolic pathways on the basis of experimentally identified information about reactions and pathways deposited in several different databases[16–18]. Knowledge-based methods use network analysis of elementary flux modes, extreme pathways and the shortest pathways to reconstruct the most efficient pathway for metabolic engineering of a host organism[19]. Once a target product is chosen, all of the plausible pathways are constructed considering reaction rules that are often based on the enzyme classification system. For example, the BNICE framework was used to reconstruct a synthetic pathway for the production of 3-hydroxypropanoate from pyruvate in *E. coli*[16]. Knowledge-based methods can also be incorporated into *in silico* metabolic models to design optimal metabolic pathways and identify genetic knockout candidates[20]. OptStrain was used to design optimal metabolic pathways for the production of hydrogen and vanillin by deleting existing reactions and inserting a minimal number of non-native reactions[21]. These achievements in *in silico* pathway prediction have opened a new avenue for metabolic engineering distinct from traditional *ad hoc* genetic engineering and hold great promise in the systematic reconstruction of synthetic pathways for the efficient production of non-natural chemicals.

Enzyme engineering and creation for synthetic pathways. What if pathway prediction methods do not return a reasonable route to the desired molecule? If enzymes involved in the pathway leading to a non-natural chemical are not known, one available approach is to create new enzymes with desired functions. A common approach is for new enzymes to be developed by altering the substrate specificities of the most plausible enzymes (that is, those that use structurally similar substrates to the selected compound or perform chemistry similar to the desired reaction) through mutagenesis and directed evolution.

For example, an *E. coli* strain has been engineered to produce a non-natural biodegradable polymer, polylactic acid (PLA)[22]. Following the screening of the most plausible enzymes that can convert lactate to PLA, propionate CoA transferase from *Clostridium propionicum*, a candidate for converting lactate to lactyl-CoA,

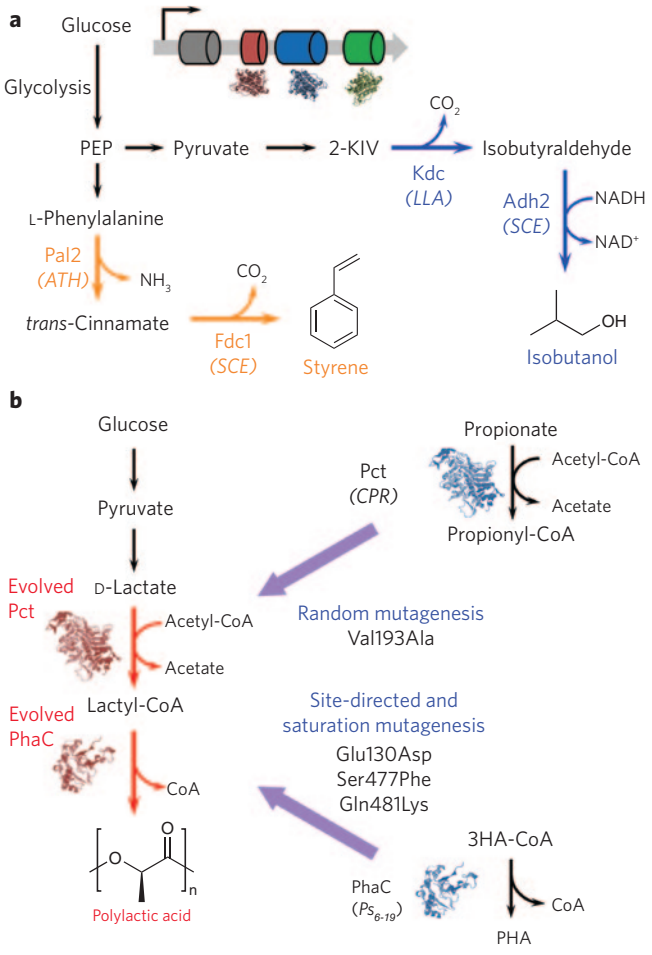

Figure 2 | Construction of synthetic metabolic pathways in a noninherent host strain. (**a**) Synthetic pathways can be constructed by overexpression of heterologous enzymes. Pathways and enzymes for the production of the natural-noninherent compound isobutanol[8] and the non-natural–noninherent compound styrene[9] are depicted. (**b**) Non-natural–created compounds can be produced through the development of new pathways involving evolved (created) enzymes having desired activities. An example shown is the generation of metabolic pathway for polylactic acid production in *E. coli*[22,81]. The modified residues in the evolved enzymes are also shown. Colors of the target chemicals follow those in **Figure 1**. *ATH*, *A. thaliana*; *CPR*, *C. propionicum*; *LLA*, *L. lactis*; $Ps_{6\text{-}19}$; *Pseudomonas* sp. MBEL 6-19; *SCE*, *S. cerevisiae*. Adh2, alcohol dehydrogenase 2; Fdc1; ferulate decarboxylase; Kdc, 2-ketoacid decarboxylase; Pal2, phenylalanine:ammonia lyase; Pct, propionate CoA transferase; PhaC, polyhydroxyalkanoate synthase. 3HA-CoA, 3-hydroxyacyl-CoA; 2-KIV, 2-ketoisovalerate; PHA, polyhydroxyalkanoate.

and PHA synthase from *Pseudomonas* sp. MBEL 6-19, a candidate for polymerizing lactyl-CoA into the growing chain of PLA, were selected as template enzymes for creating a new pathway. Introduction of the heavily evolved versions of these two enzymes allowed one-step fermentative production of PLA in *E. coli* (**Fig. 2b**). This example demonstrates that combining metabolic engineering with evolutionary protein engineering to create a pathway of previously uncharacterized function can diversify the range of both natural and non-natural chemicals, fuels and materials that can be produced.

Although some flexibility of substrate specificity enables natural enzymes to catalyze limited non-natural reactions, this strategy might not be applicable in all cases. To overcome this problem, a *de novo* design strategy, which begins with the identification of

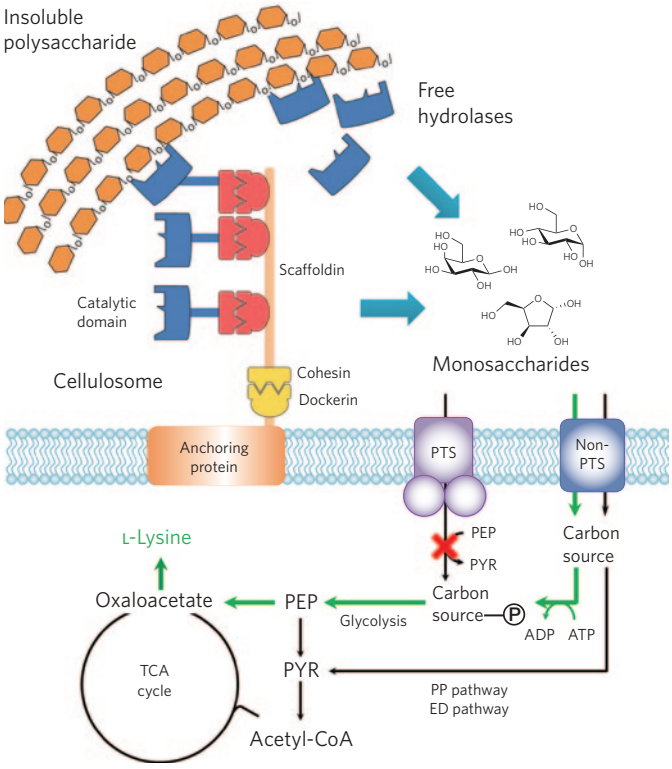

Figure 3 | Strategies for substrate utilization engineering. Complex polysaccharides that cannot be otherwise used by the wild-type strain can be depolymerized by the expression of heterologous enzymes, either in the form of an enzyme complex such as cellulosome, which is composed of catalytic domains, scaffoldin and dockerin for cellulose degradation, or in the form of noncomplex hydrolases. Monosaccharides released by depolymerization can be transported into the host cells via endogenous or heterologous transporters. Some carbon sources can be imported by either PTS or non-PTS permeases. Because PEP is used as a phosphate donor for PTS-mediated transport reactions, the use of non-PTS permeases can be advantageous for the production of some target products such as L-lysine[30], which requires an increased PEP pool (green arrows). ED, Entner-Doudoroff; PP, pentose phosphate; PYR, pyruvate; TCA, tricarboxylic acid.

the minimal active site, comprising transition states between substrates and potential catalytic residues, and then moves on to the consideration of the remainder of the enzyme as a scaffold, has also been developed[23]. Using this approach, several new enzymes catalyzing non-natural reactions, including Kemp elimination, retro-aldol reaction and the Diels-Alder reaction, were synthetically designed[24–26]. The recently developed Rosetta *de novo* design protocol simplifies the design process and largely overcomes difficulties encountered in using computational design tools to generate new catalytic functions, making it a powerful means for designing microorganisms for the biological production of non-natural chemicals[27].

Systems metabolic engineering strategies

Once a host strain is capable of producing a desired chemical, the next step is to enhance production of the target product through iterative and combinatorial applications of systems metabolic engineering strategies. These strategies can be grouped into rational-intuitive and systematic–rational-random approaches. Rational and intuitive approaches can be applied to well-defined problems for which solutions have already been well established or can be found intuitively. In contrast, systematic–rational-random approaches are undertaken when the solution is not intuitively obvious. **Table 1** provides details of how these approaches may be put into practice and considerations for their use; additional guidelines for prioritizing strategies can be found in ref. 3.

Rational and intuitive approaches in metabolic engineering

Rational and intuitive approaches are not systems-level strategies by themselves but are often used in combination to pursue multiple objectives at once. As a result, there is an advantage to using them with a systems-level perspective to (i) use a different carbon source or enhance its use; (ii) modify transporters to efficiently pump out the final product, preventing it from accumulating in the intracellular space; (iii) eliminate or reduce byproduct synthesis and enhance the target pathway flux; and (iv) reroute pathways for efficient precursor conversion to a desired target.

Engineering of carbon source utilization. Much effort has been exerted to engineer carbon substrate utilization by focusing on enhancing its uptake and utilization efficiency (upper part of **Fig. 3**)[28,29], and there has been particular focus on bulk chemical production. Also, carbon-source utilization routes are often reconstructed to enhance the production of target products. In cases where phosphoenolpyruvate (PEP) is a metabolic intermediate for the desired target product, a non-phosphotransferase system (PTS) is often more suitable than a PTS because dissipation of the precursor, PEP, can be minimized. For example, replacing the native PTS for glucose utilization of *E. coli* with the non-PTS of *Corynebacterium glutamicum* increased L-lysine yield by up to 20% (bottom part of **Fig. 3**)[30].

Transporter engineering. In most cases, target chemicals generated in the cytoplasmic space need to be excreted into the extracellular space through corresponding transporters (**Fig. 4**). This excretion serves to minimize the intracellular concentration of the target chemical, thus avoiding feedback inhibition and growth inhibition, and ultimately maximizing production of the target chemical[31–34]. At the same time, import of excreted target products should be prevented[31–34]. Recently, a new transporter engineering approach was implemented to improve tolerance against toxic target chemicals[35]. A large number of pump candidates were subjected to a competition-based selection assay in which cells with adequate efflux pump activity survived in the presence of the target biofuel and became enriched. A limonene-producing *E. coli* strain containing a new efflux pump identified by this strategy yielded twice the amount of limonene produced by the parent strain[35].

Byproduct elimination and precursor enrichment. One of the most commonly used approaches in metabolic engineering is the removal of competing pathways by gene deletion and enhancement of the targeted pathway by gene amplification to improve yield, titer and specific or overall productivity of a target product by minimizing byproduct formation and maximizing precursor concentrations. There are numerous successful examples of this simple approach, which are not further detailed here. Recently, monomers for nylons, putrescine and cadaverine were produced biologically using this conventional metabolic engineering approach. By intuitively eliminating byproducts and increasing precursor fluxes at the systems level, the engineered *E. coli* strains produced 24.2 g l^{-1} of putrescine and 9.61 g l^{-1} of cadaverine (**Fig. 4**)[33,34].

Rerouting metabolic pathways. What if the originally designed metabolic pathway leading to the formation of a desired product is not optimal because of the lower-than-expected pathway flux? The complexity of metabolic networks provides alternative routes that are often more efficient than the well-known pathways for the biosynthesis of the target product. For example, *n*-butanol can be more efficiently produced in *E. coli* by using the *Treponema denticola trans*-2-enoyl CoA reductase instead of the well-known

Table 1 | Metabolic engineering strategies for the production of natural and non-natural chemicals.

Strategy	How to use	Possible limitations	Ref.
Substrate utilization engineering	Introduction of an alternative transport system (for example, non-PTS permease instead of PTS)	Systems-level optimization of cellular demands for ATP, PEP and pyruvate needed	30
	Introduction of heterologous transporters or associated enzymes into a host that cannot use a desired inexpensive carbon source	Heterologous expression can be difficult; membrane transporters also difficult to functionally overexpress and manipulate	83
	Co-culture of several host strains that can utilize different or, in combination, complex carbon sources	Balancing the growth of multiple strains is not easy	28
	Modification of regulatory mechanisms such as removing catabolite repression to simultaneously utilize multiple carbon sources	The rates of using multiple carbon sources might be different and thus need to be optimized	84
Transporter engineering	Overexpression of product exporters, often combined with removal of product importers	Potential for unexpected phenotypes due to membrane crowding or unwanted interactions with endogenous membrane proteins	31–34
	High-throughput transporter screening	Considerable effort with no guarantee of finding efficient transporter	35
Byproduct elimination and precursor enrichment	Elimination of the competing and/or degrading pathways by gene knockout	Can increase cellular stresses or lead to severe growth retardation in some cases	31–34
	Attenuation of competing pathways by replacing promoter or start codon; preferable to gene knockout when the competing pathway is needed for cell growth or production of target	Attenuation might not be significant enough to increase the flux toward product	85
	Overexpression of enzymes to increase the precursor availability	Excess enzymes can deplete cellular resources required for cell growth; imbalanced expression can lead to accumulation of toxic intermediates	31–34
Rerouting metabolic pathways	Search for more efficient enzymes, often via combinatorial heterologous expression of enzymes from various sources	Heterologous enzymes can lack regulatory mechanisms, require different cofactors or show different activities in the strain to be engineered	36,37
	Construction of new biosynthetic pathways to use different substrates	Efficiency of newly constructed biosynthetic pathway needs to be increased	86
Cofactor optimization	Deletion or attenuation of glycolytic enzymes redirects metabolic flux through the pentose phosphate pathway to increase NADPH pool	Redirection of central metabolism might cause unwanted phenotypes such as growth retardation	38–40
	Replacement of NADH-dependent central metabolic enzymes with NADPH-dependent ones	NADPH-dependent enzymes for the desired reaction might not exist in nature	41
	Interconversion of NADH and NADPH by the introduction of transhydrogenase	Transhydrogenases such as PntAB in *E. coli* require energy for their reactions; transhydrogenases may affect proton-dependent transport reactions	42
	Enzyme evolution to alter the cofactor specificity	Considerable effort with no guarantee of finding appropriate enzymes	42–44
Simulation and omics-based identification of targets	Identification of additional gene knockout targets by *in silico* metabolic simulation (for example, elementary mode analysis, MOMA, OptKnock, OptForce and GDLS)	Genome-scale metabolic models are unable to capture the dynamic features of cellular metabolism (i.e., regulatory mechanisms and *in vivo* kinetic parameters); computational burden	20,31,32,49, 63–66,68,69
	Identification of additional gene amplification targets by *in silico* metabolic simulation (for example, OptForce, FSEOF and EMILiO)	Algorithms do not yet account for nonlinear relationships among levels of mRNA, protein, activity and flux	65–68
	In silico flux response analysis to examine the effects of particular fluxes on cell growth or production of a target	Imperfect nature of genome-scale metabolic network might result in unrealistic solution	31
	Revealing rate-limiting steps through metabolite profiling	Techniques to profile metabolome still need improvement	49
	Unraveling unknown mechanisms by analyzing omics data that can be manipulated to enhance target production	Algorithms to analyze multi-omics data in an integrated manner need to be developed	32,87
Optimization of metabolic pathways	Determination of intermediate metabolite levels by metabolite profiling	Techniques to profile metabolome still need improvement	71
	Engineering of promoter strength and gene or plasmid copy number	Optimal promoter strength and plasmid copy number must be found via trial and error due to nonlinear relationships among the levels of mRNA, protein, activity and flux	72
	Optimization of untranslated regions to achieve fine tuning of protein expression	Imperfect predictions of RNA structures limit rational design	50,51,88
	Use of synthetic scaffolds to mimic substrate channeling	Substrate trafficking may not be a rate-limiting step	76,77

(continued)

Table 1 | Metabolic engineering strategies for the production of natural and non-natural chemicals (continued).

Enzyme evolution for activity optimization	Random mutagenesis when not much information is available	Easy and inexpensive selection method is needed	22,81
	Site-directed mutagenesis and saturation mutagenesis when information on the structure-function relationship of the enzyme is available	Much effort and time needed to test all of the possible combinatorial mutations	54,89
	Codon optimization for efficient expression in heterologous hosts	Codon optimization does not necessarily improve expression	54
Metabolic evolution	Serial transfer or continuous fermentation of a rationally engineered strain to achieve desired phenotypes	Mainly useful for growth-associated production of a target product	53,78,79,90
Adaptive evolution, resequencing and reengineering	Develop a stress-tolerant strain by adaptive evolution. Compare genome, transcriptome and/or proteome analyses between mutant and parental strains to identify mutations. Reintroduce identified mutations to the parental strain to achieve desired phenotypes	Experiments require relatively long time; identification of true beneficial mutations after resequencing can be difficult	48,91,92
Multiplexed genome engineering	Introduction of a single-stranded oligonucleotide library into the mutant for high-throughput targeted mutations either iteratively (MAGE) or combined with high-throughput identification system (TRMR)	Genome sequence should be known; mainly useful for the organisms in which single-strand recombination mechanisms are well known	55,80

Clostridium acetobutylicum butyryl-CoA dehydrogenase (Bcd)[36] (**Fig. 5a**). Another notable example is the production of various fatty acids and alcohols using engineered *E. coli* strains[37]. Gonzalez and his colleagues manipulated several transcriptional regulators in *E. coli* to turn on the β-oxidation pathway in the presence of glucose (**Fig. 5b**)[37]. Then, fatty alcohols and acids with various chain lengths were produced by introducing the appropriate initiation enzyme to catalyze the Claisen condensation of acyl-CoA and the termination enzyme required for the removal of the CoA moiety from acyl-CoA. Through this new reverse β-oxidation pathway, 14 g l^{-1} of *n*-butanol, a second-generation biofuel, could be produced with a yield of 0.33 g per gram of glucose, a result comparable to that achieved using a traditional *n*-butanol producer, *C. acetobutylicum*.

Cofactor optimization. Cofactors are often required for biochemical transformations mediated by metabolic enzymes. It is important to consider the availability of necessary cofactors as well as the balance of electron-mediating organic cofactors, such as NADH and NADPH, for the production of many chemicals. These aspects can be controlled by engineering global metabolic characteristics toward the generation of a cofactor that is highly required for the production of the desired chemicals or, alternatively, by changing cofactor preferences of relevant key enzymes. The former approach includes redirection of metabolic flux through the pentose phosphate pathway[32,38–40], replacement of central metabolic enzymes that prefer other cofactors[41] and cross-conversion of NADH and NADPH. The latter approach is often implemented when the alteration of global metabolic characteristics is not easily applicable. In most cases, this approach involves screening of various heterologous enzymes[42] or application of enzyme evolution through site-directed or random mutagenesis of a target enzyme[43,44] to change cofactor specificity.

Systematic and rational-random approaches

The rational-intuitive metabolic engineering approaches described above are not sufficient to truly optimize the overall performance of the cell owing to the extreme complexity of cellular networks. Thus, system-wide analyses, such as *in silico*– and omics-based techniques (**Fig. 6a**,**b**), have been developed for and applied to various

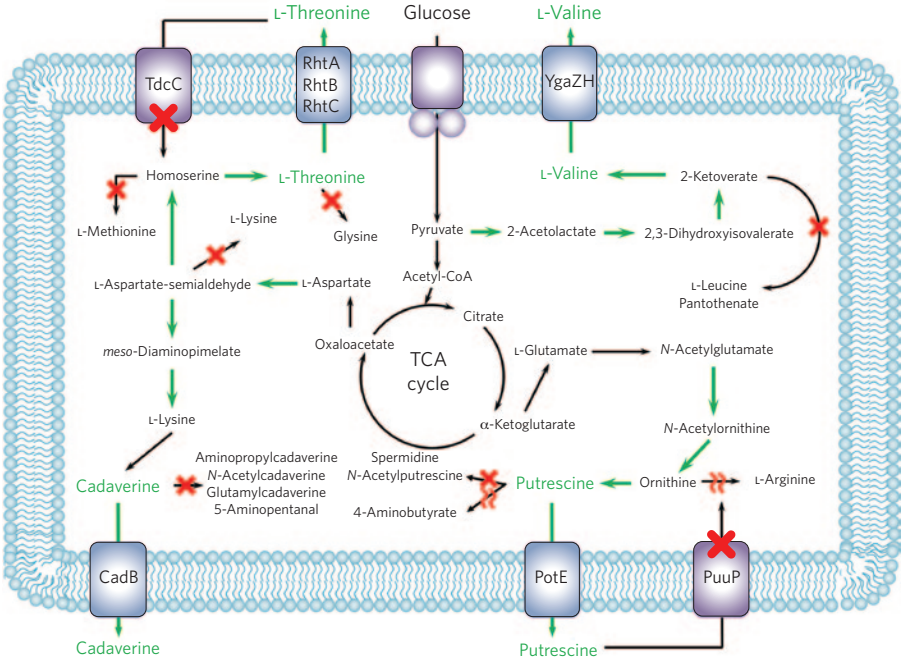

Figure 4 | Byproduct elimination, precursor enrichment and transporter engineering. These strategies can be applied for the enhanced production of L-threonine[31] (top left), L-valine[32] (top right), cadaverine[34] (bottom left) and putrescine[33] (bottom right). Strains are developed by deleting the competing pathways, target (or precursor) degradation pathways and product importers while amplifying the production pathway fluxes and product exporters. The red crosses indicate knocking out the corresponding reactions or importers, and red wavy lines indicate the reduced formation of metabolites by knocking out the major gene and leaving the minor one. The green arrows indicate increased fluxes by overexpression of the corresponding genes. CadB, a cadaverine-lysine antiporter; PotE, a putrescine-ornithine antiporter; PuuP, a putrescine importer; RhtA, RhtB and RhtC, L-threonine exporters; TdcC, an L-threonine importer; YgaZH, an L-valine exporter.

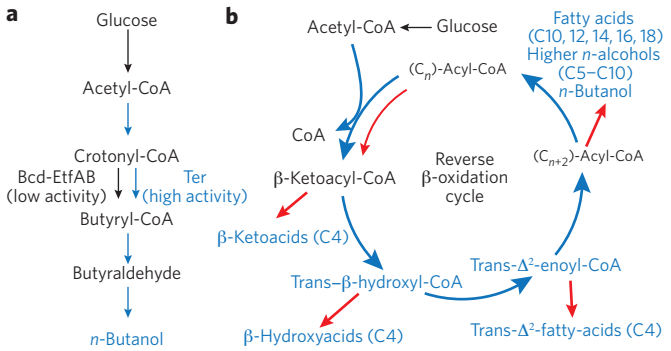

Figure 5 | Rerouting of metabolic pathways. (**a**,**b**) The rerouting of metabolic pathways for the production of *n*-butanol[36] (**a**) and higher alcohols and fatty acids[37] (**b**). As shown in **a**, *n*-butanol production could be increased by the introduction of *trans*-2-enoyl CoA reductase (Ter) instead of Bcd-EtfAB, which shows low activity in *E. coli*. Blue arrows indicate overexpression of the corresponding genes. Bcd-EtfAB, butyryl-CoA dehydrogenase combined with electron transfer flavoprotein. As shown in **b**, various fatty acids and alcohols can be formed depending on the initiation enzyme (for example, thiolase) and the termination enzyme (either thioesterase or alcohol dehydrogenase) in the reverse β-oxidation pathway, which become functional by the removal of catabolite repression. Chemicals produced by this pathway are shown. Blue arrows indicate rerouting of the pathway and increases in flux by the deletion of negative transcriptional regulators, and red arrows indicate overexpression of the corresponding genes. Both pathways shown in **a** and **b** were more efficient compared with the heterologous clostridial pathway in *E. coli*[82].

studies of metabolic engineering[31,32,45–49]. Synthetic regulation and computation-based optimization of metabolic pathways similarly represent valuable options for finely tuning metabolic fluxes at desired levels[50–52]. These strategies are complemented by approaches that not only rely on some element of randomness but also are driven by a clear rationale and can thus be termed 'rational-random' approaches, which mimic the success of natural evolution acting over long periods of time and include metabolic evolution, adaptive evolution followed by resequencing and re-engineering, massive applications of engineered global regulators and their interacting sequences, multiplexed genome engineering and activity optimization. Here, metabolic evolution refers to an iterative process in which cellular metabolism adaptively evolves after the expression of genes is purposefully altered for the production of a target chemical, whereas adaptive evolution refers to the adjustment of an organism to its environment or the process by which it enhances such fitness. All of these approaches have been developed to efficiently overcome hidden bottlenecks for strain improvement[53–55].

***In silico*– and omics-based target gene selection.** *In silico* genome-scale metabolic models and associated simulation strategies have been developed to analyze a cell in the context of global metabolic activities and thereby identify gene manipulation targets (**Fig. 6a**)[46,56,57]. These strategies are particularly valuable for gaining insight into unexpected problems, such as slow growth and low product yield, that arise after rational engineering through systems-level interrogation of cellular networks. One of the most popular *in silico* genome-scale metabolic simulation methods is constraints-based flux analysis. It allows determination of fluxes by optimization of an objective function (for example, maximum cell growth rate or maximum product formation rate) using the mass balance equations set around intracellular metabolites using the stoichiometry of metabolic reactions under pseudo–steady state assumption.

The simulation results are becoming more accurate as new algorithms are developed. They include steady state regulatory flux balance analysis, which combines the model with the transcriptional regulation network[58]; thermodynamics-based metabolic flux analysis (TMFA), which considers the thermodynamics of reactions[59]; flux balance analysis with grouping reaction constraints (FBAwGR), based on genomic properties and flux-converging patterns[60]; probabilistic regulation of metabolism (PROM), which incorporates transcriptional regulation using a probabilistic approach[61]; and flux balance analysis with membrane economics (FBA^{ME}), which examines the membrane composition of bacterial cells[62]. Algorithms for identifying gene manipulation targets have also been developed. The multiobjective memetic algorithm (MOMA) identifies gene knockout targets by the minimization of metabolic adjustment[63], whereas the bilevel optimization method OptKnock[20] couples cell growth rate with product formation rate to identify gene knockout targets. OptKnock successfully predicted gene knockout targets for enhanced 1,4-BDO production, as described above (**Fig. 6c**). Several other algorithms that have been developed are the generalized damped least-squares (GDLS) algorithm[64], which reduces the computational burden of calculation; OptForce[65], which identifies gene manipulation targets by comparing the changes of flux variabilities; OptORF[66], which integrates transcriptional regulatory networks; flux scanning based on enforced objective flux (FSEOF)[67], which allows identification of gene amplification targets; and enhancing metabolism with iterative linear optimization (EMILiO)[68], which includes reactions with individually optimized fluxes.

There have been several successful examples of strain development using these algorithms. MOMA identified gene knockout targets for the production of cubebol, a natural sesquiterpene alcohol; the engineering of identified targets successfully enhanced cubebol production to 30.1 mg l^{-1} in *S. cerevisiae*[69]. Production of L-threonine[31] and L-valine[32] could be considerably enhanced by manipulating genes identified by such simulations. FSEOF was successfully used to enhance lycopene production in *E. coli* to 283 mg l^{-1} (ref. 67). Similarly, metabolic engineering strategies for the production of adipic acid, hexamethylenediamine and 6-aminocaproic acid were suggested by *in silico* gene knockout perturbation studies using the genome-scale *E. coli* metabolic model[70]. *In silico*–based methods can also integrate omics approaches for strain improvement, as demonstrated for lysine production in *C. glutamicum* using ^{13}C metabolic flux analysis to systematically identify bottleneck reactions for lysine production (**Fig. 6b**)[49]. These achievements demonstrate that the simulation of an *in silico* genome-scale metabolic model is an impressively effective strategy for engineering the metabolism at the systems level to overcome hidden problems in the production of targeted chemicals.

Optimization and modulation of metabolic pathways. Natural metabolic pathways are tightly regulated so as to produce required metabolites in just right amounts for cell growth. However, synthetic metabolic pathways, which are commonly constructed with heterologous enzymes, are not under such regulatory control; therefore, introduction of such pathways into the cell often leads to growth retardation and causes metabolic imbalance due to accumulation of intermediates, which are often toxic. For example, DNA microarray and metabolite profiling studies have revealed that a synthetic mevalonate-based isopentenyl pyrophosphate biosynthetic pathway leads to an imbalance in carbon flux and accumulation of the toxic intermediate 3-hydroxy-3-methylglutaryl-CoA[71].

Controlling the expression of enzymes in a synthetic pathway is one of the strategies commonly used to achieve balanced fluxes. Such an approach has been used to produce taxadien-5α-ol (a precursor of paclitaxel)[72]. In this case, flux through the metabolite was optimized to avoid the production of an inhibitory indole compound by partitioning the pathway into two modules: one comprising the methylerythritol-phosphate pathway and one comprising the terpenoid-forming pathway (**Fig. 7a**). Genes for enzymes belonging

to each pathway were reconstructed as an operon and subjected to a multivariate search for optimal operon expression. Various promoters and gene copy numbers were combined, yielding 32 different combinations encompassing a wide range of module expression levels. The pathway permutations created a nonlinear taxadiene-production landscape with a global maximum of 1.02 g l^{-1} in fed-batch fermentation.

Gene expression can be finely controlled through computational optimization of untranslated regions[50,51,73] and synthetic regulators or genetic circuits. Also, translation efficiency can be estimated on the basis of mRNA sequences and their secondary structures, which can be used for consequent translational engineering. A dynamic sensor-regulator system was applied to the production of FAEE, which increased the titer by three-fold compared with that obtained with the previously engineered strain[5,74]. RNA can act as a regulatory device as well; the model-driven design of RNA devices including ribozymes and aptazymes has recently been reported, and expression of a target gene was finely controlled by the combination of these devices[75]. These approaches ensure that gene expression is controlled as desired and foster the design of optimal gene expression systems.

Enzymes belonging to the same pathway often exist in close proximity, enabling each enzyme to efficiently transfer its product to the next enzyme as a substrate. This rapid transfer of metabolic products to other enzymes enhances the efficiency of a pathway by reducing metabolite loss through outward diffusion and hence is called substrate tunneling. Recently, a scaffold-based approach that mimics the substrate tunneling effect was developed to enhance the flux through a synthetic pathway (**Fig. 7b**)[76]. In this system, heterologous enzymes involved in a synthetic mevalonate pathway were fused to a protein motif that binds a partner protein in a scaffolding protein, thus physically placing enzymes in close proximity to each other. The optimized synthetic protein scaffold increased mevalonate production up to 77-fold. In another proof-of-concept experiment, synthetic scaffolds applied to the production of glucaric acid increased the yield of glucaric acid up to three-fold[77]. This enzyme scaffold strategy thus provides a platform for the modular control of metabolic flux.

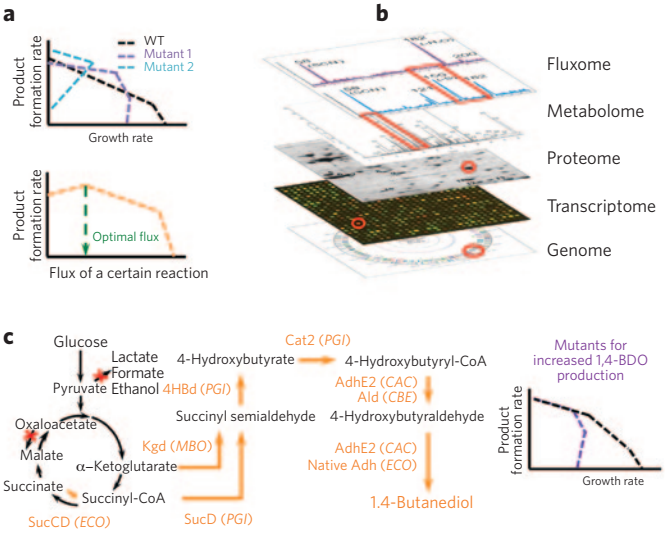

Figure 6 | *In silico*- and omics-based target gene selection and strain development based on pathway prediction and *in silico* simulation. (**a**,**b**) Identification of knockout and/or amplification target genes by *in silico* metabolic simulation (**a**) and multiple omics analyses (**b**) is depicted. As shown in **a**, many different simulations can be performed using the genome-scale metabolic model. Shown here as examples are identification of gene knockout targets by algorithms such as OptKnock (top) and determination of the optimal flux of a particular reaction that results in the highest product formation rate by flux response analysis (bottom). As shown in **b**, multi-omics analyses including genome, transcriptome, proteome, metabolome and fluxome can be performed to identify gene manipulation targets by comparative analyses. Marked in red are target genes that show significant changes under different conditions. WT, wild type. (**c**) A representative example of the development of an *E. coli* strain for the production of 1,4-BDO by systems metabolic engineering and, in particular, by using pathway prediction and *in silico* genome-scale simulation for the identification of gene manipulation targets[10]. Orange arrows indicate increasing fluxes by the overexpression of corresponding genes. Red crosses indicate the knocking out of the corresponding reactions identified by OptKnock simulation. *CAC*, *C. acetobutylicum*; *CBE*, *Clostridium beijerinckii*; *ECO*, *E. coli*; *MBO*, *Mycobacterium bovis*; *PGI*, *Porphyromonas gingivalis*. Adh, alcohol dehydrogenase; AdhE2, bifunctional alcohol–aldehyde dehydrogenase; Ald, aldehyde dehydrogenase; Cat2, 4-hydroxybutyryl-CoA transferase; 4HBd, 4-hydroxybutyrate dehydrogenase; Kgd, α-ketoglutarate decarboxylase; SucCD, succinyl-CoA synthetase; SucD, CoA-dependent succinate semialdehyde dehydrogenase.

Enzyme evolution for activity optimization. In addition to the need to find an enzyme capable of performing a desired reaction, other factors such as enzyme activity and enzyme concentration can limit the biosynthesis of target chemicals. Accordingly, optimization of enzyme activity (for example, K_m and k_{cat} values) is an important option for increasing the synthesis of a target chemical. A number of approaches have been implemented to overcome and reprogram rate-limiting or rate-controlling enzymes. Recent examples include the engineering of geranylgeranyl diphosphate synthase by codon optimization to achieve higher expression and the enzyme evolution of and replacement of amino acids at the active site of levopimaradiene synthase to improve catalytic activity[54]. Introduction of these enzymes into *E. coli* allowed increased production of levopimaradiene, a precursor of pharmaceutically important ginkgolides, by almost 2,600-fold[54].

Metabolic evolution. In cases where a suitable strategy for enhancing the production of a target chemical is absent or where simultaneous and multiplexed engineering are required, metabolic evolution can be used. By using natural selection principles, this strategy can provide insight into mechanisms of cellular adaptation and simultaneously obtain desirable phenotypes that are too complex to improve using rational approaches (**Fig. 7c**). A recent representative example of metabolic evolution is succinic acid production in *E. coli*[53,78,79]. By adopting the metabolic evolution strategy, an *E. coli* strain that produced 86.5 g l^{-1} of succinic acid could be developed from a parent strain that produced 5.8 g l^{-1}. It was later found that PEP carboxykinase had evolved to become the major carboxylating enzyme in the succinic acid–overproducing *E. coli*, and galactose permease and glucokinase were induced because of the inactivation of the glucose-specific PEP-dependent PTS. These evolved characteristics, similar to those of native succinic acid–producing rumen bacteria, suggested a new metabolic engineering approach for the production of succinic acid in *E. coli* that had not been known before applying the metabolic evolution strategy.

Adaptive evolution, resequencing and re-engineering. In some cases, a desired phenotype (such as tolerance to organic solvent) can be obtained through mimicking the natural adaptation process by exposing cells to rationally designed external conditions. The genome sequences of strains that have adaptively evolved in response to genetic or environmental perturbations can then be resequenced to identify genes responsible for the improved phenotypes. The genes identified through this reverse-engineering process can then be introduced into an engineered host strain to endow

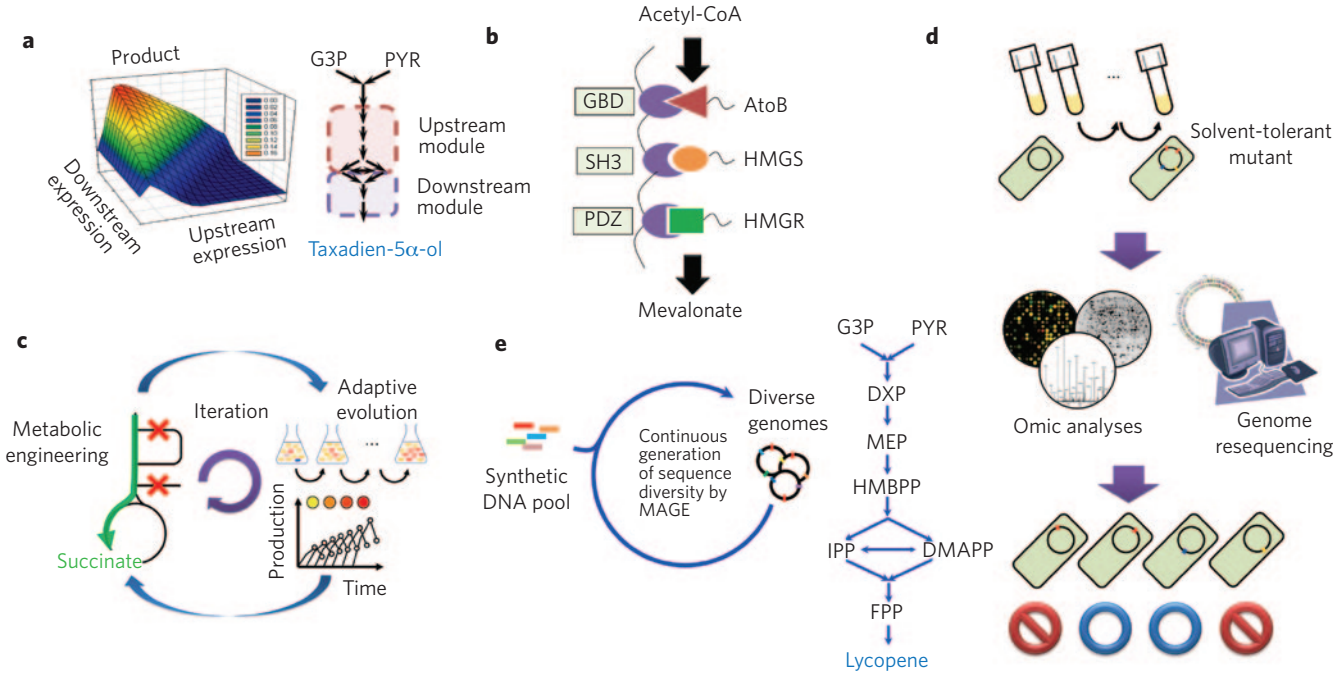

Figure 7 | Strategies for optimization and modulation of metabolic pathways and rational-random metabolic engineering. (**a**,**b**) To optimize the concentrations of metabolic intermediates that are harmful or toxic to the cells, various engineering strategies including optimization and modulation of metabolic pathways (**a**) and construction of synthetic protein scaffolds (**b**) have been successfully used. Respective examples for the production of taxadien-5α-ol[72] and mevalonate[76] are depicted. AtoB, acetoacetyl-CoA thiolase; HMGR, hydroxyl-methylglutaryl-CoA reductase; HMGS, hydroxyl-methylglutaryl-CoA synthase; G3P, glyceraldehyde 3-phosphate; GBD, GTPase binding domain; SH3, Src homology 3 domain; PDZ, post-synaptic density protein/*Drosophila* disc large tumor suppressor/zonula occludens-1 domain; PYR, pyruvate. (**c**) To obtain desirable phenotypes that are too complex to improve using rational approaches, metabolic evolution comes into play. The metabolic evolution process is composed of the iterative cycle of metabolic engineering and adaptive evolution, as demonstrated for succinate production[52,77,78]. (**d**) A mutant that has a desirable phenotype such as isobutanol tolerance[48] can be obtained by adaptive evolution, which is characterized by the use of multi-omics tools including whole-genome resequencing. The beneficial mutations identified can be reintroduced into the strain to be engineered. (**e**) In multiplexed genome engineering, such as MAGE[55], synthetic single-stranded DNA pools that contain all of the possible or desired mutations are introduced into the host strain, and numerous mutants can be simultaneously generated through homologous recombination. The desired mutant population is enriched by multiple rounds of MAGE. As a proof-of-concept experiment, an efficient lycopene-producing *E. coli* strain was developed[55]. DMAPP, dimethylallyl diphosphate; DXP, 1-deoxy-D-xylulose 5-phosphate; FPP, farnesyl pyrophosphate; HMBPP, 4-hydroxy-3-methyl-but-2-enyl pyrophosphate; IPP, isopentenyl pyrophosphate; MEP, 2-C-methyl-D-erythritol 4-phosphate.

the host cell with the desired characteristics (**Fig. 7d**). For example, isobutanol tolerance was improved in *E. coli* through adaptive evolution[48]. Genome resequencing of the isobutanol-tolerant mutant revealed one missense mutation; mutations by insertion sequence elements in 25 genes, including the *acrA*, *gatY*, *tnaA* and *yhbJ* genes; and one deletion between the *hipA* and *flxA* genes, which resulted in the loss of 62 genes including the *marCRAB* cluster. Repairing mutations in the *acrA*, *gatY*, *tnaA*, *yhbJ* genes and in the cluster comprising the *marC*, *marR*, *marA* and *marB* genes in the mutant considerably decreased isobutanol tolerance, suggesting that these mutations are key to isobutanol tolerance. Knocking out four genes and one cluster to mimic five key mutations identified by resequencing resulted in the increase of isobutanol tolerance. This example clearly demonstrates that evolutionary engineering approaches combined with systems biology produce unexpected outcomes that enhance the production of target products or the development of strains with desired phenotype.

Multiplexed genome engineering. Once engineers have collected information about possible cellular changes that would improve the overall metabolic performance of a cell, they need methods to make these changes as efficiently as possible. Multiplexed genome-engineering approaches that enable the simultaneous introduction of numerous mutations in a target organism on a genome scale have recently been developed using natural selection principles. Multiplexed automated genome engineering (MAGE), which accelerates the engineering and expression-level tuning of multiple, evolutionarily targeted genes, was developed (**Fig. 7e**)[55]. The effectiveness of this multiplexed engineering strategy was demonstrated by enhancing lycopene production by simultaneously engineering the expression of 24 target genes. The result of this genome-engineering procedure was the acquisition of a mutant strain within 3 days that showed a five-fold increase in lycopene production. A more recently developed strategy is trackable multiplex recombineering (TRMR), in which thousands of genes in a microbial genome are engineered and monitored simultaneously[80]. Using the TRMR strategy, genes affecting *E. coli* growth in the presence of various nutrients and inhibitors were identified within 1 week.

Conclusion

As these burgeoning techniques and strategies show, systems metabolic engineering is becoming an essential platform technology for the production of chemicals, fuels and materials. Integration of systems biology, synthetic biology and evolutionary engineering with traditional metabolic engineering approaches is indeed essential for developing optimized microbial cell factories.

What are the remaining challenges? As systems metabolic engineering is based on a system-wide understanding of whole-cell characteristics, a better understanding of a cell at all levels, including the metabolic, gene regulatory and signaling network levels and their interactions, needs to be pursued. The genome-scale metabolic models developed so far are by no means complete and thus do not

capture all of the cellular metabolic characteristics. Furthermore, current algorithms mainly allow analysis of cells at a steady state because of insufficient information about the dynamics of gene expression and enzyme activities governed by the regulatory network. Thus, development of more accurate genome-scale metabolic models, preferably those incorporating regulatory mechanisms, is the first priority on the to-do list of systems metabolic engineering.

Although several algorithms have been developed for designing new enzymes and pathways for the biosynthesis of non-natural chemicals, much improvement is needed in design algorithms and strategies. Although whole-genome synthesis will become easier, robust genome design principles, including metabolic and gene regulatory optimization for supporting both cell growth and product formation, need to be developed. These challenges will be overcome as systems metabolic engineering becomes further upgraded with better *in silico* genome-scale simulation algorithms, more efficient multiplex large-scale genome manipulation, synthesis of a whole artificial genome for well-defined functions, creation of artificial codons and modified ribosomes to expand the blueprint of life and very fine modulation of multiple gene transcription and translation based on the knowledge we continuously accumulate.

All of these efforts will lead to the development of engineered cells having fully reconstructed cellular genetic and metabolic networks for rapid cell growth while maintaining optimal fluxes toward the formation of a desired product, high tolerance to the product, no byproduct formation, efficient use of cost-effective carbon sources at a high yield and high overall productivity. It is expected that an increasing number of chemicals and materials will be efficiently produced from renewable resources by microorganisms developed by systems metabolic engineering.

References

1. Na, D., Kim, T.Y. & Lee, S.Y. Construction and optimization of synthetic pathways in metabolic engineering. *Curr. Opin. Microbiol.* **13**, 363–370 (2010).
2. Prather, K.L. & Martin, C.H. *De novo* biosynthetic pathways: rational design of microbial chemical factories. *Curr. Opin. Biotechnol.* **19**, 468–474 (2008).
3. 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).
4. 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).
5. Steen, E.J. *et al.* Microbial production of fatty-acid–derived fuels and chemicals from plant biomass. *Nature* **463**, 559–562 (2010).
6. Zhao, Y. *et al.* Biosynthesis of isoprene in *Escherichia coli* via methylerythritol phosphate (MEP) pathway. *Appl. Microbiol. Biotechnol.* **90**, 1915–1922 (2011).
7. Lindberg, P., Park, S. & Melis, A. Engineering a platform for photosynthetic isoprene production in cyanobacteria, using *Synechocystis* as the model organism. *Metab. Eng.* **12**, 70–79 (2010).
8. Atsumi, S., Hanai, T. & Liao, J.C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451**, 86–89 (2008).
9. McKenna, R. & Nielsen, D.R. Styrene biosynthesis from glucose by engineered *E. coli*. *Metab. Eng.* **13**, 544–554 (2011).
10. Yim, H. *et al.* Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat. Chem. Biol.* **7**, 445–452 (2011).
This publication describes a combination of the pathway prediction method—focused on chemical transformations—with other engineering strategies for the production of non-natural 1,4-BDO.
11. Cho, A., Yun, H., Park, J., Lee, S. & Park, S. Prediction of novel synthetic pathways for the production of desired chemicals. *BMC Syst. Biol.* **4**, 35 (2010).
This study reports development of a prediction tool for evaluating new synthetic pathways for the production of alcohols, 3-hydroxypropanoate and butyryl-CoA.
12. Feldman, H.J., Dumontier, M., Ling, S., Haider, N. & Hogue, C.W.C.O. A chemical ontology for identification of functional groups and semantic comparison of small molecules. *FEBS Lett.* **579**, 4685–4691 (2005).
13. McShan, D.C., Rao, S. & Shah, I. PathMiner: predicting metabolic pathways by heuristic search. *Bioinformatics* **19**, 1692–1698 (2003).
14. McShan, D.C. & Shah, I. Heuristic search for metabolic engineering: *de novo* synthesis of vanillin. *Comput. Chem. Eng.* **29**, 499–507 (2005).
15. Moriya, Y. *et al.* PathPred: an enzyme-catalyzed metabolic pathway prediction server. *Nucleic Acids Res.* **38**, W138–W143 (2010).
16. Henry, C.S., Broadbelt, L.J. & Hatzimanikatis, V. Discovery and analysis of novel metabolic pathways for the biosynthesis of industrial chemicals: 3-hydroxypropanoate. *Biotechnol. Bioeng.* **106**, 462–473 (2010).
17. Hatzimanikatis, V. *et al.* Exploring the diversity of complex metabolic networks. *Bioinformatics* **21**, 1603–1609 (2005).
18. Finley, S.D., Broadbelt, L.J. & Hatzimanikatis, V. Computational framework for predictive biodegradation. *Biotechnol. Bioeng.* **104**, 1086–1097 (2009).
19. Klamt, S. & Stelling, J. Two approaches for metabolic pathway analysis? *Trends Biotechnol.* **21**, 64–69 (2003).
20. 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).
21. Pharkya, P., Burgard, A.P. & Maranas, C.D. OptStrain: a computational framework for redesign of microbial production systems. *Genome Res.* **14**, 2367–2376 (2004).
22. Jung, Y.K., Kim, T.Y., Park, S.J. & Lee, S.Y. Metabolic engineering of *Escherichia coli* for the production of polylactic acid and its copolymers. *Biotechnol. Bioeng.* **105**, 161–171 (2010).
23. Tantillo, D.J., Chen, J. & Houk, K.N. Theozymes and compuzymes: theoretical models for biological catalysis. *Curr. Opin. Chem. Biol.* **2**, 743–750 (1998).
24. Röthlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 (2008).
25. Siegel, J.B. *et al.* Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* **329**, 309–313 (2010).
26. Jiang, L. *et al. De novo* computational design of retro-aldol enzymes. *Science* **319**, 1387–1391 (2008).
27. Richter, F., Leaver-Fay, A., Khare, S.D., Bjelic, S. & Baker, D. *De novo* enzyme design using Rosetta3. *PLoS ONE* **6**, e19230 (2011).
This paper describes the Rosetta *de novo* enzyme design protocol, using triose phosphate isomerase as an example of enzyme design.
28. Eiteman, M.A., Lee, S.A. & Altman, E. A co-fermentation strategy to consume sugar mixtures effectively. *J. Biol. Eng.* **2**, 3 (2008).
29. Ren, C., Chen, T., Zhang, J., Liang, L. & Lin, Z. An evolved xylose transporter from *Zymomonas mobilis* enhances sugar transport in *Escherichia coli*. *Microb. Cell Fact.* **8**, 66 (2009).
30. Lindner, S.N., Seibold, G.M., Henrich, A., Kramer, R. & Wendisch, V.F. Phosphotransferase system-independent glucose utilization in *Corynebacterium glutamicum* by inositol permeases and glucokinases. *Appl. Environ. Microbiol.* **77**, 3571–3581 (2011).
31. 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).
32. 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).
33. Qian, Z.G., Xia, X.X. & Lee, S.Y. Metabolic engineering of *Escherichia coli* for the production of putrescine: a four carbon diamine. *Biotechnol. Bioeng.* **104**, 651–662 (2009).
34. Qian, Z.G., Xia, X.X. & Lee, S.Y. Metabolic engineering of *Escherichia coli* for the production of cadaverine: a five carbon diamine. *Biotechnol. Bioeng.* **108**, 93–103 (2011).
35. Dunlop, M.J. *et al.* Engineering microbial biofuel tolerance and export using efflux pumps. *Mol. Syst. Biol.* **7**, 487 (2011).
36. Shen, C.R. *et al.* Driving forces enable high-titer anaerobic 1-butanol synthesis in *Escherichia coli*. *Appl. Environ. Microbiol.* **77**, 2905–2915 (2011).
37. 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).
An excellent example of metabolic engineering, incorporating disarmament of transcriptional regulation and enzyme screening for the production of aliphatic alcohols and acids, especially *n*-butanol.
38. Auriol, C., Bestel-Corre, G., Claude, J.B., Soucaille, P. & Meynial-Salles, I. Stress-induced evolution of *Escherichia coli* points to original concepts in respiratory cofactor selectivity. *Proc. Natl. Acad. Sci. USA* **108**, 1278–1283 (2011).
39. Charusanti, P. *et al.* Genetic basis of growth adaptation of *Escherichia coli* after deletion of *pgi*, a major metabolic gene. *PLoS Genet.* **6**, e1001186 (2010).
40. Chemler, J.A., Fowler, Z.L., McHugh, K.P. & Koffas, M.A.G. Improving NADPH availability for natural product biosynthesis in *Escherichia coli* by metabolic engineering. *Metab. Eng.* **12**, 96–104 (2010).
41. Martínez, I., Zhu, J., Lin, H., Bennett, G.N. & San, K.Y. Replacing *Escherichia coli* NAD-dependent glyceraldehyde 3-phosphate dehydrogenase (GAPDH) with a NADP-dependent enzyme from *Clostridium acetobutylicum* facilitates NADPH dependent pathways. *Metab. Eng.* **10**, 352–359 (2008).
42. Ma, S.M. *et al.* Optimization of a heterologous mevalonate pathway through the use of variant HMG-CoA reductases. *Metab. Eng.* **13**, 588–597 (2011).
43. Bastian, S. *et al.* Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in *Escherichia coli*. *Metab. Eng.* **13**, 345–352 (2011).
44. Ehsani, M., Fernandez, M.R., Biosca, J.A. & Dequin, S. Reversal of coenzyme

specificity of 2,3-butanediol dehydrogenase from *Saccharomyces cerevisae* and *in vivo* functional analysis. *Biotechnol. Bioeng.* **104**, 381–389 (2009).
45. Palsson, B. & Zengler, K. The challenges of integrating multi-omic data sets. *Nat. Chem. Biol.* **6**, 787–789 (2010).
46. Park, J.M., Kim, T.Y. & Lee, S.Y. Constraints-based genome-scale metabolic simulation for systems metabolic engineering. *Biotechnol. Adv.* **27**, 979–988 (2009).
47. 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).
48. Atsumi, S. *et al.* Evolution, genomic analysis, and reconstruction of isobutanol tolerance in *Escherichia coli*. *Mol. Syst. Biol.* **6**, 449 (2010).
49. Becker, J., Zelder, O., Hafner, S., Schroder, 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).
50. Na, D., Lee, S. & Lee, D. Mathematical modeling of translation initiation for the estimation of its efficiency to computationally design mRNA sequences with a desired expression level in prokaryotes. *BMC Syst. Biol.* **4**, 71 (2010).
51. Salis, H.M., Mirsky, E.A. & Voigt, C.A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27**, 946–950 (2009).
52. Solomon, K.V. & Prather, K.L. The zero-sum game of pathway optimization: emerging paradigms for tuning gene expression. *Biotechnol. J.* **6**, 1064–1070 (2011).
53. Jantama, K. *et al.* Combining metabolic engineering and metabolic evolution to develop nonrecombinant strains of *Escherichia coli* C that produce succinate and malate. *Biotechnol. Bioeng.* **99**, 1140–1153 (2008).
54. Leonard, E. *et al.* Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. *Proc. Natl. Acad. Sci. USA* **107**, 13654–13659 (2010).
This study optimizes geranylgeranyl diphosphate synthase and levopimaradiene synthase by enzyme evolution and active site mutations, respectively, for the overproduction of levopimaradiene in *E. coli*.
55. Wang, H.H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).
56. Orth, J.D., Thiele, I. & Palsson, B.O. What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).
57. Thiele, I. & Palsson, B.O. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* **5**, 93–121 (2010).
58. Shlomi, T., Eisenberg, Y., Sharan, R. & Ruppin, E. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol. Syst. Biol.* **3**, 101 (2007).
59. Henry, C.S., Broadbelt, L.J. & Hatzimanikatis, V. Thermodynamics-based metabolic flux analysis. *Biophys. J.* **92**, 1792–1805 (2007).
60. Park, J.M., Kim, T.Y. & Lee, S.Y. Prediction of metabolic fluxes by incorporating genomic context and flux-converging pattern analyses. *Proc. Natl. Acad. Sci. USA* **107**, 14931–14936 (2010).
61. Chandrasekaran, S. & Price, N.D. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* **107**, 17845–17850 (2010).
62. Zhuang, K., Vemuri, G.N. & Mahadevan, R. Economics of membrane occupancy and respiro-fermentation. *Mol. Syst. Biol.* **7**, 500 (2011).
63. Segrè, D., Vitkup, D. & Church, G.M. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. USA* **99**, 15112–15117 (2002).
64. Lun, D.S. *et al.* Large-scale identification of genetic design strategies using local search. *Mol. Syst. Biol.* **5**, 296 (2009).
65. Ranganathan, S., Suthers, P.F. & Maranas, C.D. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput. Biol.* **6**, e1000744 (2010).
66. Kim, J. & Reed, J.L. OptORF: optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Syst. Biol.* **4**, 53 (2010).
67. Choi, H.S., Lee, S.Y., Kim, T.Y. & Woo, H.M. *In silico* identification of gene amplification targets for improvement of lycopene production. *Appl. Environ. Microbiol.* **76**, 3097–3105 (2010).
68. Yang, L., Cluett, W.R. & Mahadevan, R. EMILiO: a fast algorithm for genome-scale strain design. *Metab. Eng.* **13**, 272–281 (2011).
69. Asadollahi, M.A. *et al.* Enhancing sesquiterpene production in *Saccharomyces cerevisiae* through *in silico* driven metabolic engineering. *Metab. Eng.* **11**, 328–334 (2009).
70. Burk, M.J., Burgard, A., Osterhout, R.E. & Pharkya, P. Microorganisms and methods for the biosynthesis of adipate, hexamethylenediamine and 6-aminocaproic acid. US Patent 0317069 (2010)
71. Kizer, L., Pitera, D.J., Pfleger, B.F. & Keasling, J.D. Application of functional genomics to pathway optimization for increased isoprenoid production. *Appl. Environ. Microbiol.* **74**, 3229–3241 (2008).
72. Ajikumar, P.K. *et al.* Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli*. *Science* **330**, 70–74 (2010).
A multivariate-modular approach for optimizing metabolic fluxes via the methylerythritol-phosphate and terpenoid-forming pathways for enhanced production of a taxol precursor.
73. Noirel, J., Ow, S.Y., Sanguinetti, G. & Wright, P.C. Systems biology meets synthetic biology: a case study of the metabolic effects of synthetic rewiring. *Mol. Biosyst.* **5**, 1214–1223 (2009).
74. Zhang, F., Carothers, J.M. & Keasling, J.D. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat. Biotechnol.* **30**, 354–359 (2012).
75. Carothers, J.M., Goler, J.A., Juminaga, D. & Keasling, J.D. Model-driven engineering of RNA devices to quantitatively program gene expression. *Science* **334**, 1716–1719 (2011).
76. Dueber, J.E. *et al.* Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* **27**, 753–759 (2009).
A protein-scaffold strategy that mimics natural substrate tunneling, improving metabolite conversion efficiency by putting enzymes in close proximity.
77. Moon, T.S., Dueber, J.E., Shiue, E. & Prather, K.L. Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered *E. coli*. *Metab. Eng.* **12**, 298–305 (2010).
78. Jantama, K. *et al.* Eliminating side products and increasing succinate yields in engineered strains of *Escherichia coli* C. *Biotechnol. Bioeng.* **101**, 881–893 (2008).
79. Zhang, X. *et al.* Metabolic evolution of energy-conserving pathways for succinate production in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **106**, 20180–20185 (2009).
80. 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).
81. Yang, T.H. *et al.* Biosynthesis of polylactic acid and its copolymers using evolved propionate CoA transferase and PHA synthase. *Biotechnol. Bioeng.* **105**, 150–160 (2010).
82. Atsumi, S. *et al.* Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metab. Eng.* **10**, 305–311 (2008).
83. Wargacki, A.J. *et al.* An engineered microbial platform for direct biofuel production from brown macroalgae. *Science* **335**, 308–313 (2012).
84. Nichols, N.N., Dien, B.S. & Bothast, R.J. Use of catabolite repression mutants for fermentation of sugar mixtures to ethanol. *Appl. Microbiol. Biotechnol.* **56**, 120–125 (2001).
85. Becker, J., Buschke, N., Bucker, R. & Wittmann, C. Systems level engineering of *Corynebacterium glutamicum*: reprogramming translational efficiency for superior production. *Eng. Life Sci.* **10**, 430–438 (2010).
86. Huo, Y.-X. *et al.* Conversion of proteins into biofuels by engineering nitrogen flux. *Nat. Biotechnol.* **29**, 346–351 (2011).
87. Xia, X.X. *et al.* Native-sized recombinant spider silk protein produced in metabolically engineered *Escherichia coli* results in a strong fiber. *Proc. Natl. Acad. Sci. USA* **107**, 14059–14063 (2010).
88. Lale, R. *et al.* Continuous control of the flow in biochemical pathways through 5′ untranslated region sequence modifications in mRNA expressed from the broad-host-range promoter Pm. *Appl. Environ. Microbiol.* **77**, 2648–2655 (2011).
89. Zhang, K., Li, H., Cho, K.M. & Liao, J.C. Expanding metabolism for total biosynthesis of the non-natural amino acid L-homoalanine. *Proc. Natl. Acad. Sci. USA* **107**, 6234–6239 (2010).
90. Zhang, X., Jantama, K., Moore, J.C., Shanmugam, K.T. & Ingram, L.O. Production of L-alanine by metabolically engineered *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **77**, 355–366 (2007).
91. Wierckx, N.J. *et al.* Transcriptome analysis of a phenol-producing *Pseudomonas putida* S12 construct: genetic and physiological basis for improved production. *J. Bacteriol.* **190**, 2822–2830 (2008).
92. Wierckx, N.J., Ballerstedt, H., de Bont, J.A. & Wery, J. Engineering of solvent-tolerant *Pseudomonas putida* S12 for bioproduction of phenol from glucose. *Appl. Environ. Microbiol.* **71**, 8221–8227 (2005).

Acknowledgments

We thank H.U. Kim for helpful discussion. This work was supported by the Advanced Biomass Research and Development Center of Korea (ABC-2010-0029799) through the Global Frontier Research Program (GFRP) of the Ministry of Education, Science and Technology (MEST) and also by the Intelligent Synthetic Biology Center (2011-0031963) through the GFRP of MEST. We also thank the World Class University program (R32-2008-000-10142-0) of MEST for additional support.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence should be addressed to S.Y.L.