DOI 10.1002/biot.201000171 Biotechnol. J. 2010, 5, 716–725

Research Article

Microbial 1-butanol production: Identification of non-native production routes and *in silico* engineering interventions

Sridhar Ranganathan[1] *and Costas D. Maranas*[2]

[1] Huck Institutes of Life Sciences, The Pennsylvania State University, University Park, PA, USA
[2] Department of Chemical Engineering, The Pennsylvania State University, University Park, PA, USA

The potential of engineering microorganisms with non-native pathways for the synthesis of long-chain alcohols has been identified as a promising route to biofuels. We describe computationally derived predictions for assembling pathways for the production of biofuel candidate molecules and subsequent metabolic engineering modifications that optimize product yield. A graph-based algorithm illustrates that, by culling information from BRENDA and KEGG databases, all possible pathways that link the target product with metabolites present in the production host are identified. Subsequently, we apply our recent OptForce procedure to pinpoint reaction modifications that force the imposed product yield in *Escherichia coli*. We demonstrate this procedure by suggesting new pathways and genetic interventions for the overproduction of 1-butanol using the metabolic model for *Escherichia coli*. The graph-based search method recapitulates all recent discoveries based on the 2-ketovaline intermediate and hydroxybutyryl-CoA but also pinpointes one novel pathway through thiobutanoate intermediate that to the best of our knowledge has not been explored before.

Received 17 May 2010
Revised 16 June 2010
Accepted 17 June 2010

Supporting information available online

Keywords: Biofuels · Butanol · Metabolic engineering · Metabolic pathways · Strain optimization

1 Introduction

Increasing demands for renewable energy and environmental concerns have stimulated an interest towards the production of second generation biofuels from renewable sources [1]. For the past few decades, bio-ethanol was considered as a substitute for transportation fuels. More recently, long-chain alcohols (C3–C5) have also emerged as biofuel alternatives because of their higher energy density and ease of storage [2]. Microorganisms from diverse environments naturally produce ethanol during fermentation. However, the natural synthesis of higher alcohols is not as commonplace except within certain *Clostridia* strains [3, 4]. One possible production alternative for 1-butanol and 1-propanol is to use native pathways in *Clostridium acetobutylicum* [5–9]. An alternative approach is to integrate non-native pathways into standard microbial production hosts (i.e., *Escherichia coli* or yeast) by exploiting the conversion of key intermediary amino acids into long-chain alcohols [10, 11]. In this regard, numerous efforts have been made in the recent past to clone and express *Clostridia* genes (butyryl-CoA dehydrogenase, *bcd*) responsible for the production of 1-butanol in *E. coli* [12–14]. Homologs and isoenzymes of *bcd* from *Megasphaera elsdenii* [15, 16] and crotonoyl-CoA reductase (*ccr*) from *Streptomyces coelicolor* [17] have been tested. Recently, enzymes catalyzing the final steps of the Ehrlich pathway [18] in yeast were recruited in *E. coli* to convert 2-ketoacids into 1-butanol and isobutanol [19]. The global aim of converting biomass to energy has led to an increased interest in transferring non-native meta-

Correspondence: Professor Costas D. Maranas, Department of Chemical Engineering, The Pennsylvania State University, University Park, PA 16802, USA
E-mail: costas@engr.psu.edu
Fax: +1-814-865-7846

Abbreviations: **PFL**, pyruvate formate lyase; **PGI**, phosphate isomerase; **PGM**, phosphoglycerate mutase; **PPC**, phosphoenol pyruvate carboxylase; **PPCSCT**, propanoyl CoA:succinyl CoA transferase; **SUCOAS**, succinyl CoA synthetase

bolic pathways and enzymes into industrial production hosts such as *E. coli* [20, 21] or *Saccharomyces cerevisiae* [22].

An important goal of this research requires extending the metabolic confines of microbial hosts by recruiting non-native biosynthetic pathways. So far, studies concerning the incorporation of heterologous pathways relied largely on human intuition and literature reports followed by experimentation [23, 24]. Nowadays, rapidly expanding compilations of biotransformations such as KEGG [25] and BRENDA [26] are increasingly being prospected to identify biosynthetic routes to long-chain alcohols. With a combined size that accounts for over 60 000 enzymatic reactions and 250 000 metabolites, these databases include reactant and product designation, stoichiometric coefficients, organism assignment, and occasional thermodynamic information for pathways [27]. Several optimization and graph-based methods have been employed to computationally assemble novel biochemical routes from these sources. Given a set of reactions (*i.e.*, Universal database) the OptStrain [28] procedure uses a mixed-integer linear optimization representation to identify the minimal number of reactions to be added (*i.e.*, knock-ins) into a genome-scale metabolic model to enable the production of the new molecule. However, the developed universal database, at that time, was limited to only approximately 4000 reaction entries. The combinatorial nature of the problem poses a significant challenge to the OptStrain methodology as the number of reaction database entries increase from a few to tens of thousands. At the expense of not enforcing stoichiometric balances graph-based algorithms have inherently better-scaling properties for exhaustively identifying all min-path reaction entries that link a source with a target metabolite. For instance, Ma and Zeng [29] employed the shortest pathway algorithm for reactions in databases for reconstructing genome-scale metabolic models. Hatzimanikatis *et al.* [30] introduced a graph-based heuristic approach (BNICE) to identify all possible biosynthetic routes from a given substrate to a target chemical by hypothesized enzymatic reaction rules. Recently, the BNICE framework was used to identify novel metabolic pathways for the synthesis of 3-hydroxypropionate in *E. coli* [31]. Based on a similar approach, a new scoring algorithm [32] was introduced to evaluate and compare novel pathways generated using enzyme-reaction rules. The identified pathways may involve interconversions for which no enzymatic activity has been isolated before. While this could shed light to truly novel production avenues, it may be more time-consuming to implement. In addition, several techniques such as PathMiner [33], PathComp [34], Pathway Tools [35, 36], MetaRoute [37], PathFinder [38] and UM-BBD Pathway Prediction System [39] are in use to search for bioconversion routes in reaction databases. Most of these methods, so far, have been employed to aid metabolic pathway reconstructions by matching putative enzymes with reference pathways, while their contribution towards strain optimization has so far been limited.

Here we introduced a min-path graph procedure for overcoming the complexity associated with exhaustively identifying all possible ways of linking a source with a target metabolite. The procedure is designed to remain tractable even when reaction database entries reach hundred of thousands. The first step involved the incorporation of reaction and metabolite entries from both KEGG [25] and BRENDA [26] databases into a single repository. A customized min-path algorithm [40] was then employed to compute all possible pathways that enable the bio-production of a target alcohol molecule. We further scrutinized the identified pathways by first incorporating them into the genome-scale metabolic model of the production host microorganism, and subsequently examining their maximum theoretical yields, number of enzymatic steps needed and cofactor availability. We demonstrated our integrated framework by exploring pathways from pyruvate (produced in *E. coli*) to 1-butanol. We then selectively added one or more of these pathways to the latest genome-scale metabolic model of *E. coli*, *i*AF1260 [41] and used our recent OptForce [42] procedure to predict metabolic interventions (*i.e.*, up-/down-regulations and knockouts).

2 Materials and methods

The graph-based procedure discussed here is aimed at elucidating all possible biochemical routes from compounds found in the metabolic network of a desirable production host to a target molecule of interest. Alternatively, the procedure can also be used to track native routes that may increase productivity over known synthesis pathways by restricting the reaction entries to the ones present in the metabolic model of the production host. To provide the search procedure with known metabolic routes, we downloaded the most up-to-date version of the KEGG database [25] and extracted approximately 9000 reactions and 16 000 metabolites. Unfortunately, the KEGG database does not contain complete production pathways of long-chain alcohols. We therefore added a few hundred reaction entries from the BRENDA database

Biotechnol. J. 2010, 5, 716–725

[26] that are relevant to biofuel production to restore the metabolic connectivity to long-chain alcohols. It is important to note that we did not globally reconcile the entire KEGG database with BRENDA database (containing ~250 000 metabolites and 67 191 reactions). Instead, for all reactions in BRENDA associated with the synthesis of the target alcohol, we manually recorded identifiers for all the reactants, products and stoichiometric coefficients and integrated them with the KEGG entries into a single database.

Sorting out the naming inconsistencies for compounds was the most time-consuming step. To accomplish this, we made use of available synonym data from PubChem [43] to arrive at unique metabolite identifiers. Reactions with generic descriptions (*e.g.*, metabolites named as "alcohol", "aldehyde", *etc.*) for reactant/product compounds, with unknown stoichiometry or that involve macromolecules (*e.g.*, RNAP) were excluded. The integrated database used in this work spans 9921 reactions and 17 013 metabolites from both BRENDA and KEGG (see Text S1).

We used the min-path procedure as depicted in Fig. 1 to trace all possible paths between a source and a target metabolite. We first computationally transformed the information contained within the stoichiometric coefficients (S_{ij}) that track participation of metabolites in reactions into a directed metabolite-to-metabolite graph (N_{ii}) where nodes represent metabolites. A directed arc with a weight of one exists between two nodes if one or more reactions in the database allow the direct bioconversion from one metabolite to the other. If no such reaction exists, then a very large cost value was assigned to signify that their direct interconversion is disallowed. Whenever no information was available on the directionality of reactions we assumed that the reaction could operate in both directions. Small molecules (*e.g.*, water, carbon dioxide) and cofactors (*e.g.*, NADP, ATP) are involved in a large number of reactions and thus can link reaction steps that do not share any additional metabolites. We therefore excluded all such associated directed arcs before employing the shortest path algorithm. A list of all small metabolites and cofactors is compiled in the supplementary Text S1. Next we computed all *k*-shortest "loopless" pathways [40] between a source and a target alcohol molecule. We started from the shortest path (k =1) and exhaustively sampled the combinatorial space of alternative pathways by subsequently eliminating arcs, one at a time, belonging to the shortest pathway. Given a target number of pathways (K), we incremented the value of '*k*' (from 1 to K) whenever we identified a new pathway between the source and

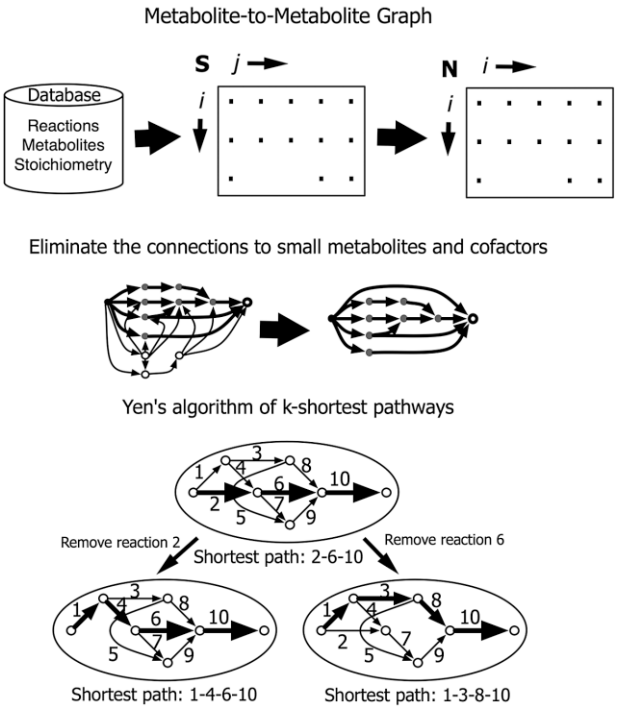

Figure 1. Graph-based procedure to min novel pathways from reaction databases using Yen's shortest path algorithm.

the target metabolite. We recomputed the shortest paths until we recorded the remaining "K – 1" shortest possible metabolic linkages to the target molecule.

We next evaluated the multiple identified pathways based on criteria such as maximum theoretical yield, number of reaction steps needed and cofactor requirements. Given a choice of a pathway to be added, we used our recent OptForce procedure [42] to identify additional strain manipulations (knockouts, up/down-regulations for fluxes) that guarantee a pre-specified yield for the alcohol molecule. The OptForce procedure uses metabolic flux measurements available for the wild-type strain and identifies which fluxes must depart from the original ranges to ensure the overproduction target for the desired alcohol molecule. Notably, the OptForce procedure [42] identifies the necessary changes (*i.e.*, MUST sets) by contrasting the maximum range of flux variability for the wild-type strain (characterized by flux measurements) against the flux ranges consistent with the overproducing network. Based on these necessary network changes, we combinatorially identified the minimal set of engineering interventions that result in a new flux distribution consistent with an overproducing strain of host microbe. All lexicographic searches needed to integrate database entries were performed using Python (version 2.4.3)

and the algorithm for the identification of shortest paths was coded using C++ on a 2.6 GHz AMD Opteron Processor with 32 GB of ECC RAM.

3. Results and discussion

3.1 Pathways from pyruvate to 1-butanol

We demonstrated our min-path procedure by identifying all synthesis routes using KEGG and BRENDA database entries for producing 1-butanol from pyruvate. We first selected promising pathways and subsequently integrated them with the genome-scale metabolic model of *E. coli*, *i*AF1260 [41]. Using OptForce [42] we next pinpointed metabolic engineering strategies for overproduction. Traditionally, two distinct synthesis routes have been employed in *E. coli* for the production of 1-butanol. The first pathway involves a fermentative transformation of pyruvate and acetyl-CoA to 1-butanol by the action enzymes from *C. acetobutylicum* [21]. The second pathway takes advantage of enzymes with broad-range substrate specificity to convert natural amino acids in *E. coli* into ketoacid precursors [19, 20] and eventually 1-butanol. In both pathways, pyruvate acts as an important precursor and a branching metabolite for butanol synthesis [44]. The fate of pyruvate at the end of glycolysis depends on the engineering strategies imparted to the production host. Therefore, here we selected pyruvate as a source metabolite in exploring pathways to 1-butanol (sink metabolite).

Figure 2 illustrates all identified pathways from pyruvate to 1-butanol using the integrated reaction database. With the exception of the thiobutanoate pathway (present in the BRENDA database), all other pathways involved butanoyl-CoA and 1-butanal as shared intermediates that are converted to 1-butanol using secondary alcohol dehydrogenase (*adhE*) from *C. acetobutylicum*. The min-path procedure recapitulated both the fermentative and ketoacid pathways for 1-butanol synthesis (shown in dotted lines). In addition, the algorithm uncovered a number of possible transformations to butanoyl-CoA involving intermediate metabolites that are produced in *E. coli*. For example, pyruvate can be converted into acetyl-CoA using pyruvate dehydrogenase natively present in *E. coli*. However, the conversion from acetyl-CoA to butanoyl-CoA is not favored because 1-butanal produced along the pathway is used up as a co-reactant along other reactions in the same pathway. This severely reduces the flux of the 1-butanol to less than 10 mmol/gDW h, which is about ten times less than the yields from existing pathways [2, 19]. Similarly, pathways involving methylmalate and methylbutanoate as intermediates require cofactors, which in turn, ad-

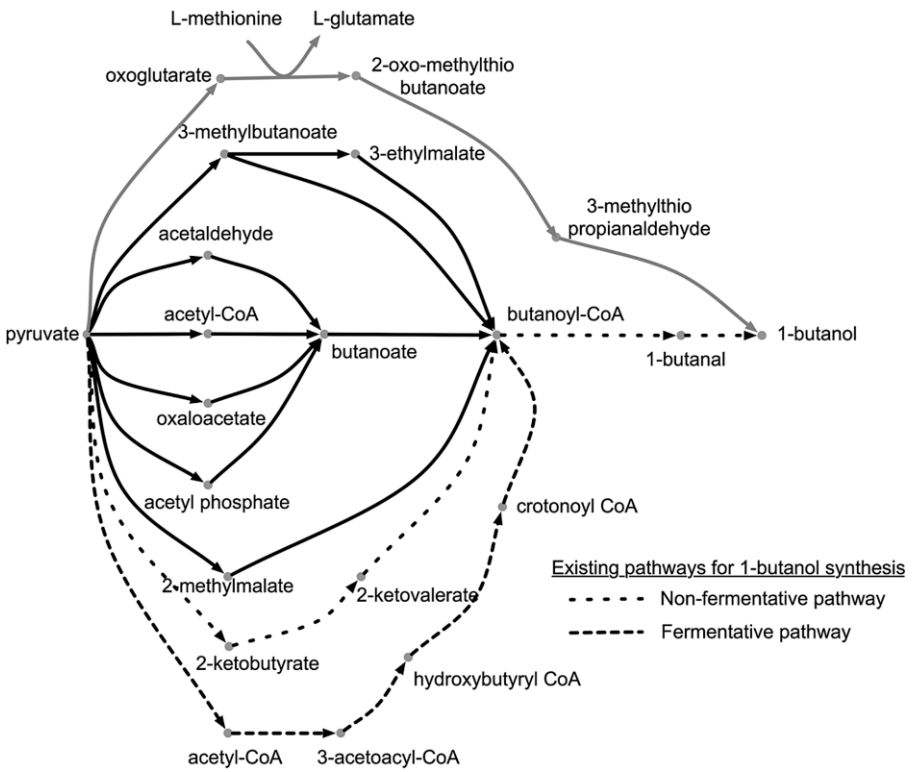

Figure 2. Pathways identified from pyruvate to 1-butanol using the graph-based procedure. Widely spaced dotted arrows represent the ketoacid pathway and the closely spaced arrows represent the fermentative pathways for 1-butanol synthesis. The thiobutanoate pathway is shown in gray.

Table 1. Maximum flux values for 1-butanol using the identified pathways from pyruvate

List of identified pathways between pyruvate and 1-butanol and intermediate reactions	Maximum flux value for 1-butanol (in mmol/gDW h)
Thiobutanoate pathway:	
L-Glutamate + pyruvate ↔ 2-ketoglutarate + L-alanine	74.8
2-Ketoglutarate + L-methionine → 2-keto methylthiobutyrate + L-glutamate	
2-Keto methylthiobutyrate → 3-methylthiopropanal + CO_2	
3-Methylthiopropanal + NADPH → 1-butanol + NADP	
Fermentative pathway:	
CoA + NAD + pyruvate → Acetyl-CoA + CO_2 + NADH	100
2 Acetyl-CoA ↔ acetoacetyl-CoA + CoA	
Acetoacetyl-CoA + NADPH + H+ ↔ 3-hydroxybutanoyl-CoA + NADP	
3-Hydroxybutanoyl-CoA ↔ crotonoyl-CoA + H_2O	
Crotonoyl-CoA + NADH + H+ ↔ butanoyl-CoA + NAD+	
Butanoyl-CoA + NADPH + H+ ↔ 1-butanal + CoA + NADP+	
1-Butanal + NADH + H+ ↔ 1-butanol + NAD+	
Ketoacid pathway:	
Pyruvate → L-threonine (isoleucine, serine and threonine metabolism)	100
L-Threonine ↔ oxobutanoate + NH_3	
Oxobutanoate + acetyl-CoA + NAD ↔ 2-ketovalerate + CO_2 + NADH	
2-Ketovalerate ↔ 1-butanal + CO_2	
1-Butanal + NADH + H+ ↔ 1-butanol + NAD+	
Methylmalate pathway:	
Acetyl-CoA + pyruvate + H_2O ↔ 2-methylmalate + CoA	32
Acetyl-CoA + 2-methylmalate ↔ acetate + butanoyl-CoA + CO_2+ H_2O	
Butanoyl-CoA + NADPH + H+ ↔ 1-butanal + CoA + NADP+	
1-Butanal + NADH + H+ ↔ 1-butanol + NAD+	
Other pathways:	
Pyruvate ↔ acetaldehyde + CO_2	9.3
Acetaldehyde + CoA + NAD+ ↔ acetyl-CoA + NADH + H+	
ATP + butanoic acid + CoA ↔ AMP + diphosphate + butanoyl-CoA	

versely reduce the yield of 1-butanol. Upon integrating these reactions in the metabolic model of *E. coli*, we estimated that the maximum theoretical yield of 1-butanol synthesis was only around 32 mmol/gDW h. The maximum butanol flux values and cofactor usages of these pathways are shown in Table 1.

The thiobutanoate pathway recruits a decarboxylase and a reductase enzyme and defines a novel synthesis route distinct from the two existing pathways. Instead of using dehydrogenases to convert butyraldehyde into 1-butanol, the new pathway proceeds with the transamination of methionine into 2-oxomethylthiobutanoate and eventually into 1-butanol. Notably, a native transaminase (EC 2.6.1.42) enzyme in *E. coli* is known to catalyze the conversion of L-methionine to L-glutamate with 2-ketoglutarate as a co-reactant [45]. The intermediate product, 2-methylthiobutanoate, is subsequently decarboxylated (EC 4.1.1.72) to 3-methylthiopropanal. This conversion is native in *Lactococcus lactis* [46]. Subsequently, 3-methylthiopropanal is reduced (EC 1.1.1.265) to 1-butanol by a reductase present in yeast [47]. It is important to note that the decarboxylase reaction removes a considerable amount carbon in the form of carbon dioxide, reducing the yield of 1-butanol by ~22% in comparison to the ketoacid pathway.

Next, we integrate these reactions in the *i*AF1260 metabolic model of *E. coli* and used OptForce [42] to identify metabolic interventions to meet an imposed overproduction target. The identified results were contrasted against the ones derived when the ketoacid pathway was integrated into the *E. coli* model. In both the case studies, the initial strain was first characterized by estimating the maximal range of flux variability using the intracellular flux measurements [48] available for the wild-type strain of *E. coli*, BW25113. The OptForce employed a bilevel optimization procedure to first identify the reaction fluxes that must increase or decrease (MUST sets) outside the wild-type flux

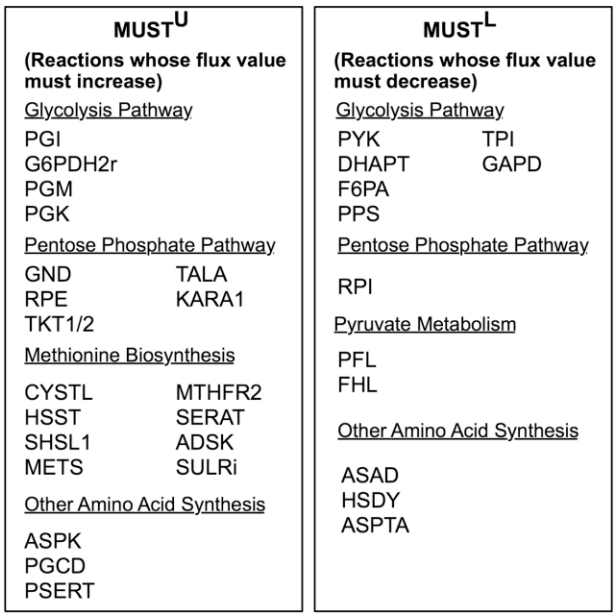

Figure 3. MUST^{U} and MUST^{L} set of reactions for 1-butanol synthesis in *E. coli* using the thiobutanoate pathway.

ranges to meet the overproduction target. A minimal set of direct interventions (*i.e.*, knock-up/down/outs) that guarantee a pre-specified yield for 1-butanol was next extracted from the MUST sets. All abbreviations for reactions and metabolites adhere to the *i*AF1260 metabolic model conventions.

3.2 Case 1: 1-Butanol synthesis using thiobutanoate pathway

Figure 3 lists the identified MUST set of reactions considering one reaction at a time. The yield for 1-butanol was set at 95% of its theoretical maximum, while allowing the production of 5% biomass to support growth. The thiobutanoate pathway branches away from 2-ketoglutarate along the oxidative arm of the TCA cycle. To increase the pool of oxaloacetate available for the TCA cycle, the fluxes of reactions in the glycolytic pathway [phosphate isomerase (PGI), phosphoglycerate mutase (PGM), phosphoglycerate kinase, phosphoenol pyruvate carboxylase (PPC), *etc.*] increase beyond their initial ranges. Many reactions in the pentose phosphate pathway (*e.g.*, GND, TKT1/2, TALA, *etc.*) were also classified in the MUST^{U} sets. The increase in the fluxes for these reactions replenishes the glycolytic intermediary metabolites. Since methionine is required as an important precursor for 1-butnaol pathway, reactions in methionine biosynthesis (*e.g.*, CYSTL, METS, MTHFR2, CYSS) were also members of the MUST^{U} set. The fluxes of reactions leading to competing by-products, pyruvate kinase (PYK) and pyruvate formate lyase (PFL) decrease below their initial ranges. Since biomass production is reduced to 5% of its theoretical maximum, reactions in amino acid biosynthesis that are directly coupled to growth appear in the MUST^{L} sets.

As expected, more complex flux changes are revealed in the network of MUST^{UU}, MUST^{UL} and MUST^{LL} sets shown in Fig. 4. These results underscore the importance of increasing the flux through the oxidative arm of the TCA cycle (FUM, *etc.*) or at the same time negating the drain towards by-products such as acetate and ethanol. Additionally, in the MUST^{UU} set, the flux of propanoyl CoA:succinyl CoA transferase (PPCSCT) or the flux of succinyl CoA synthetase (SUCOAS) must increase. Both of these fluxes are in close proximity to 2-ketoglutarate, which is an important branching metabolite in the TCA cycle for the thiobutanoate pathway. We carried out this hierarchical classification by considering three reactions at a time (Fig. 5). The increase in fluxes for IPPMI, IMPC and AIRC3 further boosts the synthesis of precursors for methionine through amino acid biosynthetic pathways.

Notably, the MUST set of reactions represent the changes that must take place in the metabolic network for overproduction that can be directly or indirectly imparted by metabolic interventions. OptForce identifies the minimal set of reaction interventions (culled from the MUST sets) that forces the target yield for 1-butanol. Figure 6a shows the FORCE set of reactions for overproducing 1-butanol in *E. coli* using the thiobutanoate pathway. Up-regulating one of the two glycolytic fluxes, glu-

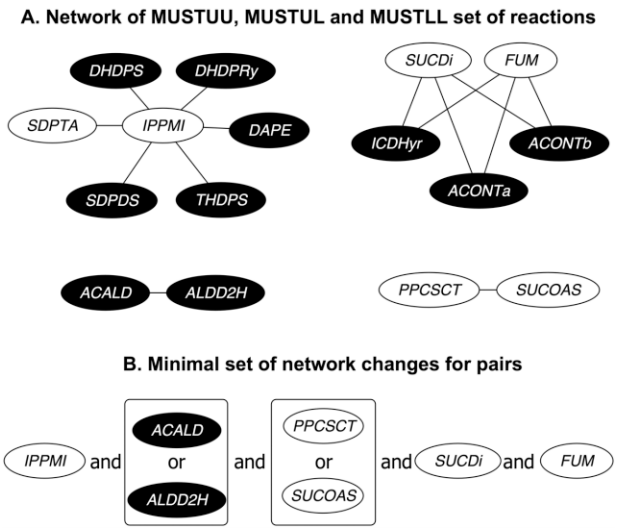

Figure 4. MUST^{UU}, MUST^{UL} and MUST^{LL} set of reactions for 1-butanol synthesis using the thiobutanoate pathway. Black ovals represent reaction flux down-regulations while white ovals denote up-regulations.

Biotechnol. J. 2010, 5, 716–725

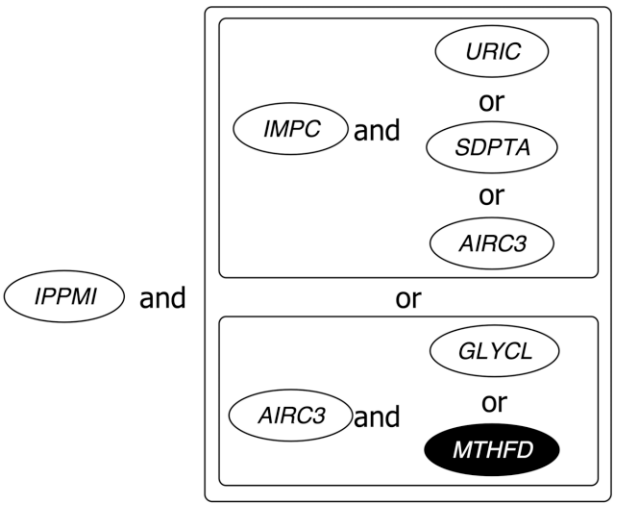

Figure 5. Minimal set of network changes for triples (*i.e.*, $MUST^{UUU}$, $MUST^{UUL}$, $MUST^{ULL}$, *etc.*). Reactions whose fluxes must increase are shown in white ovals while reactions whose fluxes decrease are shown in black ovals.

cose-6-PGI or PGM, replenishes phosphoenol pyruvate available for the anaplerotic conversion to oxaloacetate. The up-regulation for PPC results in increasing the amount of oxaloacetate for the TCA cycle. Increase in fluxes of PPCSCT or SUCOAS ensure the availability of 2-ketoglutarate for transamination along the thiobutanoate pathway. In addition, the FORCE sets also include knockouts for PFL to reduce the drain towards by-products (acetate and ethanol) and methylenetetrahydrofolate dehydrogenase (MTHFD) to prevent the drain of L-methionine away from the thiobutanoate pathway. These coordinated sets of interventions lead to a guaranteed yield for 1-butanol of 73 mmol/ gDW h.

3.3 Case 2: 1-Butanol using ketoacid pathway

Figure 6b contrasts the metabolic pathways and branching points for the ketoacid and thiobutanoate pathways on a metabolic map of *E. coli*, respectively. While the thiobutanoate pathway branches out from a TCA cycle intermediate, pyruvate serves as an important precursor for 1-butanol produced via the ketoacid pathway. We integrated the reactions along this pathway to *i*AF1260 metabolic model of *E. coli* and applied our OptForce procedure to predict the MUST sets and, subsequently, the FORCE sets. Figure 6b shows the FORCE set of eight engineering interventions for 1-butanol synthesis in *E. coli* using the ketoacid pathway. Here, OptForce suggested the up-regulation in the fluxes of reactions that convert key amino acids to 1-butanol precursors [*i.e.*, serine deaminase (SERD) and methylglyoxal synthase (MGSA)]. Down-regulation of PFL and lactate dehydrogenase (LDH) is needed to reduce carbon drain, occurring presumably due to the proximity of the ketoacid pathway to the synthesis routes for natural fermentation products (acetate, ethanol, formate, lactate, *etc.*). Additionally, down-regulation of TCA cycle reactions, fumarate reductase (FRD3) and aconitase (ACONTa/b) also appear to be essential network changes to ensure overproduction.

A notable difference between the two cases is the down-regulation of phosphogluconate dehydrogenase (GND) using the ketoacid pathway. While the flux of GND must increase for the thiobutanoate pathway (*i.e.*, member of $MUST^{U}$ set), OptForce suggests that its flux must be reduced to facilitate 1-butanol synthesis when using the ketoacid pathway. In addition, while PGI and PGM were identified as up-regulations for the thiobutanoate pathway no glycolytic reactions were up- regulated in the FORCE set for the ketoacid route. Since the ketoacid pathway branches out from precursors synthesized at the end of glycolytic pathway, OptForce indicates that the depletion of carbon can be minimized through a number of down-regulations for competing pathways without the need of overexpressing glycolytic enzymes. However, in the thiobutanoate case, the anaplerotic PPC is required to replenish oxaloacetate and to sustain an increased flux through the TCA cycle.

4 Concluding remarks

We have presented a graph-based min-path procedure that combines metabolic information from online databases (KEGG and BRENDA) to identify all possible biochemical synthesis routes to target biofuel candidates. The results for 1-butanol pathways reveal several new heterologous synthesis routes that can be computationally evaluated for overexpression and cloning experiments. Our algorithm was able to identify existing pathways (ketoacid and fermentative pathways) used for 1-butanol production. Existing strains of *E. coli* [2, 19, 21, 49] for the synthesis of 1-butanol have metabolic pathways engineered to harness either pyruvate or acetyl-CoA available at the end of glycolysis. These two metabolites serve as entry points for non-native functionalities associated with 1-butanol synthesis. In addition, acetyl-CoA can be formed from pyruvate using the pyruvate dehydrogenase reaction. We hence selected pyruvate as the starting metabolite to identify the synthesis routes to 1-butanol. Interestingly, we observed that the results also suggested several native synthesis routes to

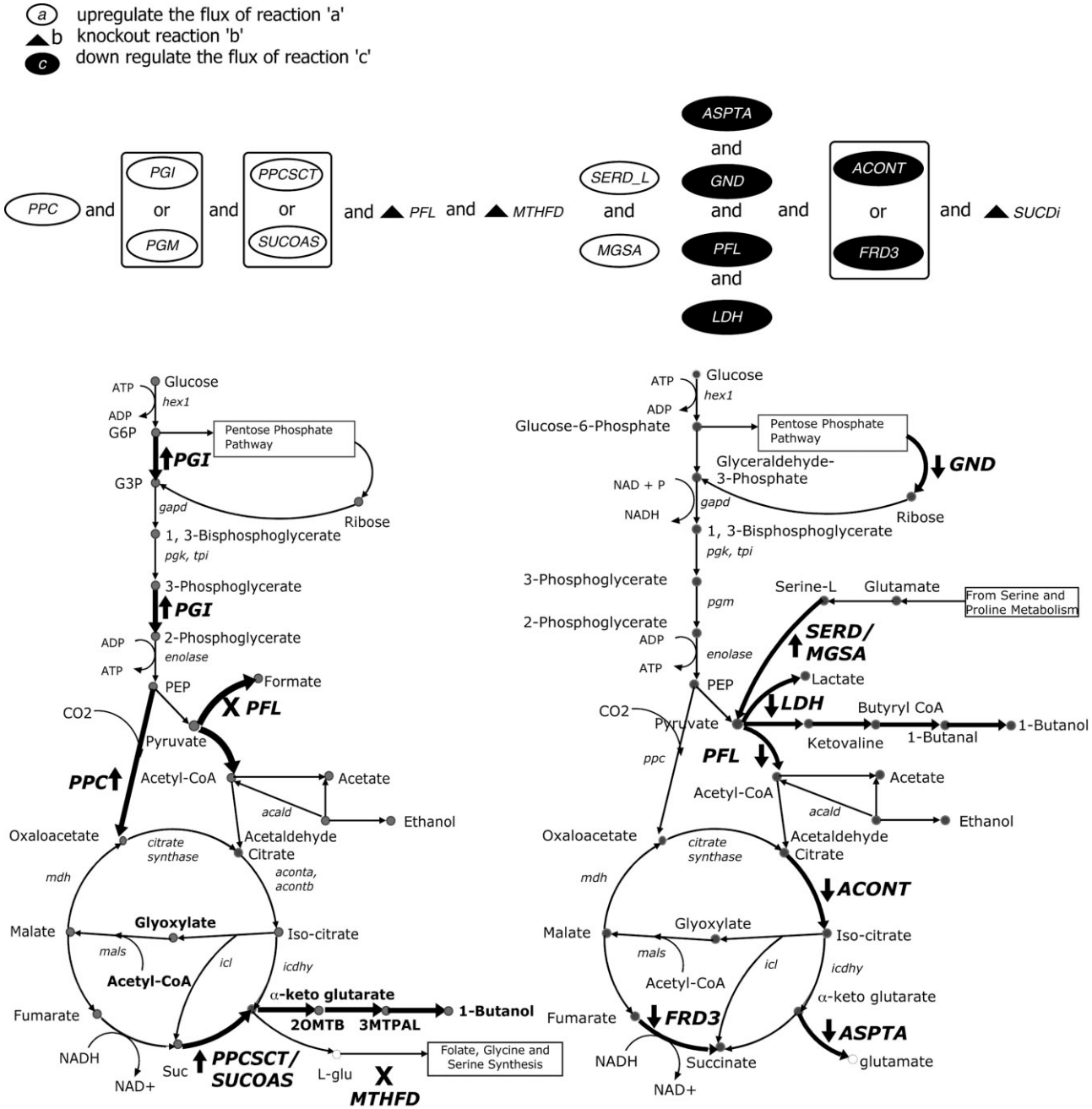

Figure 6. The FORCE set of reactions for 1-butanol synthesis using the thiobutanoate (left) and the ketoacid (right) pathways. All the reaction interventions are shown in bold. Up-regulations are denoted with (↑) symbol, down-regulations are denoted with (↓) symbol and the knockouts are shown with (×) symbol.

precursors of 1-butanol in *E. coli*. For example, seven pathways from pyruvate to butanoyl-CoA involved intermediate metabolites produced by naturally occurring enzymes in *E. coli*. However, the yield of 1-butanol using these pathways was limited. In addition, the algorithm also uncovered a new alternative route to 1-butanol synthesis through the thiobutanoate pathway. Although, the decarboxylation of methylthiobutanoate reduced 1-butanol production, the computationally derived yield was comparable to the existing strains [2, 19, 21].

The results suggested by our OptForce procedure [42] revealed the differing nature of metabolic interventions required to overproduce 1-butanol using the thiobutanoate and ketoacid pathway. Recruiting the thiobutanoate pathway for 1-butanol

overproduction required up-regulations for glycolytic fluxes (PGI, PGM). On the other hand, the ketoacid precursors were made available to 1-butanol synthesis by knocking down competing pathways (PFL, ACONTa/b, *etc.*). The flux changes observed in the MUST sets for the two cases also showcased contrasting patterns. For example, for the thiobutanoate pathway, the fluxes of the pentose phosphate pathway increased so that alternative routes for glutamate and other amino acids are maintained to support growth. However, none of the reactions from pentose phosphate pathway appeared in the FORCE sets, on the contrary, the OptForce procedure indicated that the fluxes of phosphogluconate dehydrogenase (GND) must be down-regulated on using the ketoacid pathway to synthesize 1-butanol.

Several interventions that were identified in the FORCE sets have been used in existing strains to produce 1-butanol. For example, recent strategies to delete host competing pathways encoded by the genes *ldhA, frdBC, pta, pfl* and *adhE* [2, 19, 21] have resulted in a threefold increase in the yield of 1-butanol. In addition, enhancing glycolytic fluxes by overexpressing NADH-regenerating enzymes were implemented in an *E. coli* strain [49] that yielded 580 mg/L 1-butanol. In addition to the existing interventions, the OptForce procedure also uncovered new knockouts and up-regulations that coordinate an increased synthesis of 1-butanol. For example, the up-regulation of glycolytic fluxes and PPC increase the amount of oxaloacetate for the TCA cycle. However, to effectively utilize the transamination pathway, OptForce suggested up-regulations for PPCSCT and SUCOAS that are in close proximity to the branching thiobutanoate pathway.

Using the procedure detailed here all possible metabolic routes to any target compound can be enumerated. Alternatively, the graph-based procedure can be used to identify alternative synthesis routes found entirely within the production host by selectively exploring pathways that are native. Currently, the procedure uses all the biotransformations found in the KEGG database [25, 50] and a selected set of reactions from the BRENDA [26] database. The min-path search procedure remains tractable for much larger compilations of reactions/metabolites. Notably, the interventions proposed by OptForce pertain to the reactions. A complete mapping between the reactions and the genes is required for projecting the results at the gene level.

The authors would like to thank Patrick F. Suthers and Anthony P. Burgard for helpful suggestions and discussions. This work was supported by the United States Department of Energy (DOE) Grant DE-FG02-05ER25684.

The authors have declared no conflict of interest.

5 References

[1] Stephanopoulos, G., Challenges in engineering microbes for biofuels production. *Science* 2007, *315*, 801–804.

[2] Atsumi, S., Liao, J. C., Metabolic engineering for advanced biofuels production from *Escherichia coli*. *Curr. Opin. Biotechnol.* 2008, *19*, 414–419.

[3] Janssen, P. H., Propanol as an end product of threonine fermentation. *Arch. Microbiol.* 2004, *182*, 482–486.

[4] Jones, D. T., Woods, D. R., Acetone-butanol fermentation revisited. *Microbiol. Rev.* 1986, *50*, 484–524.

[5] Nair, R. V., Papoutsakis, E. T., Expression of plasmid-encoded aad in *Clostridium acetobutylicum* M5 restores vigorous butanol production. *J. Bacteriol.* 1994, *176*, 5843–5846.

[6] Formanek, J., Mackie, R., Blaschek, H. P., Enhanced butanol production by *Clostridium beijerinckii* BA101 grown in semidefined P2 medium containing 6 percent maltodextrin or glucose. *Appl. Environ. Microbiol.* 1997, *63*, 2306–2310.

[7] Lin, Y. L., Blaschek, H. P., Butanol production by a butanol-tolerant strain of *Clostridium acetobutylicum* in extruded corn broth. *Appl. Environ. Microbiol.* 1983, *45*, 966–973.

[8] Lee, J. Y., Jang, Y. S., Lee, J., Papoutsakis, E. T., Lee, S. Y., Metabolic engineering of *Clostridium acetobutylicum* M5 for highly selective butanol production. *Biotechnol. J.* 2009, *4*, 1432–1440.

[9] Sillers, R., Chow, A., Tracy, B., Papoutsakis, E. T., Metabolic engineering of the non-sporulating, non-solventogenic *Clostridium acetobutylicum* strain M5 to produce butanol without acetone demonstrate the robustness of the acid-formation pathways and the importance of the electron balance. *Metab. Eng.* 2008, *10*, 321–332.

[10] Hanai, T., Atsumi, S., Liao, J. C., Engineered synthetic pathway for isopropanol production in *Escherichia coli*. *Appl. Environ. Microbiol.* 2007, *73*, 7814–7818.

[11] Yan, Y., Liao, J. C., Engineering metabolic systems for production of advanced fuels. *J. Ind. Microbiol. Biotechnol.* 2009, *36*, 471–479.

[12] Boynton, Z. L., Bennett, G. N., Rudolph, F. B., Cloning, sequencing, and expression of genes encoding phosphotransacetylase and acetate kinase from *Clostridium acetobutylicum* ATCC 824. *Appl. Environ. Microbiol.* 1996, *62*, 2758–2766.

[13] Fontaine, L., Meynial-Salles, I., Girbal, L., Yang, X. *et al.*, Molecular characterization and transcriptional analysis of adhE2, the gene encoding the NADH-dependent aldehyde/alcohol dehydrogenase responsible for butanol production in alcohologenic cultures of *Clostridium acetobutylicum* ATCC 824. *J. Bacteriol.* 2002, *184*, 821–830.

[14] Wiesenborn, D. P., Rudolph, F. B., Papoutsakis, E. T., Thiolase from *Clostridium acetobutylicum* ATCC 824 and its role in the synthesis of acids and solvents. *Appl. Environ. Microbiol.* 1988, *54*, 2717–2722.

[15] Becker, D. F., Fuchs, J. A., Banfield, D. K., Funk, W. D. *et al.*, Characterization of wild-type and an active-site mutant in *Escherichia coli* of short-chain acyl-CoA dehydrogenase from *Megasphaera elsdenii*. *Biochemistry* 1993, *32*, 10736–10742.

[16] O'Neill, H., Mayhew, S. G., Butler, G., Cloning and analysis of the genes for a novel electron-transferring flavoprotein from *Megasphaera elsdenii*. Expression and characterization of the recombinant protein. *J. Biol. Chem.* 1998, *273*, 21015–21024.
[17] Wallace, K. K., Bao, Z. Y., Dai, H., Digate, R. *et al.*, Purification of crotonyl-CoA reductase from *Streptomyces collinus* and cloning, sequencing and expression of the corresponding gene in *Escherichia coli*. *Eur. J. Biochem.* 1995, *233*, 954–962.
[18] Sentheshanuganathan, S., The mechanism of the formation of higher alcohols from amino acids by *Saccharomyces cerevisiae*. *Biochem. J.* 1960, *74*, 568–576.
[19] Atsumi, S., Hanai, T., Liao, J. C., Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* 2008, *451*, 86–89.
[20] Shen, C. R., Liao, J. C., Metabolic engineering of *Escherichia coli* for 1-butanol and 1-propanol production via the keto-acid pathways. *Metab. Eng.* 2008, *10*, 312–320.
[21] Atsumi, S., Cann, A. F., Connor, M. R., Shen, C. R. *et al.*, Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metab. Eng.* 2008, *10*, 305–311.
[22] Steen, E. J., Chan, R., Prasad, N., Myers, S. *et al.*, Metabolic engineering of *Saccharomyces cerevisiae* for the production of n-butanol. *Microb. Cell Fact.* 2008, *7*, 36.
[23] Bode, H. B., Muller, R., The impact of bacterial genomics on natural product research. *Angew. Chem. Int. Ed. Engl.* 2005, *44*, 6828–6846.
[24] Wenzel, S. C., Muller, R., Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways. *Curr. Opin. Biotechnol.* 2005, *16*, 594–606.
[25] Kanehisa, M., Araki, M., Goto, S., Hattori, M. *et al.*, KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 2008, *36*, D480–484.
[26] Chang, A., Scheer, M., Grote, A., Schomburg, I., Schomburg, D., BRENDA, AMENDA and FRENDA the enzyme information system: New content and tools in 2009. *Nucleic Acids Res.* 2009, *37*, D588–592.
[27] Bader, G. D., Cary, M. P., Sander, C., Pathguide: A pathway resource list. *Nucleic Acids Res.* 2006, *34*, D504–506.
[28] Pharkya, P., Burgard, A. P., Maranas, C. D., OptStrain: A computational framework for redesign of microbial production systems. *Genome Res.* 2004, *14*, 2367–2376.
[29] Ma, H., Zeng, A. P., Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics* 2003, *19*, 270–277.
[30] Hatzimanikatis, V., Li, C., Ionita, J. A., Henry, C. S. *et al.*, Exploring the diversity of complex metabolic networks. *Bioinformatics* 2005, *21*, 1603–1609.
[31] 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.* 2010, *106*, 462–473.
[32] Cho, A., Yun, H., Park, J. H., Lee, S. Y., Park, S., Prediction of novel synthetic pathways for the production of desired chemicals. *BMC Syst. Biol.* 2010, *4*, 35.
[33] McShan, D. C., Rao, S., Shah, I., PathMiner: Predicting metabolic pathways by heuristic search. *Bioinformatics* 2003, *19*, 1692–1698.
[34] Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K. F. *et al.*, From genomics to chemical genomics: New developments in KEGG. *Nucleic Acids Res.* 2006, *34*, D354–357.
[35] Karp, P. D., Paley, S., Romero, P., The Pathway Tools software. *Bioinformatics* 2002, *18 Suppl 1*, S225–232.
[36] Karp, P. D., Paley, S. M., Krummenacker, M., Latendresse, M. *et al.*, Pathway Tools version 13.0: Integrated software for pathway/genome informatics and systems biology. *Brief Bioinform.* 2009.
[37] Blum, T., Kohlbacher, O., MetaRoute: Fast search for relevant metabolic routes for interactive network navigation and visualization. *Bioinformatics* 2008, *24*, 2108–2109.
[38] Goesmann, A., Haubrock, M., Meyer, F., Kalinowski, J., Giegerich, R., PathFinder: Reconstruction and dynamic visualization of metabolic pathways. *Bioinformatics* 2002, *18*, 124–129.
[39] Ellis, L. B., Roe, D., Wackett, L. P., The University of Minnesota Biocatalysis/Biodegradation Database: The first decade. *Nucleic Acids Res.* 2006, *34*, D517–521.
[40] Yen, J. Y., Finding K shortest loopless paths in a network. *Management Science* 1971, *17*, 712–716.
[41] Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M. *et al.*, A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* 2007, *3*, 121.
[42] Ranganathan, S., Suthers, P. F., Maranas, C. D., OptForce: An optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput. Biol.* 2010, *6*, e1000744.
[43] Wang, Y., Xiao, J., Suzek, T. O., Zhang, J. *et al.*, PubChem: A public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.* 2009, *37*, W623–633.
[44] Clomburg, J. M., Gonzalez, R., Biofuel production in *Escherichia coli*: The role of metabolic engineering and synthetic biology. *Appl. Microbiol. Biotechnol. 86*, 419–434.
[45] Kagamiyama, H., Hayashi, H., Branched-chain amino-acid aminotransferase of *Escherichia coli*. *Methods Enzymol.* 2000, *324*, 103–113.
[46] Smit, B. A., van Hylckama Vlieg, J. E., Engels, W. J., Meijer, L. *et al.*, Identification, cloning, and characterization of a Lactococcus lactis branched-chain alpha-keto acid decarboxylase involved in flavor formation. *Appl. Environ. Microbiol.* 2005, *71*, 303–311.
[47] Perpete, P., Collin, S., Contribution of 3-methylthiopropionaldehyde to the worty flavor of alcohol-free beers. *J. Agric. Food Chem.* 1999, *47*, 2374–2378.
[48] Shimizu, K., Metabolic flux analysis based on ^{13}C-labeling experiments and integration of the information with gene and protein expression patterns. *Adv. Biochem. Eng. Biotechnol.* 2004, *91*, 1–49.
[49] Nielsen, D. R., Leonard, E., Yoon, S. H., Tseng, H. C. *et al.*, Engineering alternative butanol production platforms in heterologous bacteria. *Metab. Eng.* 2009, *11*, 262–273.
[50] Kanehisa, M., The KEGG database. *Novartis Found. Symp.* 2002, *247*, 91–101; discussion 101–103, 119–128, 244–152.

