Resource

OptStrain: A computational framework for redesign of microbial production systems

Priti Pharkya, Anthony P. Burgard, and Costas D. Maranas[1]

Department of Chemical Engineering, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

This paper introduces the hierarchical computational framework OptStrain aimed at guiding pathway modifications, through reaction additions and deletions, of microbial networks for the overproduction of targeted compounds. These compounds may range from electrons or hydrogen in biofuel cell and environmental applications to complex drug precursor molecules. A comprehensive database of biotransformations, referred to as the Universal database (with >5700 reactions), is compiled and regularly updated by downloading and curating reactions from multiple biopathway database sources. Combinatorial optimization is then used to elucidate the set(s) of non-native functionalities, extracted from this Universal database, to add to the examined production host for enabling the desired product formation. Subsequently, competing functionalities that divert flux away from the targeted product are identified and removed to ensure higher product yields coupled with growth. This work represents an advancement over earlier efforts by establishing an integrated computational framework capable of constructing stoichiometrically balanced pathways, imposing maximum product yield requirements, pinpointing the optimal substrate(s), and evaluating different microbial hosts. The range and utility of OptStrain are demonstrated by addressing two very different product molecules. The hydrogen case study pinpoints reaction elimination strategies for improving hydrogen yields using two different substrates for three separate production hosts. In contrast, the vanillin study primarily showcases which non-native pathways need to be added into *Escherichia coli*. In summary, OptStrain provides a useful tool to aid microbial strain design and, more importantly, it establishes an integrated framework to accommodate future modeling developments.

[Supplemental material is available online at www.genome.org. The Universal database can be found at http://fenske.che.psu.edu/Faculty/CMaranas/pubs.html.]

A fundamental goal in systems biology is to elucidate the complete "palette" of biotransformations accessible to nature in living systems. This goal parallels the continuing quest in biotechnology to construct microbial strains capable of accomplishing an ever-expanding array of desired biotransformations. These biotransformations are aimed at products that range from simple precursor chemicals (Nakamura and Whited 2003; Causey et al. 2004) or complex molecules such as carotenoids (Misawa et al. 1991), to electrons in biofuel cells (Liu et al. 2004) or batteries (Bond et al. 2002; Bond and Lovley 2003), to even microbes capable of precipitating heavy metal complexes in bioremediation applications (Finneran et al. 2002; Lovley 2003; Methe et al. 2003). Recent developments in molecular biology and recombinant DNA technology have ushered in a new era in the ability to shape the gene content and expression levels for microbial production strains in a direct and targeted fashion (Stephanopoulos 2002). The astounding range and diversity of these newly acquired capabilities and the scope of biotechnological applications imply that now more than ever we need modeling and computational aids to identify a priori the optimal sets of genetic modifications for strain optimization projects.

The recent availability of genome-scale models of microbial organisms has provided the pathway reconstructions necessary for developing computational methods aimed at identifying strain engineering strategies (Bailey 2001). These models, already

available for *Helicobacter pylori* (Schilling et al. 2002), *Escherichia coli* (Edwards and Palsson 2000; Reed et al. 2003), *Saccharomyces cerevisiae* (Forster et al. 2003), and other microorganisms (Van Dien and Lidstrom 2002; David et al. 2003; Valdes et al. 2003), provide successively refined abstractions of the microbial metabolic capabilities. An automated process to expedite the construction of stoichiometric models from annotated genomes (Segre et al. 2003) promises further to accelerate the metabolic reconstructions of several microbial organisms. At the same time, individual reactions are deposited in databases such as KEGG, EMP, MetaCyc, UM-BBD, and many more (Selkov Jr. et al. 1998; Overbeek et al. 2000; Karp et al. 2002; Ellis et al. 2003; Kanehisa et al. 2004; Krieger et al. 2004), forming encompassing and growing collections of the biotransformations for which we have direct or indirect evidence of existence in different species. Already many thousands of such reactions have been deposited; however, unlike organism-specific metabolic reconstructions (Edwards and Palsson 2000; Schilling et al. 2002; Forster et al. 2003; Reed et al. 2003), these compilations include reactions from not a single but many different species in a largely uncurated fashion. This means that currently there exists an ever-expanding collection of microbial models and at the same time ever more encompassing compilations of non-native functionalities. This newly acquired plethora of data has brought to the forefront several computational and modeling challenges that form the scope of this article. Specifically, how can we systematically select from the thousands of functionalities cataloged in various biological databases, the appropriate set of pathways/genes to recombine into existing production systems such as *E. coli* so as to endow them with the desired new functionalities? Subsequently, how

can we identify which competing functionalities to eliminate to ensure high product yield as well as viability?

Existing strategies and methods for accomplishing this goal include database queries to explore all feasible bioconversion routes from a substrate to a target compound from a given list of biochemical transformations (Seressiotis and Bailey 1988; Mavrovouniotis et al. 1990). More recently, elegant graph theoretic concepts (e.g., P-graphs [Fan et al. 2002] and k-shortest paths algorithm [Eppstein 1994]) were pioneered to identify novel biotransformation pathways based on the tracing of atoms (Arita 2000, 2004), enzyme function rules, and thermodynamic feasibility constraints (Li et al. 2004). Also, an interesting heuristic search approach that uses the enzymatic biochemical reactions found in the KEGG database (Kanehisa et al. 2004) to construct a connected graph linking the substrate and the product metabolites was recently proposed (McShan et al. 2003). Most of these approaches, however, generate linear paths that link substrates to final products without ensuring that the rest of the metabolic network is balanced and that metabolic imperatives on cofactor usage/generation and energy balances are met.

In this paper, we introduce a hierarchical optimization-based framework, OptStrain, to identify stoichiometrically balanced pathways to be generated upon recombination of non-native functionalities into a host organism to confer the desired phenotype. Candidate metabolic pathways are identified from an ever-expanding array of thousands (currently 5734) of reactions pooled together from different stoichiometric models and publicly available databases such as KEGG (Kanehisa et al. 2004). Note that the identified pathways satisfy maximum yield considerations whereas the choice of substrates can be treated as optimization variables. Important information pertaining to the cofactor/energy requirements associated with each pathway is deduced enabling the comparison of candidate pathways with respect to the aforementioned criteria. Production host selection is examined by successively minimizing the reliance on heterologous genes while satisfying the performance targets identified above. A gene set that encodes for all the enzymes needed to catalyze the identified non-native functionalities is then compiled by accounting for isozymes and multi-subunit enzymes. Subsequently, gene deletions are identified (Burgard et al. 2003; Pharkya et al. 2003) in the augmented host networks to improve product yields by removing competing functionalities that decouple biochemical production and growth objectives. The breadth and scope of OptStrain are demonstrated by addressing in detail two different product molecules (i.e., hydrogen and vanillin) that lie at the two extremes in terms of product molecule size. Briefly, computational results in some cases match existing strain designs and production practices whereas in others they pinpoint novel engineering strategies.
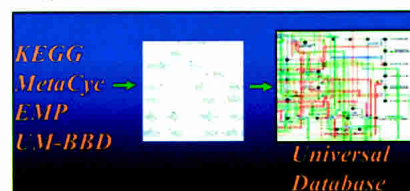
The OptStrain procedure

The first challenge addressed in this paper is to develop a systematic computational framework to identify which functionalities to add to an organism-specific metabolic network (e.g., *E. coli* [Edwards and Palsson 2000; Reed et al. 2003], *S. cerevisiae* [Forster et al. 2003], *Clostridium acetobutylicum* [Papoutsakis 1984; Desai et al. 1999], etc.) to enable a desired biotransformation. Our group has already contributed toward this objective on a much smaller scale (Burgard and Maranas 2001). Using this work as a starting point here, we aim to pinpoint gene additions identified from a Universal database composed of ~4000 elementally bal-

anced reactions as well as to investigate multiple hosts and substrate choices (see Supplemental material at www.genome.org and http://fenske.che.psu.edu/Faculty/CMaranas/pubs.html). Note that the gene additions are identified by fulfilling both criteria of maximal product yield and minimum usage of non-native reactions. Because of the extremely large size of the compiled database and the presence of multiple and sometimes conflicting objectives that need to be simultaneously satisfied, we developed the OptStrain procedure illustrated in Figure 1. Each step introduces different computational challenges arising from the specific structure and size of the optimization problems that need to be solved.
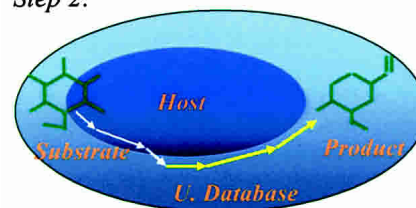
Step 1. Automated downloading and curation of the reactions in our Universal database to ensure stoichiometric balance.

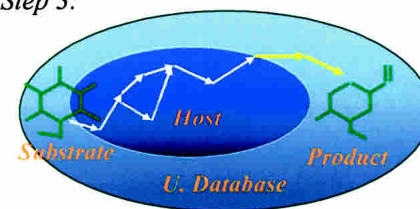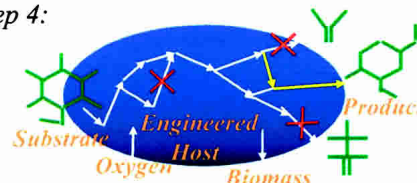**Step 2.** Calculation of the maximum theoretical yield of the

Figure 1. Pictorial representation of the OptStrain procedure. Step 1 involves the curation of database(s) of reactions to compile the Universal database, which comprises only elementally balanced reactions. Step 2 identifies a maximum-yield path enabling the desired biotransformation from a substrate (e.g., glucose, methanol, xylose) to product (e.g., hydrogen, vanillin) without any consideration for the origin of reactions. Note that the white arrows represent native reactions of the host and the yellow arrows denote non-native reactions. Step 3 minimizes the reliance on non-native reactions, and Step 4 incorporates the non-native functionalities into the microbial host's stoichiometric model and applies the OptKnock procedure to identify and eliminate reactions competing with the targeted product. The red ×s pinpoint the deleted reactions.

product given a substrate choice without restrictions on the reaction origin (i.e., native or non-native).

Step 3. Identification of a stoichiometrically balanced pathway(s) that minimizes the number of non-native functionalities in the examined production host given the maximum theoretical yield and the optimum substrate(s) found in Step 2. Alternative pathways that meet both criteria of maximum yield and minimum number of non-native reactions are generated along with comparisons between different host choices. Information pertaining to the cofactor/energy usage associated with each pathway is also determined at this stage. Finally, one or multiple gene sets that ensure the presence of the targeted biotransformations by encoding for the appropriate enzymes are derived at this stage.

Step 4. Incorporation of the identified non-native biotransformations into the stoichiometric models, if available, of the examined microbial production hosts. The OptKnock framework is next used (Burgard et al. 2003; Pharkya et al. 2003) on these augmented models to suggest gene deletions that ensure the production of the desired product becomes an obligatory byproduct of growth by "shaping" the connectivity of the metabolic network.

Curation of the database

The first step of the OptStrain procedure begins with the downloading and curation of reactions acquired from various sources in our Universal database. Specifically, given the fact that new reactions are incorporated in the KEGG database on a monthly basis, we have developed customized scripts using Perl (Brown 1999) to download all reactions in the database automatically on a regular basis and convert them into a format readable by the GAMS (Brooke et al. 1998) optimization environment. A different script is then used to parse the number of atoms of each element in every compound. The number of atoms of each type among the reactants and products of all reactions is calculated, and reactions that are elementally unbalanced are excluded from consideration. In addition, compounds with an unspecified number of repeat units [e.g., *trans*-2-Enoyl-CoA represented by $C_{25}H_{39}N_7O_{17}P_3S(CH_2)_n$] or unspecified alkyl groups R in their chemical formulas are removed from the downloaded sets. This step enables the automated downloading of reactions present in genomic databases and the subsequent verification of their elemental balance abilities forming large-scale sets of functionalities to be used as recombination targets.

Determination of the maximum yield

Once the reaction sets are determined, the second step is geared toward determining the maximum theoretical yield of the target product from a range of substrate choices, without restrictions on the number or origin of the reactions used. The maximum theoretical product yield is obtained for a unit uptake rate of substrate by maximizing the sum of all reaction fluxes producing minus those consuming the target metabolite, weighted by the stoichiometric coefficient of the target metabolite in these reactions. The maximization of this yield subject to stoichiometric constraints and transport conditions yields a linear programming (LP) problem (see Appendix for mathematical formulation). Given the computational tractability of LP problems, even for many thousands for reactions, a large number of different substrate choices can thoroughly be explored here.

Identification of the minimum number of non-native reactions for a host organism

The next step in OptStrain uses the knowledge of the maximum theoretical yield to determine the minimum number of non-native functionalities that need to be added into a specific host organism network. Mathematically, this is achieved by first introducing a set of binary variables y_j that serve as switches to turn the associated reaction fluxes v_j on or off.

$$v_j^{\min} \cdot y_j \le v_j \le v_j^{\max} \cdot y_j$$

Note that the binary variable y_j assumes a value of 1 if reaction j is active and a value of 0 if it is inactive. This constraint is imposed only on reactions associated with genes heterologous to the specified production host. The parameters $v_j^{\min}$ and $v_j^{\max}$ are specified to be very low and high values unattainable by the reaction flux v_j. This leads to a Mixed Integer Linear Programming (MILP) model for finding the minimum number of genes to be added into the host organism network while meeting the yield target for the desired product. This formulation (see Appendix for details) enables the exploration of trade-offs between the required numbers of heterologous genes versus the maximum theoretical product yield and also the iterative identification of all alternate optimal solutions using integer cut constraints. The end result of this step is a set of distinct pathways and corresponding gene complements that provide a ranked list of all alternatives for the efficient conversion of the substrate(s) into the desired product.

Incorporating the non-native reactions into the host organism's stoichiometric model

Upon identification of the appropriate host organism, the analysis proceeds with an organism-specific stoichiometric model augmented by the set of the identified non-native reactions. However, simply adding genes to a microbial production strain will not necessarily lead to the desired overproduction because microbial metabolism is primed to be as responsive as possible to the imposed selection pressures (e.g., outgrow its competition). These survival objectives are typically in direct competition with the overproduction of targeted biochemicals. To combat this, we use our previously developed bilevel computational framework, OptKnock (Burgard et al. 2003; Pharkya et al. 2003) to eliminate those functionalities that uncouple the cellular fitness objective, typically exemplified as the biomass yield, from the maximum yield of the product of interest.

Results

Computational results for microbial strain optimization focus on the production of hydrogen and vanillin. The hydrogen production case study underscores the importance of investigating multiple substrates and microbial hosts to pinpoint the optimal production environment as well as the need to eliminate competing functionalities. In contrast, in the vanillin study, identifying the smallest number of non-native reactions is found to be the key challenge for strain design. A common database of reactions, as outlined in Step 1, was constructed for both examples by pooling together metabolic pathways from the methylotroph *Methylobacterium extorquens* AM1 (Van Dien and Lidstrom 2002) and the KEGG database (Kanehisa et al. 2004) of reactions.

Hydrogen production case study

An efficient microbial hydrogen production strategy requires the selection of an optimal substrate and a microbial strain capable of forming hydrogen at high rates. First we solve the maximum yield LP formulation (Step 2) using all cataloged reactions that are balanced with respect to hydrogen, oxygen, nitrogen, sulfur, phosphorus, and carbon (~3000 reactions) as recombination candidates. Note that OptStrain allows for different substrate choices such as pentose and hexose sugars as well as acetate, lactate, malate, glycerol, pyruvate, succinate, and methanol. The highest hydrogen yield obtained for a methanol substrate is equal to 0.126 g/g substrate consumed, which is not surprising given that the hydrogen-to-carbon ratio for methanol is the highest at four to one. A comparison of the yields for some of the more efficient substrates is shown in Figure 2. We decided to explore methanol and glucose further, motivated by the high yield on methanol and the favorable costs associated with the use of glucose.

The next step in the OptStrain procedure entails the determination of the minimum number of non-native functionalities for achieving the theoretical maximum yield in a host organism. We examine three different uptake scenarios: (1) glucose as the substrate in *E. coli* (an established production system), (2) glucose in *C. acetobutylicum* (a known hydrogen producer), and (3) methanol in *M. extorquens* (a known methanol consumer).

E. coli

The MILP framework (described in Step 3) correctly verifies that with glucose as the substrate no non-native functionalities are required by *E. coli* for hydrogen production. Interestingly, hydrogen production is possible through either the ferredoxin hydrogenase reaction (E.C.# 1.12.7.2), which reduces protons to form hydrogen, or via the hydrogen dehydrogenase reaction (E.C.# 1.12.1.2), which converts NADH into NAD$^+$ while forming hydrogen through proton association. Subsequently, the upper and lower limits of maximum hydrogen formation are explored for the *E. coli* stoichiometric model (Reed et al. 2003) as a function of biomass formation rate (i.e., growth rate) for both aerobic and anaerobic conditions and a basis glucose uptake rate of 10 mmol/gDW per hour (see Fig. 3). Notably, the maximum theoretical hydrogen yield is higher under aerobic conditions. However, only under anaerobic conditions is hydrogen formed at maximum growth (see point A in Fig. 3), leading to a growth-coupled production mode. Note that hydrogen production takes place through the formate hydrogen lyase reaction, which converts formate into hydrogen and carbon dioxide under anaerobic con-
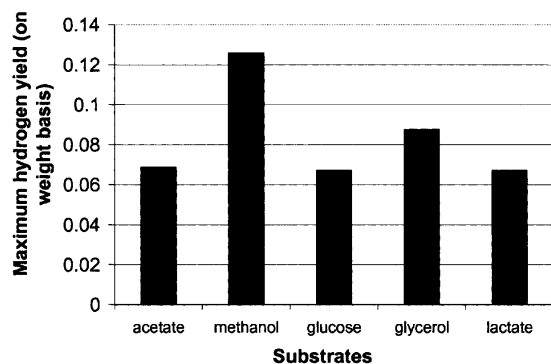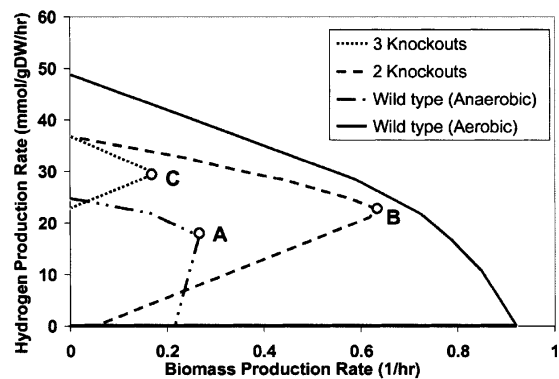

Figure 3. Hydrogen production envelopes as a function of the biomass production rate of the wild-type *E. coli* network under aerobic and anaerobic conditions as well as the two-reaction and three-reaction deletion mutant networks. The basis glucose uptake rate is fixed at 10 mmol/gDW per hour. These curves are constructed by finding the maximum and minimum hydrogen production rates at different rates of biomass formation. Point A denotes the required theoretical hydrogen production rate at the maximum biomass formation rate of the wild-type network under anaerobic conditions. Points B and C identify the theoretical hydrogen production rates at maximum growth for the two mutant networks, respectively, after fixing the corresponding carbon dioxide transport rates at the values suggested by OptKnock.

ditions, in agreement with experimental observations (Nandi and Sengupta 1998).

Moving to phenotype restriction to curtail byproduct formation (Step 4), we explored whether the production of hydrogen in the wild-type *E. coli* network could be enhanced by removing functionalities from the network that were in direct or indirect competition with hydrogen production. To this end, we used the OptKnock framework to pinpoint gene deletion strategies that couple hydrogen production with growth. Here we highlight two of the identified strategies shown in Table 1. The first (double deletion) removes both enolase (E.C.# 4.2.1.11) and glucose-6-phosphate dehydrogenase (E.C.# 1.1.1.49). The removal of the enolase reaction strongly promotes hydrogen formation by directing the glycolytic flux toward the 3-phosphoglycerate branching point into the serine biosynthesis pathway. Subsequently, serine participates in a series of reactions in one-carbon metabolism to form 10-formyltetrahydrofolate, which eventually is converted to formate and tetrahydrofolate. The dehydrogenase elimination prevents the shunting of glucose-6-phosphate flux into the pentose phosphate pathway. The second strategy, a three-reaction deletion study, involves the removal of ATP synthase (E.C.# 3.6.3.14), α-ketoglutarate dehydrogenase, and acetate kinase (E.C.# 2.7.2.1). The removal of the first reaction enhances proton availability, whereas the other two deletions ensure that maximum carbon flux is directed toward pyruvate, which is then converted into formate through pyruvate formate lyase. Formate is catabolized into hydrogen and carbon dioxide through formate hydrogen lyase. Computationally derived (not measured) flux distributions for both strategies are shown in Figure 4.

A comparison of the hydrogen production limits as a function of growth rate for both the wild-type and mutant networks is shown in Figure 3. The transport rates of carbon dioxide for the mutant networks are fixed at the values suggested by OptKnock (see Table 1), thus setting the operational imperatives (Pharkya et al. 2003). Note that whereas the two-reaction deletion mutant has a theoretical hydrogen production rate of 22.7 mmol/gDW

Figure 2. Maximum hydrogen yield on a weight basis for different substrates.

Table 1. Deletion mutants for enhanced hydrogen production in *E. coli*

No. of knockouts	ID	Reaction	Enzyme	Growth rate (L/h)	Secretion per hour (mmol/gDW)	
					CO$_2$	H$_2$
2	A	1. 2PG $\leftrightarrow$ H$_2$O + PEP 2. G6P + NADP $\leftrightarrow$ 6PGL + H$^+$ + NADPH	Enolase Glucose-6-phosphate	0.227	32.7	22.8
3	B	1. ADP + 4 H$^+$ + PI $\leftrightarrow$ ATP + 3 H$^+$ + H$_2$O 2. AC + ATP $\leftrightarrow$ ACTP + ADP 3. AKG + CoA + NAD $\rightarrow$ CO$_2$ + NADH + SUCCoA	ATP synthase Acetate kinase 2-Oxogluterate dehydrogenase	0.174	40.9	29.5

A basis glucose uptake rate of 10 mmol/gDW per hour is assumed.

per hour (0.025 g/g glucose) at the maximum growth rate (point B in Fig. 3), the three-reaction deletion mutant produces a maximum of 29.5 mmol/gDW per hour (0.033 g/g glucose) (point C in Fig. 3) at the expense of a reduced maximum growth rate. Interestingly, in both mutant networks, maximum hydrogen production requires the uptake of oxygen. This is in contrast to the wild-type case in which the lack of oxygen is preferred for hydrogen formation. Notably, it has been reported (Nandi and Sengupta 1996) that although formate hydrogen lyase can only be induced in the absence of oxygen, its catalytic activity is not affected in aerobic environments. This will have to be accounted for in any experimental study conducted on the basis of these results.

C. acetobutylicum

Ample literature evidence has identified the organisms of the *Clostridium* species as natural hydrogen production systems (Kataoka et al. 1997; Nandi and Sengupta 1998; Das and Veziroglu 2001; Chin et al. 2003). The reduction of protons into hydrogen through ferredoxin hydrogenase (E.C.# 1.12.7.2) is the key associated reaction. Not surprisingly, using OptStrain (Step 3), we verified that no non-native reactions were required for hydrogen production (Papoutsakis and Meyer 1985) in *C. acetobultylicum* with glucose as a substrate. We next explored, as in the *E. coli* case, whether hydrogen production could be enhanced by judiciously removing competing functionalities using the OptKnock framework. To this end, we used the stoichiometric model for *C. acetobutylicum* developed by Papoutsakis and coworkers (Papoutsakis 1984; Desai et al. 1999). OptKnock suggested the deletion of the acetate-forming and butyrate-transport reactions.

This deletion strategy is reasonable in hindsight upon considering the energetics of the entire network. Specifically, in the wild-type case, the formation and secretion of each butyrate molecule requires the consumption of two NADH molecules, thus reducing the hydrogen production capacity of the network (see Fig. 5). However, if butyrate is not secreted, but is instead recycled to form acetone and butyryl CoA, then butyryl CoA can again be converted to butyrate without any NADH consumption. This is evident in the flux distribution for the mutant network (see Fig. 5). The double deletion mutant has a theoretical hydrogen yield of 3.17 mol/mol glucose (0.036 g/g glucose) at the expense of slightly lower growth rate (point C in Fig. 6). Notably, in this case, biomass formation and hydrogen production are tightly coupled, in contrast to that in the wild-type network, where a range (1.38–2.96 mmol/gDW per hour) of hydrogen formation rates is possible (line AB in Fig. 6) at the maximum growth rate. Experimental results (Nandi and Sengupta 1998)

indicate that only up to 2 mol of hydrogen can be produced per mol of glucose anaerobically in *Clostridium*. In fact, it has been reported that inhibitory effects of butyrate directly on hydrogen production and indirect effects of acetate on growth inhibition (Chin et al. 2003) are responsible for the observed low hydrogen yields. Interestingly, the suggested reaction eliminations directly circumvent these inhibition bottlenecks.

M. extorquens AM1

Moving from glucose to methanol as the substrate, we next investigated hydrogen production in *M. extorquens* AM1, a facultative methylotroph capable of surviving solely on methanol as a carbon and energy source (Van Dien and Lidstrom 2002). The organism has been well studied (Anthony 1982; Chistoserdova et al. 1998; Korotkova et al. 2002; Van Dien et al. 2003; Chistoserdova et al. 2004), and recently, a stoichiometric model of its central metabolism was published (Van Dien and Lidstrom 2002). Using Step 3 of OptStrain, we identified that only a single reaction needs to be introduced into the metabolic network of *M. extorquens* to enable hydrogen production. Two such candidates are hydrogenase (E.C.# 1.12.7.2), which reduces protons to hydrogen, or alternatively, N$_5$, N$_{10}$-methenyltetrahydromethanopterin hydrogenase, which catalyzes the following transformation:

E.C.# 1.12.98.2: 5,10-methylenetrahydromethanopterin $\leftrightarrow$ 5,10-methenyltetrahydromethanopterin + H$_2$.
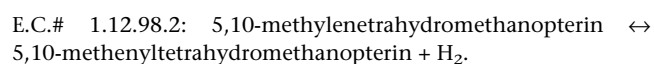
The need for an additional reaction is expected because the central metabolic pathways in the methylotroph, as abstracted in Van Dien and Lidstrom (2002), do not include any reactions that convert protons into hydrogen such as the hydrogenases found in *E. coli* and the anaerobes of the *Clostridium* species. Therefore, it is not surprising that, to the best of our knowledge, no one has achieved hydrogen production using methylotrophs such as *Pseudomonas AMI* and *P. methylica* (Nandi and Sengupta 1998). The identified reaction additions provide a plausible explanation for this outcome by pinpointing the lack of a mechanism to convert the generated protons to hydrogen.

Vanillin production case study

Vanillin is an important flavor and aroma molecule. The low yields of vanilla from cured vanilla pods have motivated efforts for its biotechnological production. In this case study, we identify metabolic network redesign strategies for the de novo production of vanillin from glucose in *E. coli*. Using OptStrain, we first determined the maximum theoretical yield of vanillin from
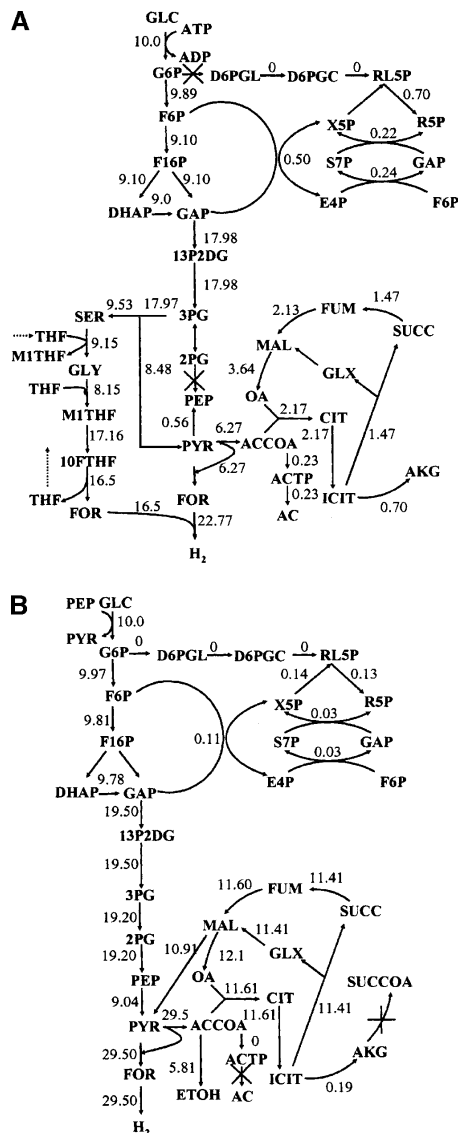
A

B

Figure 4. Calculated flux distributions at the maximum growth rates in the (*A*) two and (*B*) three deletion *E. coli* mutant networks for overproducing hydrogen. A basis glucose uptake rate of 10 mmol/gDW per hour was assumed.

glucose to be 0.63 g/g glucose by considering ~4000 candidate reactions balanced with respect to all elements but hydrogen (Step 2). We next identified that the minimum number of non-native reactions that must be recombined into *E. coli* to endow it with the pathways necessary to achieve the maximum yield is three (Step 3). Numerous alternative pathways, differing only in their cofactor usage, which satisfy both the optimality criteria of yield and minimality of recombined reactions, were identified. We then calculated the maximum theoretical yields of each of these gene addition strategies upon their incorporation into the *E. coli* stoichiometric model (Reed et al. 2003). Notably, for all these strategies, the yields are almost identical even though the stoichiometric model enforces a global balancing on cofactor usage. Therefore, for the sake of economy of presentation, only the following gene addition is discussed:
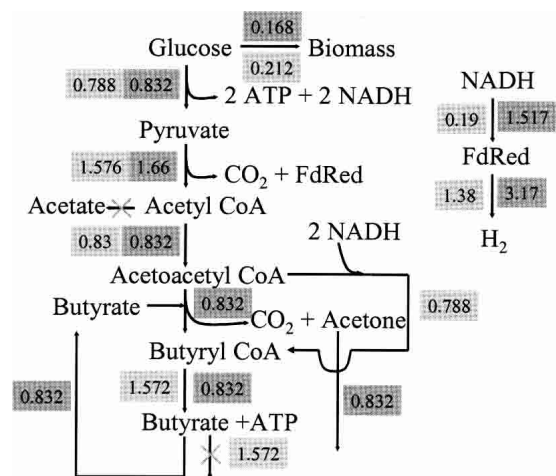

Figure 5. Calculated flux distributions at the maximum growth rates for the wild-type (light gray) and the two-reaction deletion mutant (dark gray) *C. acetobutylicum* networks. The ×s denote reactions that were selected for elimination in the mutant network. The wild-type network flux values are for the minimum hydrogen production scenario, corresponding to point A in Figure 6.

(1) E.C.# 1.2.1.46: Formate + NADH + H$^+$ ↔ formaldehyde + NAD$^+$ + H$_2$O,
(2) E.C.# 1.2.3.12: 3,4-dihydroxybenzoate (or protocatechuate) + NAD$^+$ + H$_2$O + formaldehyde ↔ vanillate + O$_2$ + NADH, and
(3) E.C.# 1.2.1.67: vanillate + NADH + H$^+$ ↔ vanillin + NAD$^+$ + H$_2$O.

Interestingly, these steps are essentially the same as those used in the experimental study by Li and Frost (1998) for the conversion of glucose into vanillin using recombinant *E. coli* cells and the biocatalyst aryl aldehyde dehydrogenase extracted from *Neurospora crassa*, demonstrating that OptStrain can recover existing engineering strategies. Note, however, that the reported experimental yield of 0.15 g/g glucose is far from the maximum theoretical yield (i.e., 0.63 g/g glucose) of the network indicating the potential for considerable improvement.
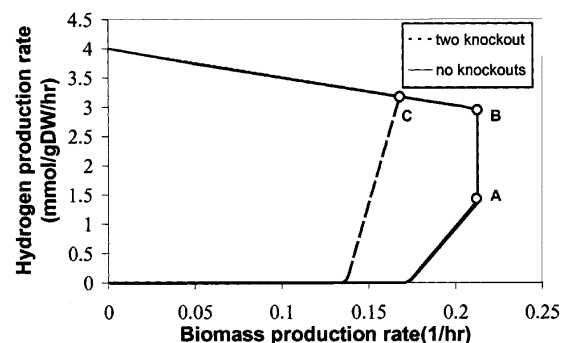

Figure 6. Hydrogen formation limits of the wild-type (solid) and mutant (dotted) *Clostridium acetobutylicum* metabolic network for a basis glucose uptake rate of 1 mmol/gDW per hour. Line AB denotes different alternate yield solutions that are available to the wild-type network at maximum biomass production rates. Point C pinpoints the hydrogen yield of the mutant network at maximum growth. This can be contrasted with the reported experimental hydrogen yield (2 mol/mol glucose) in *C. acetobutylicum* (Nandi and Sengupta 1998).
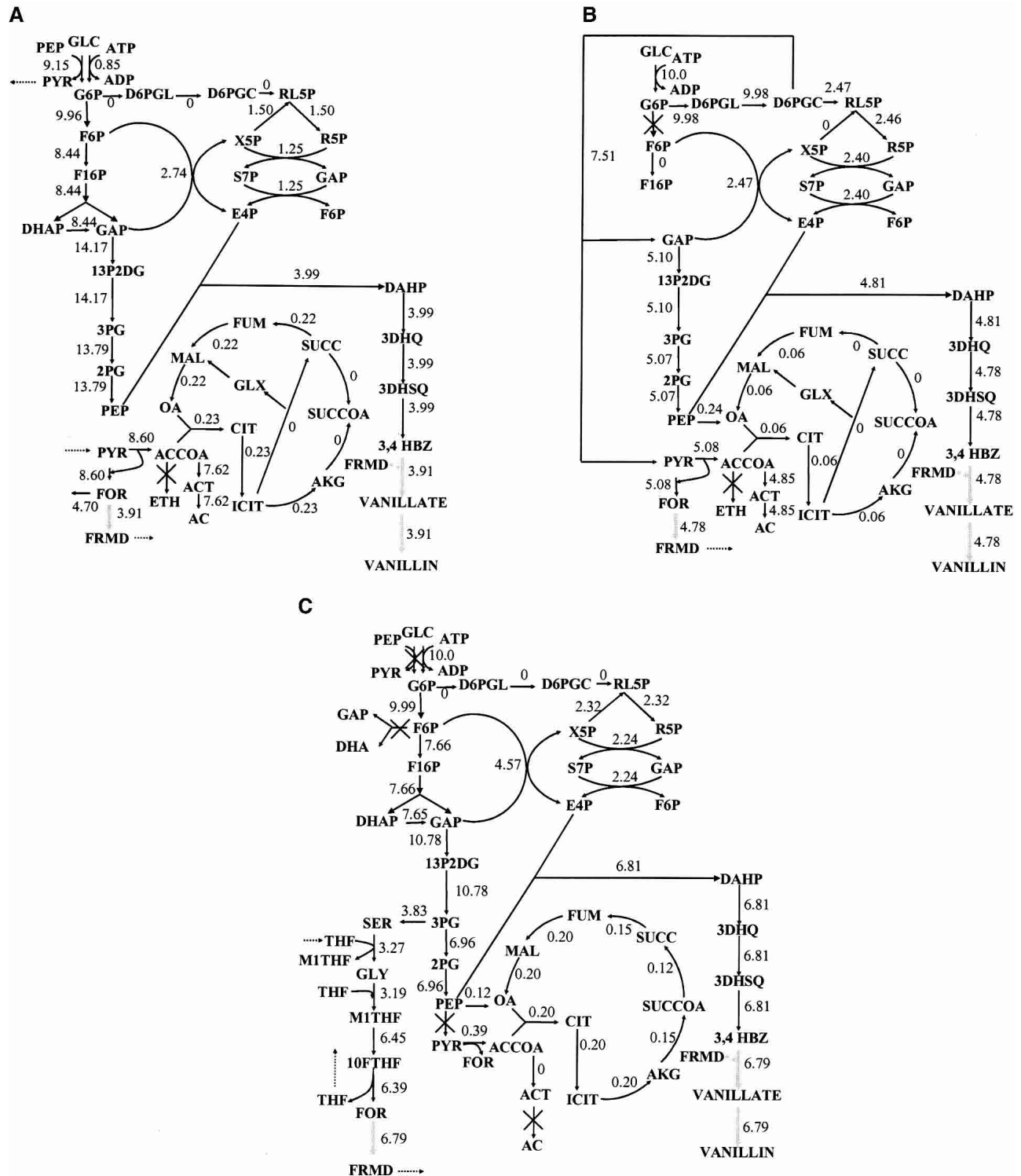
Figure 7. Calculated flux distributions at the maximum growth rates in the (*A*) one, (*B*) two, and (*C*) four deletion *E. coli* mutant networks for overproducing vanillin. Non-native reactions are denoted by the thicker gray arrows. A basic glucose uptake rate of 10 mmol/gDW per hour was assumed.

This motivates examining whether it is possible to reach higher yields of vanillin by systematically pruning the metabolic network using OptKnock (Step 4). Here the genome-scale model of *E. coli* metabolism, augmented with the three functionalities identified above, is integrated into the OptKnock framework to determine the set(s) of reactions whose deletion would force a

strong coupling between growth and vanillin production. The highest vanillin-yielding single, double, and quadruple knockout strategies are discussed next for a basis glucose uptake rate of 10 mmol/gDW per hour. In all cases, anaerobic conditions are selected by OptKnock as the most favorable for vanillin production. Flux distributions corresponding to the proposed knockout strategies are shown in Figure 7. It is worth emphasizing that, in general, the deletion strategies identified by OptStrain are dependent on the specific gene addition strategy fed into Step 4 of OptStrain. Accordingly, we tested whether alternative and possibly better, deletion strategies would accompany some of the other candidate addition strategies alluded to above. For the vanillin case study, we found the deletion suggestions and anticipated vanillin yields at maximal growth to be quite similar regardless of the gene addition strategy used.

The first deletion strategy identified by OptStrain suggests removing acetaldehyde dehydrogenase (E.C.# 1.2.1.10) to prevent the conversion of acetyl-CoA into ethanol. Vanillin production in this network, at the maximum biomass production rate of $0.205\ h^{-1}$, is 3.9 mmol/gDW per hour or 0.33 g/g glucose based on the assumed uptake rate of glucose. In this deletion strategy, flux is redirected through the vanillin precursor metabolites, phosphoenolpyruvate (PEP) and erythrose-4-phosphate (E4P), by blocking the loss of carbon through ethanol secretion. The second (double) deletion strategy involves the additional removal of glucose-6-phosphate isomerase (E.C.# 5.3.1.9) essentially blocking the upper half of glycolysis. These deletions cause the network to place a heavy reliance on the Entner-Doudoroff pathway to generate pyruvate and glyceraldehyde-3-phosphate (GAP), which undergoes further conversion into PEP in the lower half of glycolysis. Fructose-6-phosphate (F6P), produced through the nonoxidative part of the pentose phosphate pathway, is subsequently converted to E4P. Vanillin production, at the expense of a reduced maximum growth rate of $0.06\ h^{-1}$, is increased to 4.78 mmol/gDW per hour or 0.40 g/g glucose. A substantially higher level of vanillin production is predicted in the four-reaction deletion mutant network without imposing a high penalty on the growth rate. This strategy leads to the production of 6.79 mmol/gDW per hour of vanillin or 0.57 g/g glucose at the maximum

growth rate of $0.052\ h^{-1}$. The OptKnock framework suggests the deletion of acetate kinase (E.C.# 2.7.2.1), pyruvate kinase (E.C.# 2.7.1.40), the PTS transport mechanism, and fructose 6-phosphate aldolase. The first three deletions prevent leakage of flux from PEP and redirect it instead to vanillin synthesis. The elimination of fructose 6-phosphate aldolase prevents the direct conversion of F6P into GAP and dihydroxyacetone (DHA). Note that both F6P and GAP are used to form E4P in the nonoxidative branch of the pentose phosphate pathway. DHA can be further reacted to form dihydroxyacetone phosphate (DHAP) with the consumption of a PEP molecule. Thus, elimination of fructose 6-phosphate aldolase prevents the utilization of both F6P and PEP, which are required for vanillin synthesis. Furthermore, a surprising network flux redistribution involves the employment of a group of reactions from one-carbon metabolism to form 10-formyltetrahydrofolate, which is subsequently converted to formaldehyde. Figure 8 compares the vanillin production envelopes, obtained by maximizing and minimizing vanillin formation at different biomass production rates for the wild-type and mutant networks. These deletions endow the network with high levels of vanillin production under any growth conditions.

Discussion

The OptStrain framework is aimed at systematically suggesting how to reshape whole-genome-scale metabolic networks of microbial systems for the overproduction of not only small but also complex molecules. We have so far examined several different products (e.g., 1,3-propanediol, inositol, pyruvate, electron transfer, etc.) using a variety of hosts (i.e., *E. coli*, *C. acetobutylicum*, *M. extorquens*). The two case studies, hydrogen and vanillin, discussed earlier show that OptStrain can address the range of challenges associated with strain redesign allowing for the generation of multiple redesign strategies to be screened by experts and evaluated experimentally. At the same time, it is important to emphasize that the validity and relevance of the results obtained with the OptStrain framework are dependent on the level of completeness and accuracy of the reaction databases and microbial metabolic models considered. We have identified numerous instances of unbalanced reactions, especially with respect to hydrogen atoms, and ambiguous reaction directionality in the reaction databases that we mined. Careful curation of the downloaded reactions preceded all of our case studies. Whenever the elemental balance of a reaction could not be restored, the reaction was removed from consideration. We expect that this step will become less time-consuming as automated tools for reaction database testing and verification (Segre et al. 2003) are becoming available. Furthermore, the purely stoichiometric representation of metabolic pathways in microbial models used can lead to unrealistic flux distributions by not accounting for kinetic barriers and regulatory interactions (e.g., allosteric regulation). To alleviate this, we are currently working toward incorporating regulatory information in the form of Boolean constraints (Covert and Palsson 2002) into the stoichiometric model of *E. coli* and the use of kinetic expressions on an as-needed basis (Tomita et al. 1999; Varner and Ramkrishna 1999; Castellanos et al. 2004). Despite these simplifications, OptStrain has already provided useful insight into microbial host redesign in many cases and, more importantly, established for the first time an integrated framework open to future modeling improvements.
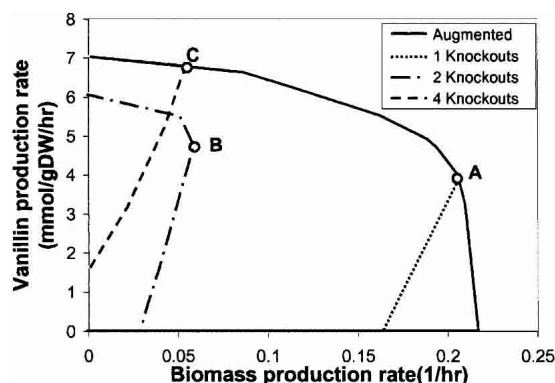

Figure 8. Vanillin production envelope of the augmented *E. coli* metabolic network for a basic 10 mmol/gDW per hour uptake rate of glucose. Points A, B, and C denote the maximum growth points associated with the one, two, and four reaction deletion mutant networks, respectively. Note that the level "augmented" refers to the *E. coli* wild-type network augmented with the three non-native reactions. An anaerobic mode of growth is suggested in all cases.

Acknowledgments

Appendix

Mathematical formulation

The redesign of microbial metabolic networks to enable enhanced product yields by using the OptStrain procedure requires the solution of multiple types of optimization problems. The first optimization task (Step 2) involves the determination of the maximum yield of the desired product in a metabolic network comprised of a set $\mathbb{N} = \{1, \ldots, N\}$ of metabolites and a set $\mathbb{M} = \{1, \ldots, M\}$ of reactions. The linear programming (LP) problem for maximizing the yield on a weight basis of a particular product $\mathbf{P}$ (in the set $\mathbb{N}$) from a set $\Re$ of substrates is formulated as:

$$\underset{v_j}{\textbf{Max}} \quad MW_i \cdot \sum_{j=1}^{M} S_{ij} v_j, \quad i = \mathbf{P}$$

$$\textbf{subject to} \quad \sum_{j=1}^{M} S_{ij} v_j \geq 0, \qquad \forall\, i \in \mathbb{N},\, i \notin \Re \quad (1)$$

$$\sum_{i \in \Re} \left(MW_i \cdot \sum_{j=1}^{M} S_{ij} v_j \right) = -1 \quad (2)$$

where MW_i is the molecular weight of metabolite i, v_j is the molar flux of reaction j, and S_{ij} is the stoichiometric coefficient of metabolite i in reaction j. In our work, the metabolite set $\mathbb{N}$ was comprised of ~4800 metabolites, and the reaction set $\mathbb{M}$ consisted of >5700 reactions. The inequality in constraint 1 allows only for secretion and prevents the uptake of all metabolites in the network other than the substrates in $\Re$. Constraint 2 scales the results for a total substrate uptake flux of one unit of mass. The reaction fluxes v_j can either be irreversible (i.e., $v_j \geq 0$) or reversible, in which case they can assume either positive or negative values. Reactions that enable the uptake of essential-for-growth compounds such as oxygen, carbon dioxide, ammonia, sulfate, and phosphate are also present.

In Step 3 of OptStrain, the minimum number of non-native reactions needed to meet the identified maximum yield from Step 2 is found. First the Universal database reactions that are absent in the examined microbial host's metabolic model are flagged as "non-native." This gives rise to the following Mixed Integer Linear Programming (MILP) problem:

$$\underset{v_j, y_j}{\textbf{Min}} \quad \sum_{j \in M_{non\text{-}native}} y_j$$

$$\textbf{subject to} \quad \sum_{j=1}^{M} S_{ij} v_j \geq 0, \qquad \forall\, i \in \mathbb{N},\, i \notin \Re \quad (1)$$

$$\sum_{i \in \Re} \left(MW_i \cdot \sum_{j=1}^{M} S_{ij} v_j \right) = -1, \quad (2)$$

$$MW_i \cdot \sum_{j=1}^{M} S_{ij} v_j \geq Yield^{t\,arg\,et}, \quad i = \mathbf{P} \quad (3)$$

$$v_j \leq v_j^{max} \cdot y_j, \qquad \forall\, j \in M_{non\text{-}native} \quad (4)$$

$$v_j \geq v_j^{min} \cdot y_j, \qquad \forall\, j \in M_{non\text{-}native} \quad (5)$$

$$y_j \in \{0,1\}, \qquad \forall\, j \in M_{non\text{-}native} \quad (6)$$

The set $\mathbb{M}_{non\text{-}native}$ comprises the non-native reactions for the examined host and is a subset of the set $\mathbb{M}$. MILP constraints 1 and 2 are identical to those in the product yield maximization problem. MILP constraint 3 ensures that the product yield meets the maximum theoretical yield, Yield^{target}, calculated in Step 2. The binary variables y_j in constraints 4 and 5 serve as switches to turn reactions on or off. A value of 0 for y_j forces the corresponding flux v_j to be 0, and a value of 1 enables it to take on nonzero values. The parameters v_j^{min} and v_j^{max} can either assume very low and very high values, respectively, or they can be calculated by minimizing and maximizing every reaction flux v_j subject to stoichiometric constraints.

Alternative pathways that satisfy both optimality criteria of maximum yield and minimum non-native reactions are obtained by the iterative solution of the MILP formulation upon the accumulation of additional constraints referred to as integer cuts. Integer cut constraints exclude from consideration all sets of reactions previously identified. For example, if a previously identified pathway uses reactions 1, 2, and 3, then the following constraint prevents the same reactions from being simultaneously considered in subsequent solutions: $y_1 + y_2 + y_3 \leq 2$. More details can be found in an earlier paper by Burgard and Maranas (2001).

Step 4 of OptStrain identifies which reactions to eliminate from the network augmented with the non-native functionalities, using the OptKnock framework developed previously (Burgard et al. 2003; Pharkya et al. 2003). The objective of this step is to constrain the phenotypic behavior of the network so that growth is coupled with the formation of the desired biochemical, thus curtailing byproduct formation. The envelope of allowable targeted product yields versus biomass yields is constructed by solving a series of linear optimization problems that maximize and then minimize biochemical production for various levels of biomass formation rates available to the network. More details on the optimization formulation can be found in Pharkya et al. (2003). All the optimization problems were solved in the order of minutes to hours using CPLEX 7.0 accessed via the GAMS (Brooke et al. 1998) modeling environment on an IBM RS6000-270 workstation.

References

Anthony, C. 1982. *The biochemistry of methylotrophs*. Academic Press, New York.

Arita, M. 2000. Metabolic construction using shortest paths. *Simulation Practice Theory* **8:** 109–125.

———. 2004. The metabolic world of *Escherichia coli* is not small. *Proc. Natl. Acad. Sci.* **101:** 1543–1547.

Bailey, J.E. 2001. Complex biology with no parameters. *Nat. Biotechnol.* **19:** 503–504.

Bond, D.R. and Lovley, D.R. 2003. Electricity production by *Geobacter sulfurreducens* attached to electrodes. *Appl. Environ. Microbiol.* **69:** 1548–1555.

Bond, D.R., Holmes, D.E., Tender, L.M., and Lovley, D.R. 2002. Electrode-reducing microorganisms that harvest energy from marine sediments. *Science* **295:** 483–485.

Brooke, A., Kendrick, D., Meeraus, A., and Raman, R. 1998. *GAMS: A user's guide*. GAMS Development Corp., Washington, D.C.

Brown, M. 1999. *Perl programmer's reference*. Osborne/McGraw-Hill, Berkeley, CA.

Burgard, A.P. and Maranas, C.D. 2001. Probing the performance limits of the *Escherichia coli* metabolic network subject to gene additions or deletions. *Biotechnol. Bioeng.* **74:** 364–375.

Burgard, A.P., Pharkya, P., and Maranas, C.D. 2003. Optknock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* **84:** 647–657.

Castellanos, M., Wilson, D.B., and Shuler, M.L. 2004. A modular minimal cell model: Purine and pyrimidine transport and metabolism. *Proc. Natl. Acad. Sci.* **101:** 6681–6686.

Causey, T.B., Shanmugam, K.T., Yomano, L.P., and Ingram, L.O. 2004.

Engineering *Escherichia coli* for efficient conversion of glucose to pyruvate. *Proc. Natl. Acad. Sci.* **101:** 2235–2240.

Chin, H.L., Chen, Z.S., and Chou, C.P. 2003. Fedbatch operation using *Clostridium acetobutylicum* suspension culture as biocatalyst for enhancing hydrogen production. *Biotechnol. Prog.* **19:** 383–388.

Chistoserdova, L., Vorholt, J.A., Thauer, R.K., and Lidstrom, M.E. 1998. C1 transfer enzymes and coenzymes linking methylotrophic bacteria and methanogenic Archaea. *Science* **281:** 99–102.

Chistoserdova, L., Laukel, M., Portais, J.C., Vorholt, J.A., and Lidstrom, M.E. 2004. Multiple formate dehydrogenase enzymes in the facultative methylotroph *Methylobacterium extorquens* AM1 are dispensable for growth on methanol. *J. Bacteriol.* **186:** 22–28.

Covert, M.W. and Palsson, B.O. 2002. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J. Biol. Chem.* **277:** 28058–28064.

Das, D. and Veziroglu, T.N. 2001. Hydrogen production by biological process: A survey of literature. *Intl. J. Hydrogen Energy* **26:** 13–28.

David, H., Akesson, M., and Nielsen, J. 2003. Reconstruction of the central carbon metabolism of *Aspergillus niger*. *Eur. J. Biochem.* **270:** 4243–4253.

Desai, R.P., Nielsen, L.K., and Papoutsakis, E.T. 1999. Stoichiometric modeling of *Clostridium acetobutylicum* fermentations with non-linear constraints. *J. Biotechnol.* **71:** 191–205.

Edwards, J.S. and Palsson, B.O. 2000. The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci.* **97:** 5528–5533.

Ellis, L.B., Hou, B.K., Kang, W., and Wackett, L.P. 2003. The University of Minnesota Biocatalysis/Biodegradation Database: Post-genomic data mining. *Nucleic Acids Res.* **31:** 262–265.

Eppstein, D. 1994. Finding the k shortest paths. In *Proceedings of 35th IEEE Symposium on Foundations of Computer Science*, pp. 154–165. Santa Fe.

Fan, L.T., Bertok, B., and Friedler, F. 2002. A graph-theoretic method to identify candidate mechanisms for deriving the rate law of a catalytic reaction. *Comput. Chem.* **26:** 265–292.

Finneran, K.T., Housewright, M.E., and Lovley, D.R. 2002. Multiple influences of nitrate on uranium solubility during bioremediation of uranium-contaminated subsurface sediments. *Environ. Microbiol.* **4:** 510–516.

Forster, J., Famili, I., Fu, P., Palsson, B.O., and Nielsen, J. 2003. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* **13:** 244–253.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32:** D277–D280.

Karp, P.D., Riley, M., Saier, M., Paulsen, I.T., Collado-Vides, J., Paley, S.M., Pellegrini-Toole, A., Bonavides, C., and Gama-Castro, S. 2002. The EcoCyc Database. *Nucleic Acids Res.* **30:** 56–58.

Kataoka, N., Miya, A., and Kiriyama, K. 1997. Studies of hydrogen production by continuous culture system of hydrogen-producing anaerobic bacteria. *Wat. Sci. Tech.* **36:** 41–47.

Korotkova, N., Chistoserdova, L., and Lidstrom, M.E. 2002. Poly-β-hydroxybutyrate biosynthesis in the facultative methylotroph *Methylobacterium extorquens* AM1: Identification and mutation of gap11, gap20, and phaR. *J. Bacteriol.* **184:** 6174–6181.

Krieger, C.J., Zhang, P., Mueller, L.A., Wang, A., Paley, S., Arnaud, M., Pick, J., Rhee, S.Y., and Karp, P.D. 2004. MetaCyc: A multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.* **32:** D438–D442.

Li, K. and Frost, J.W. 1998. Synthesis of vanillin from glucose. *J. Am. Chem. Soc.* **120:** 10545–10546.

Li, C., Ionita, J.A., Henry, C.S., Jankowski, M.D., Hatzimanikatis, V., and Broadbelt, L.J. 2004. Computational discovery of biochemical routes to specialty chemicals. *Chem. Engin. Sci.* (in press).

Liu, H., Ramnarayanan, R., and Logan, B.E. 2004. Production of electricity during wastewater treatment using a single chamber microbial fuel cell. *Environ. Sci. Tech.* **38:** 2281–2285.

Lovley, D.R. 2003. Cleaning up with genomics: Applying molecular biology to bioremediation. *Nat. Rev. Microbiol.* **1:** 35–44.

Mavrovouniotis, M., Stephanopoulos, G., and Stephanopoulos, G. 1990. Computer-aided synthesis of biochemical pathways. *Biotechnol. Bioeng.* **36:** 1119–1132.

McShan, D.C., Rao, S., and Shah, I. 2003. PathMiner: Predicting metabolic pathways by heuristic search. *Bioinformatics*

19: 1692–1698.

Methe, B.A., Nelson, K.E., Eisen, J.A., Paulsen, I.T., Nelson, W., Heidelberg, J.F., Wu, D., Wu, M., Ward, N., Beanan, M.J., et al. 2003. Genome of *Geobacter sulfurreducens*: Metal reduction in subsurface environments. *Science* **302:** 1967–1969.

Misawa, N., Yamano, S., and Ikenaga, H. 1991. Production of β-carotene in *Zymomonas mobilis* and *Agrobacterium tumefaciens* by introduction of the biosynthesis genes from *Erwinia uredovora*. *Appl. Environ. Microbiol.* **57:** 1847–1849.

Nakamura, C.E. and Whited, G.M. 2003. Metabolic engineering for the microbial production of 1,3-propanediol. *Curr. Opin. Biotechnol.* **14:** 454–459.

Nandi, R. and Sengupta, S. 1996. Involvement of anaerobic reductases in the spontaneous lysis of formate by immobilized cells of *Escherichia coli*. *Enzyme Microb. Tech.* **19:** 20–25.

———. 1998. Microbial production of hydrogen: An overview. *Crit. Rev. Microbiol.* **24:** 61–84.

Overbeek, R., Larsen, N., Pusch, G.D., D'Souza, M., Selkov Jr., E., Kyrpides, N., Fonstein, M., Maltsev, N., and Selkov, E. 2000. WIT: Integrated system for high-throughput genome sequence analysis and metabolic reconstruction. *Nucleic Acids Res.* **28:** 123–125.

Papoutsakis, E. 1984. Equations and calculations for fermentations of butyric acid bacteria. *Biotechnol. Bioeng.* **26:** 174–187.

Papoutsakis, E. and Meyer, C. 1985. Equations and calculations of product yields and preferred pathways for butanediol and mixed-acid fermentations. *Biotechnol. Bioeng.* **27:** 50–66.

Pharkya, P., Burgard, A.P., and Maranas, C.D. 2003. Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock. *Biotechnol. Bioeng.* **84:** 887–899.

Reed, J.L., Vo, T.D., Schilling, C.H., and Palsson, B.O. 2003. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.* **4:** R54.

Schilling, C.H., Covert, M.W., Famili, I., Church, G.M., Edwards, J.S., and Palsson, B.O. 2002. Genome-scale metabolic model of *Helicobacter pylori* 26695. *J. Bacteriol.* **184:** 4582–4593.

Segre, D., Zucker, J., Katz, J., Lin, X., D'Haeseleer, P., Rindone, W.P., Kharchenko, P., Nguyen, D.H., Wright, M.A., and Church, G.M. 2003. From annotated genomes to metabolic flux models and kinetic parameter fitting. *OMICS* **7:** 301–316.

Selkov Jr., E., Grechkin, Y., Mikhailova, N., and Selkov, E. 1998. MPW: The Metabolic Pathways Database. *Nucleic Acids Res.* **26:** 43–45.

Seressiotis, A. and Bailey, J.E. 1988. MPS: An artificially intelligent software system for the analysis and synthesis of metabolic pathways. *Biotechnol. Bioeng.* **31:** 587–602.

Stephanopoulos, G. 2002. Metabolic engineering by genome shuffling. *Nat. Biotechnol.* **20:** 666–668.

Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T.S., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J.C., et al. 1999. E-CELL: Software environment for whole-cell simulation. *Bioinformatics* **15:** 72–84.

Valdes, J., Veloso, F., Jedlicki, E., and Holmes, D. 2003. Metabolic reconstruction of sulfur assimilation in the extremophile *Acidithiobacillus ferrooxidans* based on genome analysis. *BMC Genomics* **4:** 51.

Van Dien, S.J. and Lidstrom, M.E. 2002. Stoichiometric model for evaluating the metabolic capabilities of the facultative methylotroph *Methylobacterium extorquens* AM1, with application to reconstruction of C(3) and C(4) metabolism. *Biotechnol. Bioeng.* **78:** 296–312.

Van Dien, S.J., Strovas, T., and Lidstrom, M.E. 2003. Quantification of central metabolic fluxes in the facultative methylotroph *Methylobacterium extorquens* AM1 using ^{13}C-label tracing and mass spectrometry. *Biotechnol. Bioeng.* **84:** 45–55.

Varner, J. and Ramkrishna, D. 1999. Mathematical models of metabolic pathways. *Curr. Opin. Biotechnol.* **10:** 146–150.

Web site references

http://fenske.che.psu.edu/Faculty/CMaranas/pubs.html; Universal reaction database.

OptStrain: A computational framework for redesign of microbial production systems

Priti Pharkya, Anthony P. Burgard and Costas D. Maranas

Supplemental Material	http://genome.cshlp.org/content/suppl/2004/11/02/14.11.2367.DC1
References	This article cites 49 articles, 14 of which can be accessed free at: http://genome.cshlp.org/content/14/11/2367.full.html#ref-list-1
License	
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**