Metabolic Engineering 14 (2012) 672–686

Contents lists available at SciVerse ScienceDirect

Metabolic Engineering

journal homepage: www.elsevier.com/locate/ymben

Minireview

Mathematical optimization applications in metabolic networks

Ali R. Zomorrodi [a], Patrick F. Suthers [a], Sridhar Ranganathan [b], Costas D. Maranas [a,*]

[a] Department of Chemical Engineering, The Pennsylvania State University, University Park, PA, United States
[b] Huck Institutes of Life Sciences, The Pennsylvania State University, University Park, PA, United States

ARTICLE INFO

Article history:
Received 22 June 2012
Received in revised form
31 August 2012
Accepted 14 September 2012
Available online 28 September 2012

Keywords:
Mathematical optimization
Metabolic model
Metabolic network analysis
Metabolic network redesign

ABSTRACT

Genome-scale metabolic models are increasingly becoming available for a variety of microorganisms. This has spurred the development of a wide array of computational tools, and in particular, mathematical optimization approaches, to assist in fundamental metabolic network analyses and redesign efforts. This review highlights a number of optimization-based frameworks developed towards addressing challenges in the analysis and engineering of metabolic networks. In particular, three major types of studies are covered here including exploring model predictions, correction and improvement of models of metabolism, and redesign of metabolic networks for the targeted overproduction of a desired compound. Overall, the methods reviewed in this paper highlight the diversity of queries, breadth of questions and complexity of redesigns that are amenable to mathematical optimization strategies.

1. Introduction

Genome-scale metabolic models and other metabolic network descriptions provide convenient ways of summarizing and codifying information known about the metabolism of an organism. The last decade has witnessed a rapid increase in the number of reconstructed genome-scale metabolic models for a wide range of microorganisms including eukaryotic, prokaryotic and archaeal species (Aziz et al., 2008; Feist et al., 2009; Meyer et al., 2008; Thiele and Palsson, 2010). These models enable the mathematical representation of the bio-transformations and metabolic processes occurring within the organism. Such models can thus be analyzed and probed using a growing toolbox of mathematical optimization based methods. These include a large number of methodologies involved in phenotype predictions such as identifying flux bounds, flux coupling, elucidation of objective functions, predictions of gene/reaction essentiality and synthetic lethality and metabolic flux analysis. Many of these analyses require only the network as input, whereas others, such as metabolic flux analysis (MFA), require experimental data for metabolic flux elucidation.

Genome-scale metabolic models, even for well-characterized microorganisms, are inherently incomplete due to incorrect or poor annotations, missing reactions/pathways, incorrect or missing regulatory constraints and inaccurate formulation of the biomass reaction. These imperfections of metabolic network models are manifested by the presence of gaps, incorrect growth phenotype predictions, inaccurate flux predictions and thermodynamically infeasible cycles (Nigam and Liang, 2007; Orth and Palsson, 2010; Satish Kumar et al., 2007; Schellenberger et al., 2011a; Yang et al., 2005). Network gaps prevent metabolic flow in one or more reactions due to the lack of a connecting path with the rest of metabolism (Satish Kumar et al., 2007). Annotation-based draft reconstructions typically involve hundreds of reactions that cannot carry any metabolic flux due to the presence of network gaps. It is thus important to first identify and subsequently attempt to restore flow to these reactions to enable the function of physiologically relevant pathways, debottleneck the biomass equation, and provide production/down-conversion (or import/export) routes for all metabolites in the system.

Gene mutant growth phenotype predictions serve as a key quality metric for metabolic models. Metabolic capabilities of a given microorganism may be over-estimated by its metabolic model if the model predicts growth for a mutant strain whereas experiments show no growth, or they can be under-estimated when the model predicts that one or more biomass components cannot be synthesized (i.e., no growth) even though experiments demonstrate growth. In the former case, some of the biochemical capabilities of the model must be eliminated or placed under a regulatory control whereas in the latter additional functionalities need to be added (as in gap filling process). Therefore, a related metabolic model correction task involves restoration of consistency with grow/no grow experiments (Kumar and Maranas, 2009). Mismatches between the predicted reaction fluxes using approaches such as flux balance analysis (FBA) and *in vivo* flux measurements (e.g., for growth rates, substrate uptake and

* Corresponding author. Fax: +814 865 7846.
E-mail address: costas@psu.edu (C.D. Maranas).

1096-7176/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ymben.2012.09.005

byproduct secretion rates and intracellular flux data using ^{13}C labeling experiments) serve as another way of identifying metabolic model limitations. The inaccurate flux predictions may arise due to missing or falsely present (e.g., poorly annotated) reactions or pathways in the model. Thus another relevant model refinement task involves reconciliation of *in silico* flux predictions and *in vivo* flux measurements by identifying the functionalities that need to be added or removed to/from the model.

Many genome-scale metabolic reconstructions also contain, to varying degrees, a number of thermodynamically infeasible reactions/pathways. For example, a loop such as $A \rightarrow B \rightarrow C \rightarrow A$ is thermodynamically infeasible as it can operate in perpetuity with zero driving force thus violating the second law of thermodynamics. Such loops manifest their presence in metabolic models by carrying unbounded metabolic fluxes when using FBA even for finite substrate inputs, and can be problematic particularly when they allow for ATP production or electron exchange across different electron carrier pools. Identification and removal of these loops can lead to more physiologically relevant predictions by metabolic models. Metabolic model pathologies described above may cause significant errors in the estimation of maximum theoretical yields, identification of alternate production pathways or redox balancing. Because of the relative large number of these pathologies (gaps, missing or incorrect functionalities and regulatory constraints and thermodynamically infeasible pathways in draft metabolic models) and the diversity of ways these can be remedied, computational approaches that attempt to refine and correct metabolic models are becoming imperative. In particular, optimization-based methods provide a straightforward way of identifying alternative hypotheses for bridging the gaps, reconciliation of the growth and flux prediction inconsistencies, and restoration of the thermodynamic feasibility by minimally perturbing the original network while preserving all correct model predictions. It is important to note that optimization-based predictions should be interpreted as mere suggestions for fixing gaps rather than as an unsupervised tool for completing metabolic models. In addition, the system-wide impact of additions to the model must be carefully assessed as in many cases they may cause other network pathologies (e.g., formation of thermodynamically infeasible cycles).

On another front, a key objective in metabolic engineering is to improve the production of metabolites with commercial value by modifying the genetic setup of microorganisms i.e., knock outs/ins/ups/downs. These manipulations can be achieved by using recombinant DNA techniques that allows one to delete, alter or add a new genetic function to the microbial host. Disrupting the function of a gene by site-directed mutagenesis (Park et al., 2007) or homologous recombination (Yu et al., 2000) enables the shutting-down of undesirable metabolic pathways in a microbial production host. In addition to gene knockouts, adding non-native genes or pathways to microbial hosts can lead to new capabilities and/or increased biochemical yields. For example, the heterologous expression of pyruvate carboxylase (*pyc*) from *Rhizobium etli* and *Lactococcus lactis* was shown to increase succinate synthesis by adding a new anaplerotic reaction that converts phosphoenolpyruvate to citric acid cycle intermediates such as succinate (Thakker et al., 2011; Wang et al., 2006, 2009). Alternatively, the addition of new pathways to a microorganism can expand the range of native metabolism by enabling the production of a new biochemical compound. Examples include production of 1-butanol (Dietrich et al., 2010; Liu and Khosla, 2010; Shen et al., 2011) and 1,4-butanediol (Yim et al., 2011) in *Escherichia coli*. In addition to genetic interventions that affect the coded amino acids, efforts have targeted codon usage as a mechanism for achieving a desired level of gene expression (Dong et al., 1995; Makrides, 1996). Furthermore, with the advent of synthetic biology tools such as the ribosome binding site calculator (Salis et al., 2009), it is possible to fine tune enzyme levels in metabolic pathways to maximize throughput.

Increasing demands for the sustainable and economically optimized synthesis of bio-products, energy requirements and environmental concerns necessitated the design of microbial strains that can produce valuable bio-chemicals (e.g. biofuels) at near theoretical maximum yields. Motivated by this need, computational approaches for identifying targets for genetic intervention have started to emerge in response to recent advances in the reconstruction of metabolic models. In particular, optimization-based approaches have been used extensively to guide all the aforementioned engineering strategies by efficiently identifying the best candidates for knockouts/ins/ups/downs in a genome-scale metabolic network leading to the overproduction of a desired chemical. These computational methods have been employed for a wide range of biotechnological and biomedical applications including identification of promising biochemical routes for overproduction of value-added chemicals from microbes (Alper et al., 2005; Atsumi et al., 2008a; Atsumi and Liao, 2008a; Bond and Lovley, 2003; Burgard et al., 2003; Misawa et al., 1991; Nakamura and Whited, 2003; Oliveira et al., 2005; Pharkya and Maranas, 2006; Sauer et al., 2008; Scott et al., 2007), understanding disease metabolism (Bosma, 2003; Danpure, 2006; Zelezniak et al., 2010) and pinpointing drug targets (Jamshidi and Palsson, 2007; Lee et al., 2005).

In this review we highlight in detail mathematical optimization applications aimed at (i) querying model predictions, (ii) correcting metabolic models and (iii) suggesting ways for redesigning metabolic networks in response to overproduction demands. All these optimization applications are relevant to metabolic engineering by assessing and correcting models of metabolism to support model-guided strain design.

2. Mathematical optimization terminology and taxonomy

Each optimization problem is composed of an objective function and a set of constraints. The objective (or fitness) function is mathematical description of the desired property of the system that should be maximized or minimized. Constraints are a set of equations and/or inequalities describing the space of all eligible possibilities for the problem of interest from which an optimal solution (i.e., maximizing/minimizing the objective function) is selected. An optimization problem may contain continuous variables, discrete (integer) variables or both. The former is used to describe the continuous properties of the system (e.g., concentrations, reaction fluxes, etc.) whereas the latter is employed to capture discrete decisions (e.g., how many reactions should be added to a metabolic model). Binary variables are specific types of integer variables, which can only take a value of zero or one. This type of variables has been extensively used in a wide array of applications involving a 'yes/no' decision making process (e.g., to knockout or not knockout a reaction in a metabolic model). If the objective function and all constraints in an optimization problem are linear, it is referred to as a linear program (LP), whereas if there is at least one non-linear term (e.g., multiplication of two variables, powers of variables, exponential of a variable, etc.) it is termed a nonlinear program (NLP). In addition, if an optimization problem contains both continuous and integer (or binary) variables it is called a mixed-integer (linear or nonlinear) program (MILP or MINLP). For MILP or MINLP problems a rank-ordered list of all alternate optimal (and sub-optimal) solutions can be enumerated by accumulated constraints that exclude from consideration the previously found solutions (i.e., integer cuts). Another group of optimization problems, which has been used extensively in the analysis of metabolic networks, is bilevel

programs. These are nested optimization problems where one (or more) of the constraints is another optimization problem. Bilevel programs can involve both continuous and integer variables and may contain linear or nonlinear constraints and objective function.

3. Using optimization to explore model predictions

In this section, we highlight a number of optimization methods that are used to examine metabolic model predictions (see also Table 1). These algorithms and methods are designed to be tractable for genome-scale metabolic models spanning hundreds to thousands of reactions and metabolites. They cover a wide range of topics, including assessing the maximum theoretical yields implied by the model, classifying the relations between reactions, examining the impact of genetic modifications, and quantifying fluxes given experimental data.

3.1. Flux balance analysis

A central optimization task in metabolic networks is flux balance analysis (FBA) (Fell and Small, 1986; Savinell and Palsson, 1992; Varma and Palsson, 1993)). The most attractive feature of FBA is its ability to make quantitative predictions about a metabolic network without any need for detailed kinetic descriptions and given only the stoichiometry of reactions. The only necessary inputs for FBA are the metabolic model (i.e., the network stoichiometry), a biological objective and the growth and environmental conditions (substrate availability). The fundamental assumption underlying FBA is that the system is at steady-state. The steady-state mass balance equation for each metabolite and environmental and growth conditions are mathematically described in the form of constraints for the optimization problem. Given that the system of equations describing the steady-state mass balances is usually underdetermined (i.e., more reactions than metabolite) one needs to select from among infinite feasible flux distributions the ones satisfying a desired objective function. A maximization principle is thus used as a surrogate for the true (and always unknown) totality of interactions. Typically this objective function is the maximization of the flux through the biomass formation reaction (Varma and Palsson, 1994). This results in a linear programming (LP) formulation that can be readily solved using a variety of software packages such as GAMS or MATLAB, or metabolic modeling frameworks such as the constrained-based modeling and analysis (COBRA) toolbox (Becker et al., 2007; Schellenberger et al., 2011b). The COBRA toolbox is available at opencobra.sourceforge.net. A useful primer for FBA was recently published (Orth et al., 2010).

FBA is somewhat limited in its predictive power unless additional constraints are appended to the optimization problem. For example, constraints can be added to disable specific reactions that are inactive due to the regulatory constraints under a given environmental or growth condition. This can be done by setting the bounds on the affected fluxes to zero. Similarly, substrate availability can be imposed by setting the bounds on the uptake fluxes corresponding to substrates that are not available to zero. By extending this systematic approach to the entire metabolic network, techniques such as regulatory FBA (rFBA) (Covert et al., 2001), steady-state regulatory FBA (SR-FBA) (Shlomi et al., 2007), gene inactivity moderated by metabolism and expression (GIMME) (Becker and Palsson, 2008), and probabilistic regulation of metabolism (PROM) (Chandrasekaran and Price, 2010) have been developed to integrate regulatory information with the metabolic network. Alternatively, phenotype phase planes (Edwards et al., 2001) have been used to map out the optimal metabolic flux distributions as they vary with two fluxes (such as oxygen uptake and carbon substrate uptake fluxes) and highlight different patterns of pathway utilization. The scope of FBA has been also recently extended to multi-species microbial systems (Bizukojc et al., 2010; Bordbar et al., 2011; Lewis et al., 2010; Stolyar et al., 2007; Zomorrodi and Maranas, 2012).

3.2. Objective function elucidation

As noted earlier, for FBA an objective function is needed and typically the maximization of biomass is used as the objective

Table 1
Optimization algorithms to explore model predictions.

Name	Task	Accessibility	Reference
Flux balance analysis (FBA)	Explore metabolic capabilities of an organism	Various, including MATLAB (via COBRA toolbox)	Fell and Small (1986), Savinell and Palsson (1992), Varma and Palsson (1993)
rFBA	Extend FBA to account for regulation	N/A	Covert et al. (2001)
SR-FBA	Extend FBA to account for regulation	N/A	Shlomi et al. (2007)
GIMME	Identify inactive reactions in a metabolic model under a particular condition or in a specific cell type	GAMS	Becker and Palsson (2008)
PROM	Integrate transcriptional regulatory networks and constraint-based modeling	N/A	Chandrasekaran and Price (2010)
MOMA	Calculate fluxes in response to genetic modifications	N/A	Segre et al. (2002)
ROOM	Calculate fluxes in response to genetic modifications	N/A	Shlomi et al. (2005)
ObjFind	Identify hypothesized objective functions consistent with experimental flux measurements	N/A	Burgard and Maranas (2003)
BOSS	Identify hypothesized objective functions consistent with experimental flux measurements	N/A	Gianchandani et al. (2008)
FVA	Determine minimum and maximum flux values for each reaction in the network	N/A	Mahadevan and Schilling (2003)
fastFVA	Determine minimum and maximum flux values for each reaction in the network	N/A	Gudmundsson and Thiele (2010)
FCF	Identify flux couplings in a reaction network	GAMS	Burgard et al. (2004)
FFCA	Identify flux couplings in a reaction network	MATLAB	David et al. (2011)
F2C2	Identify flux couplings in a reaction network	MATLAB	Larhlimi et al. (2012)
SL Finder	Identify synthetic lethals in a genome-scale model	GAMS	Suthers et al. (2009b)
OptMeas	Determine measurement sets that enable flux elucidation	C	Chang et al. (2008)
13C-FLUX	MFA calculations	N/A	Wiechert et al. (2001)
COBRA toolbox	FBA and MFA calculations, essentiality and synthetic lethality analysis (and more)	MATLAB	Schellenberger et al. (2011b)

(Varma and Palsson, 1994). Other studies have examined the use of alternative objective functions. For example, minimization of metabolic adjustments (MOMA) (Segre et al., 2002) minimizes the sum of the squared differences between the original wild-type and altered flux distributions due to gene/reaction knockouts leading to a convex quadratic optimization formulation with linear constraints. Alternatively, regulatory on/off minimization (ROOM) (Shlomi et al., 2005) minimizes the number of changes in fluxes in response to the altered flux and thus requires the definition of binary variables yielding a mixed integer linear programming (MILP) formulation.

A number of different studies have evaluated and compared various hypotheses for the cellular objective function (Feist and Palsson, 2010; Knorr et al., 2007; Ow et al., 2009; Pramanik and Keasling, 1997; Savinell and Palsson, 1992; Schuetz et al., 2007). Alternatively, others have developed optimization-based algorithms to systematically identify or predict a physiologically relevant objective function using experimental flux data (Burgard and Maranas, 2003; Gianchandani et al., 2008). Among these is ObjFind, which is a bilevel optimization procedure used to identify hypothesized objective functions that are most consistent with experimental flux measurements (Burgard and Maranas, 2003). More recently, the biological objective solution search (BOSS) was introduced that allows for discovery of objectives with previously unknown stoichiometry (Gianchandani et al., 2008). To this end, it generates a putative stoichiometric objective reaction and adds this reaction to the existing set of stoichiometric constraints of the metabolic model. It then maximizes flux of this putative objective reaction, while minimizing the difference between the computed flux distribution and available experimental flux data. The stoichiometric coefficients of the objective reaction are varied and the results are clustered, and the most populous cluster is picked as the objective function.

3.3. Flux variability analysis

Linear optimization can be used not only for identifying the maximum biomass (or product) yield but also to identify metabolic flux variability. Using flux variability analysis (FVA) (Mahadevan and Schilling, 2003) the minimum and maximum flux values for each reaction in the network can be determined. This corresponds to the smallest hyper-rectangle that entirely contains the feasible region. Additional constraint(s) can be imposed in flux variability analysis as well (Mahadevan and Schilling, 2003). For example, a constraint can be added to ensure that the biomass flux is at its maximal or at some sub-optimal value (e.g., 90% of maximum). These new optimization problems remain linear and can simply be iterated over all fluxes in the network. The computational time involved in this analysis can be improved by taking advantage of the fact that the feasible region for this set of problems does not change (the objective function is the only change), as is done in fastFVA (Gudmundsson and Thiele, 2010). In addition to exploring alternate optima (Mahadevan and Schilling, 2003), FVA has been also used to investigate network redundancy and flexibility (Thiele et al., 2010).

3.4. Flux coupling analysis

Metabolite balance constraints imply that not all fluxes in a metabolic network can vary independently. One of the first methods for determining these relationships is flux coupling analysis (Burgard et al., 2004). Reactions that cannot carry any flux under steady-state conditions (i.e., its only feasible solution is zero) are referred to as blocked. They can be identified in a straightforward manner by simply maximizing the flux through the reaction and denoting whether it can assume a non-zero value. Non-blocked reactions in a network can engage with a number of different relations with one another such as fully, partially, or directionally coupled. Reactions are fully coupled if a non-zero value for one fixes the other at exactly one non-zero value (i.e., the ratio of the two fluxes is always constant). Reactions are partially coupled if there is some variability in their ratio spanning only non-zero values. Reactions are directionally coupled if a non-zero flux for one implies that the other is also non-zero (but not the reverse). Other reaction pairs are referred to as uncoupled. The flux coupling finder (FCF) (Burgard et al., 2004) solves a sequence of LP problems to determine the coupling of reactions. During the analysis of the results, reactions that are mutually fully or partially coupled can be arranged into coupled reaction sets forming sub-networks (Marashi et al., 2012). It is important to note that flux coupling is sensitive to missing reactions (Marashi and Bockmayr, 2010). FCF splits reversible reactions into forward and backward reactions. Larhlimi and Bockmayr (2006) showed that reversibility-type pruning could be performed to reduce the number of computations as reversible reactions could be coupled together only in certain cases. The concept of elementary flux patterns (EFP), found via a mixed-integer linear program, can also be used to characterize flux coupling relationships, however it is not able to distinguish partial and full coupling. A number of recent methods such as feasibility-based flux coupling analysis (FFCA) (David et al., 2011) and the fast flux coupling calculator (F2C2) (Larhlimi et al., 2012) seek to improve the computation time for performing flux coupling. Tools for analyzing flux coupling relations can be downloaded from http://maranas.che.psu.edu/software (FCF), http://www.bioinformatics.org/ffca/ (FFCA), and https://sourceforge.net/projects/f2c2/files/ (F2C2).

3.5. Essentiality and synthetic lethality analysis

The impact of deleting a reaction on the growth phenotype of the network can be simulated by adding new constraints that set the lower and upper bounds of each affected reaction to zero and maximizing biomass production. Reaction deletions yielding a maximum biomass production of less than a pre-specified viability threshold are considered lethal knockouts and the associated reaction is called essential (for biomass formation). Burgard et al. (2001) developed an algorithm based on MILP, where the concept of essential reactions was used to identify the minimal set of reactions required to support growth under a given uptake condition. Binary variables were used to determine if a reaction should be active and objective function of the optimization problem was to minimize sum of the binary variables.

In order to directly examine the impact of knocking out genes (rather than reactions), gene–protein reaction (GPR) associations are needed. These GPR associations represent which genes produce the required enzyme for a specific reaction and denote if the coded protein is part of an enzyme complex or is an isozyme. The effect of inactivation of a gene on the flux level can be imposed by adding new linear constraints to the optimization problem describing the GPR associations. This enables the assessment of the impact of gene knockouts on biomass production (e.g., essential genes elucidation) and flux distribution in the network. The concept of essentiality can be further extended to multiple gene (or reaction) knockouts. In particular, synthetic lethals (SL) refer to pairs of non-essential genes whose simultaneous deletion is lethal (Guarente, 1993; Novick et al., 1989). Although SL reaction/gene pairs can be identified by brute-force through looping over all reactions/genes, for higher-order SLs (e.g., triples, quadruples, etc.) the computational burden becomes prohibitive. The challenge in exhaustively identifying higher-order SLs lies in the combinatorial complexity of the underlying mathematical

problem and the large size of genome-scale metabolic models. Efforts towards addressing this challenge include an *in silico* multiple knockout investigation (Deutscher et al., 2006) of the *i*FF708 yeast metabolic network (Forster et al., 2003) and a computational approach based on ideas from game theory (Deutscher et al., 2008). Alternatively, Behre et al. (2008) extended their study on single knockouts (Wilhelm et al., 2004) by introducing a generalized framework for analyzing structural robustness of metabolic networks with respect to multiple knockouts based on the concept of elementary flux modes. More recently, the synthetic lethality finder (SL finder) makes use of a bilevel optimization framework that utilizes FBA to completely identify all multi-reaction/gene lethal knockouts using genome-scale models (Suthers et al., 2009b). The inner problem adjusts the fluxes to achieve maximum biomass production, subject to network stoichiometry, reaction/gene deletions imposed by the outer problem and other possible growth and environmental constraints. The outer problem, on the other hand, aims at finding synthetic gene/reaction eliminations (captured by using binary variables) that lower the maximum biomass production below a pre-specified cutoff. In practice, the problem is solved as a single MILP by recasting the inner maximization as a set of additional constraints. A GAMS implementation of the SL finder can be downloaded from http://maranas.che.psu.edu/software.

3.6. Metabolic flux elucidation using labeling data

Optimization applications in metabolic networks described so far are based on linear networks representations. Metabolic flux analysis (MFA) (Vallino and Stephanopoulos, 1993) seeks to determine the fluxes that can best describe the observed distribution of label isotopes as measured using GC/MS and/or MS/NMR. This results in a nonlinear model that links metabolic fluxes with isotope enrichment ratios. Nonlinear optimization is thus typically used to minimize the difference between the experimental data and the predicted labeling patterns. Unlike FBA-based methods, which only need metabolic network connectivity information, MFA requires a description of the fate of atoms going from reactants to products for each reaction in order to make its predictions. MFA has been performed on large-scale models (Suthers et al., 2007) and atom mappings have been generated for genome-scale metabolic models (Ravikirthi et al., 2011). MFA has also extensively been used to analyze mammalian cells (Ahn and Antoniewicz, 2011) and plants (Schwender, 2008). A number of different methods have arisen for representing the atom mappings as well as solving the resulting nonlinear optimization problems.

One of the first such contributions on representing atom mappings is the introduction of atom mapping matrices (AMM) (Zupke and Stephanopoulos, 1994) that track the transfer of carbon atoms from reactants to products. This concept was subsequently generalized in the form of isotopomer mapping matrices (IMM) by (Schmidt et al., 1997). The use of IMMs enables the formulation of all isotopomer mass balances of a metabolite pool into a closed-form nonlinear set of algebraic equations. The variables in these representations include the metabolic fluxes and the isotopomer distribution vectors (IDV) that quantify the fraction of each metabolite being present in a particular isotope form. A potential problem with the use of IMMs is that even for a given flux distribution the identification of the underlying IDVs yields a set of equations, which remain nonlinear. The cumomer concept (Wiechert et al., 1999) was introduced to first prove that there exists a unique IDV assignment that satisfies any given feasible flux distribution and subsequently devise an IDV identification procedure by solving a cascade of linear equations. The use of cumomer balances (i.e., cumulated isotopomer fractions) enables the description of an isotopomer network with a smaller set of transformations without any loss of information. However, the nonlinear coupling between metabolic fluxes and IDVs remains. Also, Forbes et al. (2001), introduced the isotopomer path tracing concept using AMMs to account for the transfer of carbon atoms from reactants to products, which identifies all isotopomer paths that produce an observable isotopomer. An alternative method for reducing the dimensional space of the isotopomer problem is the concept of the theoretical bondomer (van Winden et al., 2002). Here, all C–C bonds that remain intact in the labeled substrate after it enters metabolism are tracked as a single species. By using individual bondomer species instead of every carbon atom the number of variables is significantly reduced. A limitation of this method is that it requires the use of a single, uniformly labeled substrate. The elementary metabolite unit (EMU) framework reduces the number of variables necessary to calculate the mass isotopomer distribution of a measured metabolite (Antoniewicz et al., 2006). It also provides an efficient method for analyzing the labeling pattern from multiple isotopic tracers. A system of linear equations is solved in an iterative manner until the flux values are determined. More recently, the concept of fluxomers was introduced that seeks to reduce the computations even further (Srour et al., 2011).

The above modeling developments link isotopomer fractions (codified through a variety of different variable sets) and metabolic flux information into a set of algebraic equations that allowed for the straightforward identification of labeling distributions given a feasible flux distribution. The next step is to solve for the fluxes that best explain the observable data. Most approaches rely on gradient-based minimization searches that minimize the sum of the squares of the differences between measurements and observations. These include the Levenberg–Marquardt algorithm (Zhao and Shimizu, 2003), the generalized reduced gradient method (Klapa et al., 2003) and trust region methods (Yang et al., 2004). Efforts to decrease the computation time led to the development of analytical derivation techniques for the Jacobian matrix (Wittmann and Heinzle, 2002). Alternatively, heuristic algorithms such as simulated annealing employed by (Schmidt et al., 1999) to avoid being trapped in local minima. Global optimization approaches relying on branch and bound coupled with convex relaxation of the problem have been also used to quantify the fluxes (Ghosh et al., 2005; Riascos et al., 2005). Through the use of BARON (branch-and-reduce optimization navigator) (Tawarmalani and Sahinidis, 2004) a deterministic global optimization package, a small problem involving eleven fluxes was solved using the theoretical abbreviated pathway model of Forbes et al. (2001).

Assessing and quantifying the impact of measurement uncertainty on all obtained solutions is critical. First, linearized statistics were used to provide confidence intervals for the split ratio into the pentose phosphate pathway (Dauner et al., 2001; Emmerling et al., 2002; Wiechert and de Graaf, 1997). In addition, Monte-Carlo stochastic simulation has been used to examine how the elucidated fluxes change when uncertainty in the form of normally distributed noise is added to the data (Forbes et al., 2001; Wittmann and Heinzle, 2002; Zhao and Shimizu, 2003). Optimization-based techniques were also introduced that allow for non-uniform determination of the upper and lower bounds (Antoniewicz et al., 2006). Because MFA requires experimental data as inputs, it is not trivial to design experiments that will give sufficient information to determine fluxes in the network. Optimization-based techniques can be used on this front. The OptMeas method was proposed to determine measurement sets that enable flux elucidation using incidence structure analysis (Chang et al., 2008). The original implementation relied on an IDV description to track isotope labeling and was improved to use the

EMU framework (Suthers et al., 2009a), which required significantly less computation time. Recently, methods for tracer selection (Crown and Antoniewicz, 2012) and experimental design (Crown et al., 2012) were introduced. Software relying on nonlinear optimization for MFA can be downloaded from www.che.udel.edu/mranton/metran.html (METRAN) and at www.13cflux.net (13C-FLUX) (Wiechert et al., 2001). The COBRA Toolbox (Becker et al., 2007; Schellenberger et al., 2011b) is also capable of performing MFA.

3.7. Identification of new metabolic routes to products

Graph-based techniques have been used extensively to identify all possible metabolic pathways between a source and end node. For example, Hatzimanikatis et al. (2005) introduced a novel procedure termed biochemical network integrated computational explorer (BNICE) that utilizes graph-based representations of biochemical and enzyme reaction rules to generate synthesis pathways by successively applying reaction chemistry operators. The distinguishing feature of BNICE is the ability to propose novel enzymatic activities operating on sometimes unseen before compounds. Identification of novel biochemical pathways to 1,4-butanediol (Yim et al., 2011) and 1,2,4-trichlorobenzene (Finley et al., 2010) are some of the examples of successful implementations of the BNICE framework. Other graph-based techniques such as PathMiner (McShan et al., 2003), PathComp (Kanehisa et al., 2006), Pathway Tools (Karp et al., 2010), MetaRoute (Blum and Kohlbacher, 2008), PathFinder (Goesmann et al., 2002) and UM-BBD pathway prediction system (Ellis et al., 2006) have been also proposed for pathway mining. Ranganathan and Maranas (2010) introduced an adaptation of Yen's *k*-shortest path algorithm to identify all possible pathways from a given starting metabolite to a target molecule. The algorithm starts by identifying the shortest pathway between two selected compounds and subsequently explores the combinatorial space by identifying alternate shortest pathways between them. The algorithm identified known pathways from pyruvate to 1-butanol (Atsumi et al., 2008b; Atsumi and Liao, 2008b; Shen and Liao, 2008; Steen et al., 2008) in addition to new ones. Optimization based methods for pathway design allow for the incorporation of stoichiometry thereby excluding paths that cannot operate under steady-state conditions. In addition, they allow for the direct assessment of product yield, cofactor balancing and branching pathways. In earlier efforts (Beasley and Planes, 2007; Planes and Beasley, 2009) node connectivity was modeled using an MILP approach to trace paths between a source and a sink metabolite. More recently, the network stoichiometry was directly incorporated into the MILP description and the impact of not considering stoichiometry in pathway prospecting was quantified (Pey et al., 2011).

4. Using optimization to correct/improve models of metabolism

In the following we review selected optimization procedures used to curate genome-scale metabolic models by (i) pinpointing and filling network gaps, (ii) identifying and fixing growth and (iii) flux prediction inconsistencies and (iv) restoring thermodynamic feasibility. Table 2 and Fig. 1 summarize the methods described in this section.

4.1. Identifying and bridging gaps in metabolic models

4.1.1. GapFind and GapFill

GapFind is an optimization-based procedure, which identifies dead-end metabolites in a metabolic reconstruction (Satish Kumar et al., 2007). These metabolites are categorized as *root no production* or *root no consumption* if they cannot be produced or consumed by any reaction (including uptake/export reactions) in the model, respectively. The lack of flow to/from root problem metabolites is propagated in the network thus preventing additional metabolites from being produced (or consumed). These metabolites are termed downstream no-production (or upstream no-consumption) metabolites (see Fig. 1A). Even though it is straightforward to locate root problem metabolites in the model by inspection of the network topology (i.e., stoichiometric matrix), identifying downstream or upstream problem metabolites is not always an easy task (Orth and Palsson, 2010). GapFind is a mixed-integer linear program that identifies such network pathologies. For the case of no-production metabolites, a binary variable is assigned to each metabolite with a value of one signifying that there exists at least one production route and a value zero a lack thereof. GapFind maximizes the sum of these binary variables over all metabolites (subject to stoichiometry and uptake/export constraints) thereby identifying all metabolites that can (or cannot) be produced in the network. The above procedure is adjusted in a straightforward manner to identify the set of all no-consumption metabolites in the network.

Once all problem metabolites in the network are identified, another mixed-integer linear program (i.e., GapFill) bridges gaps one at a time (Satish Kumar et al., 2007). Three different gap

Table 2
Optimization algorithms to correct/improve models of metabolism.

Name	Task	Accessibility	Reference
GapFind	Identify dead-end metabolites	GAMS and MATLAB (via COBRA toolbox)	Satish Kumar et al. (2007)
GapFill	Bridge the gaps (i.e., dead-end metabolites)	GAMS and MATLAB (via COBRA toolbox)	Satish Kumar et al. (2007)
SMILEY	Reconcile growth prediction inconsistencies (single gene mutations: NGGs)	MATLAB (via COBRA toolbox)	Reed et al. (2006)
GrowMarch	Reconcile growth prediction inconsistencies (single gene mutations: NGGs and GNGs)	GAMS	Satish Kumar et al. (2007)
Modified GrowMatch	Reconcile growth prediction inconsistencies (single gene mutations: NGGs and GNGs)	N/A	Henry et al. (2009)
Extended GrowMatch	Reconcile growth prediction inconsistencies (double and higher order gene mutations)	N/A	Zomorrodi and Maranas (2010)
GeneForce	Identify and correct transcriptional regulatory rules using growth prediction inconsistencies (single gene mutations: NGGs)	N/A	Barua et al. (2010)
OMNI	Reconcile flux prediction inconsistencies	N/A	Herrgard et al. (2006)
TMFA	Eliminate thermodynamically infeasible pathways and loops	N/A	Henry et al. (2007)
ll-COBRA	Eliminate thermodynamically infeasible loops	MATLAB (via COBRA toolbox)	Schellenberger et al. (2011a)

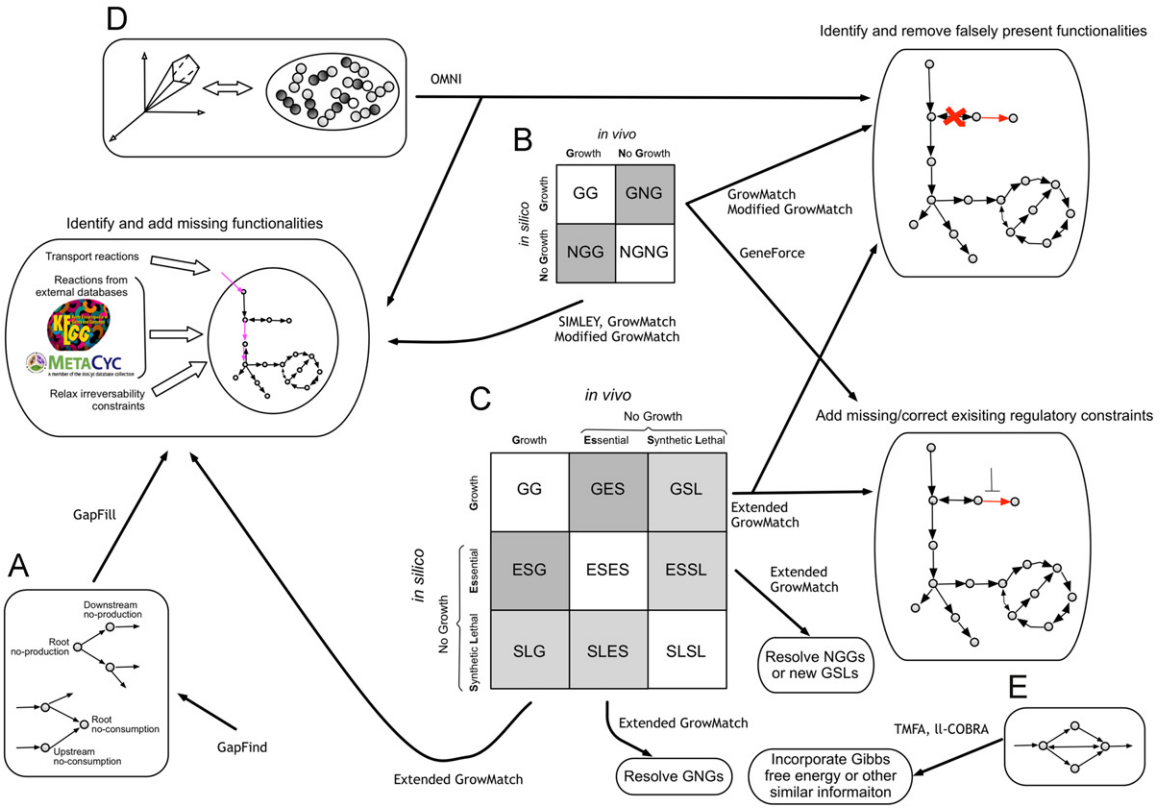

Fig. 1. Different types of metabolic model pathologies, strategies to fix them and the relevant optimization algorithms. (A) Dead-end metabolites. (B) Growth prediction inconsistencies due to single gene mutations. (C) Growth prediction inconsistencies due to multiple gene mutations. (D) Flux prediction inconsistencies. (E) Thermodynamic infeasible loops or pathways.

filling model modifications are considered: (i) relaxing the irreversibility constraint on existing reactions in the network, (ii) adding import (or export) pathway and/or inter-compartment transport reactions to the model, (iii) adding new reactions from an external database such as MetaCyc (Caspi et al., 2006) or KEGG (Kanehisa et al., 2008). Binary variables are used here to model if a specific modification should be made to the model (Satish Kumar et al., 2007) and integer-cuts are employed to identify alternative solutions. This optimization problem aims to identify the minimal number of modifications needed to fill each gap by minimizing the sum of the binary variables as the objective function. The resolution hypotheses generated by GapFill must be verified using Gibbs free energy estimates of the reactions, BLAST analysis, literature searches, etc. (Satish Kumar et al., 2007). GapFind and GapFill algorithms have been applied to both single-compartment models such as *i*JR904 model of *E. coli* (Reed et al., 2003), *i*CM925 model of *Clostridium beijerinckii* (Milne et al., 2011) and multi-compartment models such as *i*ND750 model of *Saccharomyces cerevisiae* (Duarte et al., 2004) and *i*RS1563 of zea mays (Saha et al., 2011). Prototype implementations in GAMS and MATLAB of GapFind and GapFill can be accessed at http://maranas.che.psu.edu/software.htm and COBRA toolbox (Becker et al., 2007; Schellenberger et al., 2011b), respectively.

4.2. Reconciliation of growth prediction inconsistencies

4.2.1. SMILEY

As noted earlier, flux balance analysis (FBA) can be used to predict the growth (or lack thereof) for a microorganism under a given environmental condition. These predictions can be compared with available experimental growth data to test the accuracy of the metabolic model and identify any false predictions by the model. SMILEY is the first algorithm (Reed et al., 2006) proposed to reconcile *in silico* predictions and *in vivo* observations for the case when experimental data show that a strain can grow on a specific substrate and medium, while FBA predicts that it cannot. These mismatches usually hint at reactions or functionalities that are missing in the model. Similar to GapFill (Satish Kumar et al., 2007), SMILEY is based on a MILP to identify the minimal number of reactions that need to be added to the model from a universal database such as KEGG (Kanehisa et al., 2008) to restore growth in the network. This algorithm was applied to the *i*JR904 model of *E. coli* (Reed et al., 2003), where the FBA predictions on a wide range of growth media with various carbon and nitrogen sources were compared against the available experimental growth data from Biolog (http://www.biolog.com/). The experimental verification of the generated hypotheses led to the functional assignment of eight ORFs with two new enzymatic activities and four transport functions (Reed et al., 2006). A MATLAB implementation of SMILEY can be accessed through the COBRA toolbox (Becker et al., 2007; Schellenberger et al., 2011b).

4.2.2. GrowMatch

A typical test of the accuracy of genome-scale metabolic models is to contrast the *in silico* growth phenotype of single mutant strains with available experimental gene essentiality data (Thiele and Palsson, 2010). This comparison can lead to two types of inconsistencies between model and experiment: NGG, where the model predicts that the gene deletion is lethal (*No* Growth), while experiments show a viable mutant strain (Growth), and GNG where the model predicts growth (G) while experimental observations show a lethal effect (*No* Growth) (see Fig. 1B). The SMILEY algorithm described earlier aims to fix NGG inconsistencies.

GrowMatch is another optimization-based algorithm for suggesting ways to resolve both NGG and GNG mismatches of a metabolic model (Kumar and Maranas, 2009). It reconciles NGG and GNG mismatches using two separate MILPs. NGG inconsistencies allude to the reactions and capabilities that are missing in the model. Therefore, GrowMatch employs the same three mechanisms used in GapFill (Satish Kumar et al., 2007) (i.e., relax the irreversibility constraints on existing reactions in the network, add import/export pathways and/or inter-compartment transport reactions, add external reactions from a universal database) to restore *in silico* growth for the mutant network and convert each NGG to GG (i.e., Growth–Growth) matches one at a time.

GNG inconsistencies on the other hand arise due to biochemical capabilities that are falsely present or are not appropriately constrained in the metabolic model (e.g., due to the lack of regulatory constraints). The one-by-one resolution of GNGs relies on a bilevel optimization formulation to identify the minimal number of reaction suppressions in the model (under the examined condition) that lowers the maximum biomass yield in the network below a pre-specified threshold. Binary variables are used to identify the reactions that need to be suppressed. These binary variables act as parameters for the inner problem, where the maximum biomass yield is identified. The objective function of the outer-problem is the minimization of the maximum biomass yield. Similarly to GapFill (Satish Kumar et al., 2007) and SMILEY (Reed et al., 2006) integer cuts are used to find all alternative solutions. The identified reaction suppressions by GrowMatch may suggest that these reactions should not be present in the network (e.g., due to poorly annotated genes) or they should be constrained under the examined environmental and/or growth conditions (e.g., due to regulatory constraints) in order to fix the inconsistency.

The resolution hypotheses generated by GrowMatch for NGGs and GNGs can be conditional or global (Kumar and Maranas, 2009). Global modifications are those that do not invalidate any correct model predictions, whereas conditional modifications resolve an inconsistency at the cost of creating additional mismatches. For example, resolution of a NGG may convert some NGNGs (i.e., both model and experiment imply *N*o *G*rowth) to new GNGs, or resolution of a GNG may convert some GGs (i.e., both model and experiment imply *G*rowth) to new NGGs. Therefore, GrowMatch typically rejects all conditional modifications and keeps only global ones. It is worth mentioning that even the global modifications are needed to be first verified manually using the literature resources or experimental data before incorporating into the model. We note that even though SMILEY and GrowMatch do not deal directly with dead-ends in the network, the generated hypotheses by these algorithms may lead to filling some of these gaps automatically. Caution must be exercised not to add any thermodynamically infeasible cycles to the models. A GAMS implementation of GrowMatch can be downloaded from http://maranas.che.psu.edu/software.htm.

4.2.3. Modified GrowMatch

A modified form of GrowMatch was proposed by Henry et al. (Henry et al., 2009) to identify more biologically relevant modifications to reconcile *in silico*/*in vivo* growth inconsistencies. To this end, the objective function of GrowMatch for NGGs was changed to a weighted sum of the binary variables (instead of simply their sum). The weights quantify a penalty associated with adding the corresponding reaction to the model. Addition of reactions that are not in the KEGG database, involve metabolites with unknown structures as well as reactions for which $\Delta_r G^0$ cannot be estimated, or those operating in thermodynamically unfavorable directions are more heavily penalized using these weights. Similarly, GrowMatch for resolution of GNGs is modified such that removal of an irreversible reaction from the network is penalized if it is associated with at least one gene in the model (Henry et al., 2009). Instead of relying on the concept of conditional and global modifications used in the original GrowMatch procedure, Henry et al. (2009) suggested two new optimization-based procedures called gap filling and gap generation reconciliations to improve the performance of the entire model after incorporating all suggested modifications. The gap filling reconciliation step aims to maximize the correction of NGGs, while minimizing the number of modifications to the model as well as minimizing the number of newly emerged GNGs. Alternatively, the gap generation reconciliation step maximizes the correction of GNGs with the minimum number of modifications to the model, while minimizing the addition of new NGGs. These procedures were applied to the reconstruction of the genome-scale metabolic model of *Bacillus subtilis* (*i*Bsu1103) based on the SEED annotations (Overbeek et al., 2004).

4.2.4. Extended GrowMatch

The original GrowMatch procedure relies on gene essentiality data to identify and reconcile the model/experiment mismatches. GrowMatch was recently extended to make use of the available data for higher order gene deletion experiments, (i.e., synthetic lethality data) (Zomorrodi and Maranas, 2010) as they have been suggested to provide additional layers for curation and validation of metabolic models (Harrison et al., 2007; Suthers et al., 2009b). Comparison of the predicted synthetic lethal predictions with the available double gene deletion data reveals a number of additional ways the model and experiment can disagree (see Fig. 1C). Notably, the 'no growth' phenotype in this case can be due to either essentiality (ES) or synthetic lethality (SL) of single or double gene deletions, respectively. For example, a GSL inconsistency refers to a mismatch where the model predicts that the simultaneous deletion of both genes is not lethal (i.e., Growth), while the experimental observations show that they form a synthetic lethal pair (SL). Similarly, SLES inconsistency refers to a case where, the two genes form a synthetic lethal (SL) *in silico*; however, one (or both) of the genes was reported to be essential (ES) *in vivo* (Zomorrodi and Maranas, 2010). Extended GrowMatch also targets another type of inconsistency called auxotrophy mismatches, where, the essentiality or synthetic lethality of single or double gene perturbations are in agreement with experimental data, but *in silico* predictions for supplementation rescue (i.e., auxotrophy) are inconsistent with *in vivo* observations. The optimization formulations for the resolution of GNGs and NGGs were extended to reconcile these new inconsistencies for double gene deletions. In addition, the extended GrowMatch can directly identify gene (rather than reaction) removals (suppressions) to resolve these inconsistencies.

Extended GrowMatch was applied to the multi-compartment *i*MM904 metabolic model of yeast (Mo et al., 2009) under minimal and YP media. Notably, 90 of the model correction hypotheses and 30 of the regulatory constraints that were identified by the extended GrowMatch in this study were verified and vetted using literature and other sources. Incorporation of these modifications into the *i*MM904 model (i.e., *i*AZ900) led to significant improvements in the prediction of single and double gene deletion growth phenotypes.

4.2.5. GeneForce

The extended GrowMatch procedure for the resolution of GNGs and GSLs may lead to the identification of some missing regulatory constraints by directly pinpointing gene suppressions in the network. GeneForce is another algorithm designed to identify and correct transcriptional regulatory rules incorporated into a metabolic network based on growth prediction inconsistencies (Barua et al., 2010).

In contrast to (modified/extended) GrowMatch, this procedure focuses on a specific type of NGG inconsistency, where the integrated regulatory and metabolic model predicts no growth (NG) for a specific growth medium (or in the presence of a specific gene knockout), but the metabolic model predictions without any regulatory constraints as well as experimental data show growth (G). The main idea for GeneForce is to minimally perturb the transcriptional regulatory rules to restore growth in the network. Various sets of binary variables are defined to indicate whether the conditions for expression of a gene or activity of a transcription factor are satisfied, a transcription factor is active, a gene is expressed, an enzyme is present, and a reaction is active. The regulatory rule violations are invoked by introducing a new set of binary variables called 'surrogate gene expression indicator' that can differ in value from the gene expression indicator (determined by regulatory rules) thereby allowing the utilization of a non-expressed gene in the model to readjust the flux distribution and restore growth (Barua et al., 2010). GeneForce was used to refine the transcriptional regulatory model of *E. coli* (*i*MC1010^{v1} (Covert et al., 2004)) by analyzing a large collection of *E. coli* knockout growth phenotypes. In addition, the procedure was used to identify the genes that, if over-expressed or constitutively expressed, can reconcile the growth inconsistencies under a certain growth condition for the wild-type or a mutant strain. Some of the suggested modifications were also verified experimentally. GeneForce was also applied to examine the conservation of transcriptional regulatory interactions between *E. coli* and *Salmonella typhimurium* (Barua et al., 2010).

4.3. Reconciliation of flux prediction inconsistencies

4.3.1. OMNI

Another type of experimental data that can be used to test the accuracy of metabolic models includes experimental flux measurements. Herrgard et al (2006) proposed a bilevel optimization framework called optimal metabolic network identification (OMNI) to make use of the available *in vivo* flux measurements (e.g., for growth rates, substrate uptake and byproduct secretion rates and intracellular flux data using ^{13}C labeling experiments) for model refinement (see Fig. 1D). OMNI attempts to minimize the discrepancies between the predicted fluxes using flux balance analysis (FBA) and *in vivo* observed fluxes through addition of external reactions to the model from a universal database or removal of existing reactions from the model, similarly to SMILEY (Reed et al., 2006), GapFill (Satish Kumar et al., 2007) and GrowMatch (Kumar and Maranas, 2009). However, in contrast to GrowMatch where reaction additions and removals are identified using two separate procedures, OMNI performs both these modifications in a single framework. To this end, the set of reactions is divided into two categories: fixed reactions (F) that cannot be removed from the model and reactions that can be removed from the model (D). Set F could be the set of reactions that were included in the model with high confidence, while set D contains existing reactions in the model that could be potentially removed due to inclusion with a lower confidence (e.g., only due to sequence similarity) as well as external reactions that can be added to the model from a universal database such as KEGG (Kanehisa et al., 2008) or MetaCyc (Caspi et al., 2006). OMNI was applied to evolved *E. coli* mutant strains with predicted growth rates lower than *in vivo* measurements in order to identify reactions that act as flux bottlenecks in these strains.

4.4. Identification and correction of thermodynamically infeasible pathways

4.4.1. TMFA

Pathway thermodynamics has been proven quite useful in studying the feasibility of metabolic pathways and eliminating infeasible internal flux cycles suggested by FBA solutions (see Fig. 1E) (Beard et al., 2002, 2004; Qian et al., 2003). Henry et al (2007) proposed an optimization-based procedure called thermodynamic-based metabolic flux analysis (TMFA) to integrate the thermodynamic data and constraint-based modeling of genome-scale metabolic networks. To formulate TMFA, the standard Gibbs free energy of reactions should be known (i.e., measured experimentally, or estimated). All reversible reactions are decomposed into forward and backward reactions that can only carry a non-negative flux. TMFA is a MILP, which relies on the definition of binary variables to determine if the flux of each reaction is non-zero. If a reaction carries a non-zero flux, its thermodynamic feasibility is assured by the addition of new constraints to the standard FBA problem requiring that the Gibbs free energy of that reaction must be negative. These constraints can be also adjusted to take into account the uncertainty associated with thermodynamic data. By incorporation of these constraints into the flux balance analysis, TMFA generates flux distributions that do not contain any thermodynamically infeasible reactions or pathways in the model. Furthermore TMFA can provide insights on metabolite concentrations in the network. This procedure was applied to the entire *i*JR904 model of *E. coli* (Reed et al., 2003) to determine thermodynamically feasible ranges for all fluxes as well as the ranges of the Gibbs free energy change of the reactions and activity of the metabolites (Henry et al., 2007).

4.4.2. ll-COBRA

One factor that can limit the applicability of approaches such as TMFA is that they rely on input data for metabolite concentrations and standard Gibbs free energy change of reactions (ΔG_r^0), which might not be available in many cases or could be inaccurately estimated using methods such group contribution theory (Mavrovouniotis, 1991). Schellenberger et al (2011a) recently proposed an alternative approach called loopless constrained based reconstruction and analysis (ll-COBRA) to exclude all thermodynamic infeasible solutions in flux balance analysis (FBA), flux variability analysis (FVA) and other similar approaches without a need for any input data on metabolite concentrations and Gibbs free energy change of reactions. To this end, ll-COBRA relies on a vector of continuous variables (G_r) indicating the driving forces for each reaction, which is analogous to ΔG_r of reaction in that sign(G_r)=sign(ΔG_r). Feasible thermodynamic (i.e., loopless) conditions are then imposed on FBA, FVA or similar approaches through the addition of a set of linear constraints on G_r of each internal reaction in the model. The applicability of this approach was shown for a number of genome-scale metabolic networks using FBA, FVA and Monte-Carlo sampling and in some cases the imposition of loopless condition improved the agreement between model predictions and experimental observations. ll-COBRA is accessible as part of the COBRA toolbox (Becker et al., 2007; Schellenberger et al., 2011b).

5. Using optimization to redesign metabolic networks

The current available approaches for the redesign of metabolic networks and metabolic engineering applications can be divided into three major categories: computational procedures that identify (i) gene knockouts, (ii) non-native additional pathways and (iii) a combination of knockouts, down-regulations and over-expressions leading to overproduction of a desired chemical. These methods span a wide array of optimization formulations including linear (Choi et al., 2010; Mendes and Kell, 1998), nonlinear (Kim et al., 2011; Maria et al., 2011; Patil et al., 2005), bilevel (Burgard et al., 2003; Kim and Reed, 2010; Kim et al., 2011; Pharkya et al., 2004; Pharkya and Maranas, 2006;

Ranganathan and Maranas, 2010; Ranganathan et al., 2010; Yang et al., 2011), multi-criteria (Maria et al., 2011) and evolutionary algorithms (Patil et al., 2005) as well as cybernetic-based approaches (Varner and Ramkrishna, 1999a, 1999b). In the following we briefly describe the most widely used optimization frameworks in each category as summarized in Table 3.

5.1. Identification of gene knockouts for overproduction

5.1.1. OptKnock

The challenge for overproduction is that microorganisms are primed through natural selection to counteract any externally imposed genetic or environmental perturbations by redirecting metabolic flux to restore cellular growth (Ibarra et al., 2002). Therefore, any genetic interventions must be designed in a way that is consistent with any anticipated host response towards biomass formation maximization. One of the first efforts to address this challenge was the OptKnock framework introduced by Burgard et al. (2003). OptKnock was designed to suggest gene deletion strategies that reshape the connectivity of the metabolic network in such a manner that the production of a desired chemical becomes an obligatory byproduct of cellular growth. To this end, OptKnock employs a nested (i.e., bilevel) optimization framework that involves two competing objective functions. The inner problem is designed to adjust the fluxes maximizing the biological objective (i.e., growth or minimization of metabolic adjustments). The outer problem maximizes the bioengineering objective (i.e., chemical overproduction) by selecting reaction candidates for deletions. Importantly, the gene deletions suggested by OptKnock not only reduce the drain of carbon towards competing byproducts but also ensure the availability of key biomass precursors, thus coupling the bioengineering and biological objectives with important strain stability implications. *In silico* gene deletions predicted by OptKnock for overproducing succinate, lactate, 1,3-propanediol and amino acids (Pharkya et al., 2003) have been carried out and verified by multiple research groups (Fong et al., 2005; Yim et al., 2011). A GAMS implementation of OptKnock can be downloaded from http://maranas.che.psu.edu/software.htm and a MATLAB implementation is accessible as part of the COBRA toolbox (Becker et al., 2007; Schellenberger et al., 2011b).

5.1.2. RobustKnock

A potential limitation associated with the OptKnock procedure arises when the identified mutant involves a range of possible product formation yields while the biomass formation is maximized. Because the inner linear optimization problem is solved implicitly by accumulating primal and dual constraints and setting the two objective functions equal to one another, OptKnock selects the most optimistic (i.e., highest) value for the product formation. This leads to a mutant design with the product formation possibly uncoupled from growth, which may be undesirable. To address this challenge, Tepper and Shlomi (2009) introduced a modified version of OptKnock, termed as RobustKnock, which optimizes the worst-case of the product formation while biomass production is maximized. This leads to a three-level optimization problem. The outer max–min problems aim to identify gene knockouts that maximize the minimum (i.e., guaranteed) production rate of the desired biochemical, whereas the inner problem is similar to OptKnock and maximizes the cellular objective given a set of reaction knockouts. RobustKnock was applied to the *i*JR904 metabolic model of *E. coli* (Reed et al., 2003) to suggest optimal knockout strategies for overproduction of 52 different chemicals, where the predicted double and triple knockouts for hydrogen, acetate, formate and fumarate were different from those suggested by OptKnock. A MATLAB implementation of RobustKnock is accessible at http://www.cs.technion.ac.il/~tomersh/methods.html.

5.1.3. OptGene

OptGene provides an alternative to mathematical optimization by implementing a genetic algorithm inspired procedure for exploring mutation combinations (Patil et al., 2005). Reliance on an evolutionary based procedure allows the identification of near optimal solutions of nonlinear optimization problems in generally shorter times. OptGene can also use minimization of metabolic adjustment (MOMA) (Segre et al., 2002) or regulatory on/off minimization (ROOM) (Shlomi et al., 2005) to compute the fitness of each design and is able to generate a family of alternative solutions. The applicability of OptGene was demonstrated for the overproduction of glycerol, succinate and vanillin in *S. cerevisiae*. The algorithm can be accessed as part of the open source OptFlux platform (http://www.optflux.org/) (Rocha et al., 2010).

5.1.4. Tilting of the objective function

Feist et al., (2010) introduced an alternative to RobustKnock using the concept of "tilting the objective function" to simulate the worst-case scenario for product formation in OptKnock and OptGene. This is achieved by adding in the inner objective of maximizing biomass the negative of the desired product yield

Table 3
Optimization algorithms to redesign metabolic networks.

Name	Type of optimization problem	Type of intervention	Accessibility	Reference
OptKnock	Bilevel, MILP	Knockouts	GAMS	Burgard et al. (2003)
RobustKnock	Multi-level, MILP	Knockouts	MATLAB	Tepper and Shlomi (2009)
OptGene	Evolutionary	Knockouts	Online (as part of OptFlux)	Patil et al. (2005)
Objective tilting	Bilevel, MILP	Knockouts	MATLAB (via COBRA toolbox)	Feist et al. (2010)
OptStrain	Bilevel, MILP	Addition of non-native reactions/pathways	N/A	Pharkya et al. (2004)
SimOptStrain	Bilevel, MILP	Knockouts and addition of non-native reactions/pathways	N/A	Kim et al. (2011)
BiMOMA	Bilevel, MINLP	Knockouts	N/A	Kim et al. (2011)
OptReg	Bilevel, MILP	Knockouts, upregulations and downregulations	N/A	Pharkya and Maranas (2006)
GDLS	Heuristic	Knockouts, upregulations and downregulations	N/A	Lun et al. (2009)
FSEOF	LP	Upregulations and downregulations	N/A	Choi et al. (2010)
OptORF	Bilevel, MILP	Knockouts and overexpressions (of both metabolic and regulatory genes)	N/A	Kim and Reed (2010)
OptForce	Bilevel, MILP	Knockouts, upregulations and downregulations	GAMS	Ranganathan et al. (2010)
EMILiO	Bilevel, MILP	Knockouts, upregulations and downregulations	N/A	Yang et al. (2011)

multiplied by a very small weight. This indirectly ensures that the inner problem selects the worst-case scenario for product yield as the solution. This method is easier to implement and more computationally tractable compared to RobustKnock (Tepper and Shlomi, 2009).

5.2. Enhancing microbial hosts with non-native functionalities

5.2.1. OptStrain

Databases such as KEGG (Kanehisa, 2002), MetaCyc (Caspi et al., 2006) and BRENDA (Scheer et al., 2011) offer a myriad of choices for pathways that can be added to the metabolic network of a selected microbial host; however, the central question is how to identify the best candidates to test experimentally for gene/reaction additions. One of the first procedures that addressed this challenge was the OptStrain framework (Pharkya et al., 2004) that systematically curated a universal bioreaction database (comprising of ~4000 reactions). Subsequently a bilevel optimization formulation is employed to identify the minimal number of non-native pathways that maximize the product yield. The OptStrain procedure identified gene knockouts in addition to knock-ins that further boosted the yields. The algorithm was demonstrated for the overproduction of hydrogen and vanillin from *E. coli* and *clostridia* species. An important pre-processing step for OptStrain is the curation of the universal database of biochemical pathways. In order to consider as many as possible reactions from the ever-growing compilations of biotransformations within a single standardized resource the MetRxn database has been assembled (Kumar et al., 2012). This database is freely accessible as a web resource at http://maranas.che.psu.edu/metrxn/metrxn.php.

5.2.2. SimOptStrain and BiMOMA

In a more recent effort, Kim et al. (2011) proposed two bilevel computational strain design approaches based on mixed-integer linear programming called SimOptStrain and BiMOMA. The former aims to identify simultaneous gene deletions and non-native reaction additions, whereas the latter uses minimization of metabolic adjustment as the cellular objective (instead of maximization of biomass) in the inner problem to identify gene knockouts leading to overproduction of a desired biochemical. These procedures were used to identify new metabolic engineering strategies for overproduction of succinate, glycerol, malate, serine, pyruvate and glutamate, which could not be captured previously by using sequential search and genetic algorithm techniques (Kim et al., 2011).

5.3. Identification of multiple types of genetic interventions

5.3.1. OptReg

OptReg was one of the first frameworks introduced to identify the reactions elimination, up-regulation and down-regulations that need to be imposed in order to overproduce a biochemical of interest (Pharkya and Maranas, 2006). Conceptually, flux modulations were modeled as upward or downward departures from steady-state flux values that result in an increase in the yield for the desired product. Similarly to OptKnock and OptStrain, OptReg relies on the solution of a bilevel optimization problem where the outer objective function is to maximize the synthesis of the desired chemical, however, the objective function of the inner problem maximizes the cellular objective (e.g., biomass formation) while also minimizing the network trafficking (by adding the sum of all fluxes in the objective function multiplied by the negative of a small scalar ε). Three sets of binary variables are defined for each reaction indicating if that reaction should be knocked out, up-regulated or down-regulated. OptReg was demonstrated by overproduction of ethanol in *E. coli*.

5.3.2. GDLS

The genetic design through local search (GDLS) framework (Lun et al., 2009) was introduced to provide a way of evaluating flux modulations from the perspective of fitness of recombinant strains. GDLS relies on local searches with multiple search paths to search for combinations of genetic interventions.

5.3.3. FSEOF

Flux scanning based on enforced objective flux (FSEOF) is another optimization-based procedure to identify the candidate flux up/down-regulations to achieve a desired product yield (Choi et al., 2010). A constraint representing the product formation is added to the regular biomass maximization problem. The flux through the reaction producing the desired product is then increased (enforced) gradually from its initial value (in the wild-type) to a value close (e.g., 90%) to maximum theoretical yield. The calculated values of internal fluxes are "scanned" under the enforced production formation constraints to pinpoint fluxes that increase or decrease upon enforcing the product formation. These fluxes are primary targets for up- and down-regulations, respectively. This procedure was applied to overproduction of lycopene in *E. coli*.

5.3.4. OptORF

OptORF is a bilevel optimization program that identifies optimal gene deletions and overexpressions to maximize the production of a desired compound (Kim and Reed, 2010). The objective functions of the inner and outer problem are similar to OptKnock; however, in contrast to the previous approaches that identify reaction deletions, OptORF has been designed to directly pinpoint optimal metabolic and regulatory gene deletions as well as metabolic gene overexpressions that couple the biomass production and product formation. Similarly to GeneForce, different sets of binary variables, namely gene knockout and overexpression indicators as well as surrogate gene expression indictor, are defined to impose to effect of each type of modification. The applicability of OptORF was demonstrated for the overproduction of ethanol and higher alcohols in *E. coli*.

5.3.5. OptForce

The OptForce procedure (Ranganathan et al., 2010) first characterizes the wild-type strain (or an initial strain) based on the experimentally obtained fluxes either in the form of ranges or exact values. For a pre-specified yield of a desired product, OptForce identifies a set of changes that must happen in the network by contrasting the flux ranges observed in the wild-type with those in the overproducing phenotype. Subsequently, from among these changes, OptForce identifies the minimum set of engineering interventions (knockouts/ups/downs) that must be directly imparted to achieve the desired yield. Similar to RobustKnock, the OptForce procedure simulates a worst-case scenario by solving a min–max optimization problem to identify manipulations that guarantee a minimum production yield for the compound of interest. The use of experimental data to characterize the original strain along with the solution of a worst-case scenario optimization problem distinguishes OptForce from earlier methods (such as OptReg). Experimental implementation of the metabolic interventions identified by OptForce led to the highest naringenin producing strain of *E. coli* in a laboratory scale (Xu et al., 2011) as well as increased production of fatty acids in *E. coli* (Ranganathan et al., in press).

5.3.6. EMILiO

Enhancing metabolism with iterative linear optimization (EMILiO) is a strain design algorithm based on a bilevel optimization formulation, which aims to identify optimal metabolic engineering strategies coupling growth and biochemical production (Yang et al., 2011). The inner and outer problem objectives are similar to those in OptKnock with the only difference being that the objective function of the inner problem also contains a small weighted minimization of the product formation rate (i.e., maximize $v_{biomass} - \varepsilon v_{product}$, where ε is a small number) to avoid "optimistic" solutions. EMILiO is composed of three main stages to speed up solving the bilevel optimization problem: first, an iterative linear program is formulated to identify the set of all active constraints (i.e., variable flux bounds). Second, a pruning method based on a recursive linear programming is used to identify a subset of active constraints contributing to a prespecified fraction of the maximum theoretical yield. Finally, a mixed-integer linear program (MILP) is solved to further prune the resulting subset and arrive at a minimal set of reaction modifications (knockouts/ups/downs) satisfying a target yield for the product of interest (Yang et al., 2011). EMILiO was used to suggest numerous strain design strategies for overproduction of succinate, glutamate and L-serine in *E. coli*.

6. Conclusions

In this review, we highlighted a wide range of mathematical optimization tools to address the multifaceted nature of metabolic engineering tasks. The majority of these approaches are based upon mixed-integer linear programming (MILP) that employ integer variables to capture the discrete nature of decision making required to analyze, curate and redesign metabolic networks. In addition, bilevel optimization algorithms, which were originally introduced into the field of metabolic network analysis and engineering by Burgard et al. (2003), have offered a great promise to address a wide range of challenges involving conflicting or competing objective functions. More recently, the scope of this nested optimization structure was extended to multi-species microbial systems by proposing a multi-level and multi-objective optimization framework, called OptCom, for the flux balance analysis of microbial communities and the computational assessment of the trade-offs between species- and community-level fitness driving forces (Zomorrodi and Maranas, 2012). In addition to linear programming, nonlinear optimization has been also proven useful in a number of metabolic engineering applications such as metabolic flux analysis (MFA) and kinetic modeling of metabolic networks (Mahadevan et al., 2002; Nikolaev, 2010; Pozo et al., 2011a, b; Vital-Lopez et al., 2006). A growing number of successful experimental studies guided by predictions of these optimization-based tools have started to emerge (Barua et al., 2010; Fong et al., 2005; Ranganathan et al., in press; Reed et al., 2006; Xu et al., 2011; Yim et al., 2011) thereby reinforcing the promise of mathematical optimization as a design guiding tool in metabolic engineering.

In the future, optimization challenges are expected to increase as researchers are generating metabolic (and biochemical) models of increased size and complexity to capture spatial and temporal changes or to create whole cell models (Karr et al., 2012). The pace of generation of metabolic models is facilitated by resources such as the Model Seed (Henry et al., 2010), PathwayTools (Karp et al., 2010), MetRxn (Kumar et al., 2012) and MicrobesFlux (Feng et al., 2012). At the same time, engineering intervention strategies are expanding beyond coding regions of the genome to include the rational redesign of transcription and ribosome binding sites, elimination of allosteric regulation, codon optimization, co-option of metabolite channeling, etc. (Dong et al., 1995; Makrides, 1996; Salis et al., 2009; Santos-Aberturas et al., 2011; Tao et al., 2006; Wang et al., 2012). All these new investigations provide rich modeling descriptions to optimize using an ever expanding range of optimization algorithms and (re)formulation techniques.

References

Ahn, W.S., Antoniewicz, M.R., 2011. Metabolic flux analysis of CHO cells at growth and non-growth phases using isotopic tracers and mass spectrometry. Metab. Eng. 13, 598–609.

Alper, H., Jin, Y.S., Moxley, J.F., Stephanopoulos, G., 2005. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. Metab. Eng. 7, 155–164.

Antoniewicz, M.R., Kelleher, J.K., Stephanopoulos, G., 2006. Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements. Metab. Eng. 20, 20.

Atsumi, S., Cann, A.F., Connor, M.R., Shen, C.R., Smith, K.M., Brynildsen, M.P., Chou, K.J., Hanai, T., Liao, J.C., 2008a. Metabolic engineering of *Escherichia coli* for 1-butanol production. Metab. Eng. 10, 305–311.

Atsumi, S., Hanai, T., Liao, J.C., 2008b. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature 451, 86–89.

Atsumi, S., Liao, J.C., 2008a. Directed evolution of *Methanococcus jannaschii* citramalate synthase for biosynthesis of 1-propanol and 1-butanol by *Escherichia coli*. Appl. Environ. Microbiol. 74, 7802–7808.

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

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O., 2008. The RAST server: rapid annotations using subsystems technology. BMC Genom. 9, 75.

Barua, D., Kim, J., Reed, J.L., 2010. An automated phenotype-driven approach (GeneForce) for refining metabolic and regulatory models. PLoS Comput. Biol. 6, e1000970.

Beard, D.A., Babson, E., Curtis, E., Qian, H., 2004. Thermodynamic constraints for biochemical networks. J. Theor. Biol. 228, 327–333.

Beard, D.A., Liang, S.D., Qian, H., 2002. Energy balance for analysis of complex metabolic networks. Biophys. J. 83, 79–86.

Beasley, J.E., Planes, F.J., 2007. Recovering metabolic pathways via optimization. Bioinformatics 23, 92–98.

Becker, S.A., Feist, A.M., Mo, M.L., Hannum, G., Palsson, B.O., Herrgard, M.J., 2007. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. Nat. Protoc. 2, 727–738.

Becker, S.A., Palsson, B.O., 2008. Context-specific metabolic networks are consistent with experiments. PLoS Comput. Biol. 4, e1000082.

Behre, J., Wilhelm, T., von Kamp, A., Ruppin, E., Schuster, S., 2008. Structural robustness of metabolic networks with respect to multiple knockouts. J. Theor. Biol. 252, 433–441.

Bizukojc, M., Dietz, D., Sun, J., Zeng, A.P., 2010. Metabolic modelling of syntrophic-like growth of a 1,3-propanediol producer, *Clostridium butyricum*, and a methanogenic archeon, *Methanosarcina mazei*, under anaerobic conditions. Bioprocess Biosyst. Eng. 33, 507–523.

Blum, T., Kohlbacher, O., 2008. MetaRoute: fast search for relevant metabolic routes for interactive network navigation and visualization. Bioinformatics 24, 2108–2109.

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

Bordbar, A., Feist, A.M., Usaite-Black, R., Woodcock, J., Palsson, B.O., Famili, I., 2011. A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology. BMC Syst. Biol. 5, 180.

Bosma, P.J., 2003. Inherited disorders of bilirubin metabolism. J. Hepatol. 38, 107–117.

Burgard, A.P., Maranas, C.D., 2003. Optimization-based framework for inferring and testing hypothesized metabolic objective functions. Biotechnol. Bioeng. 82, 670–677.

Burgard, A.P., Nikolaev, E.V., Schilling, C.H., Maranas, C.D., 2004. Flux coupling analysis of genome-scale metabolic network reconstructions. Genome Res. 14, 301–312.

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

Burgard, A.P., Vaidyaraman, S., Maranas, C.D., 2001. Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments. Biotechnol. Prog. 17, 791–797.

Caspi, R., Foerster, H., Fulcher, C.A., Hopkinson, R., Ingraham, J., Kaipa, P., Krummenacker, M., Paley, S., Pick, J., Rhee, S.Y., Tissier, C., Zhang, P., Karp, P.D., 2006. MetaCyc: a multiorganism database of metabolic pathways and enzymes. Nucleic Acids Res. 34, D511–D516.

Chandrasekaran, S., Price, N.D., 2010. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. Proc. Natl. Acad Sci. U.S.A. 107, 17845–17850.

Chang, Y., Suthers, P.F., Maranas, C.D., 2008. Identification of optimal measurement sets for complete flux elucidation in metabolic flux analysis experiments. Biotechnol. Bioeng. 100, 1039–1049.
Choi, H.S., Lee, S.Y., Kim, T.Y., Woo, H.M., 2010. In silico identification of gene amplification targets for improvement of lycopene production. Appl. Environ. Microbiol. 76, 3097–3105.
Covert, M.W., Knight, E.M., Reed, J.L., Herrgard, M.J., Palsson, B.O., 2004. Integrating high-throughput and computational data elucidates bacterial networks. Nature 429, 92–96.
Covert, M.W., Schilling, C.H., Palsson, B., 2001. Regulation of gene expression in flux balance models of metabolism. J. Theor. Biol. 213, 73–88.
Crown, S.B., Ahn, W.S., Antoniewicz, M.R., 2012. Rational design of 13C-labeling experiments for metabolic flux analysis in mammalian cells. BMC Syst. Biol. 6, 43.
Crown, S.B., Antoniewicz, M.R., 2012. Selection of tracers for 13C-metabolic flux analysis using elementary metabolite units (EMU) basis vector methodology. Metab. Eng. 14, 150–161.
Danpure, C.J., 2006. Primary hyperoxaluria type 1: AGT mistargeting highlights the fundamental differences between the peroxisomal and mitochondrial protein import pathways. Biochim. Biophys. Acta 1763, 1776–1784.
Dauner, M., Bailey, J.E., Sauer, U., 2001. Metabolic flux analysis with a comprehensive isotopomer model in *Bacillus subtilis*. Biotechnol. Bioeng. 76, 144–156.
David, L., Marashi, S.A., Larhlimi, A., Mieth, B., Bockmayr, A., 2011. FFCA: a feasibility-based method for flux coupling analysis of metabolic networks. BMC Bioinformatics 12, 236.
Deutscher, D., Meilijson, I., Kupiec, M., Ruppin, E., 2006. Multiple knockout analysis of genetic robustness in the yeast metabolic network. Nat. Genet. 38, 993–998.
Deutscher, D., Meilijson, I., Schuster, S., Ruppin, E., 2008. Can single knockouts accurately single out gene functions? BMC Syst. Biol. 2, 50.
Dietrich, J.A., McKee, A.E., Keasling, J.D., 2010. High-throughput metabolic engineering: advances in small-molecule screening and selection. Annu. Rev. Biochem. 79, 563–590.
Dong, H., Nilsson, L., Kurland, C.G., 1995. Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. J. Bacteriol. 177, 1497–1504.
Duarte, N.C., Herrgard, M.J., Palsson, B.O., 2004. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. Genome Res. 14, 1298–1309.
Edwards, J.S., Ibarra, R.U., Palsson, B.O., 2001. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. Nat. Biotechnol. 19, 125–130.
Ellis, L.B., Roe, D., Wackett, L.P., 2006. The University of Minnesota Biocatalysis/Biodegradation Database: the first decade. Nucleic Acids Res. 34, D517–D521.
Emmerling, M., Dauner, M., Ponti, A., Fiaux, J., Hochuli, M., Szyperski, T., Wuthrich, K., Bailey, J.E., Sauer, U., 2002. Metabolic flux responses to pyruvate kinase knockout in *Escherichia coli*. J. Bacteriol. 184, 152–164.
Feist, A.M., Herrgard, M.J., Thiele, I., Reed, J.L., Palsson, B.O., 2009. Reconstruction of biochemical networks in microorganisms. Nat. Rev. Microbiol. 7, 129–143.
Feist, A.M., Palsson, B.O., 2010. The biomass objective function. Curr. Opin. Microbiol. 13, 344–349.
Feist, A.M., Zielinski, D.C., Orth, J.D., Schellenberger, J., Herrgard, M.J., Palsson, B.O., 2010. Model-driven evaluation of the production potential for growth-coupled products of *Escherichia coli*. Metab. Eng. 12, 173–186.
Fell, D.A., Small, J.R., 1986. Fat synthesis in adipose tissue. An examination of stoichiometric constraints. Biochem. J. 238, 781–786.
Feng, X., Xu, Y., Chen, Y., Tang, Y.J., 2012. MicrobesFlux: a web platform for drafting metabolic models from the KEGG database. BMC Syst. Biol. 6, 94.
Finley, S.D., Broadbelt, L.J., Hatzimanikatis, V., 2010. In silico feasibility of novel biodegradation pathways for 1,2,4-trichlorobenzene. BMC Syst. Biol. 4, 7.
Fong, S.S., Burgard, A.P., Herring, C.D., Knight, E.M., Blattner, F.R., Maranas, C.D., Palsson, B.O., 2005. In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid. Biotechnol. Bioeng. 91, 643–648.
Forbes, N.S., Clark, D.S., Blanch, H.W., 2001. Using isotopomer path tracing to quantify metabolic fluxes in pathway models containing reversible reactions. Biotechnol. Bioeng. 74, 196–211.
Forster, J., Famili, I., Fu, P., Palsson, B.O., Nielsen, J., 2003. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. Genome Res. 13, 244–253.
Ghosh, S., Zhu, T., Grossmann, I.E., Ataai, M.M., Domach, M.M., 2005. Closing the loop between feasible flux scenario identification for construct evaluation and resolution of realized fluxes via NMR. Comput. Chem. Eng. 29, 459–466.
Gianchandani, E.P., Oberhardt, M.A., Burgard, A.P., Maranas, C.D., Papin, J.A., 2008. Predicting biological system objectives de novo from internal state measurements. BMC Bioinformatics 9, 43.
Goesmann, A., Haubrock, M., Meyer, F., Kalinowski, J., Giegerich, R., 2002. PathFinder: reconstruction and dynamic visualization of metabolic pathways. Bioinformatics 18, 124–129.
Guarente, L., 1993. Synthetic enhancement in gene interaction: a genetic tool come of age. Trends Genet. 9, 362–366.
Gudmundsson, S., Thiele, I., 2010. Computationally efficient flux variability analysis. BMC Bioinformatics 11, 489.
Harrison, R., Papp, B., Pal, C., Oliver, S.G., Delneri, D., 2007. Plasticity of genetic interactions in metabolic networks of yeast. Proc Natl. Acad Sci. U.S.A. 104, 2307–2312.
Hatzimanikatis, V., Li, C., Ionita, J.A., Henry, C.S., Jankowski, M.D., Broadbelt, L.J., 2005. Exploring the diversity of complex metabolic networks. Bioinformatics 21, 1603–1609.
Henry, C.S., Broadbelt, L.J., Hatzimanikatis, V., 2007. Thermodynamics-based metabolic flux analysis. Biophys. J. 92, 1792–1805.
Henry, C.S., DeJongh, M., Best, A.A., Frybarger, P.M., Linsay, B., Stevens, R.L., 2010. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat. Biotechnol. 28, 977–982.
Henry, C.S., Zinner, J.F., Cohoon, M.P., Stevens, R.L., 2009. iBsu1103: a new genome-scale metabolic model of *Bacillus subtilis* based on SEED annotations. Genome Biol. 10, R69.
Herrgard, M.J., Fong, S.S., Palsson, B.O., 2006. Identification of genome-scale metabolic network models using experimentally measured flux profiles. PLoS Comput. Biol. 2, e72.
Ibarra, R.U., Edwards, J.S., Palsson, B.O., 2002. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature 420, 186–189.
Jamshidi, N., Palsson, B.O., 2007. Investigating the metabolic capabilities of *Mycobacterium tuberculosis* H37Rv using the in silico strain iNJ661 and proposing alternative drug targets. BMC Syst. Biol. 1, 26.
Kanehisa, M., 2002. The KEGG database. In: Novartis Foundation Symposium vol. 247, pp. 91–101; Discussion 101–103, 119–128, 244–252.
Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., Yamanishi, Y., 2008. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 36, D480–D484.
Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., Hirakawa, M., 2006. From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res. 34, D354–D357.
Karp, P.D., Paley, S.M., Krummenacker, M., Latendresse, M., Dale, J.M., Lee, T.J., Kaipa, P., Gilham, F., Spaulding, A., Popescu, L., Altman, T., Paulsen, I., Keseler, I.M., Caspi, R., 2010. Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. Brief Bioinform. 11, 40–79.
Karr, J.R., Sanghvi, J.C., Macklin, D.N., Gutschow, M.V., Jacobs, J.M., Bolival Jr., B., Assad-Garcia, N., Glass, J.I., Covert, M.W., 2012. A whole-cell computational model predicts phenotype from genotype. Cell 150, 389–401.
Kim, J., Reed, J.L., 2010. OptORF: optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. BMC Syst Biol. 4, 53.
Kim, J., Reed, J.L., Maravelias, C.T., 2011. Large-scale bi-level strain design approaches and mixed-integer programming solution techniques. PLoS One 6, e24162.
Klapa, M.I., Aon, J.C., Stephanopoulos, G., 2003. Systematic quantification of complex metabolic flux networks using stable isotopes and mass spectrometry. Eur. J. Biochem. 270, 3525–3542.
Knorr, A.L., Jain, R., Srivastava, R., 2007. Bayesian-based selection of metabolic objective functions. Bioinformatics 23, 351–357.
Kumar, A., Suthers, P.F., Maranas, C.D., 2012. MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases. BMC Bioinformatics 13, 6.
Kumar, V.S., Maranas, C.D., 2009. GrowMatch: an automated method for reconciling in silico/*in vivo* growth predictions. PLoS Comput. Biol. 5, e1000308.
Larhlimi, A., Bockmayr, A., 2006. A new approach to flux coupling analysis of metabolic networks. In: Computational Life Sciences II, Second International Symposium, pp. 205–215.
Larhlimi, A., David, L., Selbig, J., Bockmayr, A., 2012. F2C2: a fast tool for the computation of flux coupling in genome-scale metabolic networks. BMC Bioinformatics 13, 57.
Lee, D.Y., Fan, L.T., Park, S., Lee, S.Y., Shafie, S., Bertok, B., Friedler, F., 2005. Complementary identification of multiple flux distributions and multiple metabolic pathways. Metab. Eng. 7, 182–200.
Lewis, N.E., Schramm, G., Bordbar, A., Schellenberger, J., Andersen, M.P., Cheng, J.K., Patel, N., Yee, A., Lewis, R.A., Eils, R., Konig, R., Palsson, B.O., 2010. Large-scale in silico modeling of metabolic interactions between cell types in the human brain. Nat. Biotechnol. 28, 1279–1285.
Liu, T., Khosla, C., 2010. Genetic engineering of *Escherichia coli* for biofuel production. Annu. Rev. Genet. 44, 53–69.
Lun, D.S., Rockwell, G., Guido, N.J., Baym, M., Kelner, J.A., Berger, B., Galagan, J.E., Church, G.M., 2009. Large-scale identification of genetic design strategies using local search. Mol. Syst. Biol. 5, 296.
Mahadevan, R., Edwards, J.S., Doyle 3rd, F.J., 2002. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. Biophys. J. 83, 1331–1340.
Mahadevan, R., Schilling, C.H., 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab. Eng. 5, 264–276.
Makrides, S.C., 1996. Strategies for achieving high-level expression of genes in *Escherichia coli*. Microbiol. Rev. 60, 512–538.
Marashi, S.A., Bockmayr, A., 2010. Flux coupling analysis of metabolic networks is sensitive to missing reactions. Biosystems 103, 57–66.
Marashi, S.A., David, L., Bockmayr, A., 2012. On flux coupling analysis of metabolic subsystems. J. Theor. Biol. 302, 62–69.
Maria, G., Xu, Z., Sun, J., 2011. Multi-objective MINLP optimization used to identify theoretical gene knockout strategies for *E. coli* cell. Chem. Biochem. Eng. Q. 25, 403–424.
Mavrovouniotis, M.L., 1991. Estimation of standard Gibbs energy changes of biotransformations. J. Biol. Chem. 266, 14440–14445.
McShan, D.C., Rao, S., Shah, I., 2003. PathMiner: predicting metabolic pathways by heuristic search. Bioinformatics 19, 1692–1698.
Mendes, P., Kell, D., 1998. Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. Bioinformatics 14, 869–883.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., Edwards, R.A., 2008. The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9, 386.

Milne, C.B., Eddy, J.A., Raju, R., Ardekani, S., Kim, P.J., Senger, R.S., Jin, Y.S., Blaschek, H.P., Price, N.D., 2011. Metabolic network reconstruction and genome-scale model of butanol-producing strain *Clostridium beijerinckii* NCIMB 8052. BMC Syst. Biol. 5, 130.

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

Mo, M.L., Palsson, B.O., Herrgard, M.J., 2009. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. BMC Syst. Biol. 3, 37.

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

Nigam, R., Liang, S., 2007. Algorithm for perturbing thermodynamically infeasible metabolic networks. Comput. Biol. Med. 37, 126–133.

Nikolaev, E.V., 2010. The elucidation of metabolic pathways and their improvements using stable optimization of large-scale kinetic models of cellular systems. Metab. Eng. 12, 26–38.

Novick, P., Osmond, B.C., Botstein, D., 1989. Suppressors of yeast actin mutations. Genetics 121, 659–674.

Oliveira, A.P., Nielsen, J., Forster, J., 2005. Modeling *Lactococcus lactis* using a genome-scale flux model. BMC Microbiol. 5, 39.

Orth, J.D., Palsson, B.O., 2010. Systematizing the generation of missing metabolic knowledge. Biotechnol. Bioeng. 107, 403–412.

Orth, J.D., Thiele, I., Palsson, B.O., 2010. What is flux balance analysis? Nat. Biotechnol. 28, 245–248.

Overbeek, R., Disz, T., Stevens, R., 2004. The SEED: a peer-to-peer environment for genome annotation. Commun. ACM 47, 46–51.

Ow, D.S., Lee, D.Y., Yap, M.G., Oh, S.K., 2009. Identification of cellular objective for elucidating the physiological state of plasmid-bearing *Escherichia coli* using genome-scale in silico analysis. Biotechnol. Prog. 25, 61–67.

Park, J.H., Lee, K.H., Kim, T.Y., Lee, S.Y., 2007. Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. Proc. Natl. Acad. Sci. U.S.A. 104, 7797–7802.

Patil, K.R., Rocha, I., Forster, J., Nielsen, J., 2005. Evolutionary programming as a platform for in silico metabolic engineering. BMC Bioinformatics 6, 308.

Pey, J., Prada, J., Beasley, J.E., Planes, F.J., 2011. Path finding methods accounting for stoichiometry in metabolic networks. Genome Biol. 12, R49.

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

Pharkya, P., Burgard, A.P., Maranas, C.D., 2004. OptStrain: a computational framework for redesign of microbial production systems. Genome Res. 14, 2367–2376.

Pharkya, P., Maranas, C.D., 2006. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. Metab. Eng. 8, 1–13.

Planes, F.J., Beasley, J.E., 2009. An optimization model for metabolic pathways. Bioinformatics 25, 2723–2729.

Pozo, C., Guille´n-Gosa´lbez, G., Sorribas, A., Jime´nez, L., 2011a. A spatial branch and bound framework for the global optimization of kinetic models of metabolic networks. Ind. Eng. Chem. Res. 50, 5225–5238.

Pozo, C., Marin-Sanguino, A., Alves, R., Guillen-Gosalbez, G., Jimenez, L., Sorribas, A., 2011b. Steady-state global optimization of metabolic non-linear dynamic models through recasting into power-law canonical models. BMC Syst. Biol. 5, 137.

Pramanik, J., Keasling, J.D., 1997. Stoichiometric model of *Escherichia coli* metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. Biotechnol. Bioeng. 56, 398–421.

Qian, H., Beard, D.A., Liang, S.D., 2003. Stoichiometric network theory for nonequilibrium biochemical systems. Eur. J. Biochem. 270, 415–421.

Ranganathan, S., Maranas, C.D., 2010. Microbial 1-butanol production: identification of non-native production routes and in silico engineering interventions. Biotechnol. J. 5, 716–725.

Ranganathan, S., Suthers, P.F., Maranas, C.D., 2010. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS Comput. Biol. 6, e1000744.

Ranganathan, S., Tee, T.W., Chowdhury, A., Zomorrodi, A.R., Yoon, J.M., Fu, Y., Shanks, J.V., Maranas, C.D.. An integrated computational and experimental study for overproducing fatty acids in *Escherichia coli*. Metab. Eng., http://dx.doi.org/10.1016/j.ymben.2012.08.008, in press.

Ravikirthi, P., Suthers, P.F., Maranas, C.D., 2011. Construction of an *E. coli* genome-scale atom mapping model for MFA calculations. Biotechnol. Bioeng. 108, 1372–1382.

Reed, J.L., Patel, T.R., Chen, K.H., Joyce, A.R., Applebee, M.K., Herring, C.D., Bui, O.T., Knight, E.M., Fong, S.S., Palsson, B.O., 2006. Systems approach to refining genome annotation. Proc Natl. Acad. Sci. U.S.A. 103, 17480–17484.

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

Riascos, C.A.M., Gombert, A.K., Pinto, J.M., 2005. A global optimization approach for metabolic flux analysis based on labeling balances. Comput. Chem. Eng. 29, 447–458.

Rocha, I., Maia, P., Evangelista, P., Vilaca, P., Soares, S., Pinto, J.P., Nielsen, J., Patil, K.R., Ferreira, E.C., Rocha, M., 2010. OptFlux: an open-source software platform for in silico metabolic engineering. BMC Syst. Biol. 4, 45.

Saha, R., Suthers, P.F., Maranas, C.D., 2011. Zea mays iRS1563: a comprehensive genome-scale metabolic reconstruction of maize metabolism. PLoS One 6, e21784.

Salis, H.M., Mirsky, E.A., Voigt, C.A., 2009. Automated design of synthetic ribosome binding sites to control protein expression. Nat. Biotechnol. 27, 946–950.

Santos-Aberturas, J., Payero, T.D., Vicente, C.M., Guerra, S.M., Canibano, C., Martin, J.F., Aparicio, J.F., 2011. Functional conservation of PAS-LuxR transcriptional regulators in polyene macrolide biosynthesis. Metab. Eng. 13, 756–767.

Satish Kumar, V., Dasika, M.S., Maranas, C.D., 2007. Optimization based automated curation of metabolic reconstructions. BMC Bioinformatics 8, 212.

Sauer, M., Porro, D., Mattanovich, D., Branduardi, P., 2008. Microbial production of organic acids: expanding the markets. Trends Biotechnol. 26, 100–108.

Savinell, J.M., Palsson, B.O., 1992. Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism. J. Theor. Biol. 154, 421–454.

Scheer, M., Grote, A., Chang, A., Schomburg, I., Munaretto, C., Rother, M., Sohngen, C., Stelzer, M., Thiele, J., Schomburg, D., 2011. BRENDA, the enzyme information system in 2011. Nucleic Acids Res. 39, D670–D676.

Schellenberger, J., Lewis, N.E., Palsson, B.O., 2011a. Elimination of thermodynamically infeasible loops in steady-state metabolic models. Biophys. J. 100, 544–553.

Schellenberger, J., Que, R., Fleming, R.M., Thiele, I., Orth, J.D., Feist, A.M., Zielinski, D.C., Bordbar, A., Lewis, N.E., Rahmanian, S., Kang, J., Hyduke, D.R., Palsson, B.O., 2011b. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat. Protoc. 6, 1290–1307.

Schmidt, K., Carlsen, M., Nielsen, J., Villadsen, J., 1997. Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices. Biotechnol. Bioeng. 55, 831–840.

Schmidt, K., Nielsen, J., Villadsen, J., 1999. Quantitative analysis of metabolic fluxes in *Escherichia coli*, using two-dimensional NMR spectroscopy and complete isotopomer models. J. Biotechnol. 71, 175–189.

Schuetz, R., Kuepfer, L., Sauer, U., 2007. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. Mol. Syst. Biol. 3, 119.

Schwender, J., 2008. Metabolic flux analysis as a tool in metabolic engineering of plants. Curr. Opin. Biotechnol. 19, 131–137.

Scott, E., Peter, F., Sanders, J., 2007. Biomass in the manufacture of industrial products—the use of proteins and amino acids. Appl. Microbiol. Biotechnol. 75, 751–762.

Segre, D., Vitkup, D., Church, G.M., 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. U.S.A. 99, 15112–15117.

Shen, C.R., Lan, E.I., Dekishima, Y., Baez, A., Cho, K.M., Liao, J.C., 2011. Driving forces enable high-titer anaerobic 1-butanol synthesis in *Escherichia coli*. Appl. Environ. Microbiol. 77, 2905–2915.

Shen, C.R., Liao, J.C., 2008. Metabolic engineering of *Escherichia coli* for 1-butanol and 1-propanol production via the keto-acid pathways. Metab. Eng. 10, 312–320.

Shlomi, T., Berkman, O., Ruppin, E., 2005. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. Proc. Natl. Acad. Sci. U.S.A. 102, 7695–7700.

Shlomi, T., Eisenberg, Y., Sharan, R., Ruppin, E., 2007. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. Mol. Syst. Biol., 3.

Srour, O., Young, J.D., Eldar, Y.C., 2011. Fluxomers: a new approach for 13C metabolic flux analysis. BMC Syst. Biol. 5, 129.

Steen, E.J., Chan, R., Prasad, N., Myers, S., Petzold, C.J., Redding, A., Ouellet, M., Keasling, J.D., 2008. Metabolic engineering of *Saccharomyces cerevisiae* for the production of *n*-butanol. Microb. Cell Fact. 7, 36.

Stolyar, S., Van Dien, S., Hillesland, K.L., Pinel, N., Lie, T.J., Leigh, J.A., Stahl, D.A., 2007. Metabolic modeling of a mutualistic microbial community. Mol. Syst. Biol. 3, 92.

Suthers, P.F., Burgard, A.P., Dasika, M.S., Nowroozi, F., Van Dien, S., Keasling, J.D., Maranas, C.D., 2007. Metabolic flux elucidation for large-scale models using 13C labeled isotopes. Metab. Eng. 9, 387–405.

Suthers, P.F., Chang, Y.J., Maranas, C.D., 2009a. Improved computational performance of MFA using elementary metabolite units and flux coupling. Metab. Eng. 12, 123–128.

Suthers, P.F., Zomorrodi, A., Maranas, C.D., 2009b. Genome-scale gene/reaction essentiality and synthetic lethality analysis. Mol. Syst. Biol. 5, 301.

Tao, L., Wilczek, J., Odom, J.M., Cheng, Q., 2006. Engineering a beta-carotene ketolase for astaxanthin production. Metab. Eng. 8, 523–531.

Tawarmalani, M., Sahinidis, N.V., 2004. Global optimization of mixed-integer nonlinear programs: a theoretical and computational study. Math. Program. 99, 563–591.

Tepper, N., Shlomi, T., 2009. Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. Bioinformatics 26, 536–543.

Thakker, C., Zhu, J., San, K.Y., Bennett, G., 2011. Heterologous pyc gene expression under various natural and engineered promoters in *Escherichia coli* for improved succinate production. J. Biotechnol. 155, 236–243.

Thiele, I., Fleming, R.M., Bordbar, A., Schellenberger, J., Palsson, B.O., 2010. Functional characterization of alternate optimal solutions of *Escherichia coli*'s transcriptional and translational machinery. Biophys. J. 98, 2072–2081.

Thiele, I., Palsson, B.O., 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat. Protoc. 5, 93–121.
Vallino, J.J., Stephanopoulos, G., 1993. Metabolic flux distributions in *Corynebacterium glutamicum* during growth and lysine overproduction. Biotechnol. Bioeng. 41, 633–646.
van Winden, W.A., Wittmann, C., Heinzle, E., Heijnen, J.J., 2002. Correcting mass isotopomer distributions for naturally occurring isotopes. Biotechnol. Bioeng. 80, 477–479.
Varma, A., Palsson, B.O., 1993. Metabolic capabilities of *Escherichia coli*: I. synthesis of biosynthetic precursors and cofactors. J. Theor. Biol. 165, 477–502.
Varma, A., Palsson, B.O., 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. Appl. Environ. Microbiol. 60, 3724–3731.
Varner, J., Ramkrishna, D., 1999a. Metabolic engineering from a cybernetic perspective. 1. Theoretical preliminaries. Biotechnol. Prog. 15, 407–425.
Varner, J., Ramkrishna, D., 1999b. Metabolic engineering from a cybernetic perspective. 2. Qualitative investigation of nodal architechtures and their response to genetic perturbation. Biotechnol. Prog. 15, 426–438.
Vital-Lopez, F.G., Armaou, A., Nikolaev, E.V., Maranas, C.D., 2006. A computational procedure for optimal engineering interventions using kinetic models of metabolism. Biotechnol. Prog. 22, 1507–1517.
Wang, D., Mao, Y., Ma, L., Li, Q., Li, W., Xing, J., Su, Z., 2009. Expression of heterogenous pyruvate carboxylase in *Escherichia coli* with lactose as inducer and its effect on succinate production. Sheng Wu Gong Cheng Xue Bao 25, 1338–1344.
Wang, H.H., Kim, H., Cong, L., Jeong, J., Bang, D., Church, G.M., 2012. Genome-scale promoter engineering by coselection MAGE. Nat. Methods.
Wang, Q., Wu, C., Chen, T., Chen, X., Zhao, X., 2006. Expression of galactose permease and pyruvate carboxylase in *Escherichia coli* ptsG mutant increases the growth rate and succinate yield under anaerobic conditions. Biotechnol. Lett. 28, 89–93.
Wiechert, W., de Graaf, A.A., 1997. Bidirectional reaction steps in metabolic networks: I. Modeling and simulation of carbon isotope labeling experiments. Biotechnol. Bioeng. 55, 101–117.
Wiechert, W., Mollney, M., Isermann, N., Wurzel, M., de Graaf, A.A., 1999. Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems. Biotechnol. Bioeng. 66, 69–85.
Wiechert, W., Mollney, M., Petersen, S., de Graaf, A.A., 2001. A universal framework for 13C metabolic flux analysis. Metab. Eng. 3, 265–283.
Wilhelm, T., Behre, J., Schuster, S., 2004. Analysis of structural robustness of metabolic networks. Syst. Biol. (Stevenage) 1, 114–120.
Wittmann, C., Heinzle, E., 2002. Genealogy profiling through strain improvement by using metabolic network analysis: metabolic flux genealogy of several generations of lysine-producing corynebacteria. Appl. Environ. Microbiol. 68, 5843–5859.
Xu, P., Ranganathan, S., Fowler, Z.L., Maranas, C.D., Koffas, M.A., 2011. Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. Metab. Eng. 13, 578–587.
Yang, F., Qian, H., Beard, D.A., 2005. Ab initio prediction of thermodynamically feasible reaction directions from biochemical network stoichiometry. Metab. Eng. 7, 251–259.
Yang, L., Cluett, W.R., Mahadevan, R., 2011. EMILiO: a fast algorithm for genome-scale strain design. Metab. Eng. 13, 272–281.
Yang, T.H., Wittmann, C., Heinzle, E.E., 2004. Metabolic network simulation using logical loop algorithm and Jacobian matrix. Metab. Eng. 6, 256–267.
Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J.D., Osterhout, R.E., Stephen, R., Estadilla, J., Teisan, S., Schreyer, H.B., Andrae, S., Yang, T.H., Lee, S.Y., Burk, M.J., Van Dien, S., 2011. Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. Nat. Chem. Biol. 7, 445–452.
Yu, D., Ellis, H.M., Lee, E.C., Jenkins, N.A., Copeland, N.G., Court, D.L., 2000. An efficient recombination system for chromosome engineering in *Escherichia coli*. Proc. Natl. Acad Sci. U.S.A. 97, 5978–5983.
Zelezniak, A., Pers, T.H., Soares, S., Patti, M.E., Patil, K.R., 2010. Metabolic network topology reveals transcriptional regulatory signatures of type 2 diabetes. PLoS Comput. Biol. 6, e1000729.
Zhao, J., Shimizu, K., 2003. Metabolic flux analysis of *Escherichia coli* K12 grown on 13C-labeled acetate and glucose using GC–MS and powerful flux calculation method. J. Biotechnol. 101, 101–117.
Zomorrodi, A.R., Maranas, C.D., 2010. Improving the iMM904 *S. cerevisiae* metabolic model using essentiality and synthetic lethality data. BMC Syst. Biol. 4, 178.
Zomorrodi, A.R., Maranas, C.D., 2012. OptCom: a multi-level optimization framework for the metabolic modeling and analysis of microbial communities. PLoS Comput. Biol. 8, e1002363.
Zupke, C., Stephanopoulos, G., 1994. Modeling of isotope distributions and intracellular fluxes in metabolic networks using atom mapping matrices. Biotechnol. Prog. 10, 489–498.