Special Issue: *Systems Metabolic Engineering*. Metabolic engineering combines a mix of approaches, including in silico modeling, omics studies, synthetic biology and protein engineering to improve microorganism strains for increased yields and reduced production costs of desirable chemicals. Such an achievement is exemplified on this Special Issue's cover, which shows an electron microscopy image of *Corynebacterium glutamicum* that has been engineered to produce a sustainable bio-nylon monomer from hemicellulose sugar found in the cell walls of plants. Image provided by Buschke et al.

Biotechnology Journal – list of articles published in the May 2013 issue.

Research Article

SMET: Systematic multiple enzyme targeting – a method to rationally design optimal strains for target chemical overproduction

David Flowers[1], R. Adam Thompson[2], Douglas Birdwell[3], Tsewei Wang[1] and Cong T. Trinh[1,2]

[1] Department of Chemical and Biomolecular Engineering, University of Tennessee, Knoxville, TN, USA
[2] Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, TN, and Oak Ridge National Laboratory, Oak Ridge, TN, USA
[3] Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN, USA

Identifying multiple enzyme targets for metabolic engineering is very critical for redirecting cellular metabolism to achieve desirable phenotypes, e.g., overproduction of a target chemical. The challenge is to determine which enzymes and how much of these enzymes should be manipulated by adding, deleting, under-, and/or over-expressing associated genes. In this study, we report the development of a systematic multiple enzyme targeting method (SMET), to rationally design optimal strains for target chemical overproduction. The SMET method combines both elementary mode analysis and ensemble metabolic modeling to derive SMET metrics including *l*-values and *c*-values that can identify rate-limiting reaction steps and suggest which enzymes and how much of these enzymes to manipulate to enhance product yields, titers, and productivities. We illustrated, tested, and validated the SMET method by analyzing two networks, a simple network for concept demonstration and an *Escherichia coli* metabolic network for aromatic amino acid overproduction. The SMET method could systematically predict simultaneous multiple enzyme targets and their optimized expression levels, consistent with experimental data from the literature, without performing an iterative sequence of single-enzyme perturbation. The SMET method was much more efficient and effective than single-enzyme perturbation in terms of computation time and finding improved solutions.

Received 16 DEC 2012
Revised 26 MAR 2013
Accepted 03 APR 2013

Supporting information available online

Keywords: Elementary mode analysis · Ensemble metabolic modeling · Minimal cut set · Rational strain design · Systematic multiple enzyme targeting (SMET)

Correspondence: Dr. Cong T. Trinh, University of Tennessee, 1512 Middle Dr., R432, Knoxville, TN 37920, USA
E-mail: ctrinh@utk.edu

Additional correspondence: Dr. Tsewei Wang, Department of Chemical and Biomolecular Engineering, University of Tennessee, 1512 Middle Dr., R431, Knoxville, TN 37920, USA
E-mail: twang@utk.edu

Abbreviations: cMCS, constrained minimal cut set; **EMA,** elementary mode analysis; **EM,** elementary mode; **EMM,** ensemble metabolic modeling; **FBA,** flux balance analysis; **MCS,** minimal cut set; **MMF,** minimal metabolic functionality; **SMET,** systematic multiple enzyme targeting

1 Introduction

Since the inception of metabolic engineering more than two decades ago [1–3], the field has played a significant role in optimizing microbial biocatalysts and has had a significant impact on biotechnological applications related to health [4, 5], food [2, 6], energy [7, 8], and environment [9, 10]. One of the key questions that metabolic engineers face is to identify which genes should be targeted to develop a robust and efficient strain to achieve desirable phenotypes, e.g. production of a target compound at high yields, titers, and productivities.

Several different metabolic pathway analysis tools have been developed to assist rational strain design by using constraint-based metabolic network modeling, such as flux balance analysis (FBA) [11] and elementary mode analysis (EMA) [12]. Both of these techniques are based on the structural analysis of the metabolic network without requiring the enzyme kinetic parameters, which are largely unknown in a complex metabolic network [13, 14]. These techniques primarily seek to find genetic modifications (e.g. gene knockouts) to achieve optimal metabolic networks with desirable phenotypes [15, 16].

Various FBA-derived techniques including MOMA [17], Optknock [18], OptStrain [19], OptForce [20], RobustKnock [21], OptGene [22], EMILiO [23], and OptORF [24] were developed to design optimal strains for metabolite overproduction [25–27]. These techniques formulate the problem by using objective functions such as maximum growth rate and maximum product formation, and can identify one optimal solution of genetic knockout targets that optimize cell growth and chemical overproduction [28]. The advantage of these techniques is that they can analyze large-scale metabolic networks. However, the optimal solution may not guarantee that the engineered strain will not use other (sub)optimal pathways to function [29–31].

EMA is a metabolic pathway analysis tool that can identify all feasible metabolic pathways called elementary modes (EMs) inherent to a metabolic network that contain a unique and minimal set of enzymatic reactions operating under steady state. Given complete knowledge of all EMs, the optimal metabolic network of a desirable strain can be designed by using various techniques such as the minimal metabolic functionality (MMF) [32], constrained minimal cut set (cMCS) method [33], FluxDesign [34, 35], and CASOP [36]. The MMF method was validated with experimental data for optimal production of cell biomass [37], biofuels (ethanol, isobutanol, and butanol are examples) [32, 38–42] and secondary metabolites (such as carotenoids) [43] in various organisms. However, the above methods are typically limited to moderate-scale metabolic networks because of the combinatorial explosion of computing all EMs [44].

Experimentally, an engineered strain sometimes may not achieve a desirable programmed phenotype (such as improved growth rate and/or overproduction of a target chemical) after multiple gene deletions suggested by either the FBA or EMA technique. One reason is the capacity of some native enzymes in the various reaction steps is rate-limiting and these enzymes have not yet evolved to fit the optimal redesigned cellular metabolism [26, 38, 42, 45]. A fundamental challenge is to identify key bottlenecks that prevent cells from reaching desirable phenotypes in a reprogrammed metabolic network. This kind of problem can only be addressed adequately when many kinetic parameters are available to perform perturbation analysis to identify the rate-limiting reaction steps [14].

Due to the lack of parameters describing mass-action kinetics in vivo, different ensemble metabolic modeling (EMM) techniques were developed to sample a space of kinetic parameter sets that can describe an experimentally determined steady-state phenotype (metabolic flux distribution) of a cell and predict the stability of the network under uncertainty [46–52]. Tran et al. developed a useful method to generate an ensemble of kinetic models that are anchored to the steady-state flux distribution of a wildtype strain and used the ensemble to study the effects of over- and/or under-expression of enzyme concentrations on overproduction of a target chemical [46]. The approach used the available perturbation experiments to screen for models that match the observed phenotypes. The models that were not consistent with the experimental phenotypes were screened out of the ensemble. This screening process was repeated over several experimental data sets. After a certain number of screening steps, a small ensemble of predictive models remained and became useful for guiding further enzyme choices for over- and under-expression. The approach relied on a large set of available experiments to screen out models, but might not identify the rate limiting steps nor suggest new perturbation experiments in the absence of the experimental data [53–55].

The goal of this study is to develop a systematic multiple enzyme targeting method, called SMET, that can identify which enzymes and how much of these enzymes need to be simultaneously manipulated to achieve the desirable optimal cellular metabolisms for enhanced product yields, titers, and productivities without requiring multiple sets of experimental data. The only experimental data required by SMET is the steady-state flux distribution of the wildtype. The SMET method is tested and validated for a simple network and an *E. coli* DAHP production network. The simple network is chosen to demonstrate the SMET concept while the DAHP production network describes the production of 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP), a precursor for the production of aromatic amino acids that have numerous industrial applications in the food and pharmaceutical industries [3, 56–58].

2 Materials and methods

2.1 Problem formulation

For a given metabolic network, the principal law of mass conservation of metabolites can be written as:

$$\frac{d\mathbf{c}}{d\mathbf{t}} = \mathbf{S}\mathbf{v} \quad (1)$$

where $\mathbf{c}$ is an $m \times 1$ metabolite concentration vector, $\mathbf{S}$ is an $m \times n$ stoichiometric matrix, and $\mathbf{v}$ is an $n \times 1$ reaction rate (flux) vector. In Eq. (1), the bold lowercase and upper-

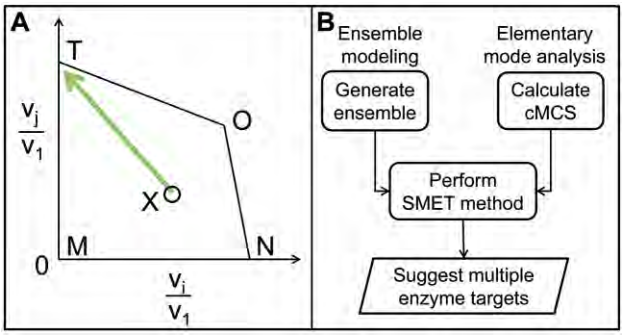

Figure 1. (**A**) A representative phenotypic space of the wildtype strain projected on 2D space of v_j/v_1 versus v_i/v_1. v_i, v_j are two different fluxes that are normalized with a reference flux v_1 such as a substrate uptake rate. (**B**) Scheme of the SMET method for systematically identifying multiple enzyme targets to manipulate for enhanced product yields, titers, and productivities.

case letters signify a vector and a matrix, respectively. Figure 1A shows an example of a steady-state phenotypic space of a metabolic network projected on the 2-dimensional space of v_j/v_1 versus v_i/v_1 and enclosed by the area connecting points M, N, O, and T. Under a defined condition, a wildtype strain can operate anywhere within this space. Assume that the desirable phenotypic state is at point T for maximizing v_j/v_1 and the phenotypic state of the wildtype strain experimentally determined is at point X. The SMET method is designed to systematically identify which enzymes and how much of these target enzymes need to be manipulated to shift the phenotypic state of the wildtype strain at point X to the desirable optimal state of the mutant strain at point T. Figure 1B shows the flow chart of the SMET method that combines both EMA and EMM to generate novel SMET metrics. These metrics suggest systematic multiple enzyme targets to perturb for enhanced product yields, titers, and productivities.

2.2 Elementary mode analysis

EMA is applied to identify all unique feasible metabolic pathways that define the phenotypic space of the network (Fig. 1A). Under steady state, the mass balance Eq. (1) can be rewritten as:

$$\mathbf{Sv} = 0 \tag{2}$$

Since $m << n$ for $\mathbf{S}$ for a typical metabolic network, Eq. (2) is an underdetermined system of homogenous equations and admits an infinite number of nontrivial solutions which all reside in the n-m dimensional null space of $\mathbf{S}$, provided $\mathbf{S}$ is full-rank.

2.2.1 Computing elementary modes (EMs)

All finite solutions of Eq. (2) that span the phenotypic space can be computed by imposing two constraints: a thermodynamic constraint $v_i > 0$ for some $i \in 1, 2, \ldots, n$ and a non-decomposability constraint. The latter constraint states that if EM_1 and EM_2 are any of two EMs, a set of non-zero flux indexes $S(EM_1)$ belonging to EM_1 is not a subset of $S(EM_2)$ and vice versa. These spanning solutions represent a set of unique and elementary pathways known as EMs [12, 59]. Metatool 5.0 was used in this study to compute all EMs of a metabolic network [60]. The theory and algorithms used to compute EMs can be found in relevant literatures [61–66].

2.2.2 Computing constrained minimal cut set (cMCS)

From a complete set of computed EMs that span the phenotypic space of a metabolic network by using EMA, it is straightforward to identify a desirable optimal phenotypic subset such as the point T of Fig. 1A corresponding to maximal yield v_j/v_1 [12]. The cMCS method is applied to identify all MCSs containing unique sets of enzymes that, when deleted, can disrupt all EMs other than those whose span includes the phenotypic subspace with maximum v_j/v_1 at point T. The software CellNetAnalyzer version 9.8 was used in this study to calculate MCSs [67].

2.3 Ensemble metabolic modeling (EMM)

From the experimentally determined steady-state flux distribution of the wildtype, EMM can be used to generate an ensemble of the kinetic parameter sets that yield the steady-state phenotypic state of the wildtype strain [46]. Each stoichiometric reaction in the metabolic network can be modeled as a series of elementary reactions, and Eq. (1) can be rewritten as:

$$\text{Model } k: \frac{\mathrm{d}\mathbf{c}_k}{\mathrm{d}\mathbf{t}} = \mathbf{S}_{\exp,k}\mathbf{v}_{\exp,k} \tag{3}$$

where $\mathbf{S}_{\exp,k}$ is the expanded stoichiometric matrix, $\mathbf{v}_{\exp,k}$ is the expanded flux vector, and the subscript k indicates the kth sampled ensemble model. Each model is a function of randomly generated enzyme fraction E and reaction reversibility R constrained by defined thermodynamic ranges (see Tran et al. [46] for details.) A Matlab script [53] was used to generate an ensemble of 1500 models for each network investigated in this study.

2.4 Systematic multiple enzyme targeting (SMET)

The SMET method was developed in this study to systematically identify multiple enzyme targets to engineer the wildtype strain to reach the desirable phenotype based on perturbation analysis. Here the perturbation analysis refers to deletion, under-, or over-expression of an individual enzyme or group of enzymes in the metabolic network and is fundamentally different from sensitivity analysis. Sensitivity analysis performs small changes of some variable values relative to steady-state conditions to predict the rate of change in a desired quantity relative to

those changes. Perturbation analysis, on the other hand, performs large changes. In the realm of Metabolic Control Analysis [68, 69], sensitivity analysis typically measures the response of one variable such as a target metabolic flux to individual infinitesimal changes in other variables such as enzyme concentrations, also known as metabolic flux control coefficients.

The SMET method involves three steps.

Step 1: Determine the flux distribution vector $\mathbf{v}_{rep}$ ($n \times 1$) that represents the most dominant simulated behavior of flux distribution vectors at the end of a chosen simulation time period. These vectors are derived from the ensemble of kinetic models generated from the original metabolic network after the introduction of MCS perturbation suggested by the cMCS analysis.

Step 2: Determine the ideal flux distribution vector $\mathbf{v}_{ideal}$ ($n \times 1$), to which the representative flux vector $\mathbf{v}_{rep}$ is to be compared. The vector $\mathbf{v}_{ideal}$ lies in the space spanned by the linear combination of the high-yield EMs resulted from the cMCS analysis.

Step 3: Compare the differences between $\mathbf{v}_{rep}$ and $\mathbf{v}_{ideal}$ element-wise to identify enzyme targets for perturbation by using two SMET metrics, the *l*- and *c*-values, and to confirm these targets by a follow-up perturbation analysis.

In steps 1 and 3, the simulated time for the enzyme perturbation was chosen arbitrarily (t = 2000 h) and long enough for the models to reach steady-states if any had existed. For metabolic networks investigated in this study, longer simulated time (t = 10000 h) was also tested and led to comparable *c*- and *l*-values for the same enzyme targets.

In step 1, two possible cases might happen after the MCS perturbation. In case 1, some ensemble models can reach steady states, indicating that the wildtype can be engineered to reach the desirable metabolic steady-state. Some ensemble models cannot reach steady states because some intermediate metabolites are abnormally accumulated and increasing the simulated time does not allow these models to reach steady states. The SMET metrics can be used to guide which limiting enzymes should be manipulated to enhance the target flux and further increase the population of kinetic models to reach steady states. It should be noted that the steady-state yields are guaranteed to be maximum after the MCS perturbation. In case 2, none of the ensemble models can achieve steady states upon the MCS knockouts, suggesting that the wildtype cannot be engineered to achieve the desirable metabolic steady-state. The SMET method must be applied to determine which enzymes and by how much of these enzymes should be perturbed to reach the optimal metabolic steady state.

2.4.1 *Determining the representative flux vector v_{rep}*

In order to find $\mathbf{v}_{rep}$, a MCS is first used as knockout targets for the perturbation analysis on the ensemble models generated for the wildtype. Most of these models or even all may not reach steady states after MCS perturbation, indicating that native enzymes of some reaction steps are limiting and cannot cope with the reprogrammed optimal cellular metabolism predicted by the cMCS analysis. At the end of the perturbation simulation time, the flux vectors from all the ensemble models are arranged as columns of the matrix $\mathbf{V}_f$ ($n \times k$). Each column of $\mathbf{V}_f$ is then normalized to unit magnitude to prevent models that predict large flux values from biasing $\mathbf{v}_{rep}$ as shown below:

$$\mathbf{v}_{fn,i} = \frac{1}{\| \mathbf{v}_{f,i} \|_2} \mathbf{v}_{f,i} \tag{4}$$

where $\mathbf{v}_{fn,i}$ denotes the ith column of $\mathbf{V}_{fn}$. The matrix $\mathbf{V}_{fn}$ is subjected to singular value decomposition (SVD) to extract $\mathbf{v}_{rep}$ as follows:

$$\mathbf{V}_{fn} = \mathbf{U}_{fn} \sum_{fn} \mathbf{W}_{fn}^{T} \tag{5}$$

$$\mathbf{v}_{rep} \triangleq \mathbf{u}_{fn,1} \tag{6}$$

where $\mathbf{U}_{fn}$ and $\mathbf{W}_{fn}$ are the left and right singular matrices of $\mathbf{V}_{fn}$, respectively. The first column of $\mathbf{U}_{fn}$ is the first singular vector ($\mathbf{u}_{fn,1}$, also the dominant vector) in the column space of $\mathbf{V}_{fn}$. Since the values across the columns of $\mathbf{V}_{fn}$ are not mean-centered, $\mathbf{u}_{fn,1}$ represents the general behavior of the flux vectors and is chosen to be $\mathbf{v}_{rep}$, provided that the effective rank of $\mathbf{V}_{fn}$ is one. The Scree plot should be examined to ensure that the first left singular vector is indeed representative of the behavior of the flux vectors of ensemble models [70].

A geometric interpretation of this computational step is useful: The normalized steady-state flux vectors can be viewed as situated on the surface of an n-dimensional unit sphere in the flux space (the column space of the normalized flux matrix $\mathbf{V}_{fn}$). The first column of $\mathbf{U}_{fn}$ is on the surface of this unit sphere and approximates the centroid of the ensemble of the normalized flux vectors if they are tightly clustered (the Euclidean distances between pairs of these vectors are much less than unity). The first column of $\mathbf{U}_{fn}$ is then representative of the cluster of flux vectors and likely to be predictive of the perturbed flux behavior of the ensemble models. Therefore, it is used to define $\mathbf{v}_{rep}$. The ratio of the square of the largest singular value to the sum of the squares of all the singular values provides a measure of how much of the variance embedded in $\mathbf{V}_{fn}$ is captured by the first column of $\mathbf{U}_{fn}$. If it is close to 100%, then it implies that the data represented by $\mathbf{V}_{fn}$ lies mostly in one-dimensional space spanned by the first column of $\mathbf{U}_{fn}$. The Scree plot provides a visualization of this ratio. The SMET method recommends a ratio greater than 90% to be used.

2.4.2 Determining the ideal flux vector $\mathbf{v}_{ideal}$

By performing gene knockouts guided by the cMCS analysis, a subset of EMs that result in desirable phenotypes with maximum target product yield remains. Each of the remaining EMs (assuming there are '*r*' of them) is independent of each other and they span an r-dimensional high-yield EM subspace. The projection of $\mathbf{v}_{rep}$ onto this subspace by using a projection matrix $\mathbf{P}_o$ yields the ideal flux vector $\mathbf{v}_{ideal}$. $\mathbf{P}_o$ is calculated by performing SVD for the high-yield EM matrix as described below:

$$\mathbf{EM}_0 = \mathbf{U}_0\boldsymbol{\Sigma}_0\mathbf{V}_0^{\mathrm{T}} \tag{7}$$

$$\mathbf{P}_0 = \mathbf{U}_0\mathbf{U}_0^{\mathrm{T}} \tag{8}$$

$$\mathbf{v}_{\mathrm{ideal}} = \mathbf{P}_0 \cdot \mathbf{v}_{\mathrm{rep}} \tag{9}$$

$\mathbf{EM}_0$ $(n \times r, n > r)$ represents the high-yield EM matrix. The SVD to be performed here is the reduced decomposition with the left singular matrix $\mathbf{U}_0$ having the same dimension as $\mathbf{EM}_0$ [71]. $\mathbf{U}_0$ has orthonormal columns but not rows and the columns of $\mathbf{U}_0$ span the same column space as that of $\mathbf{EM}_0$. $\boldsymbol{\Sigma}_0$ is a square diagonal matrix containing positive diagonal elements arranged in decreasing order. The product of $\mathbf{U}_0\mathbf{U}_0^{\mathrm{T}}$ forms a projection matrix $\mathbf{P}_0$ of rank *r* (not an identity matrix) that projects any given vector onto the column space of $\mathbf{U}_0$, which corresponds to the high-yield EM subspace. Thus, $\mathbf{v}_{ideal}$ is the minimum norm least-squares approximation to $\mathbf{v}_{rep}$ on the high-yield EM subspace.

2.4.3 Determining multiple enzyme targets for perturbation by *l*-values

By comparing $\mathbf{v}_{rep}$ and $\mathbf{v}_{ideal}$ element-wise, one can identify a rate-limiting reaction(s) with a large difference(s) between the corresponding elements of the two flux vectors. The flux ratios r_i are designed to access these differences and determine what degree each reaction flux in $\mathbf{v}_{rep}$ deviates from its value in $\mathbf{v}_{ideal}$.

$$r_i \triangleq \frac{\mathbf{v}_{\mathrm{ideal},i}}{\mathbf{v}_{\mathrm{rep},i}} \quad \text{where } i = 1,2,3,\ldots,n \tag{10}$$

For a linearly-connected pathway, a down-stream reaction *i* with a large flux ratio r_i is not necessarily rate-limiting because its immediate upstream reactions can have large flux ratios and do not supply enough flux to the down-stream reaction *i*. Therefore, to find the "true" rate-limiting reactions for perturbation, it is important to compare the flux ratios between a reaction and all immediately preceding reactions and identify the reaction along a linearly connected pathway to exhibit a large change in flux ratios. This comparison is done by evaluating *l*-values as follows:

$$l_i \triangleq \frac{r_i}{r_{i,\min r}}, \quad \text{where } i = 1,2,3,..n \tag{11}$$

where r_i is the flux ratio for reaction *i* and $r_{i,minr}$ is the minimum flux ratio of all the reactions that are immediately upstream of reaction *i*. A reaction "immediately upstream" of reaction *i* is a reaction that has at least one product which is a reactant of reaction *i*. The minimum upstream flux ratio is chosen instead of using an averaging approach to prevent multiple upstream reactions from excluding potential perturbation targets. Reactions with *l*-values significantly deviated from one (*l*-value >> 1 or *l*-value << 1) are those that are more likely to be effective enzyme targets for perturbation.

2.4.4 Determining levels of enzyme perturbation by *c*-values

When multiple enzyme targets are selected, the suggested levels of enzyme perturbation are determined by the *c*-values as follows:

$$c_i = \frac{r_i}{r_{\mathrm{input}}} \quad \text{where } i = 1,2,3,\ldots,n \tag{12}$$

where r_{input} refers to the flux ratio of the substrate uptake reaction.

2.4.5 Determining the *s*-values

To determine the deviation of the flux distributions of ensemble models from their corresponding steady-state values, the *s*-values were calculated as follows:

$$\mathbf{s}_k = \| \mathbf{S}\mathbf{v}_k \|_2 \tag{13}$$

where $\|\cdot\|_2$ represents the two-norm of a vector. The flux vector $\mathbf{v}_k$ of the *k*th kinetic model reached steady state if $s_k \approx 0$ (a threshold of <1E−3 was used) at the end of the simulation time.

2.4.6 Computational setup

The SMET method was performed to calculate the SMET metrics using Matlab version 7.11 (The MathWorks Inc., Natick, MA, 2012). The computational code was parallelized for the SMET method and written to interface with computational modules that compute EMs [60] and generate parameter sets for ensemble models [46].

2.5 Metabolic network models

2.5.1 A simple network

Two metabolic networks were investigated in this study. The first one was the simple network used to demonstrate the SMET method (Fig. 2A) [15]. This network contained nine reactions, six of which were irreversible, and nine metabolites, five of which were internal. The standard Gibbs free energies of these reactions (SGFEs) were contrived for the EMM. Reaction stoichiometry, standard Gibbs free energies, and wild-type steady-state fluxes are presented in Supporting information, Table S1. For the EMM, internal metabolites including A and P were con-

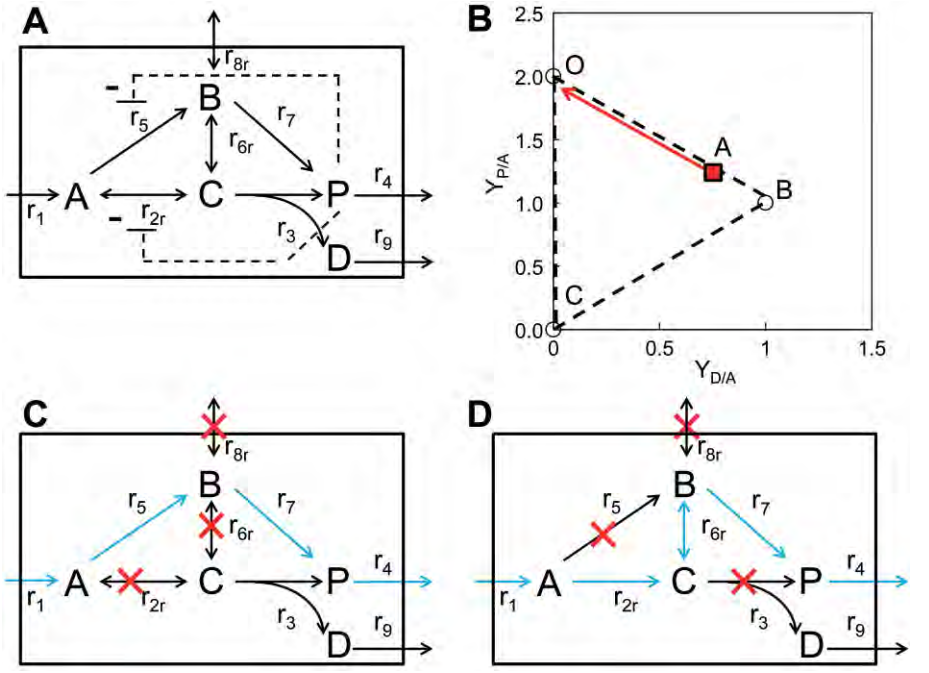

Figure 2. (**A**) The metabolic map of the simple network. (**B**) The phenotypic space inherent to the simple metabolic network. (**C**) Optimal pathway 1. The reaction knockouts (e.g. r_{2r}, r_{6r}, r_{8r}) that eliminated all EMs other than the optimal pathway 1 were shown as red Xs. (**D**) Optimal pathway 2. The reaction knockouts (e.g. r_3, r_5, r_{8r}) that eliminated all EMs other than the optimal pathway 2 were shown as red Xs.

strained to vary between 0.01 and 100 times their wild-type steady-state concentrations. The metabolite P exerts enzymatic feedback inhibition on r_{2r} and r_5.

2.5.2 *DAHP production network*

The DAHP production network was previously described [54] (Supporting information, Fig. S1) and chosen for this study because (1) it was well-studied experimentally, enabling validation of the prediction capability of the SMET method and (2) it allowed direct comparison between results using the SMET method and a previously described ensemble modeling interpretation [54]. The network contained 38 reactions, 28 of which were reversible and 36 metabolites, 31 of which were internal. Supporting information, Table S2 lists the full names and abbreviations of enzymes and metabolites present in the network, reaction stoichiometry, wild-type steady-state fluxes, standard Gibbs free energies, and regulatory metabolites [54].

3 Results

3.1 Case study 1: The simple network

3.1.1 *Elementary mode analysis (EMA) for identifying the phenotypic space of the network*

There were a total of 9 EMs inherent to the simple network (Fig. 2A). The phenotypic space of the simple network projected on the 2D space of $Y_{P/A}$ versus $Y_{D/A}$ was shown in Fig. 2B. This space was enclosed by the area connecting points C, O, and B. Considering only mass-balance and reversibility constraints, the wildtype expression could be located anywhere in the area COB. When some reaction fluxes were measured for the wildtype under a defined condition [15], the metabolic flux distribution that defined the metabolic state of the wildtype could be determined as shown by point A (Fig. 2B). The most optimal pathway achieving the maximum yield $Y_{P/A}$ = 2 mol/mol was located at point O. Therefore, it is desirable to constrain the metabolism of the wildtype to the optimal one by multiple reaction deletions.

3.1.2 *Constrained minimal cutset (cMCS) for identifying MCS knockouts to achieve the desirable phenotypic state*

By using the cMCS method [33], two optimal metabolic pathways could be identified to convert A to P at the maximum yield $Y_{P/A}$ of 2 mol/mol (Fig. 2C and D). The first optimal metabolic pathway contained reactions r_1, r_4, r_5, and r_7 and could be constrained to operate by deleting reactions r_{2r}, r_{6r}, and r_{8r}. The second optimal metabolic pathway included r_1, r_{2r}, r_4, r_{6r}, and r_7 and could be constrained to operate by disrupting reactions r_3, r_5, and r_{8r}. Both pathways could be retained by only knocking out reactions r_3 and r_{8r} and constituted the high-yield $\mathbf{EM}_0$ (see Section 2; Fig. 2C and D). From the knowledge of the metabolic flux distribution determined for the wildtype (e.g. point A, Fig. 2B), it is critical to evaluate whether the wildtype can be engineered after MCS knockouts to reach the designed optimal metabolic pathway (e.g. point O) to achieve maximum $Y_{P/A}$. This problem could only be addressed adequately with kinetic metabolic network modeling as demonstrated by the SMET method.

3.1.3 Generation of ensemble models anchored to the steady-state flux distribution of the wildtype

The EMM was implemented to generate ensemble models whose kinetic parameters were within the space that achieved the steady-state flux distribution of the wildtype. By using the ensemble models generated, it is possible to systematically analyze the effects of single enzyme perturbations on desirable phenotypes such as high product yields, titers, and productivities. However, this approach is inefficient and computationally expensive, especially for a large metabolic network. For a single enzyme perturbation with either over- or down-expression, it is often not obvious to conclude which enzymes and how much these enzymes should be manipulated to achieve the desirable phenotype (e.g. $Y_{P/A}$ = 2 mol/mol; Supporting information, Figs. S2 and 3), especially when the experimental perturbation data for comparison is not available. This approach to search for the optimal phenotype becomes more daunting if multiple enzymes are targeted due to a large search space (enzyme targets, over-/down-expression, and the quantitative level of perturbation). For instance, if three enzymes belonging to three reactions of the 9-reaction simple network were targeted for perturbation with the same level of over- or down-expression, a total of 672 perturbations would be carried out. The SMET method was developed to perform selective simultaneous perturbation of multiple enzyme targets instead of iterative screening perturbation of single enzymes, which reduces computation time significantly and enhances prediction capability.

3.1.4 SMET method for identifying rate-limiting reaction steps for enhanced product yields and productivities

The cMCS analysis suggested that the deletion of r_3 and r_{8r} would collapse the complete set of all EMs of the simple network to two optimal pathways with a desirable maximum yield of 2 mol P/ mol A. By using the ensemble models anchored to the steady-state flux distribution of the wildtype, the SMET method simulated the MCS knockouts of r_3 and r_{8r} to determine whether the wildtype metabolism could be engineered to reach the desirable metabolic state (e.g. point O, Fig. 2B).

The MCS perturbation only shows that about 13% of the kinetic models approached steady states to achieve a maximum yield of 2 mol/mol (Fig. 4A). Table 1 shows $\mathbf{v}_{rep}$, $\mathbf{v}_{ideal}$, *l*- and *c*-values for reactions in the network, generated by the SMET method. Based on the Scree plot of the SVD of the normalized simulated flux matrix, the first left singular vector captured 96% of the variance embedded in the flux matrix, indicating that $\mathbf{v}_{rep}$ was representative of the flux distribution predictions of the ensemble models (Supporting information, Fig. S4B). The l-values indicated r_{2r}, r_7, and r_5 as the ideal over-expression targets. The *c*-values suggested that r_{2r}, r_7, and r_5 should be overexpressed by scaling the enzyme levels by factors of 1.9, 1.3, and 1.2, respectively.

Table 1. The SMET metrics for the simple network. It should be noted that $\mathbf{v}_{rep}$ is not mass-balanced because some ensemble models do not reach steady states. Reactions not shown have null fluxes.

Enzyme	$\mathbf{v}_{rep,i}$	$\mathbf{v}_{ideal,i}$	r_i	l_i	c_i
r_{2r}	−0.0908	−0.1764	1.9430	1.9430	1.9430
r_7	0.3184	0.4215	1.3238	1.3051	1.3238
r_5	0.5103	0.5979	1.1716	1.1716	1.1716
r_1	0.4215	0.4215	1.0000	1.0000	1.0000
r_4	0.6494	0.8430	1.2981	0.9805	1.2981
r_{6r}	0.1739	0.1764	1.0144	0.5221	1.0144

($\mathbf{v}_{rep,i}$, representative flux vector; $\mathbf{v}_{ideal,i}$, ideal flux distribution vector; $\mathbf{r}_i$, flux ratio for reaction **i**; $\mathbf{l}_i$, see Eq. (11); $\mathbf{c}_i$, see Eq. (12)

Based on the SMET prediction, over-expression of r_{2r}, r_7, and r_5 following the MCS perturbation was simulated and expected to increase the fraction of kinetic models reaching steady states. To demonstrate the effectiveness of these multiple enzyme targets selected for perturbation, perturbation analysis was also performed for the individual enzyme targets. Figures 3A–E and 4A–E showed the flux and yield distributions of all ensemble models, respectively. Not all of the 1500 models reached steady-states except for those that approached the maximum yield $Y_{P/A}$ = 2 mol/mol as constrained by MCS knockouts (Figs. 3F and 4F). Some kinetic models did not reach steady states because they had a high accumulation of metabolites B and C. Over-expression of r_{2r}, r_7, and r_5 with the MCS perturbation significantly increased the fraction of kinetic models reaching steady states (from ~13% to ~24%, based on the *s*-values) and maximum yields ($Y_{P/A}$ = 2 mol/mol) as suggested by the SMET analysis (Fig. 4E). A fraction of steady state models also had higher absolute fluxes than that of the wildtype (1.25 mol/g DCW/h), indicating that the perturbation suggested by the SMET analysis could improve the productivities and titers (Figs. 3F and 4F).

3.2 Case study 2: The DAHP production network

3.2.1 Identifying and constraining the optimal metabolic state of the DAHP production network with maximum $Y_{DAHP/GLC}$ by using EMA and cMCS

By applying EMA, a total of 26 EMs were identified. Of these EMs, one EM exhibited the maximum theoretical yield of 0.86 mol DAHP/mol glucose. From the cMCS analysis, a total of 100 MCSs were found that repressed all EMs except the maximum $Y_{DAHP/GLC}$ EM. The MCS containing knockout genes *pfl*, *ppc*, *pyk*, and *zwf* was chosen out of the 100 MCSs because (1) it minimized the number of "dead-end" reactions of a pathway that ended in metabolites that were not reactants of any reaction and (2) some of these knockout genes were experimentally implemented to increase DAHP production [72, 73]. It should be noted that simulation of other MCSs identified similar

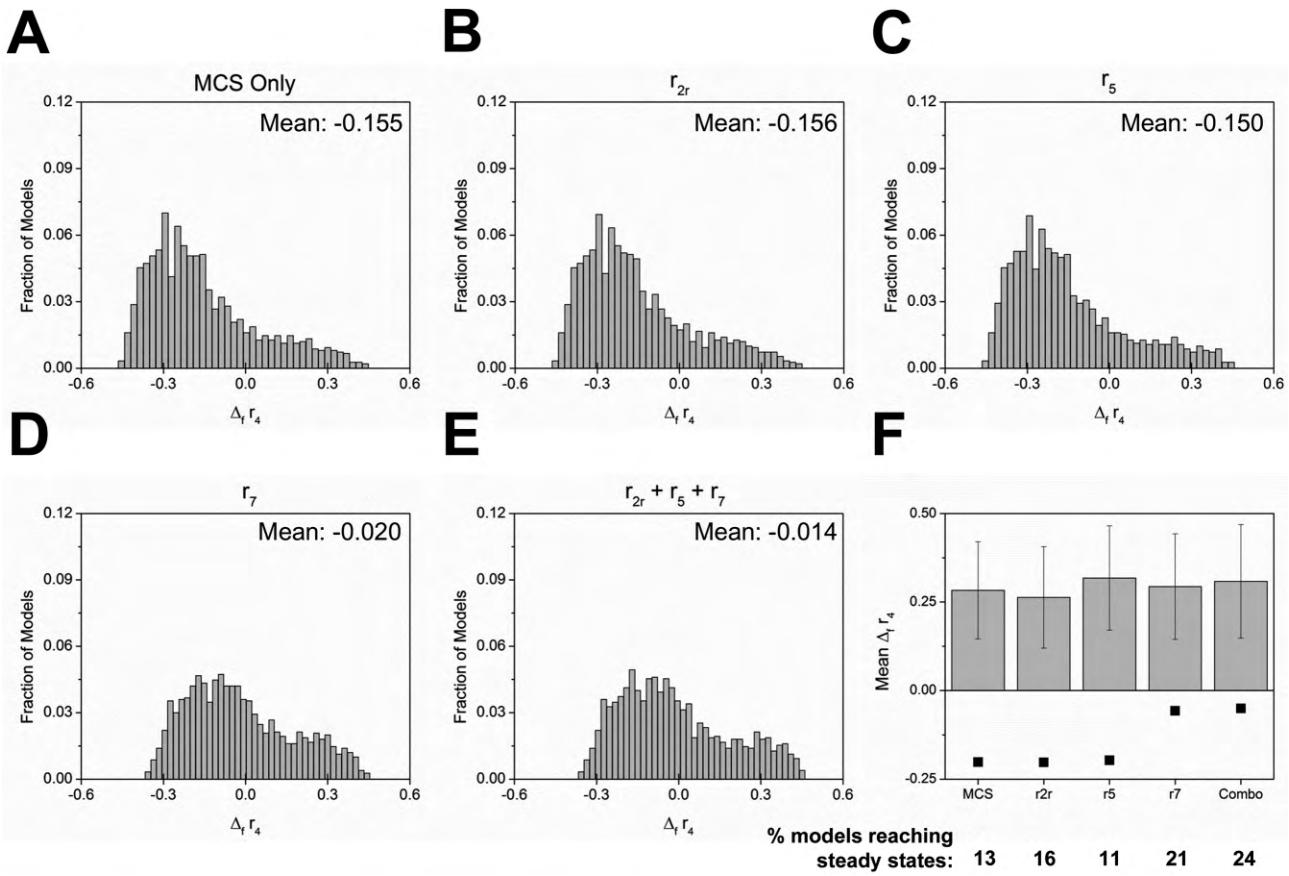

Figure 3. Distributions of fractional changes in r_4 fluxes from the wildtype flux resulting from perturbation analysis of the simple network following MCS knockouts. (**A**) MCS perturbation; (**B–E**) MCS plus r_{2r}, r_5, r_7, and $r_{2r} + r_5 + r_7$ perturbations, respectively; (**A–E**) show the distributions of fractional changes in r_4 for all ensemble models; (**F**) shows the mean fractional change for kinetic models which reached steady states after enzyme perturbations (gray bars, steady state %) compared to the mean for the entire population of models (black squares). In (**B–E**), r_{2r} over-expressions were set by a factor of 1.9, r_5 set to 1.2, and r_7 set to 1.3. Fractional change $\Delta_f r_4$ is defined as $(r_4 - r_{4,wt})/r_{4,wt}$.

enzyme targets based on the calculated SMET metrics (results not shown).

EMM was then applied to generate a set of 1,500 kinetic models anchored to the steady-state flux distribution of the wildtype strain. Consistent with the observation made for the analysis of the simple network, a systematic single enzyme perturbation on these kinetic models was carried out (although it was computationally expensive) and was found to be ineffective in achieving the maximum $Y_{DAHP/GLC}$ and the highest flux r_{DAHP} (Supporting information, Fig. S5). This iterative approach suggested practically no information about the enzyme targets for perturbation, especially in the absence of experimental perturbation data used for observed phenotype comparison as described by other studies [46, 54, 55].

3.2.2 *Systematically identifying multiple enzyme targets to achieve maximum Y_{DAHP} yield and enhance $\mathbf{v}_{DAHP}$ flux by the SMET analysis*

The MCS perturbation for the DAHP network shows that none of the kinetic models reached steady states at the end of simulation (Fig. 6A) due to a significant accumulation of pyruvate. This result implies the enzyme capacities for some reactions in the wildtype strain might have become limiting and hindered attainment of the optimal metabolic state with the highest $Y_{DAHP/GLC}$ of 0.86 mol/mol. The SMET analysis was then performed on the DAHP production network. The calculated $\mathbf{v}_{rep}$, $\mathbf{v}_{ideal}$, *l*-, and *c*-values are shown in Table 2, with reactions sorted by descending *l*-values. The first four reactions (Pps, Tal, AroG, and Tkt2) were selected to be the over-expression targets because they had *l*-values significantly larger than one and formed a natural cluster well separated from those trailing after.

The suggested enzyme over-expressions were simulated in four groups with the levels based on the *c*-values

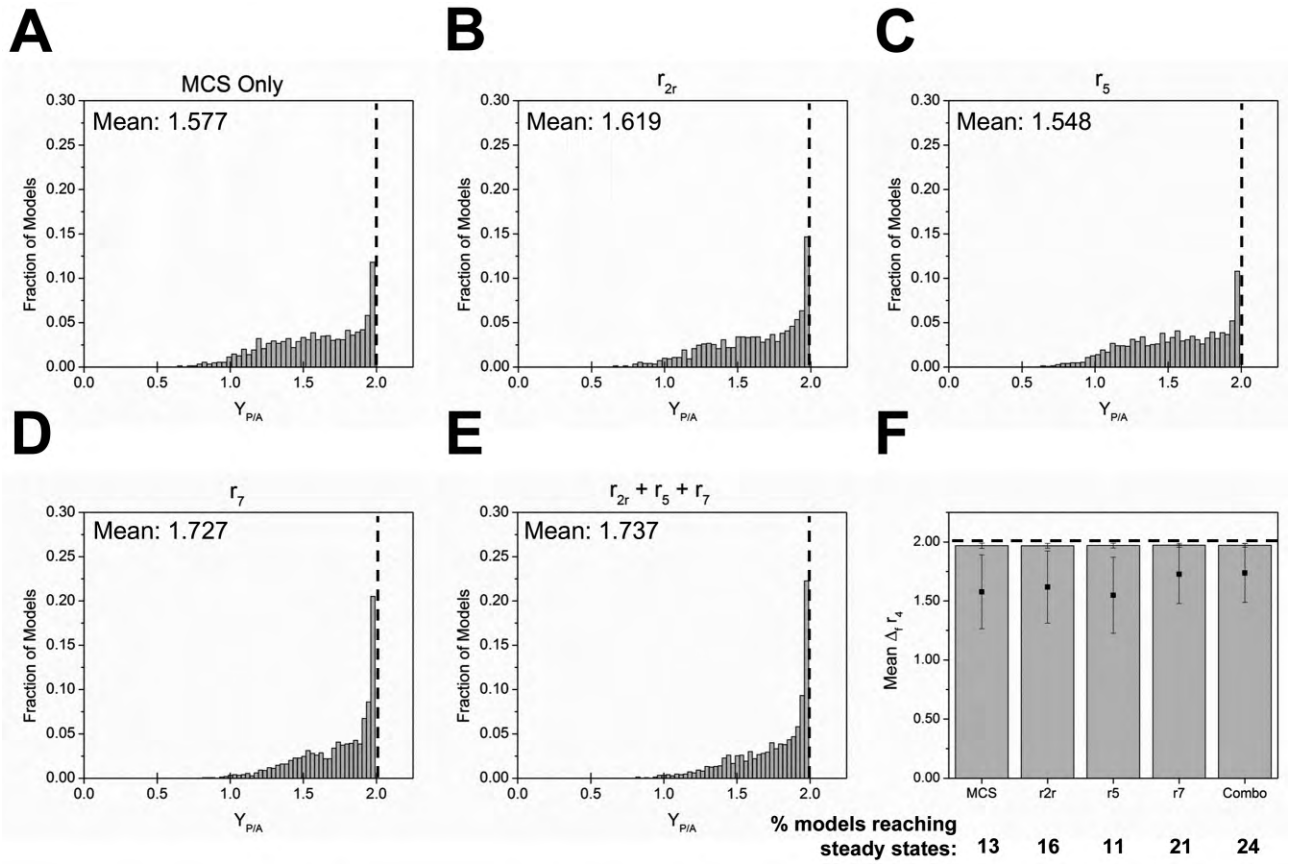

Figure 4. Distributions of yields $Y_{P/A}$ resulting from perturbation analysis of the simple network following MCS knockouts. (**A**) MCS perturbation; (**B**–**E**) MCS plus r_{2r}, r_5, r_7, and $r_{2r} + r_5 + r_7$ perturbations, respectively; (**A**–**E**) show the distributions of yields $Y_{P/A}$ for all ensemble models; (**F**) shows the mean yields $Y_{P/A}$ for kinetic models that reached steady states after enzyme perturbations (gray bars, steady state %) compared to the mean yields for the entire model population (black squares). In (**B**–**E**), r_{2r} over-expressions were set by a factor of 1.9, r_5 set to 1.2, and r_7 set to 1.3. The dashed lines in all panels represent maximum theoretical yield of 2 mol P/mol A. The wildtype yield $Y_{P/A}$ was 1.25 mol/mol.

(Table 2). Group I included Pps, Tkt2, and Tal. Group II included the top four suggested enzymes, which are Pps, Tkt2, Tal, and AroG. Since Tkt1 and Tkt2 reactions were catalyzed by the same transketolase, Group III that consisted of all enzymes in Group I and Tkt1 were also investigated to determine whether additional Tkt1 over-expression had any effect. Group IV over-expressed only Pps, Tkt1, and Tkt2, which were the same experimentally determined over-expression set suggested by the previously described ensemble modeling interpretation [54].

The DAHP flux and yield distributions for all ensemble models were shown in Figs. 5 and 6. The result shows that some kinetic models reached steady states for all over-expression groups (Figs. 5B–E and 6B–E). Groups I and III exhibited very similar yield and flux distribution profiles, indicating that over-expression of Tkt1 along with Tkt2 were unlikely to have a significant impact on the flux and yield distributions. The enzyme over-expression of Group II clearly resulted in the highest yields for the most models among all four groups, consistent with the enzyme targets manipulated experimentally to achieve the high yield of DAHP close to theoretical limit [57]. Figures 5F and 6F also show that SMET could directly identify the simultaneous over-expression of enzyme targets in Group II that could increase the fraction of ensemble models (1.1%) to achieve steady states while approaching the maximum yield with higher absolute desirable fluxes.

3.2.3 *Iterative SMET analysis*

The Group II over-expression data was chosen for a follow-up, second-round SMET analysis. The result shows that the first top four enzymes with the highest *I*-values for further perturbation were the same four as before but their *I*-values (slightly greater than one) were not nearly so large as those computed in the first SMET iteration (Supporting information, Table S3). The fifth (Rpe) and sixth (Ei) enzyme suggestions had comparable *I*-values of slightly larger than one.

Table 2. The SMET metrics for the first round of SMET analysis for the DAHP network

Enzyme	$v_{rep,i}$	$v_{ideal,i}$	r_i	l_i	c_i
Pps	0.0041	0.2132	51.975	64.556	64.765
Tal	0.0072	0.0609	8.4911	9.1677	10.581
AroG	0.0393	0.1828	4.6520	7.8121	5.7967
Tkt2	–0.0321	–0.1218	3.7953	4.4578	4.7292
Ei	0.2648	0.2132	0.8051	1.3520	1.0032
Rpe	–0.0134	–0.0609	4.5415	1.1966	5.6590
Pfk	0.1646	0.1523	0.9255	1.0870	1.1532
Pgi	0.2504	0.2132	0.8514	1.0574	1.0609
Tkt1	0.0180	0.0609	3.3908	1.0119	4.2251
Eiibc	0.2648	0.2132	0.8051	1.0032	1.0032
Eno	0.3069	0.1828	0.5955	1.0020	0.7420
Fba	0.1644	0.1523	0.9263	1.0009	1.1543
Gpm	0.3075	0.1828	0.5943	1.0001	0.7405
Pgk	0.3075	0.1828	0.5943	1.0000	0.7405
Rec_{atp}	0.1614	–0.1828	–1.1322	1.0000	–1.4108
Rec_{nadh}	0.3054	0.1828	0.5984	1.0000	0.7456
$Glucose_{in}$	0.2657	0.2132	0.8025	1.0000	1.0000
Tpi	0.1644	0.1523	0.9262	0.9999	1.1541
Eiia	0.2649	0.2132	0.8048	0.9998	1.0029
Hpr	0.2649	0.2132	0.8050	0.9998	1.0030
$Dahp_{out}$	0.0393	0.1828	4.6499	1.0000	5.7941
Rpi	0.0182	0.0609	3.3508	0.7378	4.1753
Gap	0.3075	0.1828	0.5943	0.6416	0.7405

($\mathbf{v}_{rep,i}$, representative flux vector; $\mathbf{v}_{ideal,i}$, ideal flux distribution vector; $\mathbf{r}_i$, flux ratio for reaction **i**; $\mathbf{l}_i$, see Eq. (11); $\mathbf{c}_i$, see Eq. (12)

From simulation of the top four, five, and six over-expressions at levels suggested by the *c*-values in the follow-up SMET analysis, respectively, the distribution curve for DAHP fluxes are more spread out and shifted to the right with larger mean values as additional enzymes were over-expressed (Supporting information, Figs. S6 and 7). For all three sets of enzyme over-expressions, the fractions of ensemble models that reached steady states could achieve the maximum yield of 0.86 mol DHAP/mol glucose (Supporting information, Fig. S7E) and improved fluxes (Supporting information, Fig. S6E).

In both the simple and DHAP networks, it should be emphasized that the down-expression of these enzymes did not increase the number of kinetic models that reached steady states and improve absolute fluxes (results not shown) as expected because *l*-values are only slightly lower than 1 (Tables 1 and 2 and Supporting information, Table S3).

4 Discussion

The SMET method was developed to systematically and directly identify multiple enzyme targets for gene deletion, over- and down-expression to achieve an optimal desirable metabolic state such as overproduction of a target chemical. The input for the SMET method included a metabolic network, standard Gibbs free energy of reactions, and a steady-state flux distribution of the wildtype as the only required experimental data. The output of the SMET method provided a set of SMET metrics consisting of the *l*- and *c*-values for the reactions in the metabolic network. These reactions were ranked from the highest to lowest based on the *l*-values. Those that have the *l*-values significantly different from one were chosen for enzyme perturbation. The *l*- and *c*-values suggest what enzymes and by how much of these enzymes, respectively, need to be manipulated to achieve the optimal metabolic state. The significant advantage of using the SMET method is to avoid long and tedious computational burden of having to screen a large number of enzyme combinations for perturbation in the network, especially when experimental data is not available for perturbation.

The SMET method was tested and validated by analyzing the simple and DHAP metabolic networks. It has been experimentally reported that deleting competing pathways and increasing fluxes to synthesize the precursor metabolites-phosphoenolpyruvate (PEP) and erythrose-4-phosphate (E4P) are important for enhanced DHAP production. Solely based on the steady-state metabolic flux distribution of the wildtype, the SMET analysis suggests simultaneous gene deletion targets (*ppc* encoding phosphoenolpyruvate carboxylase, *pyk* encoding pyruvate kinase, *zwf* encoding glucose-6-phosphate-1-dehydrogenase, and *pfl* encoding pyruvate formate lyase) and simultaneous enzyme over-expression targets (Pps, Tkt, Tal, and AroG). Very interestingly, these simultaneous enzyme targets for perturbation are consistent with experimental data [57, 72]. Deletion of Ppc [73], Pyk [72], and Pfl increases the PEP pool while deletion of Zwf enhances the glycolytic flux for synthesizing PEP and E4P pools. Over-expression of Pps [3] enhances the PEP pool while over-expression of Tkt [74] and Tal [75] increases the E4P pool. Over-expression of AroG [3, 57, 72] results in the improved DHAP production.

For relevant industrial and laboratory host strains such as *E. coli* and *Saccharomyces cerevisiae*, the levels of over- and/or down-expression of multiple simultaneous target enzymes suggested by the SMET method should be invaluable for strain engineering, and can be implemented by engineering proteins to change their enzyme activities [76], manipulating gene copy numbers [77], and controlling transcription and translation rates by various synthetic biology toolboxes (promoter engineering [78], gene codon optimization [79], and intergenic gene engineering [80].)

While structural metabolic network modeling approaches such as FBA and EMA have proven very powerful to identify the optimal metabolic state and suggest genetic manipulations to reach it, they cannot guarantee whether the wildtype strain can be engineered to reach that optimal metabolic state after the genetic

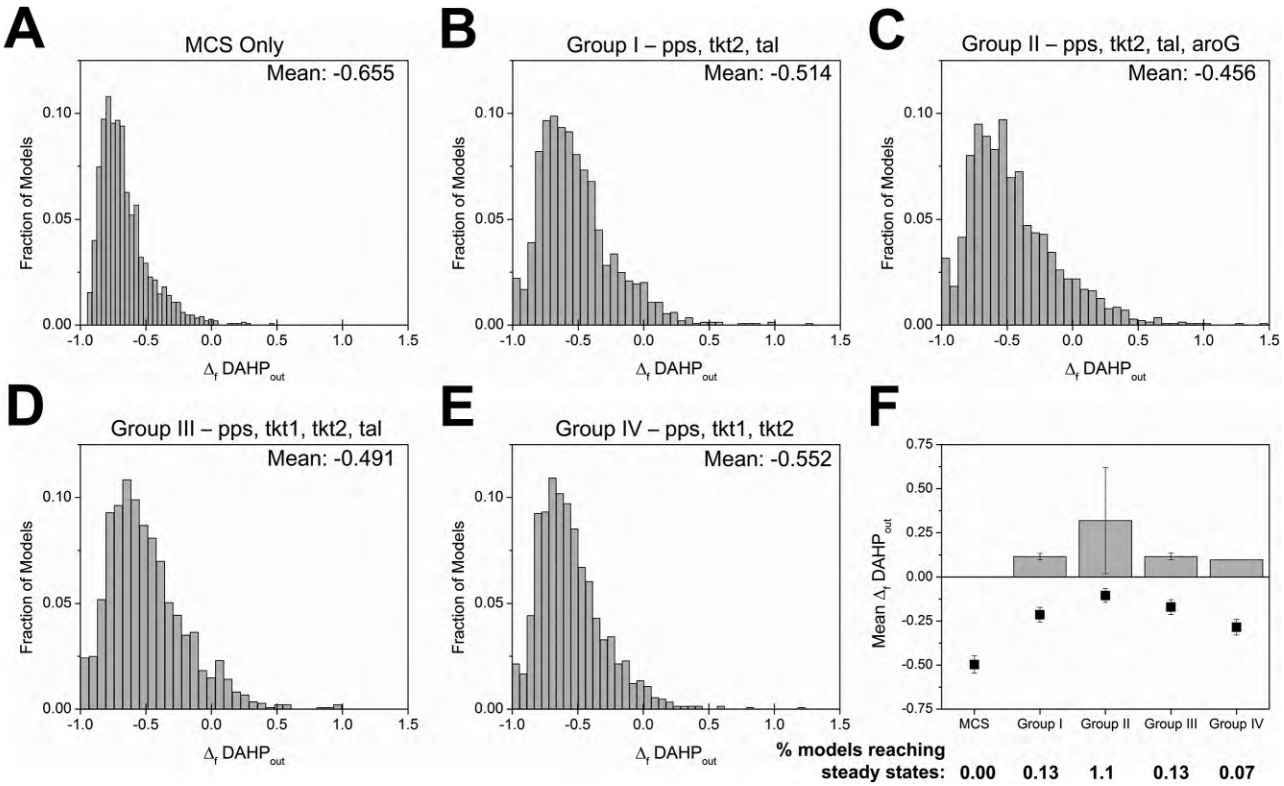

Figure 5. Distributions of fractional changes in DAHP flux relative to the wild type resulting from the perturbation of the predicted groups of targeted enzymes and MCS knockouts. (**A**) MCS perturbation; (**B**–**E**) MCS plus Group I–IV perturbations, respectively; (**A**–**E**) show the distributions of fractional changes in DAHP flux for all ensemble models; (**F**) shows mean fractional changes in DAHP flux for kinetic models that reached steady states with positive relative flux after enzyme perturbations (gray bars, steady state %), compared to the mean fractional changes for the entire population of models (black squares). In (**B**–**E**), levels of enzyme perturbation in each group were based on the c_i/c_{input} in Table 2. Fractional change $\Delta_f DAHP_{out}$ is defined as $(DAHP_{out} - DAHP_{out,wt})/DAHP_{out,wt}$. The wild-type flux was 0.26 mmol/gDCW/hr.

manipulations. These approaches were limited to predict what the reaction bottlenecks were because they did not examine the dynamics of the metabolic network transitioning from the wildtype metabolic state to the defined optimal one. The optimal metabolic state sometimes could be achieved experimentally but only after undergoing the directed metabolic pathway evolution [26, 38]. To address this challenge, the SMET method was developed to determine whether the wildtype strain could be engineered to achieve the optimal metabolic state with genetic manipulations as predicted by structural metabolic network modeling. In addition, the SMET method could identify the reaction bottlenecks of a metabolic network solely based on its steady-state flux distribution. Once these bottlenecks were compensated for by enzyme perturbation, the network could achieve its optimal metabolic state. The calculated SMET metrics suggest what enzymes and how much of these enzymes should be manipulated to achieve the target optimal metabolic state. It should be noted that even though the SMET method uses both EMA and cMCS methods to design the optimal metabolic state of the desirable mutants, other techniques based on the FBA framework can also be used with SMET.

The SMET method is quite different from the metabolic control analysis that is typically applied to determine the metabolic flux control coefficients [68, 69]. The SMET method seeks to perform multiple simultaneous enzyme perturbations instead of one at a time, and the enzyme perturbation typically involves large changes in enzyme concentrations. Indeed, this study showed that performing single enzyme perturbation is not an effective way to identify the target enzymes that can be manipulated to reach the optimal metabolic state due to a large screening space required and the robustness of the network to a single perturbation. The combinations of the number of enzymes to be manipulated and the levels of enzymes required for either over- or down-expression lead to computationally infeasible approaches.

The SMET method is also different from the other methods that analyze ensemble models by performing sequential single enzyme perturbations [46, 54, 55]. The latter analysis approach relies on the large set of experimental data to screen out the models that do not match

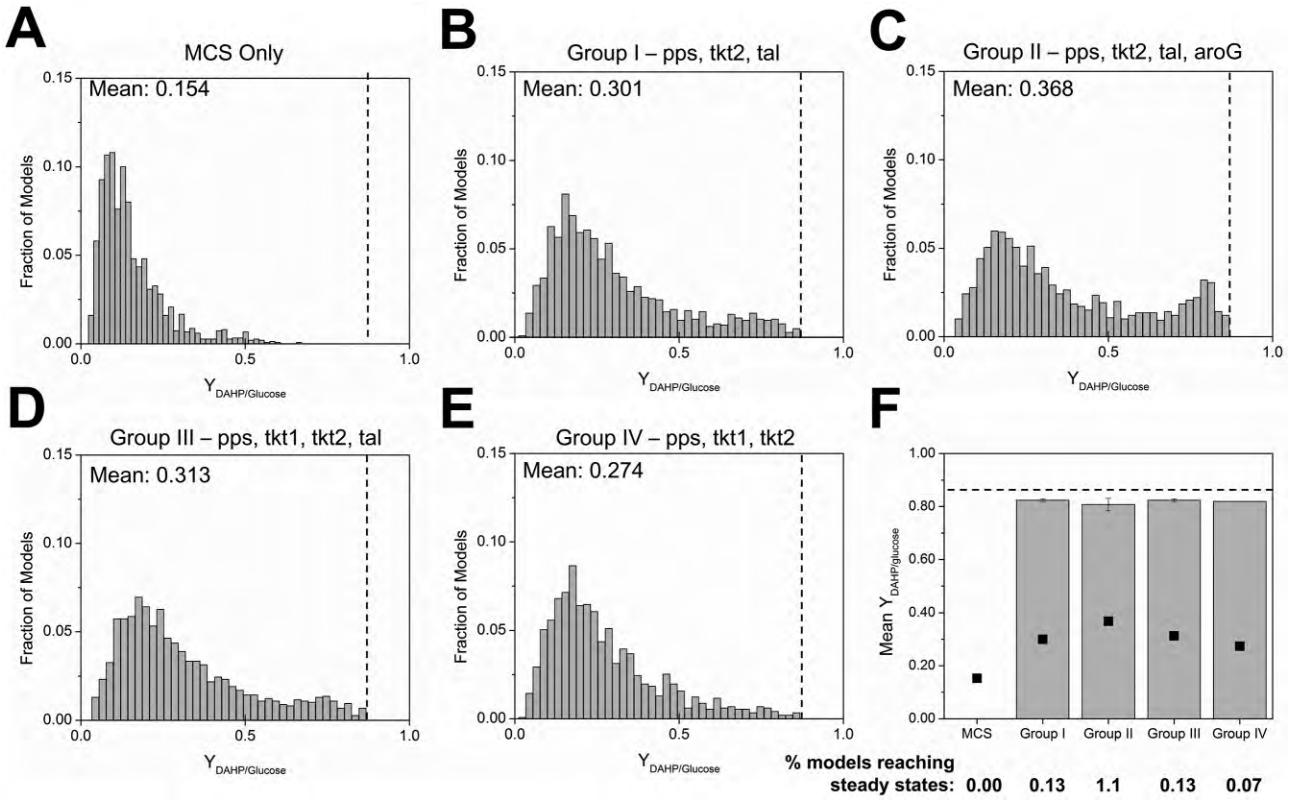

Figure 6. Distributions of yields resulting from the perturbation of the predicted group of targeted enzymes and MCS knockouts. (**A**) MCS perturbation; (**B**–**E**) MCS plus Group I–IV perturbations, respectively; (**A**–**E**) show the distributions of yields for all ensemble models. In (**B**–**E**), levels of enzyme perturbation in each group were based on the *c*-values in Table 2, (**F**) shows mean yields for kinetic models with positive relative flux that reached steady states after enzyme perturbations (grey bars, steady state %), compared to the mean yields for the entire population of models (black squares). In all panels, the dashed line represents maximum theoretical yield of 0.86 mol DAHP/mol glucose. The wild-type yield was 0.2 mol/mol.

the observed experimental phenotypes. After extensive screening, the approach seeks to retain a subset of kinetic models that would be applicable for single enzyme perturbation. This approach however does not guide what experiments to perform starting from the wildtype in the absence of available experimental data. On the contrary, the SMET method requires only the experimentally determined steady-state flux distribution of the wildtype to identify its metabolic state and suggests what enzymes and by how much of these enzymes need to be manipulated to achieve the optimal metabolic state.

As with any other metabolic network modeling approach, the prediction accuracy of the SMET method depends on the metabolic network and determination of the steady state flux distribution of the wildtype. The ^{13}C-labeling metabolic flux analysis is a useful tool to determine the steady-state flux distribution of the wildtype and validate the network accuracy [81]. This study has also found that computational time for the SMET method can become demanding for performing multiple simultaneous enzyme perturbations, especially when the size of the metabolic network increases. This problem can be addressed with task parallel computation as implemented in this study. It is anticipated that the SMET method should become useful for rational strain design to achieve novel programmed phenotypes.

We acknowledged the Newton HPC Program at the University of Tennessee, Knoxville for using the supercomputing machines. C.T.T. acknowledged the laboratory start-up, SEERC, and JDRD seed funds from the University of Tennessee, Knoxville and a grant (DE-PS02-06ER64304) from the Bioenergy Science Center (BESC), a DOE-Funded Bioenergy Center. D.F., T.W., D.B., C.T.T. designed experiments. D.F. and R.A.T. performed the experiments. D.F., R.A.T., T.W., D.B., and C.T.T. analyzed the data. C.T.T. wrote the paper. All authors read, edited, and approved the paper.

The authors declare no conflict of interest.

5 References

[1] Bailey, J. E., Toward a science of metabolic engineering. *Science* 1991, *252*, 1668–1675.

[2] Stephanopoulos, G., Vallino, J., Network rigidity and metabolic engineering in metabolite overproduction. *Science* 1991, *252*, 1675–1681.

[3] Patnaik, R., Liao, J. C., Engineering of *Escherichia coli* central metabolism for aromatic metabolite production with near theoretical yield. *Appl. Environ. Microbiol.* 1994, *60*, 3903–3908.

[4] Martin, V. J., Pitera, D. J., Withers, S. T., Newman, J. D., Keasling, J. D., Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* 2003, *21*, 796–802.

[5] Ajikumar, P. K., Xiao, W.-H., Tyo, K. E. J., Wang, Y. et al., Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli*. *Science* 2010, *330*, 70–74.

[6] Sauer, U., Hatzimanikatis, V., Bailey, J. E., Hochuli, M. et al., Metabolic fluxes in riboflavin-producing *Bacillus subtilis*. *Nat Biotech* 1997, *15*, 448–452.

[7] Steen, E. J., Kang, Y., Bokinsky, G., Hu, Z. et al., Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* 2010, *463*, 559–562.

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

[9] Brim, H., Osborne, J. P., Kostandarithes, H. M., Fredrickson, J. K. et al., *Deinococcus radiodurans* engineered for complete toluene degradation facilitates Cr(VI) reduction. *Microbiology* 2006, *152*, 2469–2477.

[10] Heitzer, A., Malachowsky, K., Thonnard, J. E., Bienkowski, P. R. et al., Optical biosensor for environmental on-line monitoring of naphthalene and salicylate bioavailability with an immobilized bioluminescent catabolic reporter bacterium. *Appl. Environ. Microbiol.* 1994, *60*, 1487–1494.

[11] Varma, A., Palsson, B., Metabolic flux balancing: Basic concepts, scientific and practical use. *Nat. Biotech.* 1994, *12*, 994–998.

[12] Schuster, S., Fell, D. A., Dandekar, T., A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.* 2000, *18*, 326–332.

[13] Wang, L., Hatzimanikatis, V., Metabolic engineering under uncertainty – II: Analysis of yeast metabolism. *Metab. Eng.* 2006, *8*, 142–159.

[14] Wang, L., Birol, I., Hatzimanikatis, V., Metabolic control analysis under uncertainty: Framework development and case studies. *Biophys. J.* 2004, *87*, 3750–3763.

[15] Trinh, C. T., Wlaschin, A., Srienc, F., Elementary mode analysis: A useful metabolic pathway analysis tool for characterizing cellular metabolism. *Appl. Microbiol. Biotechnol.* 2009, *81*, 813–826.

[16] Lewis, N. E., Nagarajan, H., Palsson, B. O., Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. *Nat Rev Micro* 2012, *10*, 291–305.

[17] Segre, D., Vitkup, D., Church, G. M., Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. USA* 2002, *99*, 15112–15117.

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

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

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

[21] Tepper, N., Shlomi, T., Predicting metabolic engineering knockout strategies for chemical production: Accounting for competing pathways. *Bioinformatics* 2010, *26*, 536–543.

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

[23] Yang, L., Cluett, W. R., Mahadevan, R., EMILiO: A fast algorithm for genome-scale strain design. *Metab. Eng.* 2011, *13*, 272–281.

[24] Kim, J., Reed, J., OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Syst. Biol.* 2010, *4*, 53.

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

[26] Fong, S. S., Burgard, A. P., Herring, C. D., Knight, E. M. et al., In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol. Bioeng.* 2005, *91*, 643–648.

[27] Asadollahi, M. A., Maury, J., Patil, K. R., Schalk, M. et al., Enhancing sesquiterpene production in *Saccharomyces cerevisiae* through in silico driven metabolic engineering. *Metab. Eng.* 2009, *11*, 328–334.

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

[29] Reed, J. L., Palsson, B. Ø., Genome-scale in silico models of *Escherichia coli* have multiple equivalent phenotypic states: Assessment of correlated reaction subsets that comprise network states. *Genome Res.* 2004, *14*, 1797–1805.

[30] Mahadevan, R., Schilling, C. H., The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* 2003, *5*, 264–276.

[31] Lee, S., Phalakornkule, C., Domach, M. M., Grossmann, I. E., Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Comput. Chem. Eng.* 2000, *24*, 711–716.

[32] Trinh, C. T., Unrean, P., Srienc, F., Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. *Appl. Environ. Microbiol.* 2008, *74*, 3634–3643.

[33] Hädicke, O., Klamt, S., Computing complex metabolic intervention strategies using constrained minimal cut sets. *Metab. Eng.* 2011, *13*, 204–213.

[34] Melzer, G., Esfandabadi, M., Franco-Lara, E., Wittmann, C., Flux design: In silico design of cell factories based on correlation of pathway fluxes to desired properties. *BMC Syst. Biol.* 2009, *3*, 120.

[35] Becker, J., Zelder, O., Häfner, S., Schröder, H., Wittmann, C., From zero to hero – Design-based systems metabolic engineering of *Corynebacterium glutamicum* for L-lysine production. *Metab. Eng.* 2011, *13*, 159–168.

[36] Hädicke, O., Klamt, S., CASOP: A computational approach for strain Optimization aiming at high productivity. *J. Biotechnol.* 2010, *147*, 88–101.

[37] Trinh, C. T., Carlson, R., Wlaschin, A., Srienc, F., Design, construction and performance of the most efficient biomass producing *E. coli* bacterium. *Metab. Eng.* 2006, *8*, 628–638.

[38] Trinh, C. T., Srienc, F., Metabolic engineering of *Escherichia coli* for efficient conversion of glycerol to ethanol. *Appl. Environ. Microbiol.* 2009, *75*, 6696–6705.

[39] Trinh, C. T., Li, J., Blanch, H. W., Clark, D. S., Redesigning *Escherichia coli* metabolism for anaerobic production of isobutanol. *Appl. Environ. Microbiol.* 2011, *77*, 4894–4904.

[40] Unrean, P., Srienc, F., Metabolic networks evolve towards states of maximum entropy production. *Metab. Eng.* 2011, *13*, 666–673.

[41] Trinh, C. T., Elucidating and reprogramming *Escherichia coli* metabolisms for obligate anaerobic n-butanol and isobutanol production. *Appl. Microbiol. Biotechnol.* 2012, *95*, 1083–1094.

[42] Unrean, P., Srienc, F., Predicting the adaptive evolution of *Thermoanaerobacterium saccharolyticum*. *J. Biotechnol.* 2012, *158*, 259–266.
[43] Unrean, P., Trinh, C. T., Srienc, F., Rational design and construction of an efficient *Escherichia coli* for production of diapolycopendioic acid. *Metab. Eng.* 2010, *12*, 112–122.
[44] Klamt, S., Stelling, J., Combinatorial complexity of pathway analysis in metabolic networks. *Mol. Biol. Rep.* 2002, *29*, 233–236.
[45] Fong, S. S., Palsson, B. O., Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. *Nat. Genet.* 2004, *36*, 1056–1058.
[46] Tran, L. M., Rizk, M. L., Liao, J. C., Ensemble modeling of metabolic networks. *Biophys. J.* 2008, *95*, 5606–5617.
[47] Miskovic, L., Hatzimanikatis, V., Production of biofuels and biochemicals: In need of an ORACLE. *Trends Biotechnol.* 2010, *28*, 391–397.
[48] Steuer, R., Gross, T., Selbig, J., Blasius, B., Structural kinetic modeling of metabolic networks. *Proc. Natl. Acad. Sci. USA* 2006, *103*, 11868–11873.
[49] Khazaei, T., McGuigan, A., Mahadevan, R., Ensemble modeling of cancer metabolism. *Front. Physiol.* 2012, *3,* 135.
[50] Grimbs, S., Selbig, J., Bulik, S., Holzhütter, H.-G., Steuer, R., The stability and robustness of metabolic states: Identifying stabilizing sites in metabolic networks. *Mol. Syst. Biol.* 2007, *3,* 146.
[51] Schellenberger, J., Palsson, B. Ø., Use of randomized sampling for analysis of metabolic networks. *J. Biol. Chem.* 2009, *284*, 5457–5461.
[52] Alves, R., Savageau, M. A., Systemic properties of ensembles of metabolic networks: application of graphical and statistical methods to simple unbranched pathways. *Bioinformatics* 2000, *16*, 534–547.
[53] Tan, Y., Lafontaine Rivera, J. G., Contador, C. A., Asenjo, J. A., Liao, J. C., Reducing the allowable kinetic space by constructing ensemble of dynamic models with the same steady-state flux. *Metab. Eng.* 2011, *13*, 60–75.
[54] Rizk, M. L., Liao, J. C., Ensemble modeling for aromatic production in *Escherichia coli*. *PLoS ONE* 2009, *4*, e6903.
[55] Contador, C. A., Rizk, M. L., Asenjo, J. A., Liao, J. C., Ensemble modeling for strain development of L-lysine-producing *Escherichia coli*. *Metab. Eng.* 2009, *11*, 221–233.
[56] Bongaerts, J., Krämer, M., Müller, U., Raeven, L., Wubbolts, M., Metabolic engineering for microbial production of aromatic amino acids and derived compounds. *Metab. Eng.* 2001, *3*, 289–300.
[57] Patnaik, R., Spitzer, R. G., Liao, J. C., Pathway engineering for production of aromatics in *Escherichia coli*: Confirmation of stoichiometric analysis by independent modulation of AroG, TktA, and Pps activities. *Biotechnol. Bioeng.* 1995, *46*, 361–370.
[58] Báez, J. L., Bolívar, F., Gosset, G., Determination of 3-deoxy-D-arabino-heptulosonate 7-phosphate productivity and yield from glucose in *Escherichia coli* devoid of the glucose phosphotransferase transport system. *Biotechnol. Bioeng.* 2001, *73*, 530–535.
[59] Schuster, S., Hilgetag, C., Woods, J. H., Fell, D. A., Elementary modes of functioning in biochemical networks, in: Cuthbertson, R., Holcombe, M., Paton, R. (Eds.), *Computation in cellular and molecular biological systems*, World Scientific, Singapore 1994, pp. 151–165.
[60] von Kamp, A., Schuster, S., Metatool 5.0: Fast and flexible elementary modes analysis. *Bioinformatics* 2006, *22*, 1930–1931.
[61] Klamt, S., Gagneur, J., von Kamp, A., Algorithmic approaches for computing elementary modes in large biochemical reaction networks. *Syst. Biol.* 2005, *152*, 249–255.
[62] Gagneur, J., Klamt, S., Computation of elementary modes: A unifying framework and the new binary approach. *BMC Bioinformatics*. 2004, *5*, 175.
[63] Terzer, M., Stelling, J., Large scale computation of elementary flux modes with bit pattern trees. *Bioinformatics* 2008, *24*, 2229–2235.
[64] Schuster, S., Hilgetag, C., Woods, J. H., Fell, D. A., Reaction routes in biochemical reaction systems: Algebraic properties, validated calculation procedure and example from nucleotide metabolism. *J. Math. Biol.* 2002, *45*, 153–181.
[65] Wagner, C., Nullspace approach to determine the elementary modes of chemical reaction systems. *J. Phys. Chem. B* 2004, *108*, 2425–2431.
[66] Jevremovic, D., Trinh, C. T., Srienc, F., Sosa, C. P., Boley, D., Parallelization of nullspace algorithm for the computation of metabolic pathways. *Parallel Comput.* 2011, *37*, 261–278.
[67] Klamt, S., von Kamp, A., An application programming interface for CellNetAnalyzer. *Biosystems* 2011, *105*, 162–168.
[68] Fell, D. A., Increasing the flux in metabolic pathways: A metabolic control analysis perspective. *Biotechnol. Bioeng.* 1998, *58*, 121–124.
[69] Fell, D. A., Metabolic control analysis: A survey of its theoretical and experimental development. *Biochem. J.* 1992, *286*, 313–330.
[70] Cattell, R. B., The scree test for the number of factors. *Multivariate Behav. Res.* 1966, *1*, 245–276.
[71] Golub, G. H., Van Loan, C. F., *Matrix Computations*, Johns Hopkins University Press, Baltimore, USA 1996.
[72] Gosset, G., Yong-Xiao, J., Berry, A., A direct comparison of approaches for increasing carbon flow to aromatic biosynthesis in *Escherichia coli*. *J. Indust. Microbiol.* 1996, *17*, 47–52.
[73] Miller, J. E., Backman, K. C., Oconnor, M. J., Hatch, R. T., Production of phenylalanine and organic-acids by phosphoenolpyruvate carboxylase-deficient mutants of *Escherichia coli*. *J. Indust. Microbiol.* 1987, *2*, 143–149.
[74] Draths, K., Pompliano, D., Conley, D., Frost, J. et al., Biocatalytic synthesis of aromatics from D-glucose: the role of transketolase. *J. Am. Chem. Soc.* 1992, *114*, 3956–3962.
[75] Lu, J. l., Liao, J. C., Metabolic engineering and control analysis for production of aromatics: Role of transaldolase. *Biotechnol. Bioeng.* 2000, *53*, 132–138.
[76] Leonard, E., Ajikumar, P. K., Thayer, K., Xiao, W.-H. et al., Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. *Proc. Natl. Acad. Sci. USA* 2010, *107*, 13654–13659.
[77] Smolke, C. D., Keasling, J. D., Effect of copy number and mRNA processing and stabilization on transcript and protein levels from an engineered dual-gene operon. *Biotechnol. Bioeng.* 2002, *78*, 412–424.
[78] Alper, H., Fischer, C., Nevoigt, E., Stephanopoulos, G., Tuning genetic control through promoter engineering. *Proc. Natl. Acad. Sci. USA* 2005, *102*, 12678–12683.
[79] Gustafsson, C., Govindarajan, S., Minshull, J., Codon bias and heterologous protein expression. *Trends Biotechnol* 2004, *22*, 346–353.
[80] Pfleger, B. F., Pitera, D. J., Smolke, C. D., Keasling, J. D., Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat. Biotech.* 2006, *24*, 1027–1032.
[81] Wiechert, W., 13C Metabolic flux analysis. *Metab. Eng.* 2001, *3*, 195–206.