Quantitative Biology 2014, 2(3): 100–109
DOI 10.1007/s40484-014-0033-7

RESEARCH ARTICLE

OP-Synthetic: identification of optimal genetic manipulations for the overproduction of native and non-native metabolites

Honglei Liu, Yanda Li and Xiaowo Wang*

MOE Key Laboratory of Bioinformatics, Bioinformatics Division and Center for Synthetic and Systems Biology, TNLIST/ Department of Automation, Tsinghua University, Beijing100084, China
* Correspondence: xwwang@tsinghua.edu.cn

Received September 12, 2014; Revised October 13, 2014; Accepted October 14, 2014

Constraint-based flux analysis has been widely used in metabolic engineering to predict genetic optimization strategies. These methods seek to find genetic manipulations that maximally couple the desired metabolites with the cellular growth objective. However, such framework does not work well for overproducing chemicals that are not closely correlated with biomass, for example non-native biochemical production by introducing synthetic pathways into heterologous host cells. Here, we present a computational method called OP-Synthetic, which can identify effective manipulations (upregulation, downregulation and deletion of reactions) and produce a step-by-step optimization strategy for the overproduction of indigenous and non-native chemicals. We compared OP-Synthetic with several state-of-the-art computational approaches on the problems of succinate overproduction and *N*-acetylneuraminic acid synthetic pathway optimization in *Escherichia coli*. OP-Synthetic showed its advantage for efficiently handling multiple steps optimization problems on genome wide metabolic networks. And more importantly, the optimization strategies predicted by OP-Synthetic have a better match with existing engineered strains, especially for the engineering of synthetic metabolic pathways for non-native chemical production. OP-Synthetic is freely available at:http://bioinfo.au.tsinghua.edu.cn/member/xwwang/OPSynthetic/.

Keywords: metabolic network; flux analysis; optimization

INTRODUCTION

Metabolic engineering aims to promote microorganisms into cell factories to produce valuable chemicals. However, wild type microorganisms often produce the desired chemicals with low efficiency [1]. Furthermore, many such chemicals cannot be produced through native metabolic pathways in host organisms widely used by industry like *Escherichia coli* and yeast [2]. Over the past few decades, many researchers attempted to apply genetic engineering to overproduce high-value chemicals by tuning the metabolic pathways [3–5], and introducing synthetic metabolic pathways into the heterologous host strains to produce non-native chemicals [1,6,7].

In many cases, changes to local metabolic pathways may significantly affect the flux distribution on the whole metabolic network. Due to the explosion of omics data, computational approaches have emerged to provide systematic and rational design strategies for the overproduction of biochemicals, which greatly saves time and labor during metabolic engineering [8,9]. In the last decade, computational methods have been widely used to guide metabolic engineering experiments, like the overproduction of succinate [10,11], L-valine [4], L-threonine [12], lycopene [3,13] in *E. coli*, and overproduction of ethanol in *Saccharomyces cerevisiae* [14], etc.

Most current computational methods are developed based on the steady-state metabolic flux model with stoichiometric constraints [15]. Flux balance analysis (FBA) [16], flux variability analysis (FVA) [17] and minimization of metabolic adjustment (MOMA) [18] are widely used constraint-based analyzing approaches. FBA uses the stoichiometry constraint to define a bounded solution space, and then uses biomass or other objective

functions to determine the optimal flux distribution on the network [15,19]. FVA describes the tolerable range of flux variability in near optimal and sub-optimal states. MOMA minimizes the metabolic adjustment to achieve the flux distribution in mutant strain. Based on these approaches, a number of methods have been proposed to identify the gene modifications that enhance the production of the desired metabolites [20,21]. For example, OptKnock [22], OptReg [23], OptStrain [24], and Redirector [25] use a bilevel optimization strategy to search for the genetic manipulations that best couple biomass and the desired compounds. FSEOF [26] and FVSEOF [27] identify targets by scanning changes in flux variability in response to a pre-specified objective yield. OptForce [28] uses flux variability to find minimum sets of manipulations to let the target metabolites meet a pre-specified yield. Different from OptForce, CosMos [29] can find continuous interventions and identify more strategies that guarantee objective production. To overcome the limitation of insufficient computational time and resources as the size of the metabolic network and the kinds of manipulations increase, GDLS [30] uses the local search method with multiple search paths instead of global search to find gene knockouts. GDBB [31] is another method aiming to shorten the computational time through a truncated branch algorithm. These methods were shown to be efficient when they are applied to the overproduction of succinate and acetate, that have close connection with biomass [22,31]. However, in some cases, especially when dealing with some non-native biochemical overproduction problems, desired chemical production is not directly coupled to biomass production. Therefore, no matter how to choose the manipulations, objective metabolite cannot be optimized as the byproduct of biomass.

In this work, we introduced a new optimization procedure named OP-Synthetic that searches for possible genetic manipulations (reaction upregulation, downregulation, and deletion) to overproduce the desired metabolite using a stepwise searching method. OP-Synthetic computes the capacity of the desired metabolite constrained by the suboptimal reaction range obtained by FVA. Hence, the desired metabolite is not directly coupled with the optimization of biomass, but connects with the organism's tolerable flux variability range. OP-Synthetic generates a step-by-step optimization procedure and users can choose the desired number of manipulations to balance the cost and efficiency. We compared OP-Synthetic with state-of-the-art publicly freely available programs, including OptKnock, GDLS, GDBB and OptForce, on the optimization problems of the succinate overproduction and *N*-acetylneuraminic acid synthesis pathway in *E. coli*. OP-Synthetic showed its advantage in low time complexity on large scale metabolic network optimization problems. More importantly, it could predict more effective genetic manipulations supported by the literature, especially on non-growth-coupled metabolic pathway optimization problems.

METHODS

FBA, FVA and metabolic model reduction

FBA modeling: An $m \times n$ stoichiometric matrix S is established from m metabolites and n reactions in the whole metabolic network. The ij-th element in S is the stoichiometric coefficient of metabolite i in reaction j. The flux distribution vector v with length n contains the flux value of each reaction. Each element v_i has its lower bound $v_{\min,i}$ and upper bound $v_{\max,i}$. Under the stoichiometric constraint, the metabolic network is assumed to reach a steady-state. With the objective to maximize the biomass function v_{biomass}, an FBA model is built:

$$\begin{aligned} \max \quad & v_{\text{biomass}} \\ \text{s.t.} \quad & S \cdot v = 0, \\ & v_{\min,i} \leqslant v_i \leqslant v_{\max,i} \text{ for } i = 1,\ 2,...,\ n. \end{aligned}$$

After solving the optimization problem, the flux distribution v of the wild-type strain, the maximum value of biomass Z_{biomass} and the initial yield of desired metabolite v_{obj} could be obtained [15,16].

Estimating flux variability constraints: FVA is an approach to estimate the flux variability according to the tolerance of the metabolic network. The flux distribution is suboptimal when the constraint ($v_{\text{biomass}} \geqslant x \cdot Z_{\text{biomass}}$) is added, where $0 < x < 1$, which represents the proportion of biomass production rate of the engineered strain relative to the wild type strain. FVA can get the flux variability range $[v^x_{\min,i}, v^x_{\max,i}]$ by maximizing and minimizing every reaction subject to stoichiometric constraint.

$$\begin{aligned} \max/\min \quad & v_i \\ \text{s.t.} \quad & S \cdot v = 0, \\ & v_{\text{biomass}} \geqslant x \cdot Z_{\text{biomass}}, \\ & v_{\min,i} \leqslant v_i \leqslant v_{\max,i} \text{ for } i = 1,\ 2,\ ...,\ n. \end{aligned}$$

In OP-Synthetic, we use FVA to estimate two kinds of ranges to represent different states of metabolic network. First, we use parameter *IN_wt* to represent the minimum permitted percentage of wild type strain biomass production rate which defines the sub-optimal solution space of the wild type strains. We use *OUT_ms* to represent the

minimum permitted percentage of biomass production rate in theoretical mutant strains relative to the wild type. Generally, $IN\_wt \geqslant OUT\_ms$, since after genetic manipulations, the engineered strains with higher biochemical yield may have a less biomass production rate. Then we let the range $I_{\text{bound}} = [v_{\min,i}^{IN\_wt}, v_{\max,i}^{IN\_wt}]$ represent the flux variability of the wild-type organism, and I_{bound} is derived by solving the FVA models when $x = IN\_wt$. Range $O_{\text{bound}} = [v_{\min,i}^{OUT\_ms}, v_{\max,i}^{OUT\_ms}]$ represents the flux variability of the theoretical mutant strain, and O_{bound} is derived when $x = OUT\_ms$.

Metabolic network model reduction: The reactions are defined as linked reactions when they are in the same linear metabolic pathway without branches [30]. Under the flux balance assumption, the ratio of these reaction fluxes is a fixed constant, thereby they can be considered as one super node on the metabolic network.

Based on FBA, the mathematical description for the change of a metabolite me_i which is in two reactions is,

$$\frac{\mathrm{d}me_i}{\mathrm{d}t} = S_{ij}v_j + S_{ik}v_k,$$

At the steady state,

$$\frac{S_{ij}}{S_{ik}} = -\frac{v_k}{v_j}.$$

Similar to Lun, D.S., et al [30], we combined reactions in each linear pathway into one reaction to get the reduced metabolic model. For the *E. coli* iAF1260 metabolic model [32], which had 2,382 reactions, we reduced the metabolic model to 1,458 reactions. This reduction greatly improved the computational efficiency. If one linear pathway needs to be decreased (or deleted), biologists can just decrease (or delete) the upstream reaction [33,34]. In the contrary, all reactions need to be increased to make a linear pathway produce a higher yield.

OP-Synthetic procedure

OP-Synthetic uses GNU linear programming kit (GLPK) solver to solve the optimization problems. Based on the flux range I_{bound} and O_{bound} estimated by FVA with user defined sub-optimal biomass tolerance parameters *IN_wt* and *OUT_ms*, OP-Synthetic predicts the maximal production rate of objective metabolite using stepwise searching method (Figure 1). In each step, the algorithm increases production rate from the initial value v_{obj} by an increment s, and computes the flux value v^* for each reaction using FBA. Here, s is determined by initial value v_{obj}. In our applications, we chose $s \in (0.02 \times v_{\text{obj}},\ 0.1 \times v_{\text{obj}})$. If the initial value v_{obj} equals to zero, which often occurs in non-native pathway optimization problems, we chose $s \in (0.02 \times v_i,\ 0.1 \times v_i)$, according to nonzero flux reaction i that has the shortest distance on the network from the target product. After every increase, the algorithm computes reaction flux by FBA with a fixed v_{obj}, then examines if there is any reaction v_i^* exceeds its own range $I_{\text{bound},i}$. If so, the range of reaction i which exceeds $I_{\text{bound},i}$ will be extended to $O_{\text{bound},i}$. Then manipulations of these reactions and v_{obj} are recorded, and the step number n is increased by one. Based on the initial flux distribution v, the new flux distribution v^*, and the flux variability range I_{bound}, the manipulations will be classified into three types as downregulation, upregulation and deletion (Figure 2). The desired metabolite production rate will keep rising until any reaction exceeds the flux range O_{bound}. At this time, further increasing v_{obj} will cause biomass production rate be lower than the minimum permitted value. Thus the desired metabolite now reaches the maximum yield under the constraints.

After determining the reactions that need to be manipulated, the last step is deciding the number of manipulations for reduced linear pathways. We assume that if one set of linear reactions needs to be repressed or deleted, the number of manipulations is only one. However, if the linear reactions need to be increased, all the reactions have to be manipulated.

RESULTS

The overproduction of desired metabolites usually has conflict with cell growth. In the meanwhile to overproduce the desired metabolites, we should also keep the strain to have a reasonable biomass production rate. Different from other approaches that maximally couple the desired metabolite with biomass, OP-Synthetic computes flux variability ranges of the metabolic network, then computes the maximum tolerable yield of the desired metabolite. Thus OP-Synthetic could be applied on both direct and indirect growth-coupled designs.

We applied OP-Synthetic on the optimization problems of succinate overproduction and *N*-acetylneuraminic acid production with synthetic metabolic pathway in *E. coli*. We also compared OP-Synthetic with four freely available state-of-the-art methods OptKnock, GDLS, GDBB and OptForce on both core and genome-scale *E. coli* metabolic model to evaluate its efficiency.

Parameter setting for OP-Synthetic

Parameters *IN_wt* and *OUT_ms* determine the number of manipulations that would be reported by OP-Synthetic. In principle, a lower *OUT_ms* bound makes the program to search for a higher production rate for the desired metabolite but with the cost of lower biomass production rate. A higher *IN_wt* value means the wild type organism

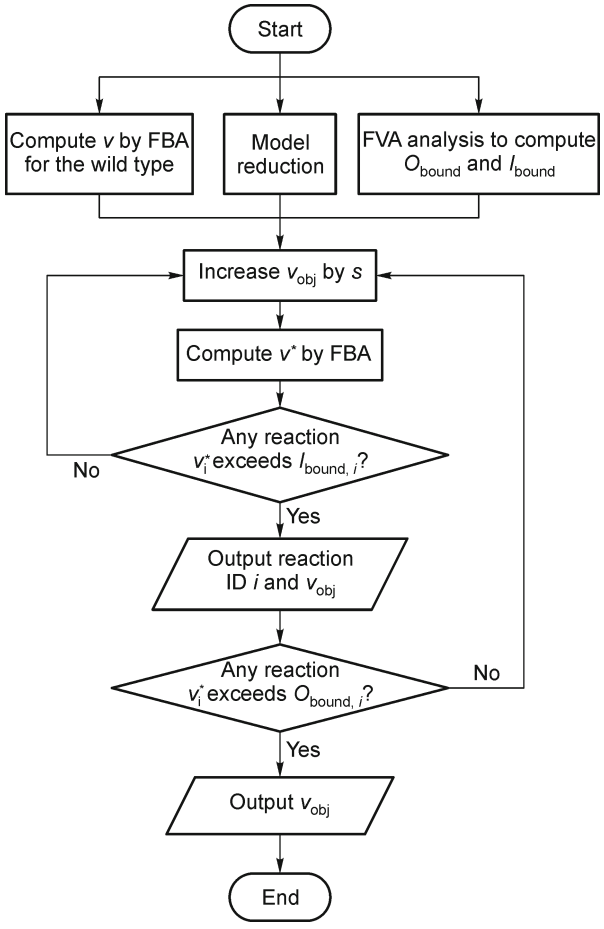

Figure 1. OP-Synthetic flowchart. *i*: reaction ID; v: the wild type flux distribution; v_i^*: the new value of reaction *i*; $I_{bound,i}$: flux variability range of reaction *i* in wild type; $O_{bound,i}$: flux variability range of reaction *i* in theoretical mutant strain; v_{obj}: objective production rate; *s*: incremental amount in stepwise searching.

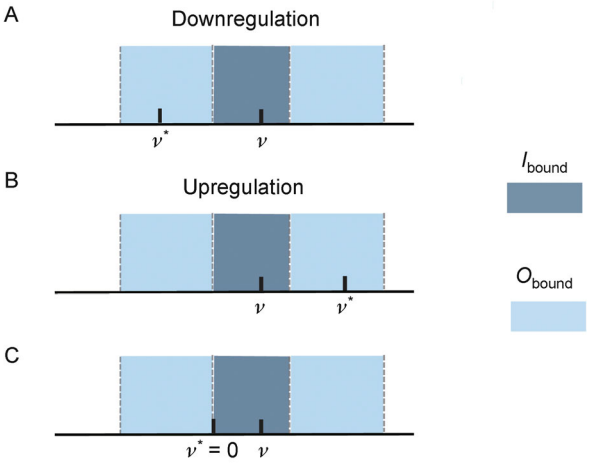

Figure 2. Manipulation type. (A) Downregulation; (B) Upregulation; (C) Reaction deletion. v represents the initial flux value obtained from FBA analysis for the wild type. v^* represents the new flux value after increasing the objective metabolite yield. The reaction calls for a decrease when the new flux value v^* is lower than the wild type flux range. On the contrary, the reaction needs to be upregulated when the new flux value v^* is higher than the wild type flux range. If the initial v is not 0, but v^* equal to 0, we regard this reaction to be a candidate knockout reaction.

could tolerate narrower flux range, and OP-Synthetic may tend to report more reactions that need to be manipulated.

To test the influence of different parameter settings, we first applied OP-Synthetic on a core *E. coli* model (http://gcrg.ucsd.edu/Downloads/EcoliCore) to find potential manipulations that can produce higher succinate production rate. This core *E. coli* model is a subset of the genome-scale metabolic network iAF1260, and contains 72 metabolites and 95 reactions. The carbon source was glucose and its maximum input value was set to 10 mmol gDW^{-1} h^{-1}. Other nutrients such as ammonia were allowed in unlimited quantities. In all cases, OP-Synthetic eliminated the oxygen uptake reaction to get an anaerobic growth condition. Secretion routes for acetate, carbon dioxide, ethanol were enabled. Detailed setting of exchange reactions is in supplementary information (Supplementary Table S1). We chose the increment s to be 0.1 mmol gDW^{-1} h^{-1}, and tested 170 different parameter combinations in total, by varying *IN_wt* from 15% to 100% and *OUT_ms* from 10% to 90% both with 5% increase at each step (Supplementary Figure S1). The constraint $IN\_wt > OUT\_ms$ was required. Although different (*IN_wt*, *OUT_ms*) combinations may have different estimated absolute production value, the final optimization strategies are very consistent at most parameter settings (Supplementary Table S2).

We also tested the influence of parameter setting on the genome-scale network iAF1260 by changing *IN_wt* from 15% to 100% and *OUT_mt* from 10% to 90%, both with 5% increase. iAF1260 has 1,668 metabolites and 2,382 reactions [32]. The predicted optimization strategies also showed high consistency (Supplementary Table S3). Fifty-three out of the 170 parameter settings have the same optimization strategies compared with (*IN_wt*, *OUT_ms*) = (90%, 40%), while most of the other parameter settings can get the results within 2 different manipulations. According to the adaptive evolution experiments [35,36], we chose the default minimum permitted percentage of wild type biomass to be 90%. To balance the biologic feasibility, we chose the minimum permitted biomass production rate of mutant strain to be no less than 40% of wild type. Further lower down the

OUT_ms value may get higher theoretical production rate, but with the cost of low cell fitness. In the following analysis, we used (*IN_wt*, *OUT_ms*) = (90%, 40%) as the default.

Comparison with OptKnock, GDLS, GDBB and OptForce on succinate overproduction

We compared OP-Synthetic with four publicly available programs OptKnock, GDLS, GDBB and OptForce for the overproduction of succinate in core *E. coli* metabolic model and genome-scale model iAF1260 under the same parameter settings.The addition of heterologous pyruvate carboxylase gene *pyc* from *Lactococcus lactis* and *Rhizobium etli* is also concerned, since it can effectively further improve the production of succinate [10,34,37,38].

When applying OP-Synthetic on iAF1260 metabolic network, five steps with five manipulations are reported (Table 1, Supplementary Figure S2). The maximum yield of succinate after five steps was 10.85 mmol gDW^{-1} h^{-1}. With the addition of gene *pyc*, OP-Synthetic gave an optimization strategy including five steps and nine manipulations. The final succinate yield reached 15.15 mmol gDW^{-1} h^{-1} (Table 1, Supplementary Figure S2).

As shown in Table 1, among the five programs, OP-Synthetic reported the maximum number of manipulations that could be supported by the existing succinate overproduction strains [10,33,37,38]. Especially for iAF1260 model with the addition of gene *pyc*, 7 out of 9 manipulations could be verified in engineered strains, while the numbers of verified manipulations are 0, 1, 1, 1 by OptKnock, GDLS, GDBB and OptForce, respectively. We also compared the stable optimization procedures predicted by the five methods in a step-by-step manner by varying the number of manipulations from 1 to 9 in iAF1260 metabolic model with the addition of gene *pyc*. OP-Synthetic can get the most manipulations that match with literature in all steps (Figure 3A).

We also compared the computation time of OP-Synthetic with the other methods. In general, the running time grows with the size of network and the number of manipulations. Of all the five methods, OptForce showed the highest time complexity, and GDBB was the fastest (Table 1). The computation time of OP-Synthetic can be mainly divided into two parts, including FVA computation time and stepwise searching time. When metabolic network size increased from core model to genome-scale model, the FVA computation time increased from 6 s to 478 s, while the running time of stepwise searching increased from 24 s to 36 s. In contrast, the running time of OptKnock changed from 21 s to 53,553 s, GDLS increased from 3 s to 49,329 s, and OptForce increased from 45,831 s to 154,583 s. Furthermore, when the number of manipulations grew, OptKnock, GDLS showed large increase in running time (Figure 3B). In all cases, OP-Synthetic showed tolerable running time (less than 10 min), and the computation time does not change too much with the increase of manipulation number. Therefore, OP-Synthetic can be easily applied on the multi-step optimization problems on large networks. We ran OptKnock and GDLS programs in COBRA Toolbox [39] using GLPK solver. Computation was performed on a computer with Intel ® Core™2 Quad CPU Q9400 @2.66 GHz, 8 Gb memory, running Windows 7 operation system.

In conclusion, OP-Synthetic finds more effective manipulations to increase the objective production with tolerable time complexity. The optimization strategy determined by OP-Synthetic matches the existing mutation strains much better than the other methods.

Engineering of an *N*-acetylneuraminicacid synthetic metabolic pathway

N-acetylneuraminic acid (NeuAc) can promote the infant brain development, and also keep the function of elderly brain. It is important for the cell recognition and signal transmission process [40,41]. There are several engineered strains to overproduce NeuAc [42,43]. In a recent work by Kang et al., two synthetic reactions including *N*-acetylglucosamine 2-epimerase (*slr1975*) and glucosamine-6-phosphate acetyl transferase (*GNA1*) were heterologously introduced into *E. coli* from *Synechocystis* sp. PCC6803 and *Saccharomyces cerevisiae* EBY100 respectively. Then they further manipulated the strain to get a yield up to 7.85 $g1^{-1}$ in fed-batch fermentation by gene knockout and mutagenesis [6].

We applied OP-Synthetic to find the optimization procedure for the overproduction of NeuAc on *E. coli* model iAF1260. The model had 2,384 reactions after adding the two synthetic reactions. The carbon source was glucose with a maximum input value of 10 mmol gDW^{-1} h^{-1}. The maximum uptake value of oxygen was set to 10 mmol gDW^{-1} h^{-1}. Other nutrients such as potassium and ammonia were allowed in unlimited quantities. Secretion routes were enabled for acetate, carbon dioxide, ethanol, etc. Increment *s* was 0.1 mmol gDW^{-1} h^{-1}, and parameters (*IN_wt*, *OUT_ms*) was set to (90%, 40%).

OP-Synthetic reported a three step optimization strategy including 14 manipulations in total (Figure 4). The final NeuAc yield is 4.00 mmol gDW^{-1} h^{-1}. Five out of eight effective manipulations in strains produced by Kang et al. were found in OP-Synthetic optimization strategy (Table 2). Four manipulations consistent with

Table 1. Comparison of OptKnock, GDLS, GDBB, OptForce and OP-Synthetic in succinate overproduction.

	Approach	Manipulation	Succinate production (mmol gDW^{-1} h^{-1})	Computation time (s)
Core *E. coli* metabolic model	OptKnock	ACALD2x, CYTBD, FBP, GLUN, GND, LDH_D, SUCDi	10.29	21
	GDLS	ACALD, GLUDy, PFL, PYK, ME2	9.92	3
	GDBB	NA	NA	NA
	OptForce	AKGDH(↑), GAPD(×), FUM(×), PFL(×), ATPM(×)	15*	45831
	OP-Synthetic[a]	(step 1) FUM(↑) (step 2) MDH(↑) (step 3) PPC(↑) (step 4) PYK(×) (step 5) ALCD2x-ETOHt2r-EX_etoh(e)(×), ACALD (×)	10.30	30
E. coli iAF1260 metabolic model	OptKnock	ALCD2x, AP5AH, FE2t3pp, GLCDPP, PRATPP	10.79	53553
	GDLS	CO2tex, LDH_D, PFL, THD2pp, PPKr	9.77	49329
	GDBB	ALCD2x, GLCptspp, GLUDy, Pit2rpp, RPE	9.42	53
	OptForce	MCOATA_f(×), PPA(×), GLYCLTDx(↑), NADH18pp(↑), TRSARr_f(↑)	15*	154583
	OP-Synthetic[a]	(step 1) FRD2(↑) (step 2) ETOHtex-ETOHt2rpp-EX_etoh(e) (×) (step 3) ACALD(×) (step 4) PYK(×) (step 5) PPC(↑)	10.85	514
E. coli iAF1260 metabolic model with addition of *pyc*	OptKnock	AP5AH, DDCAtexi, DTMPK, GLYK, MCITS, MLTG1, NOtpp, R1PK, URAtex	15.10	62483
	GDLS	ACALD, ACCOAL, EAR40x, EAR80x, Htex, ME2, PGI, PPKr, SUCOAS	12.98	9395
	GDBB	ACALD, LALDO2x, ALR4x, OAADC, GART, THRA2i, GND, ME2, THD2pp	10.69	136
	OptForce	3HAD60(×), ACS(×), O2tex_f(×), PFL(×), PGAMT_f(×), GAPD6(↑), MALS(↑), MDH3(↑), NADH18pp(↑)	15*	154592
	OP-Synthetic[a]	(step 1) ACALD(×), ETOHtex-ETOHt2rpp-EX_etoh(e) (×), ALCD2x(×) (step 2) FDR2(↑) (step 3) MDH(↑), PYC(↑) (step 4) EX_for(e)-FORtex (×), FORtppi(×) (step 5) PFL (×)	15.15	522

*The minimum guaranteed flux for succinate by OptForce. [a]The predicted yield and computation time for OP-Synthetic is the final combination of all steps.

In OP-Synthetic results, the meaning of different signals, (↑): upregulation; (↓): downregulation; (×): deletion. In the optimization strategies proposed by OptKnock, GDLS and GDBB, all the manipulations are reaction deletions. The manipulations with grey backgrounds were found in existing mutant strains with high succinate production. (Note that the reaction names are consistent with the *E. coli* iAF1260 metabolic model.)

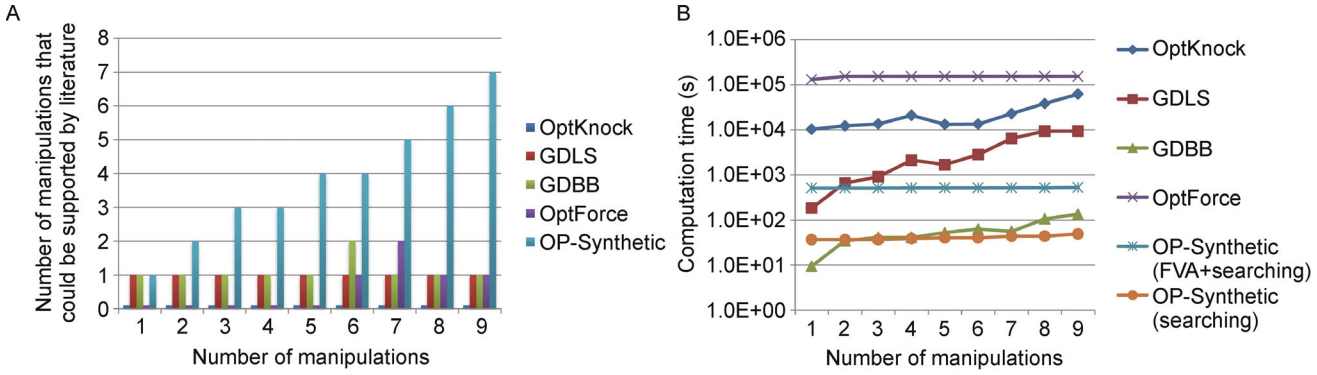

Figure 3. Comparison of the number of reactions that can be verified by literature and running time using OptKnock, GDLS, GDBB, OptForce and OP-Synthetic. *E. coli* iAF1260 genome-scale metabolic network with the addition of gene *pyc* is used. (A) The number of reactions that could be verified by literature is shown by the five methods. (B) The running time is shown under the same number of manipulations by the five methods.

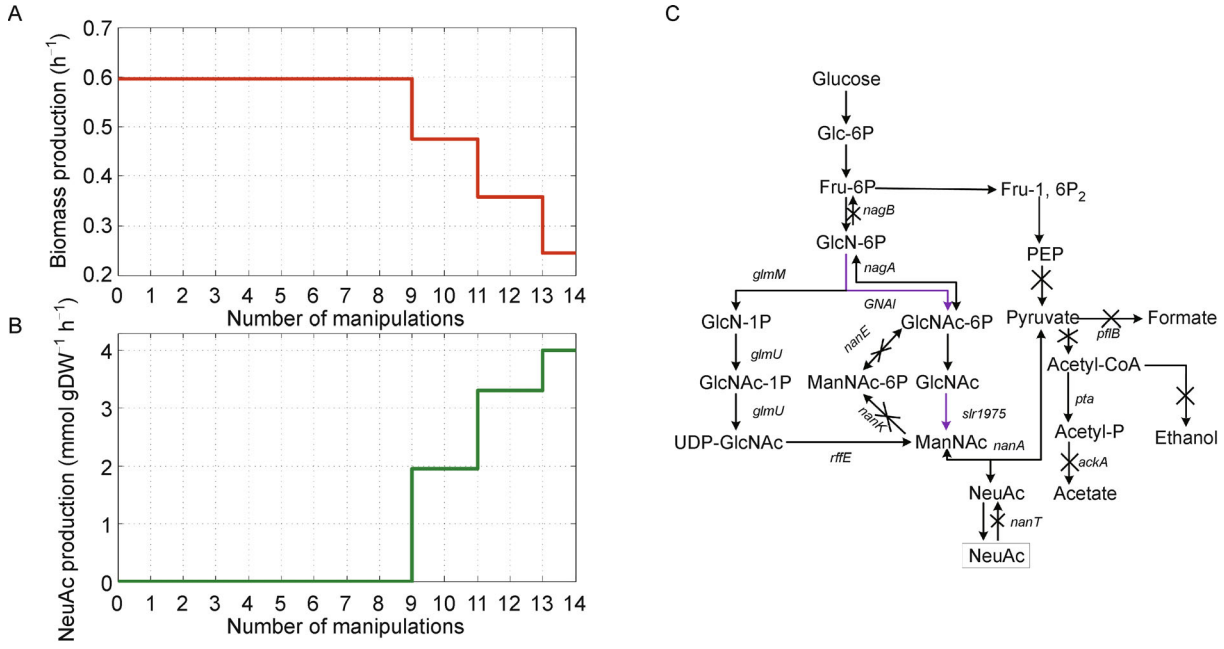

Figure 4. The optimization procedure of *N*-acetylneuraminic acid by OP-Synthetic. (A) Biomass production rate after every step. (B) NeuAc production rate after every step. (C) Part of the *E. coli* map to show the optimization strategy for the overproduction of NeuAc. The signal × represents gene deletion. The purple lines represent the two synthetic reactions *N*-acetylglucosamine 2-epimerase (*slr1975*) and glucosamine-6-phosphate acetyl transferase (*GNA1*) which are introduced from *Synechocystis* sp. PCC6803 and *S. cerevisiae* EBY100, respectively.

literature were predicted in the first step, with NeuAc production rate of 1.95 mmol gDW^{-1} h^{-1} after the first step.

Knockouts of *nanE* and *nanK* reduce the expression of *N*-acetylmannosamine 6-phosphate epimerase and *N*-acetyl-D-mannosamine kinase. As a result, more GlcNAc-6P is available to produce ManNAc, which is the precursor of NeuAc production. The deletion of gene *nagB* forces Fru-6P to be converted into GlcN-6P, which is the precursor of the NeuAc synthetic metabolic pathway. *nanT* is the secondary transporter of NeuAc. The deletion of *nanT* avoids retro-transportation, and prevents consumption of NeuAc in *E. coli* [6]. Deletion of acetate production pathway can reduce the by-product,

Table 2. The manipulations identified by OP-Synthetic for NeuAc yield optimization.

Step	Manipulations
1	G6PDA(*nagB*) (×), AMANAPEr(*nanE*)(×), EDD(×), EX_neu(*nanT*)(×), AMANK(*nanK*)(×), GLUABUTt7pp(×), ETOHtex- ETOHt2rpp- EX_etoh(e)(×), AGDC(×), ABUTt2pp(×)
2	EX_for(e)-FORtex(×), PFL(×), PYK(×)
3	Act2rpp(×), ACtex-EX_ac(e)(×)

Signal × represents gene deletion

thus more pyruvate can be gotten, which is also a precursor of NeuAc production. Though the other manipulations are not presented in literature, they can provide suggestions for further experiments from another perspective. For example, the deletion of formate and ethanol production pathways are suggested to eliminate by-product synthesis, leading to the availability of more pyruvate, and thereby more NeuAc.

We also tried to apply OptKnock, GDLS, GDBB and OptForce to the same problem on *E. coli* metabolic network with the addition of the synthetic metabolic pathways. However, no optimization strategy was found for the overproduction of NeuAc. The basic concept of bilevel optimization methods is to maximize connection between the growth rate and desired metabolite. Nevertheless, in this case, the synthetic metabolic pathway to produce NeuAc has little contribution to biomass function. There are no manipulations that can directly couple the production of NeuAc with biomass. Therefore no effective intervention can be predicted by GDLS, OptKnock, and GDBB. For OptForce, the overproduction of this synthetic metabolic pathway does not cause significant change of flux range defined by the program. That is, all the reaction flux range in wild type and overproduction metabolic network has overlapping ranges. So no optimization strategy was predicted. We further applied iterative knockout with linear MOMA [18] on iAF1260 with two synthetic reactions in this case. We could not find any knockout strategy that can lead to a none zero production of NeuAc. Thus we concluded that OP-Synthetic showed its advantage over other existing publicly available methods on non-growth-coupled metabolite NeuAc overproduction problem.

DISCUSSION

In this work, we introduced an optimization procedure called OP-Synthetic to identify step-by-step metabolic manipulations (reaction upregulation, downregulation and deletion) for overproduction of biochemical compounds. We applied OP-Synthetic on both growth-coupled and non-growth-coupled metabolic pathway optimization problems on core and genome-scale *E. coli* metabolic networks. The results showed that OP-Synthetic optimization procedure better matches the reported experimental results than existing computational methods. Additionally, since OP-Synthetic has low time complexity, it can be applied on large-scale metabolic network optimization problems.

Non-native chemical biosynthesis is an important field of metabolic engineering. When synthetic metabolic pathway is introduced into the host strain, the objective targets are often not growth-coupled. The computational strategies that maximize the connection of biomass and desired metabolite, are not suitable to predict the manipulations under this condition. Different with other optimization methods that not related on growth-coupled design, such as OptForce, FSEOF, etc., OP-Synthetic searches for the maximum tolerable yield in the flux variability range of wild type and theoretical mutant strains. OP-Synthetic can effectively handle the overproduction problem of non-growth-coupled pathways, which makes it especially useful for non-native chemical overproduction problems.

Compared with the other methods, OP-Synthetic does not have a constraint on the maximum number of allowed manipulations. Instead, it reports all the putative manipulations under the constraint of maximum tolerant biomass variance. As OP-Synthetic provides step-by-step manipulation predictions, biologists can choose the appropriate number of steps to balance the growth rate and the desired metabolite production. In addition, multiple strain mutations can be produced simultaneously by multiplex automated genome engineering (MAGE) [44], which may further reduce the time needed to produce the optimally engineered strains.

A potential improvement of OP-Synthetic is integrating other flux analysis methods, such as pFBA [45] and MOMA. Another is that, during optimization OP-Synthetic requires the predicted mutant strains to keep in a biology feasible state (reasonable growth rate in current version). Such framework could also be applied on other metabolic state constraints, such as high by-product yield. We are also trying to implement CRISPRi [46] based approach to evaluate the biologic feasibility of different strategies by experiments. Systematically testing different optimization strategies will provide new insight for designing better optimization prediction methods.

SUPPLEMENTARY MATERIALS

The supplementary materials can be found online with this article at DOI 10.1007/s40484-014-0033-7.

ACKNOWLEDGMENTS

We thank Dr. Monica Sleumer for helpful discussion and checking the language. This work was supported by the NBRPC grant (2012CB316503), NSFC grant (61322310, 31371341), FANEDD grant (201158 of China), and Outstanding Tutors for doctoral dissertations of S&T project in Beijing (No. 20111000304).

COMPLIANCE WITH ETHICS GUIDELINES

The authors Honglei Liu, Yanda Li and XiaowoWang declare that they have no conflict of interests.

This article does not contain any studies with human or animal subjects performed by any of the authors.

REFERENCES

1. Lee, J. W., Na, D., Park, J. M., Lee, J., Choi, S. and Lee, S. Y. (2012) Systems metabolic engineering of microorganisms for natural and non-natural chemicals. Nat. Chem. Biol., 8, 536–546
2. Prather, K. L. J. and Martin, C. H. (2008) De novo biosynthetic pathways: rational design of microbial chemical factories. Curr. Opin. Biotechnol.,19, 468–474
3. Alper, H., Miyaoku, K. and Stephanopoulos, G. (2005) Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets. Nat. Biotechnol., 23, 612–616
4. Park, J. H., Lee, K. H., Kim, T. Y. and 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. USA, 104, 7797–7802
5. Ro, D.-K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S., Kirby, J., et al. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature, 440, 940–943
6. Kang, J., Gu, P., Wang, Y., Li, Y., Yang, F., Wang, Q. and Qi, Q. (2012) Engineering of an *N*-acetylneuraminic acid synthetic pathway in *Escherichia coli*. Metab. Eng., 14, 623–629
7. Steen, E. J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., Del Cardayre, S. B. and Keasling, J. D. (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. Nature, 463, 559–562
8. Lee, J. W., Kim, T. Y., Jang, Y.-S., Choi, S. and Lee, S. Y. (2011) Systems metabolic engineering for chemicals and materials. Trends Biotechnol., 29, 370–378
9. Park, J. M., Kim, T. Y. and Lee, S. Y. (2009) Constraints-based genome-scale metabolic simulation for systems metabolic engineering. Biotechnol. Adv., 27, 979–988
10. Cox, S. J., Shalel Levanon, S., Sanchez, A., Lin, H., Peercy, B., Bennett, G. N. and San, K.-Y. (2006) Development of a metabolic network design and optimization framework incorporating implementation constraints: a succinate production case study. Metab. Eng., 8, 46–57
11. Lin, H., Bennett, G. N. and San, K.-Y. (2005) Metabolic engineering of aerobic succinate production systems in *Escherichia coli* to improve process productivity and achieve the maximum theoretical succinate yield. Metab. Eng., 7, 116–127
12. Lee, K. H., Park, J. H., Kim, T. Y., Kim, H. U. and Lee, S. Y. (2007) Systems metabolic engineering of *Escherichia coli* for *L*-threonine production. Mol. Syst. Biol., 3, 149
13. Alper, H., Jin, Y.-S., Moxley, J. F. and Stephanopoulos, G. (2005) Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. Metab. Eng., 7, 155–164
14. Bro, C., Regenberg, B., Förster, J. and Nielsen, J. (2006) In silico aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production. Metab. Eng., 8, 102–111
15. Kauffman, K. J., Prakash, P. and Edwards, J. S. (2003) Advances in flux balance analysis. Curr. Opin. Biotechnol.,14, 491–496
16. Orth, J. D., Thiele, I. and Palsson, B. Ø. (2010) What is flux balance analysis? Nat. Biotechnol., 28, 245–248
17. Mahadevan, R. and Schilling, C. H. (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab. Eng., 5, 264–276
18. Segrè, D., Vitkup, D. and Church, G. M. (2002) Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. USA, 99, 15112–15117
19. Price, N. D., Reed, J. L. and Palsson, B. O. (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat. Rev. Microbiol., 2, 886–897
20. Kim, J. and Reed, J. L. (2010) OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. BMC Syst. Biol., 4, 53
21. Yang, L., Cluett, W. R. and Mahadevan, R. (2011) EMILiO: a fast algorithm for genome-scale strain design. Metab. Eng., 13, 272–281
22. Burgard, A. P., Pharkya, P. and Maranas, C. D. (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol. Bioeng., 84, 647–657
23. Pharkya, P. and 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
24. Pharkya, P., Burgard, A. P. and Maranas, C. D. (2004) OptStrain: a computational framework for redesign of microbial production systems. Genome Res., 14, 2367–2376
25. Rockwell, G., Guido, N. J., and Church, G. M. (2013) Redirector: designing cell factories by reconstructing the metabolic objective. PLoS Comput. Biol., 9, e1002882
26. Choi, H. S., Lee, S. Y., Kim, T. Y. and Woo, H. M. (2010) In silico identification of gene amplification targets for improvement of lycopene production. Appl. Environ. Microbiol., 76, 3097–3105
27. Park, J. M., Park, H. M., Kim, W. J., Kim, H. U., Kim, T. Y. and Lee, S. Y. (2012) Flux variability scanning based on enforced objective flux for identifying gene amplification targets. BMC Syst. Biol., 6, 106
28. Ranganathan, S., Suthers, P. F. and Maranas, C. D. (2010) OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS Comput. Biol., 6, e1000744
29. Cotten, C. and Reed, J. L. (2013) Constraint-based strain design using continuous modifications (CosMos) of flux bounds finds new strategies for metabolic engineering. Biotechnol. J., 8, 595–604
30. Lun, D. S., Rockwell, G., Guido, N. J., Baym, M., Kelner, J. A., Berger, B., Galagan, J. E. and Church, G. M. (2009) Large-scale identification of genetic design strategies using local search. Mol. Syst. Biol., 5,

296

31. Egen, D. and Lun, D. S. (2012) Truncated branch and bound achieves efficient constraint-based genetic design. Bioinformatics, 28, 1619–1623
32. Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbelt, L. J., Hatzimanikatis, V. and Palsson, B. Ø. (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol. Syst. Biol., 3, 121
33. Jantama, K., Haupt, M. J., Svoronos, S. A., Zhang, X., Moore, J. C., Shanmugam, K. T. and Ingram, L. O. (2008) Combining metabolic engineering and metabolic evolution to develop nonrecombinant strains of *Escherichia coli* C that produce succinate and malate. Biotechnol. Bioeng., 99, 1140–1153
34. Sánchez, A. M., Bennett, G. N. and San, K.-Y. (2006) Batch culture characterization and metabolic flux analysis of succinate-producing Escherichia coli strains. Metab. Eng., 8, 209–226
35. Jensen, P. A. and Papin, J. A. (2011) Functional integration of a metabolic network model and expression data without arbitrary thresholding. Bioinformatics, 27, 541–547
36. Ibarra, R. U., Edwards, J. S. and Palsson, B. O. (2002) *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature, 420, 186–189
37. Sánchez, A. M., Bennett, G. N. and San, K. Y. (2005) Efficient succinic acid production from glucose through overexpression of pyruvate carboxylase in an *Escherichia coli* alcohol dehydrogenase and lactate dehydrogenase mutant. Biotechnol. Prog., 21, 358–365
38. Sánchez, A. M., Bennett, G. N. and San, K.-Y. (2005) Novel pathway engineering design of the anaerobic central metabolic pathway in *Escherichia coli* to increase succinate yield and productivity. Metab. Eng., 7, 229–239
39. Schellenberger, J., Que, R., Fleming, R. M. T., Thiele, I., Orth, J. D., Feist, A. M., Zielinski, D. C., Bordbar, A., Lewis, N. E., Rahmanian, S., et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat. Protoc., 6, 1290–1307
40. Schauer, R. (2000) Achievements and challenges of sialic acid research. Glycoconj. J., 17, 485–499
41. Wang, B. (2009) Sialic acid is an essential nutrient for brain development and cognition. Annu. Rev. Nutr., 29, 177–222
42. Ishikawa, M. and Koizumi, S. (2010) Microbial production of *N*-acetylneuraminic acid by genetically engineered *Escherichia coli*. Carbohydr. Res., 345, 2605–2609
43. Tao, F., Zhang, Y., Ma, C. and Xu, P. (2011) One-pot bio-synthesis: N-acetyl-D-neuraminic acid production by a powerful engineered whole-cell catalyst. Sci. Rep., 1, 142
44. Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R. and Church, G. M. (2009) Programming cells by multiplex genome engineering and accelerated evolution. Nature, 460, 894–898
45. Lewis, N. E., Hixson, K. K., Conrad, T. M., Lerman, J. A., Charusanti, P., Polpitiya, A. D., Adkins, J. N., Schramm, G., Purvine, S. O., Lopez-Ferrer, D., et al. (2010) Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. Mol. Syst. Biol., 6, 390
46. Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P. and Lim, W. A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell, 152, 1173–1183

