BMC
Systems Biology

METHODOLOGY ARTICLE

Open Access

Flux variability scanning based on enforced objective flux for identifying gene amplification targets

Jong Myoung Park[1,2], Hye Min Park[1], Won Jun Kim[1], Hyun Uk Kim[1,2], Tae Yong Kim[1,2] and Sang Yup Lee[1,2,3*]

Abstract

Background: In order to reduce time and efforts to develop microbial strains with better capability of producing desired bioproducts, genome-scale metabolic simulations have proven useful in identifying gene knockout and amplification targets. Constraints-based flux analysis has successfully been employed for such simulation, but is limited in its ability to properly describe the complex nature of biological systems. Gene knockout simulations are relatively straightforward to implement, simply by constraining the flux values of the target reaction to zero, but the identification of reliable gene amplification targets is rather difficult. Here, we report a new algorithm which incorporates physiological data into a model to improve the model's prediction capabilities and to capitalize on the relationships between genes and metabolic fluxes.

Results: We developed an algorithm, flux variability scanning based on enforced objective flux (FVSEOF) with grouping reaction (GR) constraints, in an effort to identify gene amplification targets by considering reactions that co-carry flux values based on physiological omics data *via* "GR constraints". This method scans changes in the variabilities of metabolic fluxes in response to an artificially enforced objective flux of product formation. The gene amplification targets predicted using this method were validated by comparing the predicted effects with the previous experimental results obtained for the production of shikimic acid and putrescine in *Escherichia coli*. Moreover, new gene amplification targets for further enhancing putrescine production were validated through experiments involving the overexpression of each identified targeted gene under condition-controlled batch cultivation.

Conclusions: FVSEOF with GR constraints allows identification of gene amplification targets for metabolic engineering of microbial strains in order to enhance the production of desired bioproducts. The algorithm was validated through the experiments on the enhanced production of putrescine in *E. coli*, in addition to the comparison with the previously reported experimental data. The FVSEOF strategy with GR constraints will be generally useful for developing industrially important microbial strains having enhanced capabilities of producing chemicals of interest.

Keywords: Flux variability scanning based on enforced objective flux, Grouping reaction constraints, Putrescine, *Escherichia coli*

* Correspondence: leesy@kaist.ac.kr
[1]Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical and Biomolecular Engineering (BK21 program), Center for Systems and Synthetic Biotechnology, Institute for the BioCentury, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Republic of Korea
[2]BioInformatics Research Center, KAIST, Daejeon 305-701, Republic of Korea
Full list of author information is available at the end of the article

Background

One of the most ambitious goals in metabolic engineering is the design of biological systems based on *in silico* predictions using mathematical models. The advent of high-throughput technologies and the completion of genome sequencing for many organisms have led to an explosion of systems-wide biological data [1,2]. Genome-scale stoichiometric models of the increasing number of microorganisms and mammalian cells have been developed at the moment [3,4]. Some of such models have been used to identify gene knockout targets for the efficient production of important industrial chemicals, including amino acids [5,6] and chemicals that are conventionally derived from petroleum [7-9]; other such models have been used to identify drug targets in pathogens [10-12]. In modeling and simulation approaches, target reactions whose knockout is predicted to overproduce the chemical of interest can be easily tested experimentally by deleting the corresponding genes in the microbial host.

Increasing the expression levels of the relevant genes has also been successfully employed for the overproduction of target chemicals [13,14]. To avoid unnecessarily massive experiments to be performed, several computational algorithms have been devised in an effort to reveal the relationship between metabolic reactions and the biological properties of interest [15-27]; however, the identification of gene amplification targets is more complicated than the identification of gene knockout targets; hence, correlations among the genes, mRNAs, transcriptional or translational regulations, proteins, and metabolic fluxes must be carefully examined. Genome-scale metabolic models that rely on constraints-based flux analysis without additional physiological information are limited in their ability to describe the complex nature of biological systems, particularly biological phenomena beyond metabolism. Several systematic methods have been developed to overcome such limitations: flux variability analysis (FVA) [17,19-21], flux coupling analysis [16-18], flux sensitivity analysis [15], flux response analysis [26], OptReg [22], genetic design through local search [25], OptForce [27], and flux scanning based on enforced objective flux (FSEOF) [23]. In particular, FSEOF is a method that first scans and searches for variations in the metabolic fluxes in response to the enforced fluxes directed towards a target product. Reactions were then selected as amplification targets, the flux values of which increased in accordance with the enforced fluxes toward the production of a target chemical. This method was experimentally validated by identifying amplification targets that improved the production of lycopene in *Escherichia coli* [23]. These approaches demonstrated that incorporating physiological constraints during the model simulation are critical to identifying trustworthy gene amplification targets, but much improvement is still needed [24,28]. One of the major problems is the existence of a too large flux solution space in optimization problems.
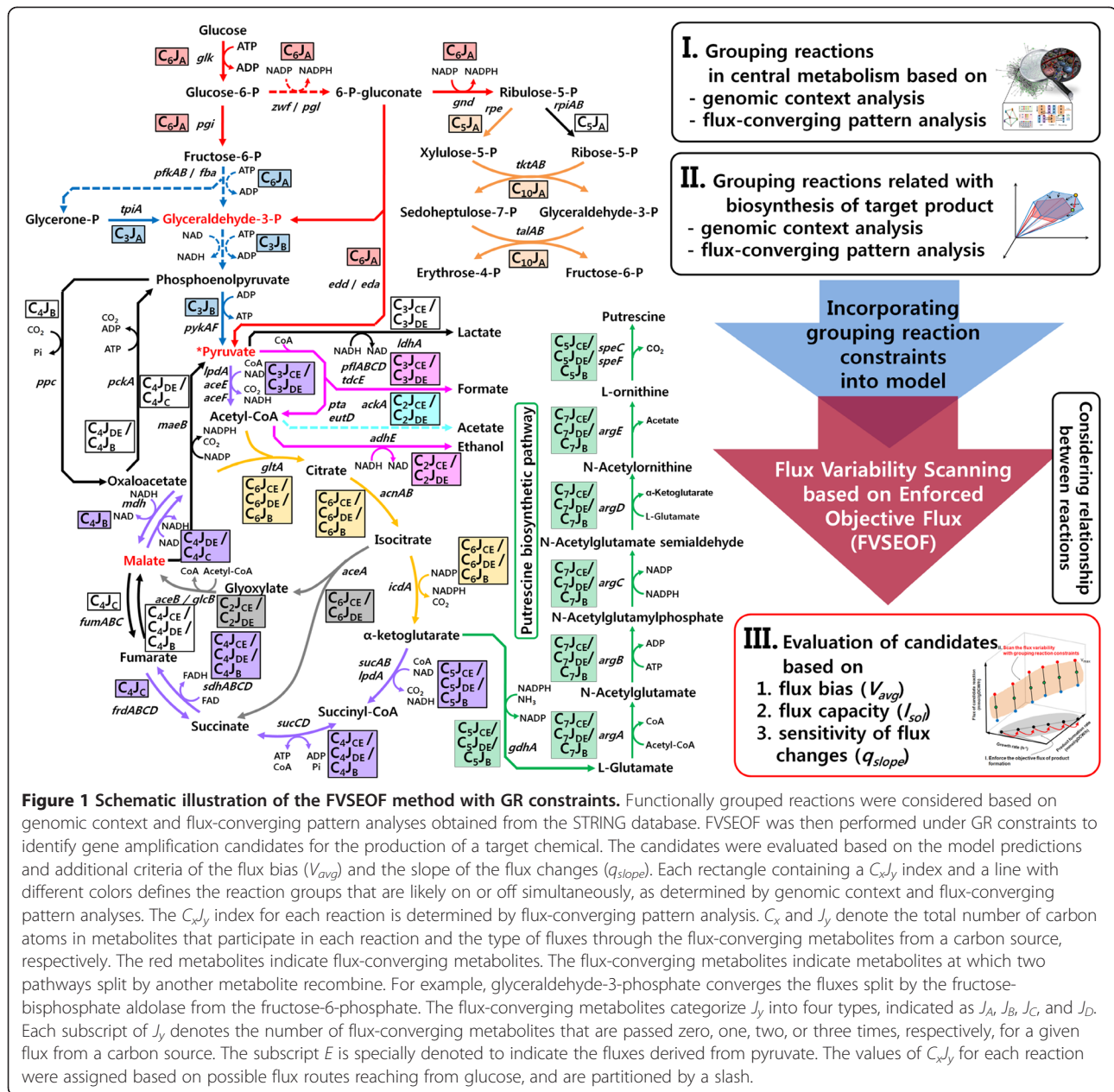
In this study, in order to systematically handle the large flux solution spaces, as also revealed in the implementation of FSEOF [23], we considered functionally grouped reactions that simultaneously carry fluxes based on unique features of microbial genomes. Considering such functionally grouped reactions helps reducing the number of and selecting multiple solutions existing for each optimal objective value, enabling to identify more reliable gene amplification targets when combined with FSEOF. Grouped reactions were previously revealed by genomic context and flux-converging pattern analyses as promising constraints [28]. Genomic context analysis interrogates conserved neighborhood, gene fusion, and co-occurrence using a STRING database with the goal of suggesting groups of reaction fluxes that are most likely correlated in their on/off activities [28,29]. Flux-converging pattern analysis further limits the range of possible flux values in a metabolic reaction by examining the number of carbon atoms in metabolites that participate in the reactions and the converging patterns of fluxes from a carbon source (see Methods and Figure 1) [28]. Consequently, flux balance analysis (FBA) with constraints controlling simultaneous on/off activity ($C_{on/off}$) and the flux scale (C_{scale}) of the metabolic reactions accurately predicted flux distributions in gene knockout mutant strains [28].

Based on these analyses, the grouping reaction (GR) constraints that constrain reactions to co-carry fluxes altogether regardless of the condition were incorporated into the *E. coli* genome-scale metabolic model. The model then facilitated the scanning of changes in the variability among metabolic fluxes using FVA in response to the enforced enhancement of the fluxes toward a target chemical. This newly developed method, called flux variability scanning based on enforced objective flux (FVSEOF) with GR constraints, was employed in this study to identify gene amplification targets for the production of target chemicals. FVSEOF with GR constraints was first validated based on amplification targets reported for the production of shikimic acid and putrescine in *E. coli*, and then further validated by actually engineering *E. coli* for the enhanced production of putrescine based on new amplification targets.

Methods

E. coli genome-scale metabolic model

EcoMBEL979 was used throughout this study [30], which is a slightly modified version of the genome-scale *E. coli* metabolic network model, *i*JR904 [31]. EcoMBEL979 contains 814 metabolites (144 extracellular metabolites and 670 intermediates) and 979 metabolic reactions, along with a biomass equation derived from the *E. coli* biomass composition [32].

Figure 1 Schematic illustration of the FVSEOF method with GR constraints. Functionally grouped reactions were considered based on genomic context and flux-converging pattern analyses obtained from the STRING database. FVSEOF was then performed under GR constraints to identify gene amplification candidates for the production of a target chemical. The candidates were evaluated based on the model predictions and additional criteria of the flux bias (V_{avg}) and the slope of the flux changes (q_{slope}). Each rectangle containing a C_xJ_y index and a line with different colors defines the reaction groups that are likely on or off simultaneously, as determined by genomic context and flux-converging pattern analyses. The C_xJ_y index for each reaction is determined by flux-converging pattern analysis. C_x and J_y denote the total number of carbon atoms in metabolites that participate in each reaction and the type of fluxes through the flux-converging metabolites from a carbon source, respectively. The red metabolites indicate flux-converging metabolites. The flux-converging metabolites indicate metabolites at which two pathways split by another metabolite recombine. For example, glyceraldehyde-3-phosphate converges the fluxes split by the fructose-bisphosphate aldolase from the fructose-6-phosphate. The flux-converging metabolites categorize J_y into four types, indicated as J_A, J_B, J_C, and J_D. Each subscript of J_y denotes the number of flux-converging metabolites that are passed zero, one, two, or three times, respectively, for a given flux from a carbon source. The subscript E is specially denoted to indicate the fluxes derived from pyruvate. The values of C_xJ_y for each reaction were assigned based on possible flux routes reaching from glucose, and are partitioned by a slash.

Constraints-based flux analysis

The stoichiometric relationships among the metabolites and the reactions of the *E. coli* genome-scale metabolic model were balanced under the pseudo-steady state assumption. The balanced reaction model was almost always underdetermined in calculations of the flux distribution due to insufficient measurements of the extracellular fluxes. Thus, the unknown fluxes within the metabolic reaction network were calculated by linear programming-based optimization using an objective function that maximized the growth rate, subject to constraints pertaining to mass conservation and reaction thermodynamics [33], This optimization problem can be mathematically formulated as follows:

$$\sum_{j \in J} S_{ij}v_j = b_i$$

$$a_j \le v_j \le \beta_j$$

where S_{ij} represents the stoichiometric coefficient for metabolite i in reaction j, v_j is the flux of reaction j, J is the set of all reactions, and b_i is the net transport flux of metabolite i. If this metabolite is an intermediate, b_i is equal to zero. α_j and β_j are the lower and upper bounds of the flux of reaction j, respectively. Herein, the flux of any irreversible reaction is considered to be positive; a negative flux indicates the reverse direction of a reaction.

Grouping reaction (GR) constraints based on the genomic context and flux-converging pattern analyses

The algorithm introduced in this study, FVSEOF with GR constraints, starts with formulation of GR constraints, which are based on the genomic context and flux-converging pattern analyses (Figure 1). Briefly, genomic context and flux-converging pattern analyses aim at grouping functionally related reactions. Such functionally related reactions were constrained to be on or off simultaneously (Figure 1) [28]. First, reactions were grouped using STRING database that performs genomic context analysis, including conserved neighborhood, gene fusion, and co-occurrence [28,29]. Simultaneous on/off constraint ($C_{on/off}$) can be described as follows:

$$y(v_1) = y(v_2)$$

$$y(v_1) \cdot \alpha_1 \leq v_1 \leq y(v_1) \cdot \beta_1$$

$$y(v_2) \cdot \alpha_2 \leq v_2 \leq y(v_2) \cdot \beta_2$$

where $y(v_1)$ and $y(v_2)$ indicate binary variables (on or off) of a certain reaction 1 and 2, respectively.

Each reaction is then given a $C_x J_y$ index, determined by flux-converging pattern analysis. C_x and J_y denote the total number of carbon atoms in metabolites that participate in each reaction and the number of passing flux-converging metabolites, respectively. Here, it should be noted that cofactors were not considered because the flux scales are controlled by the carbon number of primary metabolites, not cofactors, according to ^{13}C-based flux analysis [28]. For J_y, the flux-converging metabolites indicate metabolites at which two pathways split by another metabolite converge. J_y has four types, including J_A, J_B, J_C, and J_D, depending on the characteristics of flux-converging metabolites. Subscript of J_y denotes the passing number of flux-converging metabolites, counting zero, one, two, or three times for the flux coming from a carbon source. In some cases, the subscript E is placed next to the subscripts of A, B, C, or D to indicate the fluxes derived from pyruvate, which causes more complex changes in flux distributions. The values of $C_x J_y$ for each reaction were assigned based on possible flux routes reaching from glucose, and are partitioned by a slash. Based on this analysis, another constraint C_{scale}, indicating the flux scale of a reaction, can be given to the metabolic reactions. First, terms used to describe the flux scale of the reaction are as follows:

$$C_{xy} J_{yj}$$

$$x_j = \frac{N_{C,R_j}}{2}$$

where C_{xj} indicates the carbon number involved in a reaction j, J_{yj} the number of the passing of the flux through the flux-converging metabolite near reaction j, and $N_{C,Rj}$ the total number of carbon of primary metabolites without cofactors in reaction j.

If reaction 1 and 2 were predicted to be in the same functional unit according to the genomic context analysis, and their $C_{x1} J_{y1}$ and $C_{x2} J_{y2}$ are equivalent, C_{scale} is applied to these two reactions, which is defined as follows:

$$\sqrt{\frac{\left(|v_1^n| - \frac{|v_1^n| + |v_2^n|}{2}\right)^2 + \left(|v_2^n| - \frac{|v_1^n| + |v_2^n|}{2}\right)^2}{2}} \leq \delta$$

where v_1^n and v_2^n are the normalized flux of reaction 1 and 2, obtained by dividing each reaction flux by the carbon source uptake rate, such as glucose. δ is the constant defining the flux level of reactions in this functional unit; the value of δ is recommended as 0.3.

Flux variability scanning based on enforced objective flux (FVSEOF) with grouping reaction (GR) constraints

Once grouping reaction constraints are defined, FVSEOF with GR constraints is subsequently performed as follows (Figure 2). First, the initial or theoretical minimum ($v_{target\ product}^{initial}$) and theoretical maximum ($v_{target\ product}^{max}$) of the target product formation rates were calculated; these were implemented by setting the objective function as minimizing and maximizing the target product formation rate using constraints-based flux analysis with GR constraints. This can be formulated as follows:

$$\text{Min/Max } Z(v_{target\ product}) = v_{target\ product}^{initial} \text{ or } v_{target\ product}^{max}$$

$$\text{Subject to} \qquad \sum_{j \in J} S_{ij} v_j = b_i$$

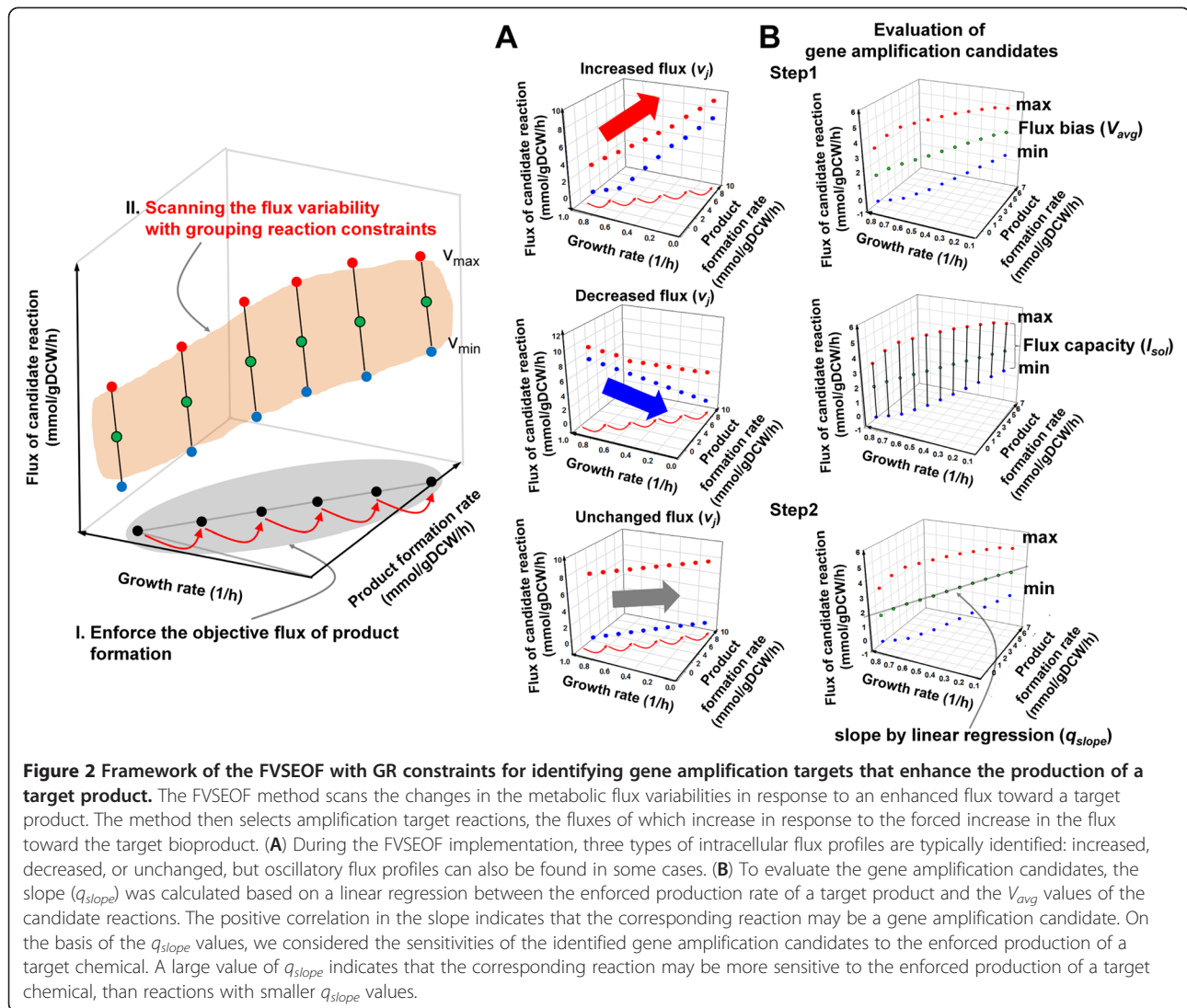$$l_i \leq \sum_{j \in J} S_{ij} v_j \leq u_i$$

$$\alpha_j \leq v_j \leq \beta_j$$

$$\beta_j = 1000 \text{ mmol} \cdot \text{g DCW}^{-1} \cdot \text{h}^{-1}$$

$$\alpha_j = -1000 \text{ mmol} \cdot \text{g DCW}^{-1} \cdot \text{h}^{-1}$$

$$v_{carbon}^{uptake} = 10 \text{ mmol} \cdot \text{g DCW}^{-1} \cdot \text{h}^{-1}$$

where $v_{target\ product}^{initial}$ indicates the initial or minimal point of the flux value constrained for the target bioproduct, while $v_{target\ product}^{max}$ indicates the maximal flux value for the bioproduct. l_i and u_i are the lower and upper bound for the net transport flux of metabolite i, respectively, and v_{carbon}^{uptake} is the carbon source uptake rate.

Figure 2 Framework of the FVSEOF with GR constraints for identifying gene amplification targets that enhance the production of a target product. The FVSEOF method scans the changes in the metabolic flux variabilities in response to an enhanced flux toward a target product. The method then selects amplification target reactions, the fluxes of which increase in response to the forced increase in the flux toward the target bioproduct. (**A**) During the FVSEOF implementation, three types of intracellular flux profiles are typically identified: increased, decreased, or unchanged, but oscillatory flux profiles can also be found in some cases. (**B**) To evaluate the gene amplification candidates, the slope (q_{slope}) was calculated based on a linear regression between the enforced production rate of a target product and the V_{avg} values of the candidate reactions. The positive correlation in the slope indicates that the corresponding reaction may be a gene amplification candidate. On the basis of the q_{slope} values, we considered the sensitivities of the identified gene amplification candidates to the enforced production of a target chemical. A large value of q_{slope} indicates that the corresponding reaction may be more sensitive to the enforced production of a target chemical, than reactions with smaller q_{slope} values.

Second, the cell growth rate, $Z(v_{biomass})$, was maximized while gradually increasing the target product formation rate from its initial (or minimal) flux value to its near theoretical maximum: $v_{target\ product}^{enforced} = v_{target\ product}^{initial} + \frac{k}{n}\left(v_{target\ product}^{max} - v_{target\ product}^{initial}\right)$, $K = \{k|k = 1, 2, \cdots, n-1\}$ ($n{\geq}10$) [23]. The $v_{target\ product}^{enforced}$ is an additional constraint provided during this stage of the constraints-based flux analysis; it starts with the initial value $v_{target\ product}^{initial}$ plus one n^{th} of the difference between the $v_{target\ product}^{max}$ and $v_{target\ product}^{initial}$, and is increased to a value adjacent to $v_{target\ product}^{max}$ in k steps.

Third, FVA was carried out with GR constraints by maximizing or minimizing the fluxes of all intracellular reactions, $Z(v_{intracellular\ reaction})$, with additional constraints: the enforced production rate of the target

bioproduct, which varied from its initial to maximum values in 10 steps, and 95% optimal cell growth rate, $v_{biomass} = 0.95 \cdot Z(v_{biomass})^{opt}$, for each step. The attainable flux ranges of intracellular reactions for each step were subsequently subjected to the targeting criteria introduced in the following section.

FVSEOF with GR constraints was calculated using mixed integer nonlinear programming with the DICOPT solver, subject to the constraints including GR constraints, mass conservation and reaction thermodynamics.

Flux bias, its slope and flux capacity as targeting criteria
Flux bias (V_{avg}), its slope (q_{slope}) and flux capacity (l_{sol}) were employed as targeting criteria for the initial set of gene amplification targets predicted from FVSEOF with GR constraints (Figure 2 and 3). Among them, V_{avg} and l_{sol} were determined as follows in order to effectively
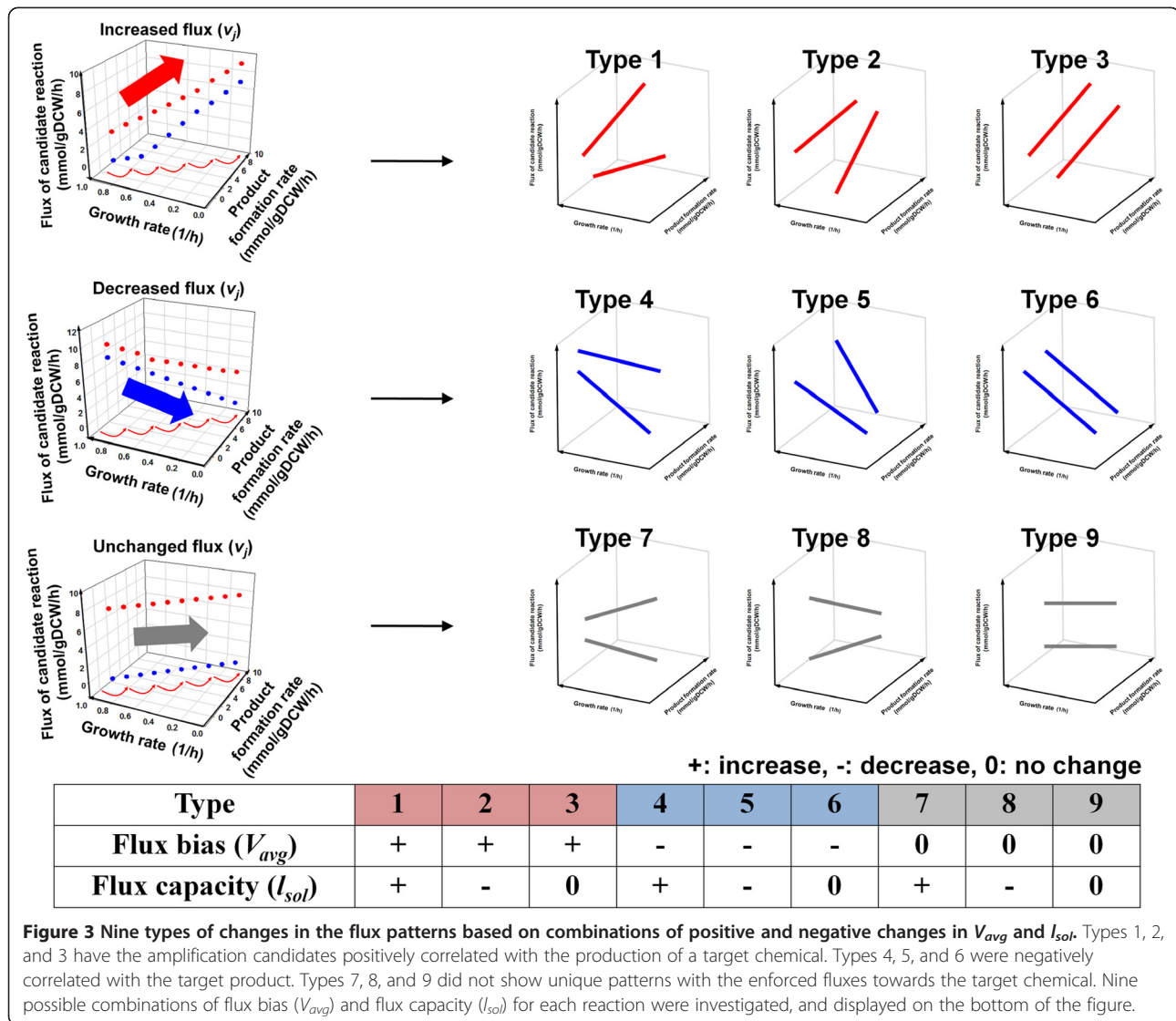
Figure 3 Nine types of changes in the flux patterns based on combinations of positive and negative changes in V_{avg} and l_{sol}. Types 1, 2, and 3 have the amplification candidates positively correlated with the production of a target chemical. Types 4, 5, and 6 were negatively correlated with the target product. Types 7, 8, and 9 did not show unique patterns with the enforced fluxes towards the target chemical. Nine possible combinations of flux bias (V_{avg}) and flux capacity (l_{sol}) for each reaction were investigated, and displayed on the bottom of the figure.

investigate the changes of flux variabilities for genetic perturbations [28]:

$$V_{avg} = \frac{V'_{max} + V'_{min}}{2}$$

$$l_{sol} = V'_{max} - V'_{min}$$

The V'_{max} and V'_{min} indicate the maximal and minimal flux values for a reaction under the given condition. The l_{sol} indicates the difference between the maximal and minimal flux values for a reaction. q_{slope} was calculated using linear regression of the flux values for a reaction towards the gradually maximized product formation rate.

Bacterial strains and plasmids

The *E. coli* strains used in this study are listed in the Additional file 1. The XQ52 strain, a putrescine producer, was used as a base strain [34]. *E. coli* TOP10 was used for gene cloning studies. The plasmid p15SpeC containing a strong *tac* promoter was used as an expression vector. The plasmid p15SpeC was constructed from the pTac15K plasmid by cloning the *speC* gene (encoding ornithine decarboxylase in the putrescine biosynthetic pathway) into the site between the *EcoRI* and *SacI* restriction enzyme sites of pTac15K. The plasmid contained a kanamycin resistance selective marker. Cells were grown in Luria–Bertani (LB) broth or on plates containing appropriate antibiotics at 37°C for the construction of strains and plasmids. Antibiotics were added at following concentrations: 50 μg/mL ampicillin, 25 μg/mL kanamycin, and 35 μg/mL chloramphenicol.

The plasmids used in this study are listed in the Additional file 1. Polymerase chain reaction (PCR) primers for the gene cloning studies conducted here are listed in the Additional file 2. *Pfu* DNA polymerase was purchased from Solgent (Daejeon, Korea). Restriction

enzymes and T4 DNA ligase were obtained from New England Biolabs (Ipswich, MA) and Roche (Mannheim, Germany), respectively. The genomic DNA of *E. coli* W3110 was amplified to overexpress the target genes using the *Pfu* polymerase and PCR primers (Additional file 2). The PCR product was then digested with *SacI* and *XbaI*, and ligated into p15SpeC at the same restriction sites downstream of the *tac* promoter.

Fermentation

Batch cultivation was conducted at 37°C in a 6.6 L jar fermentor (Bioflo 3000; New Brunswick Scientific Co., Edison, NJ) containing 2 L R/2 medium supplemented with 10 g/L glucose and 3 g/L $(NH_4)_2SO_4$. The R/2 medium (pH 6.8) contained (per liter): 2 g $(NH_4)_2HPO_4$, 6.75 g KH_2PO_4, 0.85 g citric acid, and 0.7 g $MgSO_4 \cdot 7H_2O$. In addition, 5 mL/L of a trace metal stock solution [35] was added. The trace metal solution contained per liter of 5 M HCl: 10 g $FeSO_4 \cdot 7H_2O$, 2.25 g $ZnSO_4 \cdot 7H_2O$, 1 g $CuSO_4 \cdot 5H_2O$, 0.5 g $MnSO_4 \cdot 5H_2O$, 0.23 g $Na_2B_4O_7 \cdot 10H_2O$, 2 g $CaCl_2 \cdot 2H_2O$, and 0.1 g $(NH_4)_6Mo_7O_{24}$. One milliliter of the overnight culture was transferred into a 300 mL Erlenmeyer flask containing 50 mL of the R/2 medium at 37°C and spun at 220 rpm in a shaking incubator (JEIOTech. Co. SI-900R). After obtaining an initial OD_{600} of 0.3, the seed cultures (200 mL) were introduced into the bioreactor for batch cultivation. The culture pH was maintained at 6.8 by the addition of 6 M KOH. The dissolved oxygen concentration was maintained at 20% air saturation by automatically adjusting the agitation speed. Under the comparable batch culture conditions, the single gene-overexpressing strains based on the *E. coli* XQ52 strain harboring p15SpeC, denoted as XQ52 (p15SpeC), with each target gene were tested by flask cultivation in duplicate using R/2 medium supplemented with 10 g/L glucose at 37 °C.

Analytical procedures

Cell growth was estimated by measuring the optical density at 600 nm (OD_{600}) using an Ultrospec 3000 spectrophotometer (Amersham Biosciences, Uppsala, Sweden). Glucose concentrations were measured using a glucose analyzer (model 2700 STAT; Yellow Springs Instrument, Yellow Springs, OH, USA). The concentrations of glucose and organic acids were determined by high-performance liquid chromatography (ProStar 210; Varian, Palo Alto, CA) equipped with UV/visible light (ProStar 320; Varian, Palo Alto, CA) and refractive index (Shodex RI-71, Tokyo, Japan) detectors. A MetaCarb 87H column (300 by 7.8 mm; Varian) was eluted isocratically with 0.01 NH_2SO_4 at 60°C at a flow rate of 0.4 mL/min.

The putrescine concentration was determined by derivatizing putrescine with *o*-phthaldialdehyde (OPA; Sigma, St. Louis, MO), and the *o*-phthaldialdehyde derivative was detected by high-performance liquid chromatography (1100 Series HPLC, Agilent Technologies, Palo Alto, CA) with UV detection, as described previously [34]. The OPA derivatization reagent was prepared as described previously [34,36,37]. Following the addition of the OPA reagent, the mixture was filtered through a 0.2 mm PVDF syringe filter (Whatman, Maidstone, UK), and the filtrate was immediately injected into the HPLC. A SUPELCO C18 column (cat# 504955; 5μm, 150 mm x 4.6 mm) was operated at 25°C with a 0.8 mL/min mobile phase flow rate. The mobile phase consisted of solution A (55% methanol in 0.1 M sodium acetate, pH 7.2) and solution B (methanol). The following gradient was applied (values given in vol%): 1–6 min, 100% A; 6–10 min, linear gradient of B from 0% to 30%; 10–15 min, linear gradient of B from 30% to 50%; 15–19 min, linear gradient of B from 50% to 100%; 19–23 min, 100% B; 23–25 min, linear gradient of B from 100% to 30%; 25–28 min, linear gradient of B from 30% to 0% [34]. The derivatized putrescine was detected at a wavelength of 230 nm using a variable wavelength detector (G1314A, Agilent Technologies).

Results and discussion

FVSEOF with GR constraints

Functionally related reactions can be grouped by genomic context and flux-converging pattern analyses [28]. Several reactions appeared to be related with one another based on genomic context analysis of conserved neighborhoods, gene fusions, and co-occurrence [28,29]. Flux-converging pattern analysis narrows the range of plausible flux values for metabolic reactions by examining the number of carbon atoms in metabolites that participate in reactions and the converging patterns of fluxes from a carbon source [28]. By controlling the simultaneous on/off activity ($C_{on/off}$) and flux scale (C_{scale}) of the metabolic reactions, based on FBA with GR constraints, the flux distributions in gene knockout mutants were accurately predicted [28]. In this study, the GR constraints were further applied to reactions related to the biosynthesis of a target chemical to improve the model accuracy (Figure 1 and the Additional file 3).

FVSEOF with GR constraints was implemented as follows (Figure 2). First, the theoretical minimal and maximal flux values for the target product formation were calculated using constraints-based flux analysis by minimizing and maximizing the target product formation rate with GR constraints. Second, again with GR constraints, the cell growth rate was maximized while gradually increasing the constraint value for the target product formation rate (our objective function which is artificially enforced) from a minimal to a theoretical maximum, as calculated from the first step. Finally, FVA was conducted by maximizing and minimizing the fluxes

of all intracellular reactions under additional constraints, including GR constraints, the enforced production rate of the target chemical varied from a minimal to a maximal value, and a 95% optimal growth rate constraint for each step. The attainable flux ranges for the intracellular reactions were calculated under the imposed constraints for each of the three steps.

Criteria for selecting gene amplification targets

Initial simulation results for FVSEOF with GR constraints were filtered based on rational criteria in an effort to select only the most effective amplification targets. The most important criterion was to identify gene amplification targets, the fluxes of which increased with the flux directed toward the target chemical. This procedure was implemented with quantitative values of q_{slope}, V_{avg}, and l_{sol} (Figures 2 and 3). The flux bias (V_{avg}) and flux capacity (l_{sol}) indicate an average value for the maximal and minimal flux values and the length of the attainable flux ranges for a reaction, respectively [28]. Finally, the gene amplification candidates were evaluated by calculating the slope (q_{slope}) of the V_{avg} flux for each metabolic reaction using linear regression analysis (Figure 2). Changes in the patterns of the reaction fluxes in response to incrementally increasing fluxes toward a target product were categorized into nine types based on combinations of positive and negative changes in V_{avg} and l_{sol} in order to facilitate the identification of amplification targets (Figure 3). Types 1, 2, and 3 displayed positive correlations with the amplification candidates for the production of a target chemical; types 4, 5, and 6 displayed negative correlations with the amplification candidates. Finally, types 7, 8, and 9 displayed no clear correlations with the production of a target chemical, based on V_{avg} (Figure 3). The reaction sets that were positively correlated (type 1, 2, and 3) were initially selected as amplification candidates.

Reaction sets belonging to type 1, 2 and 3, showing positive correlations with the enforced fluxes toward a target chemical, can then be further divided into strongly and weakly positive reactions (Figure 3 and Additional file 4). This step also allows narrowing down the candidates of gene amplification targets. The strongly positive reactions display a continuously increasing V_{avg} and a positive q_{slope} in response to the enhanced production of a target chemical, whereas weakly positive reactions show the same pattern, except for the presence of a partially negative q_{slope} (Additional file 4). Certainly, strongly positive reactions deserve primary attention as potential gene amplification targets.

The potential gene amplification candidates were prioritized by considering the l_{sol} value, which indicates the length between the maximal and minimal flux values of a metabolic reaction. Among the reactions that were positively correlated with the desired product, reactions with smaller values of l_{sol} received higher priorities because these reactions were more likely to display the predicted flux values than reactions with larger values of l_{sol}. A final list of gene amplification targets obtained from the above procedure was then selected based on biological knowledge.

Implementation of FVSEOF with GR constraints for shikimic acid production in *E. coli*

FVSEOF with GR constraints was employed to identify gene amplification targets for the enhanced production of an important aromatic chemical shikimic acid in *E. coli* (Figure 2). Shikimic acid is a key metabolic intermediate in the aromatic amino acid biosynthetic pathway. Shikimic acid and its derivatives are industrially important starting compounds for the production of several chemicals, such as phenols, herbicides, antibacterial agents, and the neuramidase inhibitor Tamiflu used for the treatment of influenza infections [38,39]. FVSEOF with GR constraints predicted that 11 reaction fluxes in the glycolysis (*glk* and *pps*), pentose phosphate pathway (*rpi*, *talAB*, and *tktAB*), and the shikimic acid biosynthetic pathway (*aroB*, *aroD*, *aroE*, *aroF*, *aroG*, and *aroH*) were potential amplification targets. The amplification of *aroB*, *aroD*, *aroE*, *aroF*, *aroG*, *aroH*, *talAB*, *tktA*, *glk*, and *pps* genes [38-46], which are the amplification targets predicted by FVSEOF with GR constraints, was previously reported to enhance the production of shikimic acid. The previous FSEOF method without FVA and GR constraints could not identify *pps* gene as an amplification target. FSEOF results without FVA and GR constraints did not show notable fluxes among metabolic reactions controlled by the *pps* gene in response to the enforced shikimic acid production rate; however, the FVSEOF method with GR constraints correctly predicted the *pps* gene as one of the amplification targets beneficial for the accumulation of phosphoenolpyruvate, an important precursor for the production of shikimic acid from pyruvate. This consistency partly demonstrated the power of utilizing FVA and GR constraints for predicting reliable amplification targets by FSEOF. In practice, the overexpression of phosphoenolpyruvate synthase encoded by the *pps* gene also increased the yield of precursors for the production of shikimic acid [43,45]. Thus, this strategy enabled the successful identification of gene amplification targets for the enhanced production of a primary metabolite, shikimic acid, in *E. coli*, in accordance with previous literature reports.

Implementation of FVSEOF with GR constraints for enhanced putrescine production in *E. coli*

The general applicability of FVSEOF with GR constraints was examined by applying the method to

putrescine production in *E. coli*. Putrescine (1,4-diaminobutane) is an important industrial precursor for the synthesis of polymers, pharmaceuticals, surfactants, and certain additives [34]. We confirmed the validity of the newly predicted gene amplification targets by comparison with the genes engineered in the previously reported putrescine-producing *E. coli* XQ52 (p15SpeC) strain [34]. FVSEOF with GR constraints predicted potential gene amplification targets among the reactions involved in glycolysis (*eno, pgm, gapA, fbaAB, tpiA, pgk, pykAF,* and *glk*), TCA cycle (*icd, acnA, acnB,* and *gltA*), putrescine biosynthesis (*gdhA, argA, argB, argC, argD, argE, speC,* and *speF*), and other pathways (*ackA* and *ppc*). The predicted amplification targets (*argB, argC, argD, argE, speC,* and *speF*) involved in the putrescine biosynthetic pathway were consistent with the mutations introduced in the *E. coli* XQ52 (p15SpeC) strain, as described in the previous report [34].

The genes predicted to be relevant to the putrescine biosynthetic pathway were expected based on the pathway knowledge, and were intuitively obvious; hence, we focused on the effects associated with amplifying the predicted gene targets involved in other metabolic pathways in order to more rigorously validate FVSEOF with GR constraints. Accordingly, each of the predicted amplification targets was examined one by one by amplifying the gene dosage in the *E. coli* XQ52 (p15SpeC) strain (see Methods). Among these genes, the *glk, acnA, acnB, ackA,* and *ppc* gene, five out of the sixteen target genes, were found to attain increased putrescine yield when they were individually amplified in the *E. coli* XQ52 (p15SpeC) strain (Additional file 5). These strains, initially examined using flask cultivation, were further validated by batch cultivation at 37°C under aerobic condition (Figure 4 and see the Additional files 1 and 2). The recombinant *E. coli* XQ52 (p15SpeC) strain additionally expressing the *glk, acnA, acnB, ackA,* and *ppc* genes resulted in the production of 2.23, 1.90, 1.89, 2.04, and 2.06 g/L of putrescine, respectively, which are 20.5 % more than that (1.68 g/L) produced by the control
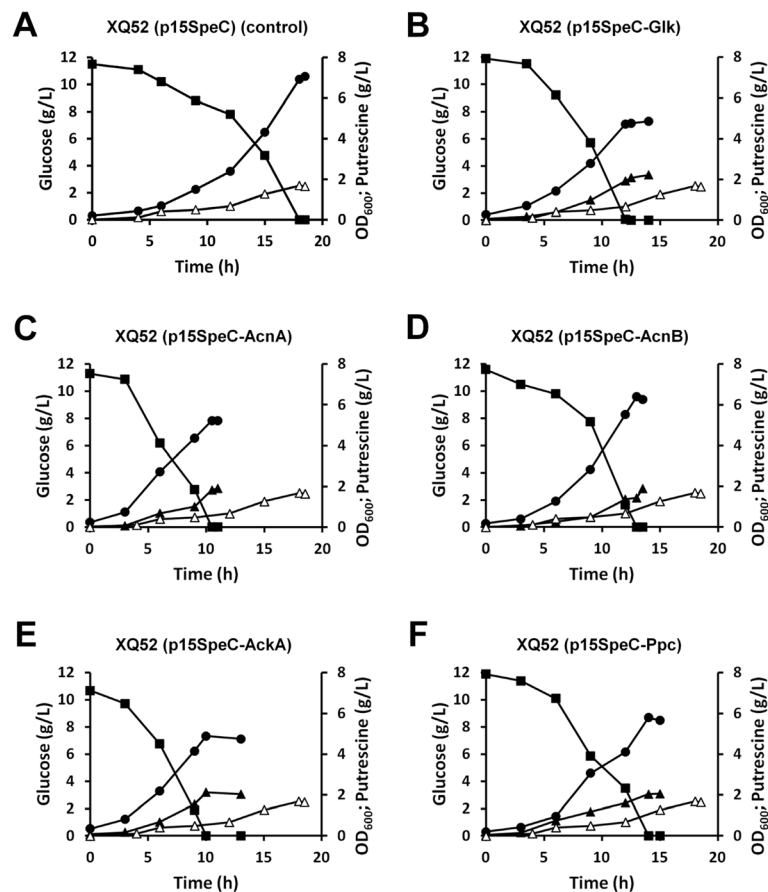

Figure 4 Batch cultivation of the control and engineered *E. coli* strains producing putrescine. (**A**) XQ52 (p15SpeC) (control), (**B**) XQ52 (p15SpeC-Glk), (**C**) XQ52 (p15SpeC-AcnA), (**D**) XQ52 (p15SpeC-AcnB), (**E**) XQ52 (p15SpeC-AckA), and (**F**) XQ52 (p15SpeC-Ppc) strains were tested for putrescine production. The symbols indicate: cell concentration measured by OD_{600} (●), glucose concentration (■), and putrescine concentration (▲) for the engineered strains in (**A**) to (**F**), and putrescine concentration (△) for the XQ52 (p15SpeC) (control) strain in (**A**).

strain on average (Figure 4). The yields of putrescine obtained with these strains were 0.223, 0.190, 0.189, 0.204, and 0.206 g putrescine per g glucose, which are again higher than that (0.168 g putrescine per g glucose) obtained with the control strain. Thus, FVSEOF with GR constraints could be successfully used to identify non-obvious gene amplification targets that enhance the production of putrescine in *E. coli*.

Other gene amplification targets identified by FVSEOF with GR constraints, which did not affect putrescine production in flask cultivation, also deserve discussion. These false-positive hits are most likely involved in biological processes that were not accurately captured in the genome-scale metabolic model. These ineffective genes, including *eno*, *pgm*, *gapA*, *fbaAB*, *tpiA*, *pgk*, and *pykAF* genes in glycolysis, and *icd* and *gltA* genes in TCA cycle might have been associated with transcriptional and translational regulations because the direct correlation between gene expressions and the metabolic fluxes was not observed. The fact that some of obvious gene amplification targets, such as *icd* gene responsible for biosynthesis of α-ketoglutarate, seem to be resistant to gene manipulations indicates that other biological variables may affect the effects of the gene amplifications. Potential variables include the plasmid copy number, gene dosage, optimal gene expression, and the gene expression method, either plasmid-based overexpression or chromosomal integration [47]. Although we improved the accuracy of the predicted gene targets by imposing GR constraints, these factors should be carefully considered in any implementation of the FVSEOF method with GR constraints [48].

Conclusions

FVSEOF with GR constraints, which is an upgraded version of the FSEOF method, allows for the *in silico* identification of fluxes to be amplified for the enhanced production of target products. This method was conducted through the analysis of trends in reaction flux variability in response to varying the flux of target chemical production from initial to maximal flux values under GR constraints. The confidence with which amplification targets are identified may be increased by incorporating physiological data. This approach involves grouping functionally related reactions based on their genomic context and flux-converging pattern analyses. The interaction data may be obtained easily from public databases, and subjected to GR constraints. FVA was also performed to overcome the problems associated with multiple solutions for an optimal objective value. FVSEOF with GR constraints was shown to suggest successful metabolic engineering strategies (in particular, gene amplification) for the production of shikimic acid and putrescine in *E. coli*. In conclusion, the strategy

reported here should be generally useful for developing industrial strains that display enhanced production of a target chemical.

Additional files

Additional file 1: Bacterial strains and plasmids used in this study.

Additional file 2: Oligonucleotides used in this study.

Additional file 3: Genomic context and flux-converging pattern analyses for shikimic acid and putrescine production in *Escherichia coli*.

Additional file 4: Analysis of flux patterns with partial variations.

Additional file 5: Putrescine production yield (g putrescine/g glucose) for the single gene-overexpressing strains based on *E. coli* XQ52 (p15SpeC) strain by flask cultivation on R/2 medium supplemented with 10 g/L glucose at 37 °C.

Abbreviations
FBA: Flux balance analysis; FSEOF: Flux scanning based on enforced objective flux; FVA: Flux variability analysis; FVSEOF: Flux variability scanning based on enforced objective flux; GR: Grouping reaction.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
JMP, TYK, and SYL generated the ideas. JMP, HMP, WJK, TYK, and SYL designed the research. JMP, HMP, and WJK performed the research. JMP and HMP performed the analytical experiments. JMP, HMP, WJK, and HUK analyzed the data. JMP, HMP, WJK, HUK, and SYL wrote the paper. All authors read and approved the final manuscript.

Author details
[1]Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical and Biomolecular Engineering (BK21 program), Center for Systems and Synthetic Biotechnology, Institute for the BioCentury, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Republic of Korea. [2]BioInformatics Research Center, KAIST, Daejeon 305-701, Republic of Korea. [3]Department of Bio and Brain Engineering and BioProcess Engineering Research Center, KAIST, Daejeon 305-701, Republic of Korea.

References
1. Lee SY, Lee DY, Kim TY: **Systems biotechnology for strain improvement.** *Trends Biotechnol* 2005, **23**:349–358.
2. Joyce AR, Palsson BO: **The model organism as a system: integrating 'omics' data sets.** *Nat Rev Mol Cell Biol* 2006, **7**:198–210.
3. Kim HU, Kim TY, Lee SY: **Metabolic flux analysis and metabolic engineering of microorganisms.** *Mol Biosyst* 2008, **4**:113–120.
4. Kim TY, Sohn SB, Kim YB, Kim WJ, Lee SY: **Recent advances in reconstruction and applications of genome-scale metabolic models.** *Curr Opin Biotechnol* 2011, **23**:617–623.
5. Lee KH, Park JH, Kim TY, Kim HU, Lee SY: **Systems metabolic engineering of *Escherichia coli* for L-threonine production.** *Mol Syst Biol* 2007, **3**:149.
6. Park JH, Lee KH, Kim TY, Lee SY: **Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and *in silico* gene knockout simulation.** *Proc Natl Acad Sci U S A* 2007, **104**:7797–7802.

7. Lee SJ, Lee DY, Kim TY, Kim BH, Lee J, Lee SY: **Metabolic engineering of** *Escherichia coli* **for enhanced production of succinic acid, based on genome comparison and** *in silico* **gene knockout simulation.** *Appl Environ Microbiol* 2005, **71**:7880–7887.

8. Bro C, Regenberg B, Forster J, Nielsen J: *In silico* **aided metabolic engineering of** *Saccharomyces cerevisiae* **for improved bioethanol production.** *Metab Eng* 2006, **8**:102–111.

9. Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, Boldt J, Khandurina J, Trawick JD, Osterhout RE, Stephen R, *et al*: **Metabolic engineering of** *Escherichia coli* **for direct production of 1,4-butanediol.** *Nat Chem Biol* 2011, **7**:445–452.

10. Jamshidi N, Palsson BO: **Investigating the metabolic capabilities of** *Mycobacterium tuberculosis* **H37Rv using the** *in silico* **strain** *i*NJ661 **and proposing alternative drug targets.** *BMC Syst Biol* 2007, **1**:26.

11. Kim HU, Kim TY, Lee SY: **Genome-scale metabolic network analysis and drug targeting of multi-drug resistant pathogen** *Acinetobacter baumannii* **AYE.** *Mol Biosyst* 2010, **6**:339–348.

12. Kim HU, Kim SY, Jeong H, Kim TY, Kim JJ, Choy HE, Yi KY, Rhee JH, Lee SY: **Integrative genome-scale metabolic analysis of** *Vibrio vulnificus* **for drug targeting and discovery.** *Mol Syst Biol* 2011, **7**:460.

13. Jensen PR, Hammer K: **The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters.** *Appl Environ Microbiol* 1998, **64**:82–87.

14. Koffas MA, Jung GY, Stephanopoulos G: **Engineering metabolism and product formation in** *Corynebacterium glutamicum* **by coordinated gene overexpression.** *Metab Eng* 2003, **5**:32–41.

15. Delgado J, Liao JC: **Inverse flux analysis for reduction of acetate excretion in** *Escherichia coli. Biotechnol Prog* 1997, **13**:361–367.

16. Burgard AP, Nikolaev EV, Schilling CH, Maranas CD: **Flux coupling analysis of genome-scale metabolic network reconstructions.** *Genome Res* 2004, **14**:301–312.

17. Puchalka J, Oberhardt MA, Godinho M, Bielecka A, Regenhardt D, Timmis KN, Papin JA: **Martins dos Santos VA: Genome-scale reconstruction and analysis of the** *Pseudomonas putida* **KT2440 metabolic network facilitates applications in biotechnology.** *PLoS Comput Biol* 2008, **4**:e1000210.

18. Bundy JG, Papp B, Harmston R, Browne RA, Clayson EM, Burton N, Reece RJ, Oliver SG, Brindle KM: **Evaluation of predicted network modules in yeast metabolism using NMR-based metabolite profiling.** *Genome Res* 2007, **17**:510–519.

19. Khannapho C, Zhao H, Bonde BK, Kierzek AM, Avignone-Rossa CA, Bushell ME: **Selection of objective function in genome scale flux balance analysis for process feed development in antibiotic production.** *Metab Eng* 2008, **10**:227–233.

20. Bushell ME, Sequeira SI, Khannapho C, Zhao H, Chater KF, Butler MJ, Kierzek AM, Avignone-Rossa CA: **The use of genome scale metabolic flux variability analysis for process feed formulation based on an investigation of the effects of the** *zwf* **mutation on antibiotic production in** *Streptomyces coelicolor. Enzyme Microb Technol* 2006, **39**:1347–1353.

21. Mahadevan R, Schilling CH: **The effects of alternate optimal solutions in constraint-based genome-scale metabolic models.** *Metab Eng* 2003, **5**:264–276.

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

23. Choi HS, Lee SY, Kim TY, Woo HM: *In silico* **identification of gene amplification targets for improvement of lycopene production.** *Appl Environ Microbiol* 2010, **76**:3097–3105.

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

25. Lun DS, Rockwell G, Guido NJ, Baym M, Kelner JA, Berger B, Galagan JE, Church GM: **Large-scale identification of genetic design strategies using local search.** *Mol Syst Biol* 2009, **5**:296.

26. Jung YK, Kim TY, Park SJ, Lee SY: **Metabolic engineering of** *Escherichia coli* **for the production of polylactic acid and its copolymers.** *Biotechnol Bioeng* 2010, **105**:161–171.

27. Ranganathan S, Suthers PF, Maranas CD: **OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions.** *PLoS Comp Biol* 2010, **6**:e1000744.

28. Park JM, Kim TY, Lee SY: **Prediction of metabolic fluxes by incorporating genomic context and flux-converging pattern analyses.** *Proc Natl Acad Sci U S A* 2010, **107**:14931–14936.

29. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, *et al*: **STRING 8–a global view on proteins and their functional interactions in 630 organisms.** *Nucleic Acids Res* 2009, **37**:D412–416.

30. Lee SY, Woo HM, Lee D-Y, Choi HS, Kim TY, Yun H: **Systems-level analysis of genome-scale microbial metabolisms under the integrated software environment.** *Biotechnol Bioproc Eng* 2005, **10**:425–431.

31. Reed JL, Vo TD, Schilling CH, Palsson BO: **An expanded genome-scale model of** *Escherichia coli* **K-12 (** *i*JR904 GSM/GPR**).** *Genome Biol* 2003, **4**:R54.

32. Neidhardt FC, Umbarger HE: **Chemical composition of** *Escherichia coli.* In *Escherichia coli and Salmonella: cellular and molecular biology. 2nd edition.* Edited by Neidhardt FC, Curtiss R. Washington, D.C.: ASM Press; 1996:13–16.

33. Park JM, Kim TY, Lee SY: **Constraints-based genome-scale metabolic simulation for systems metabolic engineering.** *Biotechnol Adv* 2009, **27**:979–988.

34. Qian ZG, Xia XX, Lee SY: **Metabolic engineering of** *Escherichia coli* **for the production of putrescine: a four carbon diamine.** *Biotechnol Bioeng* 2009, **104**:651–662.

35. Lee SY, Chang HN: **High cell density cultivation of** *Escherichia coli* **W using sucrose as a carbon source.** *Biotechnol Lett* 1993, **15**:971–974.

36. Onal A: **A review: Current analytical methods for the determination of biogenic amines in foods.** *Food Chem* 2007, **103**:1475–1486.

37. Yildirim HK, Uren A, Yucel U: **Evaluation of biogenic amines in organic and non-organic wines by HPLC OPA derivatization.** *Food Technol Biotechnol* 2007, **45**:62–68.

38. Kramer M, Bongaerts J, Bovenberg R, Kremer S, Muller U, Orf S, Wubbolts M, Raeven L: **Metabolic engineering for microbial production of shikimic acid.** *Metab Eng* 2003, **5**:277–283.

39. Johansson L, Lindskog A, Silfversparre G, Cimander C, Nielsen KF, Linder G: **Shikimic acid production by a modified strain of E. coli (W3110.shik1) under phosphate-limited and carbon-limited conditions.** *Biotechnol Bioeng* 2005, **92**:541–552.

40. Knop DR, Draths KM, Chandran SS, Barker JL, von Daeniken R, Weber W, Frost JW: **Hydroaromatic equilibration during biosynthesis of shikimic acid.** *J Am Chem Soc* 2001, **123**:10173–10182.

41. Knaggs AR: **The biosynthesis of shikimate metabolites.** *Nat Prod Rep* 2003, **20**:119–136.

42. Chandran SS, Yi J, Draths KM, von Daeniken R, Weber W, Frost JW: **Phosphoenolpyruvate availability and the biosynthesis of shikimic acid.** *Biotechnol Prog* 2003, **19**:808–814.

43. Escalante A, Calderon R, Valdivia A, de Anda R, Hernandez G, Ramirez OT, Gosset G, Bolivar F: **Metabolic engineering for the production of shikimic acid in an evolved** *Escherichia coli* **strain lacking the phosphoenolpyruvate: carbohydrate phosphotransferase system.** *Microb Cell Fact* 2010, **9**:21.

44. Lu JL, Liao JC: **Metabolic engineering and control analysis for production of aromatics: Role of transaldolase.** *Biotechnol Bioeng* 1997, **53**:132–138.

45. Patnaik R, Liao JC: **Engineering of** *Escherichia coli* **central metabolism for aromatic metabolite production with near theoretical yield.** *Appl Environ Microbiol* 1994, **60**:3903–3908.

46. Yi J, Draths KM, Li K, Frost JW: **Altered glucose transport and shikimate pathway product yields in** *E. coli. Biotechnol Prog* 2003, **19**:1450–1459.

47. Keasling JD: **Gene-expression tools for the metabolic engineering of bacteria.** *Trends Biotechnol* 1999, **17**:452–460.

48. Thiele I, Jamshidi N, Fleming RM, Palsson BO: **Genome-scale reconstruction of** *Escherichia coli*'**s transcriptional and translational machinery: a knowledge base, its mathematical formulation, and its functional characterization.** *PLoS Comput Biol* 2009, **5**:e1000312.