BMC Systems Biology

Methodology article

Integration of enzyme activities into metabolic flux distributions by elementary mode analysis

Hiroyuki Kurata*, Quanyu Zhao, Ryuichi Okuda and Kazuyuki Shimizu

Address: Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka, 820-8502, Japan

Email: Hiroyuki Kurata* - kurata@bio.kyutech.ac.jp; Quanyu Zhao - qyzhao@bio.kyutech.ac.jp; Ryuichi Okuda - e791006k@bio.kyutech.ac.jp; Kazuyuki Shimizu - shimi@bio.kyutech.ac.jp

* Corresponding author

This article is available from: http://www.biomedcentral.com/1752-0509/1/31

Abstract

Background: In systems biology, network-based pathway analysis facilitates understanding or designing metabolic systems and enables prediction of metabolic flux distributions. Network-based flux analysis requires considering not only pathway architectures but also the proteome or transcriptome to predict flux distributions, because recombinant microbes significantly change the distribution of gene expressions. The current problem is how to integrate such heterogeneous data to build a network-based model.

Results: To link enzyme activity data to flux distributions of metabolic networks, we have proposed Enzyme Control Flux (ECF), a novel model that integrates enzyme activity into elementary mode analysis (EMA). ECF presents the power-law formula describing how changes in enzyme activities between wild-type and a mutant are related to changes in the elementary mode coefficients (EMCs). To validate the feasibility of ECF, we integrated enzyme activity data into the EMCs of *Escherichia coli* and *Bacillus subtilis* wild-type. The ECF model effectively uses an enzyme activity profile to estimate the flux distribution of the mutants and the increase in the number of incorporated enzyme activities decreases the model error of ECF.

Conclusion: The ECF model is a non-mechanistic and static model to link an enzyme activity profile to a metabolic flux distribution by introducing the power-law formula into EMA, suggesting that the change in an enzyme profile rather reflects the change in the flux distribution. The ECF model is highly applicable to the central metabolism in knockout mutants of *E. coli* and *B. subtilis*.

Background

In systems biology, a mathematical approach is required to integrate heterogeneous data, such as transcriptome, proteome, metabolome, and fluxome, to build comprehensive metabolic models. Mathematical models are consistently improved or modified by accommodating new experimental data with the current models, thereby enhancing the validation of metabolic networks or the prediction of their dynamic behaviors[1,2].

Network-based pathway analysis becomes a core method for constructing a mathematical model that predicts the flux distribution for large-scale metabolic networks. Network-based analysis generally employs a constraint-based modelling approach [3], e.g., Flux Balance Analysis (FBA) that uses a stoichiometric matrix and objective function that define a network's allowable solution space. The target flux capacity is provided by optimizing specific objec-

tive functions such as cell growth, energy, or metabolite synthesis maximization[3,4].

Recent network-based metabolic pathway analysis has focused on two approaches, elementary modes (EMs) [5] and extreme pathways[6]. Both consist of a convex set of vectors used to characterize all steady-state flux distributions of a biochemical network. The elementary mode is the minimal set of enzymes that can operate at steady-state, with all the irreversible reactions operating properly. Elementary mode or extreme pathway analysis enables an understanding of a large-scale network, predicting optimal or suboptimal metabolic fluxes under constrained conditions.

Since some genetic or environmental perturbations often cause a significant change in gene expression, a network-based flux analysis requires considering not only pathway architectures but also the proteins or mRNA transcripts to predict a flux distribution. However, only a few methods have been developed that connect such heterogeneous data to network-based analyses[7,8], e.g., the constraints from absent model genes with Boolean logic were employed. A current problem is to explore the relationship between an enzyme activity profile and a metabolic flux distribution, or how to mathematically integrate such heterogeneous data to build a network-based model.

To elucidate theoretical relationships between transcriptome or enzyme activities and flux, the control-effective flux (CEF) method has previously been proposed to predict gene expression (enzyme) levels through elementary mode analysis[9], but CEF does not estimate metabolic flux distributions from transcriptome or enzyme activities data. In our limited knowledge, there are few efficient mathematical models that use an enzyme activity profile to estimate a metabolic flux distribution. The mathematical connections between enzymes and flux distributions are a very important task for a network-based model.

To link metabolic enzyme data to flux distributions, we propose Enzyme Control Flux (ECF), a non-mechanistic and static method that integrates enzyme activities into EMA. The ECF model presents a novel mathematical formula describing how the change in enzyme activities between wild-type and a mutant are related to that in the EM coefficients (EMCs). To validate the feasibility of ECF, we integrated enzyme activity data into the EMCs of the central metabolism of *E. coli* and *B. subtilis* wild-type to simulate the metabolic flux distributions of the mutants.

Results and discussion

The algorithm of ECF that links an enzyme activity profile to a metabolic flux distribution is presented as shown in Figure 1. The mathematical procedure of ECF is intelligi-

bly illustrated (see Additional file 1). The power law formula uses the change in an enzyme activity profile between wild-type and a mutant to calculate the EMCs of the mutant, thereby simulating the flux distribution of the mutant. To guarantee the generality and applicability of ECF, we simulate the flux distributions of *E. coli* and *B. subtilis* mutants by integrating the change in an enzyme activity profile into the EMCs of their wild-type.

ECF algorithm
Elementary mode analysis
Biological networks can be represented by a stoichiometric matrix (**S**). The rows of **S** correspond to metabolites in a reaction network. The columns of **S** correspond to the reactions in a network, with elements corresponding to stoichiometric coefficients of the associated reactions. At steady-state, mass balance provides the flux-balance equations:

$$\mathbf{S} \cdot \mathbf{v} = \mathbf{0}, \qquad (1)$$

where $\mathbf{v} = (v_1, v_2, ..., v_n)^t$ is the vector whose elements correspond to fluxes through the associated reactions in **S**. The set of all possible solutions can be described by a set of basis vectors. The elementary mode (EM) is the minimal set of enzymes that can operate at steady-state with all the irreversible reactions operating properly[5]. The elementary mode matrix (**P**) is uniquely determined from the stoichiometric matrix and the flux vector is provided by:

$$\mathbf{v} = \mathbf{P} \cdot \mathbf{c}, \qquad (2)$$

Figure 1 flow chart:

- EM extraction in wild type → Flux distribution for wild type $\mathbf{v} = \mathbf{P} \cdot \mathbf{c}$ — $\mathbf{c}$: An EMC vector
- Calculation of the EMC vectors for wild type based on α-spectrum → $\mathbf{r}_j$: EMC vectors optimized for wild type
- Estimation of the EMC vectors for a mutant by the power-law formula → $\mathbf{r}_j^{mutant}$: EMC vectors for a mutant
- Estimation of flux distributions of the mutant by using the EMC vectors for the mutant → Flux distributions for a mutant: $\mathbf{v}_j^{mutant} = \mathbf{P} \cdot \mathbf{r}_j^{mutant}$
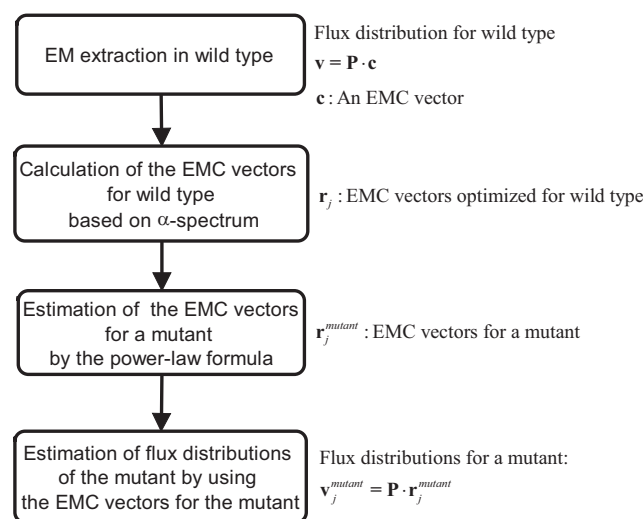
Figure 1
A flow chart for the ECF-based estimation of a flux distribution of a mutant.

where $\mathbf{c} = (\lambda_1, \lambda_2, ..., \lambda_m)^t$ is the EMC vector, n is the number of reactions, and m is the number of EMs. The ingredients of these vectors and matrix are displayed as:

$$
\begin{pmatrix} v_1 \\ v_2 \\ . \\ . \\ . \\ v_n \end{pmatrix} = \begin{pmatrix} e_{11} & e_{12} & . & . & . & . & . & e_{1m} \\ e_{21} & e_{22} & . & . & . & . & . & e_{2m} \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ e_{n1} & e_{n2} & . & . & . & . & . & e_{nm} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ . \\ . \\ . \\ . \\ . \\ \lambda_m \end{pmatrix}.
$$

(3)

The i-th column for the **P** matrix is the i-th EM vector: $\mathbf{e}_i = (e_{1i}, e_{2i}, ..., e_{ni})^t$. The flux distribution can be also represented as a superposition of the EM vectors with non-negative EMCs as follows:

$$
\mathbf{v} = \sum_{i=1}^{m} \lambda_i \mathbf{e}_i. \tag{4}
$$

EMC spectrum for wild-type
Generally the EMCs are not uniquely determined. For FBA, an objective function can be designed for phenomenological behaviors such as cell growth and product formation to determine the solution spectra of EMCs. Here we do not define any specific objective function involving phenomenological behaviors. Based on the α-spectrum [10], the allowable solution space of the EMCs is calculated by maximizing or minimizing each EMC in a given steady-state flux distribution as follows:

for $j = 1, 2, ..., m$

Maximize λ_j, subject to $\mathbf{v} = \mathbf{P} \cdot \mathbf{c}$, $\lambda_j \geq 0$ ($j = 1, 2, ..., m$),
  (5)

for $j = 1, 2, ..., m$

Minimize λ_j, subject to $\mathbf{v} = \mathbf{P} \cdot \mathbf{c}$, $\lambda_j \geq 0$ ($j = 1, 2, ..., m$).
  (6)

Linear programming is performed using Matlab (Mathworks Inc., Natick, MA) to optimize the EMC vectors for wild type. For each j in Eqs. (5, 6), as the EMC vectors (**c**), $\mathbf{r}_{2j-1} = (\gamma_{1,2j-1}, \gamma_{2,2j-1}, ..., \gamma_{m,2j-1})^t$ and $\mathbf{r}_{2j} = (\gamma_{1,2j}, \gamma_{2,2j}, ..., \gamma_{m,2j})^t$ are obtained. Consequently, *2m* of the EMC vectors for wild type: $\mathbf{r}_j = (\gamma_{1j}, \gamma_{2j}, ..., \gamma_{mj})^t$ ($j = 1, 2, ..., 2m$) are generated, where γ_{ij} is the i-th EMC value of the j-th EMC vector.

Enzyme-Control Flux (ECF)
In the ECF model, the power law formula integrates the change in an enzyme activity profile between wild-type and a mutant into the EMCs of wild-type to calculate the EMCs of the mutant, thereby simulating the flux distribution of the mutant.

First, the ECF model is constructed based on the correlation equation between the intermediate EMC vectors for a mutant:

$$
\mathbf{inter\_r}_j^{mutant} = (inter\gamma_{1,j}^{mutant}, inter\gamma_{2,j}^{mutant}, ..., inter\gamma_{m,j}^{mutant})^t \quad (j = 1, 2, ..., 2m),
$$

and the EMC vectors of wild-type $\mathbf{r}_j$. The superscript of *mutant* indicates a mutant. EMCs indicate the weight of flux through their corresponding EMs, which is affected by the enzyme activities that belong to the EM. Generally, Metabolic Control Analysis (MCA) says that control of metabolic flux is not determined by a rate-limiting step but shared in many enzyme reactions. In analogy to MCA, we assume that the fluxes through EMs are shared in the enzyme reactions that belong to the EMs. On the other hand, it is difficult to identify the degree by which each enzyme affects the EMC involving it. Thus, assuming that changes in the enzyme activities that belong to an EM affect the flux through the EM synergistically, the power-law formalism provides the correlation between $\mathbf{inter\_r}_j^{mutant}$ ($inter\gamma_{ij}^{mutant}$) and $\mathbf{r}_j(\gamma_{ij})$:

$$
\mathbf{inter\_r}_j^{mutant} = \begin{pmatrix} \gamma_{1j} \prod_{p=1}^{n} \alpha_{p1}{}^{\beta} \\ \gamma_{2j} \prod_{p=1}^{n} \alpha_{p2}{}^{\beta} \\ ... \\ \gamma_{mj} \prod_{p=1}^{n} \alpha_{pm}{}^{\beta} \end{pmatrix} \quad (j = 1, 2, ..., 2m),
$$

(7)

or

$$
inter\gamma_{ij}^{mutant} = \gamma_{ij} \prod_{p=1}^{n} \alpha_{pi}{}^{\beta}, \quad (i = 1, 2, .., m; \ j = 1, 2, ..., 2m)
$$

(8)

where β is the power factor. The symbol of α_{pi} is defined by:

$$\alpha_{pi} = \begin{cases} a_p & (e_{pi} \neq 0) \\ 1 & (e_{pi} = 0) \end{cases}. \tag{9}$$

where α_p is the relative enzyme activity of a mutant to wild-type for the *p*-th flux. The power factor of β is the unique parameter that adjusts ECF to a real metabolic system. For simplification, the power factor is set as the same value for each enzyme activity. If a given enzyme activity is zero, the EMC involving it is zero according to the power-law formula.

Second, the EMC vector for the mutant, as defined by:

$$\mathbf{r}_j^{mutant} = (\gamma_{1,j}^{mutant}, \gamma_{2,j}^{mutant}, ..., \gamma_{m,j}^{mutant})^t \quad (j = 1, 2, ..., 2m),$$

is mathematically connected to $\mathbf{inter\_r}_j^{mutant}$ under the constraint that the incoming flux (v_q) is setting to 100 (e.g., v_{19} = 100 in Figure 2) as follows:

$$\mathbf{r}_j^{mutant} = \frac{v_q}{\mathbf{u}_q \cdot (\mathbf{P} \cdot \mathbf{inter\_r}_j^{mutant})} \mathbf{inter\_r}_j^{mutant} \quad (j = 1, 2, ..., 2m). \tag{10}$$

or

$$\gamma_{ij}^{mutant} = \frac{v_q}{\mathbf{u}_q \cdot (\mathbf{P} \cdot \mathbf{inter\_r}_j^{mutant})} inter\gamma_{ij}^{mutant} \quad (i = 1, 2, .., m; \; j = 1, 2, ..., 2m), \tag{11}$$

Here, $\mathbf{u}_q$ is the *m*-dimensional raw unit vector with the *q*-th component of 1.

Finally, the flux distribution of the mutant $\mathbf{v}_j^{mutant} = (v_{1j}^{mutant}, v_{2j}^{mutant}, ..., v_{nj}^{mutant})^t$ is provided by:

$$\mathbf{v}_j^{mutant} = \mathbf{P} \cdot \mathbf{r}_j^{mutant} \quad (j = 1, 2, ..., 2m). \tag{12}$$

The ECF model does not use any kinetics to calculate the steady-state flux distributions for mutants. In this context ECF is a static or non-mechanistic model that links an enzyme activity profile to a metabolic flux distribution.

Characterization of ECF for an enzyme activity profile

To demonstrate the validity of ECF, ECF is applied to simulate the flux distribution of a mutant by integrating enzyme activity data into the EMC vector of wild type. Since a distribution occurs in the simulated flux vectors of $\{\mathbf{v}_j^{mutant} \mid j = 1, 2, ..., 2m\}$ in Eq. (12), the mean and the rel-

ative standard deviation are calculated for each flux. The mean simulated flux distribution of $\mathbf{v}_{mean}^{mutant}$ is defined by:

$$\mathbf{v}_{mean}^{mutant} = \frac{1}{2m} \sum_{j=1}^{2m} \mathbf{v}_j^{mutant} \tag{13}$$

or

$$v_{i\,mean}^{mutant} = \frac{1}{2m} \sum_{j=1}^{2m} v_{i\,j}^{mutant} \quad (i = 1, 2, ..n) \tag{14}$$

The relative model error for the *i*-th flux is defined by:

$$\text{Relative Model Error for Flux } (i) = \frac{\left| v_{i\,mean}^{mutant} - v_{i\,exp}^{mutant} \right|}{v_{i\,exp}^{mutant}} \quad (i = 1, 2, ..n) \tag{15}$$
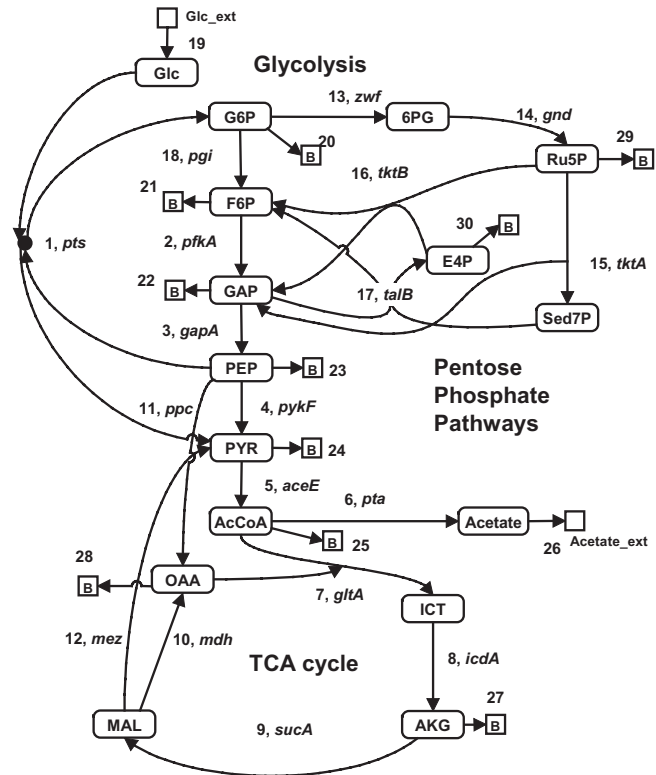

Figure 2
A schematic diagram of central metabolism for *E. coli*. The network map is made based on our previous experiments. The numbers in the figure are those of reactions shown in Table 1. The gene name is representative of the genes related to the reaction. The metabolites indicated by squares, i.e., B (biomass), Glc_ext (environmental glucose) and Acetate_ext (environmental acetate), are external. All other metabolites, enclosed by rounded rectangles, are internal.

where $v_{i\,exp}^{mutant}$ is the i-th flux for an experimental mutant. The subscript of "exp" means experimental data.

To characterize the overall accuracy of ECF, the model error, which is the difference between the estimated flux distribution and the experimental one for each optimized EM vector, is defined by:

$$\text{Model Error}(j) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(v_{ij}^{mutant} - v_{i\,exp}^{mutant})^2} \quad (j = 1, 2,..., 2m).$$

(16)

Since the model error has a distribution, the mean and the relative standard deviation are calculated. Model accuracy becomes high as the mean model error decreases.

Characterization of ECF for combinatorial enzyme activity profiles
For the metabolic systems employed in this study, enzyme activities are not given for all metabolic reactions. Thus, it raises a question: if the ECF-estimated flux distributions depend on the members of the enzymes incorporated into the model. Therefore, the flux distributions for a mutant need to be simulated for all combinations of the enzymes with measured activities.

The combination number of incorporated enzyme activities (N_c) is provided by:

$$N_c = \gamma Cx,$$

(17)

where γ is the total number of measured enzymes and x is the number of incorporated enzymes in the ECF model. For each combination the flux distributions of $\{\mathbf{v}_j^{mutant} \mid j = 1, 2,..., 2m\}$ ($j = 1, 2, ..., 2m$) are simulated. When x enzyme activities are incorporated, the number of $\{\mathbf{v}_j^{mutant} \mid j = 1, 2,..., 2m\}$ is given to:

$$L = 2m \cdot \gamma Cx.$$

(18)

Consequently, the flux distribution of a mutant is provided by:

$$\mathbf{v}_j^{mutant} = \mathbf{P} \cdot \mathbf{r}_j^{mutant} \quad (j = 1, 2,..., L),$$

(19)

The mean estimated flux of $v_{i\,mean}^{mutant}$ is defined by:

$$v_{i\,mean}^{mutant} = \frac{1}{L}\sum_{j=1}^{L} v_{ij}^{mutant} \quad (i = 1, 2,..n).$$

(20)

The model error profile with respect to each optimized EM vector is given by:

$$\text{Model Error}(j) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(v_{ij}^{mutant} - v_{i\,exp}^{mutant})^2} \quad (j = 1, 2,..., L).$$

(21)

where $v_{i\,exp}^{mutant}$ is the i-th experimental flux distribution.

Since the model error has a distribution, the mean and the relative standard deviation are calculated. Eq. (21) corresponds to Eq. (16) at $\gamma = x$.

Application example of ECF to a pykF knockout mutant model
Estimation of EMCs in wild-type
The *pykF*(-) knockout mutant in *E. coli* is employed to explain the algorithm of ECF (Figure 2 and Table 1)[11] and to demonstrate the feasibility of ECF by applying it to estimation of the flux distribution of the mutant. Elementary mode analysis generated 73 EMs from 30 reactions of the central metabolic network by using FluxAnalyzer [12]. The flux distributions in wild type and the mutant were measured (Table 2). One hundred and forty-six (= 73 × 2) sets of the EMC vectors for wild type $\mathbf{r}_j = (\gamma_{1,j},\ \gamma_{2,j},\ ...,\ \gamma_{m,j})^t$ ($j = 1, 2, ..., 146$) were optimized by Eqs. (5,6). The resulting EMC vectors showed similar spectrums, i.e., their resultant spectrums do not greatly vary with changes in optimization trials (see Additional file 2). Here we characterized the profiles for the EMC vectors. Figure 3a shows the allowable solution spectrum for each EM, where the spectrum for each EMC contains 146 data points. To explore representative values of each EMC, we calculated the mean (m) and standard deviation (σ) for the EMC spectrum and plotted the relative standard deviation (σ/m) with respect to m, as shown in Figure 3b. For small mean values of the EMCs, the relative standard deviations are large, revealing that many data points are scattered apart. In contrast, for large mean values the relative standard deviations are much smaller, showing that most data points are located around the mean.

Flux estimation in a pykF knockout mutant
In this estimation, the factor of β is set to 1, which will be validated in the later section (Validation of the unique parameter of β). Notice that a β factor of 1 does not reflect any linear relationship between flux and each enzyme activity, as shown in Eqs. (10, 11). In order for ECF to simulate the flux distribution in the central metabolic pathways of the *pykF*(-) knockout mutant, we integrated the enzyme activity data (Table 2) into the EMC vectors for wild type according to the power law formalism of Eqs. (7, 8, 9), simulating the EMC vectors for the mutant $\mathbf{r}_j^{mutant} = (\gamma_{1,j}^{mutant},\gamma_{2,j}^{mutant},...,\gamma_{m,j}^{mutant})^t$ ($j = 1, 2, ..., 146$)

Table 1: Reactions from the *E. coli* central metabolic network

Reaction	Gene name	Enzyme	Chemical reaction
1	*ptsH, ptsI* *crr, glk*	Glucose phosphotransferase system	Glc + PEP => G6P + PYR
2	*pfkA, pfkB* *fba* *tpi*	Phosphofructokinase Fructose-16-bisphosphatate aldolase Triphosphate isomerase	F6P => 2 GAP
3	*gapA* *pgk* *gpmA*	Glycelaldehyde-3-phosphate dehydrogenase Phosphoglycerate kinase Phosphoglycerate mutase I	GAP => PEP
4	*pykF* *pykA*	Pyruvate kinase I	PEP => PYR
5	*aceE* *aceF* *lpdA*	Pyruvate dehydrogenase	PYR => AcCoA
6	*pta, ackA*	Acetyl-CoA synthetase	AcCoA => Acetate
7	*gltA* *acnA, acnB*	Citrate synthase Aconitase A	AcCoA + OAA => ICT
8	*icdA*	Isocitrate dehydrogenase	ICT => AKG
9	*sucA, sucB, lpdA* *sucC, sucD* *sdhC* *fumA, fumB, fumC*	2-ketoglutarate dehydrogenase Succinyl-CoA synthetase Succinate dehydrogenase Fumarase	AKG => MAL
10	*mdh*	Malate dehydrogenase	MAL => OAA
11	*ppc*	Phosphenolpyruvate carboxylase	PEP => OAA
12	*mez*	Malic enzyme	MAL => PYR
13	*zwf*	Glucose-6-phosphate-1-dehydrogenase	G6P => 6PG
14	*gnd*	6-phosphoglycononate dehydrogenase	6PG => Ru5P
	rpiA	Ribose-5-phosphate isomerase A	Ru5P => R5P
	rpe	Ribose phosphate 3-epimerase	Ru5P => X5P
15	*tktA*	Transketolase I	X5P + R5P => GAP + Sed7P
16	*tktB*	Transketolase II	X5P + E4P => F6P +GAP
17	*talB*	Transaldolase B	GAP + Sed7P => F6P + E4P
18	*pgi*	Phosphoglucoisomerase	G6P => F6P
19		Glucose uptake	Glc_ext => Glc
20		Materials are used for biomass synthesis	G6P => Biomass
21		Materials are used for biomass synthesis	F6P => Biomass
22		Materials are used for biomass synthesis	GAP => Biomass
23		Materials are used for biomass synthesis	PEP =-> Biomass
24		Materials are used for biomass synthesis	PYR => Biomass
25		Materials are used for biomass synthesis	AcCoA => Biomass
26		Membrane transport reaction	Acetate => Acetate_ext
27		Materials are used for biomass synthesis	AKG => Biomass
28		Materials are used for biomass synthesis	OAA => Biomass
29		Materials are used for biomass synthesis	R5P => Biomass
30		Materials are used for biomass synthesis	E4P => Biomass

Abbreviation: AcCoA Acetyl-CoA, AKG α-ketoglutarate, E4P erythrose-4-phosphate, F6P fructose-1,6-bisphosphate, Glc glucose, GAP glyceraldehydes-3-phosphate, G6P glucose-6-phosphate, ICT isocitrate, Mal malate, OAA oxaloacetate, PEP phosphoenolpyruvate, 6PG 6-phosphogluconate, Ru5P ribulose-5-phosphate, X5P, xylulose 5-phosphate, R5P ribose-5-phosphate, Sed7P sedoheptulose-7-phosphate, Glc_ext Glucose external, Acetate_ext Acetate external.

by Eqs. (10, 11). The EMC vectors for the mutant were employed to simulate the flux distribution for the mutant $\mathbf{v}_j^{mutant} = (\nu_{1j}^{mutant}, \nu_{2j}^{mutant}, ..., \nu_{7j}^{mutant})^t$ ($j = 1, 2, ..., 146$). Figure 4a shows the spectrums of the estimated flux distributions in the mutant, where the mean $\mathbf{v}_{mean}^{mutant}$ and the standard deviation are plotted for each flux. The resultant solution spectrums were broad.

First, to explore a representative value for such broad spectrums, we plotted the relative standard deviation of the estimated flux with respect to the mean as shown in Figure 4b. For very small mean values of the estimated fluxes (< 2), the relative standard deviations were large, revealing that much of the flux data were scattered. In contrast, for large values the relative standard deviations were much smaller, indicating that most flux data were located around the mean. These findings suggest that the mean flux with a relatively large value can be representative for the estimated flux. To further confirm if the mean is representative, we plotted the frequency distribution with respect to flux (see Additional file 3). The frequency density is very high around the mean and only a few flux points are scattered away from the mean. Thus, the mean value is regarded as the estimated flux.

Second, to characterize the precision of ECF we compared the mean estimated fluxes with the experimental ones for

Table 2: Experimental data of flux and enzyme activities in wild type and the *pykF*(-) knockout mutant

	Flux in the wild type	Flux in the mutant	Relative enzyme activities in the *pykF*(-) mutant
1	100	100	0.57
2	83	65	0.57
3	163	151	1
4	30	1	0.14
5	107	103	1
6	20	1	0.37
7 (*gltA*)	87	81	0.45
8	87	81	1
9	78	73	1
10	75	52	1.47
11	17	44	2.3
12	3	21	3.5
13	34	79	1.7
14	34	79	1.44
15	11	25	1
16	8	22	1
17	11	25	1
18	65	20	0.59
19	100	100	1
20	1	1	1
21	1	2	1
22	11	1	1
23	16	6	1
24	26	19	1
25	1	21	1
26	19	1	1
27	9	8	1
28	5	15	1
29	4	7	1
30	3	3	1

The cells were cultivated at a dilution rate of 0.1 h^{-1}. The glucose uptake flux was normalized to 100. The relative enzyme activities in the mutant were normalized by those in the wild type. Enzymes were extracted from the cultured cells and their activities were measured *in vitro*.

the *pykF*(-) mutant. The relative model error for each flux (Eq. (15)) is plotted with respect to the experimental flux for the mutant as shown in Figure 4c. As the experimental flux of the mutant increases, the relative model error for flux decreases, showing that ECF simulates high flux values accurately. Out of 30 reactions, ECF estimates 20 fluxes with a low relative model error for flux of less than 0.5.

Third, we provide a physiological validation for the estimated flux distribution. To investigate which EMCs are enhanced in the *pykF*(-) knockout mutant, we calculated the part of Eqs. (7, 8), $\varepsilon_i = \prod_{p=1}^{n} a_{pi}{}^{\beta}$, where β is 1 and ε_t is 1 for wild-type. The ε_t values of the EMs that contain the reactions of 11 (*ppc*) and 12 (*mez*) are greater than 1, whereas the ε_t values of the EMs that do not have them are
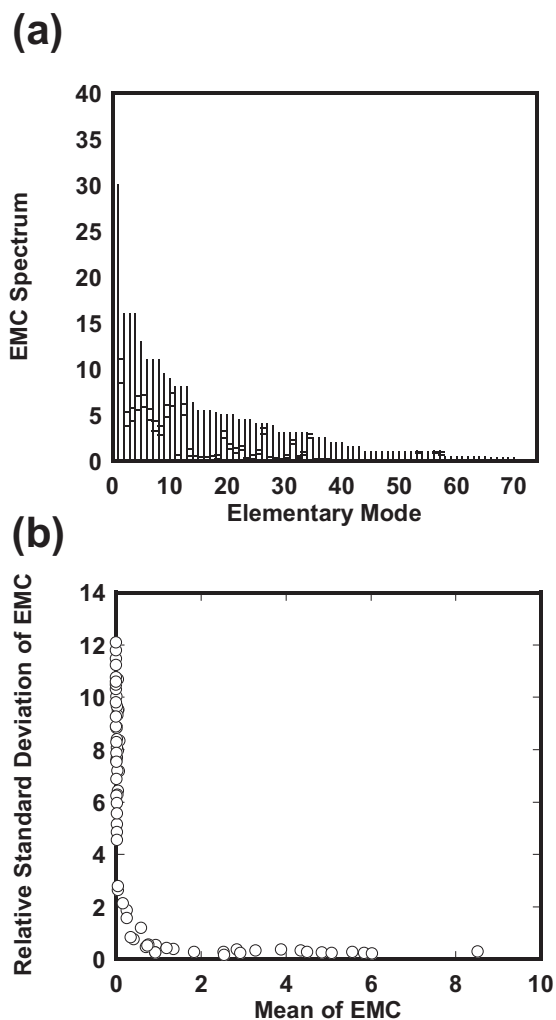

Figure 3
Estimated EMCs in the *pykF*(-) knockout mutant. (a) Spectrum for the EMCs with respect to elementary modes. The EMCs with respect to the (*2j-1* and *2j*)-th optimization trial are calculated by maximizing or minimizing the *j*-th EMC (*j* = *1*, *2*, ..., *73*). EMs are sorted according to their maximum values. The vertical lines indicate the allowable ranges for the EMCs. Each line has two transverse dashes at the right: The lower one indicates the mean (*m*) and the upper one is the sum of the mean and the standard deviation (σ). (b) Relative standard deviation of the estimated EMC spectrum. The relative standard deviations (σ/m) are plotted with respect to the means (*m*). Note that the mean EMCs are not zero.

less than 1. These show the fluxes that contain *ppc* and *mez* are enhanced, which are consistent with the fact that the *pykF*-knockout-mediated blockage of the PEP to PYR pathway greatly activates both *ppc* and *mez* enzymes to supply PYR through alternative pathways[11].
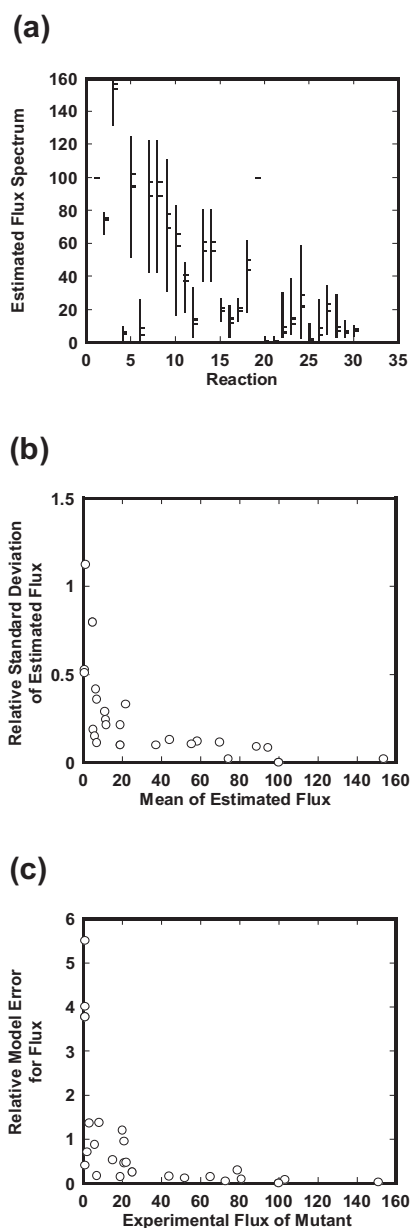
(a)

(b)

(c)

Figure 4
Characterization of each estimated flux in the *pykF*(-) knockout mutant. (a) Spectrum for the ECF-simulated metabolic flux in central metabolic pathways of the *pykF*(-) mutant. One hundred forty-six sets of EMCs optimized for wild-type were employed to estimate the flux distribution profiles. The vertical lines indicate the allowable ranges for the estimated flux. Each line has two transverse dashes at the right: The lower one indicates the mean (*m*) and the upper one is the sum of the mean and the standard deviation (σ). (b) Relative standard deviations for the estimated flux spectrum. The relative standard deviations (σ/*m*) are plotted with respect to the means (*m*). **(c)** Relative model error for flux. The relative model error for each flux is plotted with respect to the experimental flux for the mutant.

Effect of the number of integrated enzyme activities
To investigate how the number of integrated enzyme activities affects the estimated fluxes in the *pykF*(-) mutant, we calculated the model error profile defined by Eq. (21) for all combinations of the enzymes as shown in Figure 5. The spectrum for the model error is described with respect to the number of integrated enzyme activities. The allowable ranges of the model error were broad, but the mean model error for all the enzyme combinations decreased as the number of integrated enzymes increased. To characterize the relationships between the model error and the number of integrated enzymes, we calculated the standard deviation for each number of integrated enzymes and plotted the frequency distribution with respect to the model error (see Additional file 4). The standard deviations were less than the means and the frequency distribution indicated a high density around the mean. The mean can be regarded as an indicator of the model error profile. The model error decreased with an increase in the number of integrated enzymes, indicating that enzyme data is effectively used by ECF.

We investigated the mechanism of how the allowable model error showed a broad range. We compared the combinations for the incorporated enzymes between the mutants showing underestimated and overestimated model errors, clarifying that the members of the incorporated enzymes and the reaction 7 (*gltA*) are critical. The model error was relatively small when multiple enzymes responsible for branching reactions were incorporated together (data not shown). In contrast, when no enzyme at the branching reactions was integrated, the model error was relatively large. These findings suggest that incorporation of branching reactions can be effective in enhancing estimation accuracy, although it does not change the model error remarkably.

Next, we investigated the relationship between enzyme activities and their associated fluxes. While the flux of the *gltA* reaction in the mutant was almost the same as that in wild-type, the enzyme activity for the mutant was less than a half of that for wild-type. The mutants that incorporate *gltA* show relatively large errors, whereas the mutants without it indicate small errors. Thus, the incorporation of *gltA* is suggested to cause the model error to increase. To confirm it, we excluded the enzyme activity of *gltA* and simulated the model error with respect to the number of integrated enzymes. The allowable ranges of the model error became narrow significantly (data not shown). In the case of *gltA* being integrated, some metabolic factors, such as changes in the concentration of AcCoA, may be involved to compensate the decreased *gltA* activity.
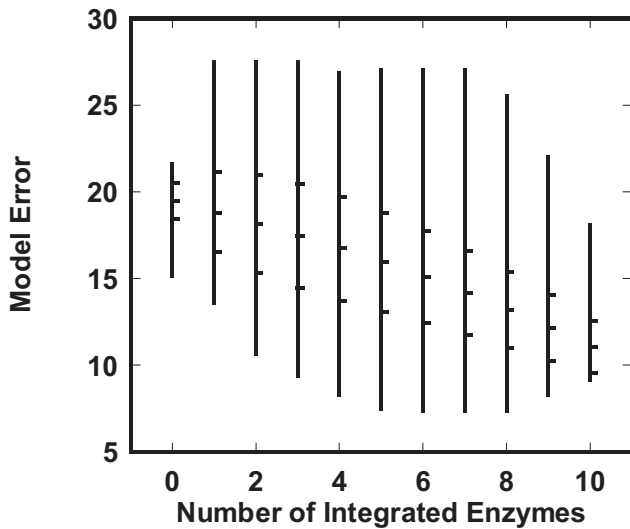
Figure 5
Effect of the number of integrated enzyme activities on the model error in the *pykF*(-) knockout mutant. The model error was calculated for all combinations of the enzymes with measured activities. The vertical lines indicate the regions between the minimum and maximum values. Each line has three transverse dashes at the right: The middle dash is the mean of the model error. The upper and lower dashes indicate $m + \sigma$ and $m - \sigma$, respectively (m and σ are the mean and the standard deviation of the model error, respectively).

Validation of the unique parameter of β
Here, we demonstrate that a β factor of 1 is a best choice. As shown in Figure 6, we investigated how a change in the β factor affected the mean and relative standard deviation of the model error profile of Eq. (21). The β factor determines the degree by which the change in an enzyme profile affects the flux distribution for a mutant. A β factor of 0 indicates that no enzyme data is used by ECF (Eqs. (7, 8)). In a factor range from 0.5 to 2, the mean model error decreases as the number of incorporated enzymes increases (Figure 6a). At a factor value of 4, the mean model error does not decrease monotonically with respect to the number of incorporated enzymes and the mean model error becomes larger than any other value of β. The mean model error becomes very low at a β factor of 1. In Figure 6b, the relative standard deviation increases as the β factor increases. Since accurate estimation requires that both the mean and the relative standard deviation are small, a β factor of 1 is a reasonable value.

Generality and applicability of ECF
To guarantee the generality and applicability of ECF, we demonstrate that (i) a β factor of 1 is consistent for available mutants, and (ii) the mean model error decreases with an increase in the number of integrated enzymes, by
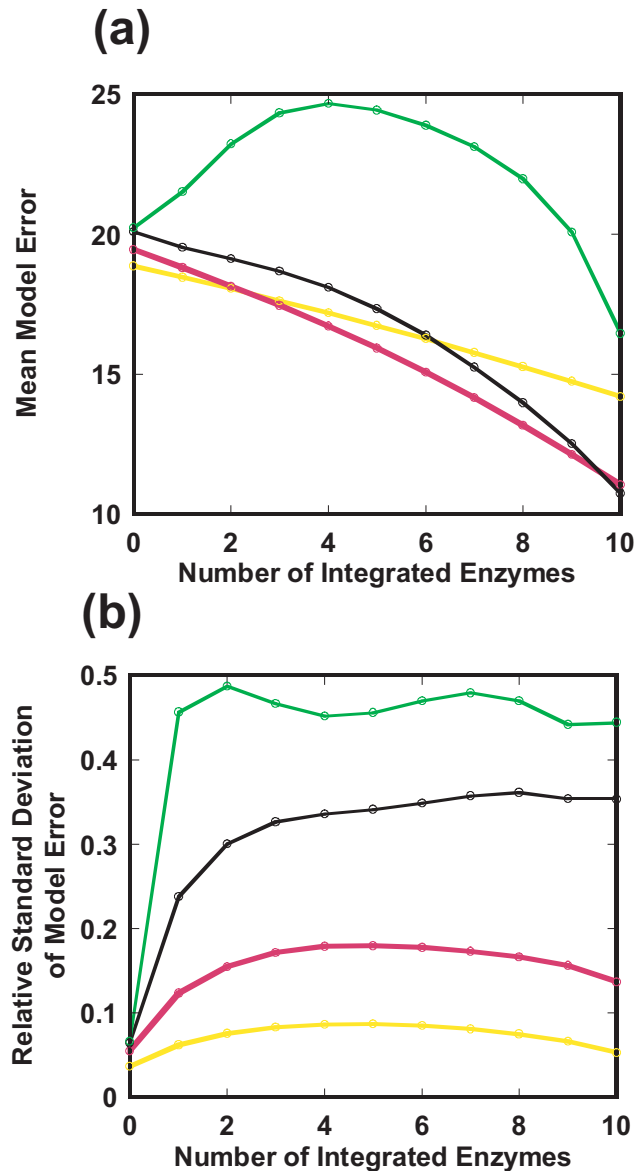

Figure 6
Optimization of the β factor in the *pykF* (-) knockout mutant. The model error profile is simulated with respect to the number of integrated enzymes when the β factor is changed as 0.5 (yellow), 1 (red), 2 (black), 4 (green). (a) Mean model error. (b) Relative standard deviation of the model error. The mean for each β is not consistent when the number of incorporated enzymes is zero. This is caused by a non-zero enzyme activity for *pykF* in the mutant.

using *E. coli ppc*[13], *cra*, *fnr* (data not shown), *gnd*[14], *pgi*[15], and *zwf*[14,16] knockout mutants, and an *als*-overexpressing and *pta* (-) knockout mutant of *Bacillus subtilis* (see Additional file 5)[17].

As shown in Figure 7, we investigated how a change in the β factor affected the mean and relative standard deviation of the model error of Eq. (21). The model accuracy for the *pgi*, *gnd*, and *zwf* knockout mutants was not significant due to the small number of incorporated enzymes ($n \le 4$) (data not shown). However, for a relatively large number ($n \ge 6$) of integrated enzymes, the mean model error of the *ppc*, *cra*, and *fnr* knockout mutants and the *als*-overexpressing and *pta* (-) knockout mutant decreased greatly. For these mutants a β factor of 1 is a best choice, which minimizes the mean model error with a relatively small value of the relative standard deviation. Furthermore, the mean model error for a β factor of 1 clearly decreases with an increase in the number of integrated enzymes, indicating that an enzyme activity profile is effectively used by ECF. The comparison between the estimated and experimental fluxes for each reaction is shown in Figure 8, where all available enzyme data are used by ECF and a β factor of 1 is employed. A small value of the relative model error for flux means consistency with experimental data. For a relatively large value of experimental flux, the simulated flux by ECF is rather consistent with the experimental one of these mutants as well as that of the *pykF* knockout mutant (Figure 4c).

Conclusion

It is important to explore some relationships between an enzyme activity profile and a metabolic flux distribution for rationally designing organism production systems or for understanding the physiological dynamics within a cell. The ECF model is a non-mechanistic model to mathematically link enzyme data to metabolic flux. Assuming that the flux through each EM is synergistically affected by all enzyme activities that belong to the EM, the power-law formula is used to integrate all enzyme activities into EMA. The ECF model successfully links enzyme activities to the flux distributions of *E. coli* and *B. subtilis* mutants, indicating that many enzyme activities affect the flux through each EM simultaneously. In other words, the ECF model is highly applicable to the metabolic models whose flux distribution would be determined by the effects of multiple enzyme activities rather than by a few rate-limiting reactions. In addition, the incorporation of branching reactions is suggested to play a significant role in enhancing model accuracy.

In the power-law formula the β factor determines the degree by which the change in an enzyme profile affects the flux distribution for a mutant. A β factor from 0.5 to 1 is effective in decreasing the model error for all available mutants. A β factor of 1 is a best choice in all experimental mutants, but it does not mean that the change in an elementary mode flux is linear to an enzyme activity because the relationship between a flux and enzymes is complicated than expected from a β factor of 1 as shown in Eqs.

(10, 11). A β factor of 1 is not the exact optimal value but an approximate or representative one. The important thing would be to notice that the ECF model decreases the model error at an appropriate value of β rather than to insist biological basis of a β factor of 1.

In the metabolic engineering field, powerful analyses had been proposed to simulate the change in flux distributions in response to environmental or genetic changes. For example, network rigidity had extensively been investigated to reveal how the rigidity of some principal nodes at branch reactions is generated by complex enzyme regulations including allosteric regulation[18]. Such a study reduces the analysis of large-scale metabolic networks to that of the principal nodes, intensively analyzing local kinetics around them to improve flux distributions. The principal nodes showing strong rigidity play a major role in determining the flux distribution in response to different stimulus. On the other hand, the ECF model adopts a different way from such principal node analysis. The ECF model neither considers any allosteric kinetics nor reduces large-scale network analysis to local one. Despite such a plain idea, ECF predicts how the change in enzyme profiles affects the flux distribution. This indicates that the change in an enzyme profile plays a major role in determining flux distributions or it rather reflects the change in the flux distribution.

Finally we show two limitations of ECF. One is that metabolic control analysis (MCA) is not applied to analysis for the ECF model. The other is the limitation of network size. Generally MCA is used to quantify flux control of each reaction by estimating the logarithmic gain of a flux with respect to changes in an enzyme activity [19-21]. In the ECF model, the power factor in Eqs. (7, 8) seems to be a logarithmic gain of the flux through an EM with respect to an enzyme activity. However, MCA may not be applied to the ECF model, because ECF is a non-mechanistic model that neither considers detailed enzyme kinetics, such as allosteric binding of inhibitors and activators, nor the concentrations of substrates. The size of networks analyzed by ECF is limited. Since the number of EMs increases exponentially with the network size, it is hard to analyze large-scale networks with more than hundreds of reactions. A few methods, which divide the network into subsystems by redefining internal and external metabolites [22] or improve the algorithm deriving EMs from a stoichiometric matrix[23], have been proposed to reduce calculation complexity, but they have not fully established yet.

Methods
Metabolic model
For more than a decade, the metabolism of *E. coli* has served as a testing ground for network analysis methods.
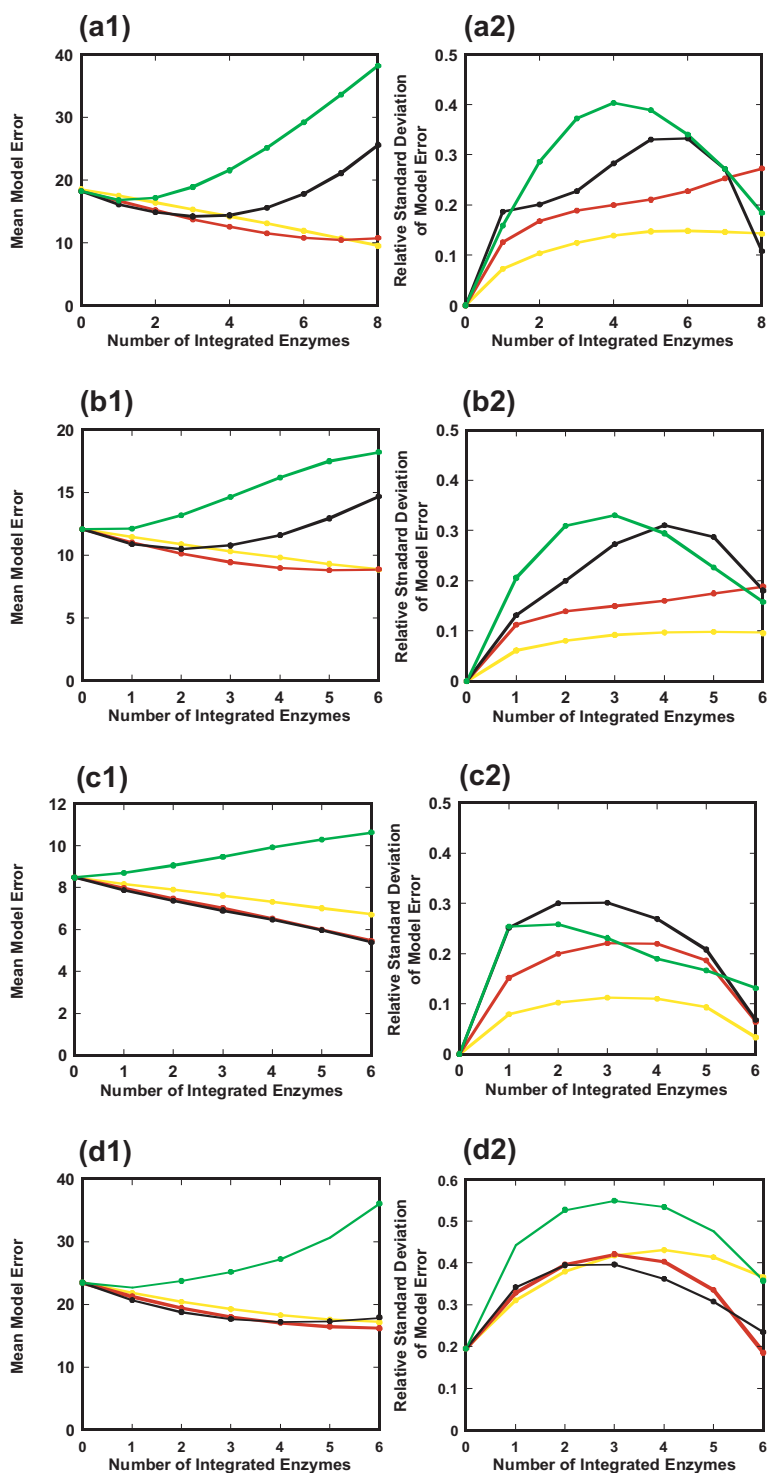
Figure 7
Applicability and generality of ECF. Effect of the number of integrated enzymes and the β factor on the model error profile is investigated for various mutants to demonstrate the applicability of the ECF model with a β factor of 1. (a) *ppc*(-) knockout mutant, (b) *cra*(-)knockout mutant, (c) *fnr*(-) knockout mutant in *E. coli*. (d) *als*-overexpressing and *pta* (-) knockout mutant in *B. subtilis*. The mean (a1, b1, c1, d1) and relative standard deviation (a2, b2, c2, d2) of the model error are simulated with respect to the number of integrated enzymes when the β factor is changed as 0.5 (yellow), 1 (red), 2 (black), 4 (green).
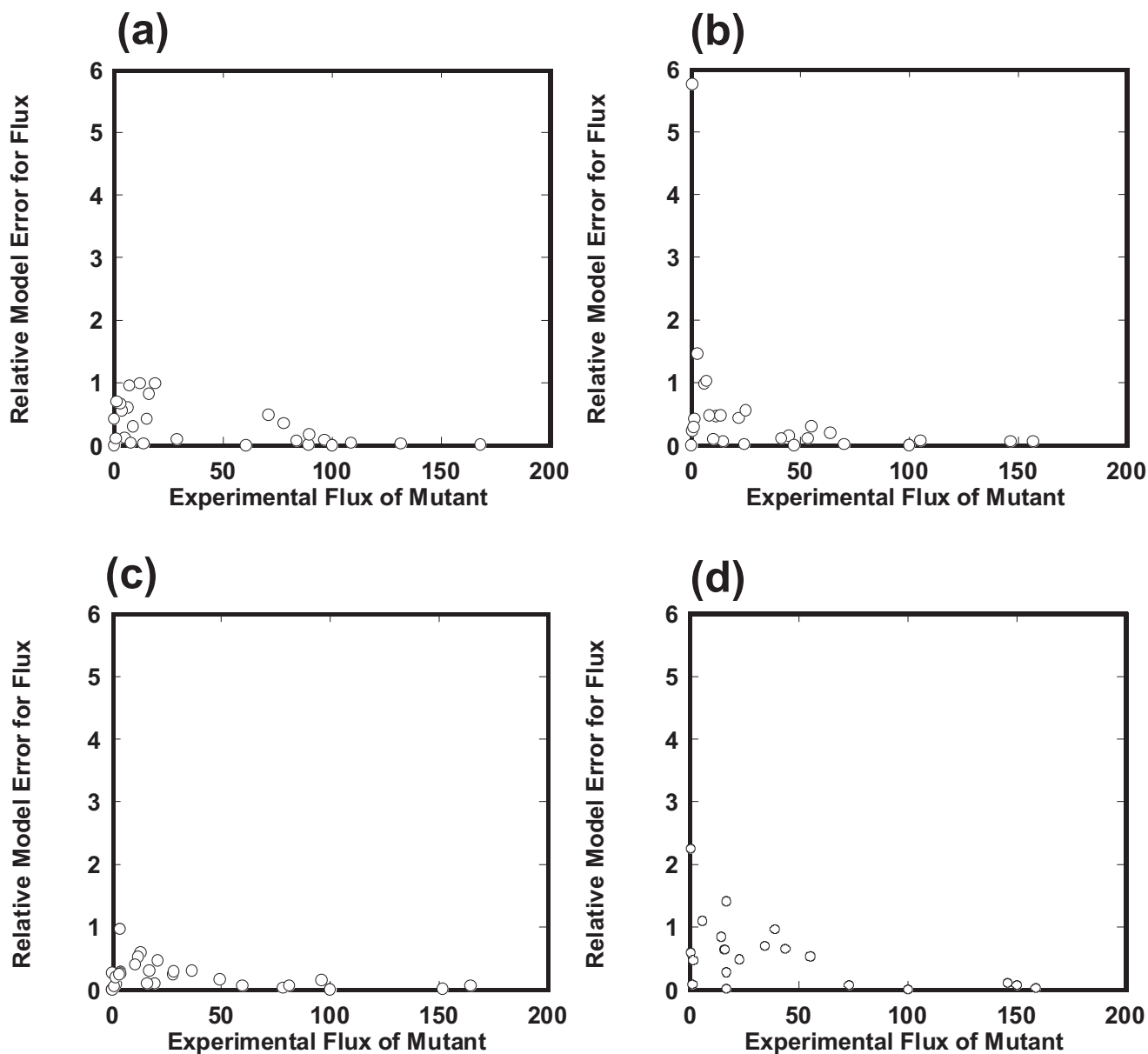
Figure 8
Consistency of the estimated flux with experimental one for each reaction. We compared the estimated flux with experimental data for each reaction: (a) *ppc* (-) knockout mutant, (b) *cra* (-) knockout mutant, and (c) *fnr* (-) knockout mutant in *E. coli*. (d) *als*-overexpressing and *pta* (-) knockout mutant of *B. subtilis*. A β factor of 1 is used. The relative model error for flux is plotted with respect to its experimental flux. A small value of the relative model error for flux indicates that the simulated flux is consistent with the experimental one.

The studies of this metabolic network have been facilitated by the extensive availability of *E. coli* biochemical data. In the present analysis, we simplified the central metabolic networks of glycolysis, the tricarboxylic acid cycle (TCA), and the pentose phosphate pathway (Figure 2 and Table 1), which together contain 16 metabolites and 30 reactions. The strains used were wild-type (*E. coli* K-12) and the knockout mutants (*pykF*, *ppc*, *cra*, *fnr*, *gnd*, *pgi*, and *zwf*), which were constructed by deletion of the corresponding gene from *E. coli* derivative BW25113[24]. In these mutants, the internal metabolic flux had been estimated using results from ^{13}C-labeled glucose experiments, and the key enzymes involving the branching reactions (G6P, PYR, AcCoA, OAA, etc) were measured in the main pathway of the TCA cycle and glycolysis [11,13-16].

Implementation
CADLIVE, which is freely available from http://www.cadlive.jp, is employed to draw the *E. coli* metabolic network map and describe its corresponding reactions in XML format (SANAC)[25,26]. The generated XML file is converted into a FluxAnalyzer readable format[12]. FluxAnalyzer identified the stoichiometric matrix and elementary modes. The programs for optimization and ECF are written in Matlab (Mathworks Inc., Natick, MA).

Authors' contributions
HK and KS designed the research. HK, ZQ, and RO performed the simulations and analyzed the results. All authors have read and approved the final manuscript.

Additional material

Additional file 1
A plain instruction for ECF. Supplementary text.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1752-0509-1-31-S1.pdf]

Additional file 2
Supplementary figure 1. Estimation of 73 EMCs in central metabolic pathways of wild type.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1752-0509-1-31-S2.pdf]

Additional file 3
Supplementary figure 2. Frequency distributions for the estimated flux in the pykF*(-) knockout mutant.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1752-0509-1-31-S3.pdf]

Additional file 4
Supplementary figure 3. Frequency distributions for the model error in the pykF*(-) knockout mutant.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1752-0509-1-31-S4.pdf]

Additional file 5
Supplementary figure 4 and tables 1 and 2. Data for the B. subtilis *metabolic network.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1752-0509-1-31-S5.pdf]

References
1. Borodina I, Nielsen J: **From genomes to in silico cells via metabolic networks.** *Curr Opin Biotechnol* 2005, **16**:350-355.
2. Wiback SJ, Mahadevan R, Palsson BO: **Using metabolic flux data to further constrain the metabolic solution space and predict internal flux patterns: the Escherichia coli spectrum.** *Biotechnol Bioeng* 2004, **86**:317-331.
3. Price ND, Reed JL, Palsson BO: **Genome-scale models of microbial cells: evaluating the consequences of constraints.** *Nat Rev Microbiol* 2004, **2**:886-897.
4. Papin JA, Stelling J, Price ND, Klamt S, Schuster S, Palsson BO: **Comparison of network-based pathway analysis methods.** *Trends Biotechnol* 2004, **22**:400-405.
5. Schuster S, Dandekar T, Fell DA: **Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering.** *Trends Biotechnol* 1999, **17**:53-60.
6. Schilling CH, Letscher D, Palsson BO: **Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective.** *J Theor Biol* 2000, **203**:229-248.
7. Akesson M, Forster J, Nielsen J: **Integration of gene expression data into genome-scale metabolic models.** *Metab Eng* 2004, **6**:285-293.
8. Schwarz R, Musch P, von Kamp A, Engels B, Schirmer H, Schuster S, Dandekar T: **YANA - a software tool for analyzing flux modes, gene-expression and enzyme activities.** *Bmc Bioinformatics* 2005, **6**:135.
9. Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED: **Metabolic network structure determines key aspects of functionality and regulation.** *Nature* 2002, **420**:190-193.
10. Wiback SJ, Mahadevan R, Palsson BO: **Reconstructing metabolic flux vectors from extreme pathways: defining the alpha-spectrum.** *J Theor Biol* 2003, **224**:313-324.
11. Siddiquee KA, Arauzo-Bravo MJ, Shimizu K: **Effect of a pyruvate kinase (pykF-gene) knockout mutation on the control of gene expression and metabolic fluxes in Escherichia coli.** *FEMS Microbiol Lett* 2004, **235**:25-33.
12. Klamt S, Stelling J, Ginkel M, Gilles ED: **FluxAnalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps.** *Bioinformatics* 2003, **19**:261-269.
13. Peng L, Arauzo-Bravo MJ, Shimizu K: **Metabolic flux analysis for a ppc mutant Escherichia coli based on 13C-labelling experiments together with enzyme activity assays and intracellular metabolite measurements.** *FEMS Microbiol Lett* 2004, **235**:17-23.
14. Zhao J, Baba T, Mori H, Shimizu K: **Global metabolic response of Escherichia coli to gnd or zwf gene-knockout, based on 13C-labeling experiments and the measurement of enzyme activities.** *Appl Microbiol Biotechnol* 2004, **64**:91-98.
15. Kabir MM, Shimizu K: **Gene expression patterns for metabolic pathway in pgi knockout Escherichia coli with and without phb genes based on RT-PCR.** *J Biotechnol* 2003, **105**:11-31.
16. Hua Q, Yang C, Baba T, Mori H, Shimizu K: **Responses of the central metabolism in Escherichia coli to phosphoglucose iso-**

merase and glucose-6-phosphate dehydrogenase knockouts. *J Bacteriol* 2003, **185:**7053-7067.

17. Zhu Y, Chen X, Chen T, Zhao X: **Enhancement of riboflavin production by overexpression of acetolactate synthase in a pta mutant of Bacillus subtilis.** *FEMS Microbiol Lett* 2007, **266:**224-230.

18. Stephanopoulos G, Vallino JJ: **Network rigidity and metabolic engineering in metabolite overproduction.** *Science* 1991, **252:**1675-1681.

19. Kacer H, Burns JA: **The control of flux.** *Symp Soc Exp Biol* 1973, **27:**65-104.

20. Heinrich R, Rapoport TA: **A linear steady state treatment of enzymatic chains. General properties, control and effector strength.** *Eur J Biochem* 1974, **42:**89-95.

21. Stephanopoulos GN, Aristos AA, Nielsen J: **Metabolic Engineering Principles and Methodologies.** San Diego, Academic Press; 1998.

22. Schuster S, Pfeiffer T, Moldenhauer F, Koch I, Dandekar T: **Exploring the pathway structure of metabolism: decomposition into subnetworks and application to Mycoplasma pneumoniae.** *Bioinformatics* 2002, **18:**351-361.

23. Urbanczik R, Wagner C: **An improved algorithm for stoichiometric network analysis: theory and applications.** *Bioinformatics* 2005, **21:**1203-1210.

24. Datsenko KA, Wanner BL: **One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.** *Proc Natl Acad Sci U S A* 2000, **97:**6640-6645.

25. Kurata H, Masaki K, Sumida Y, Iwasaki R: **CADLIVE Dynamic Simulator: Direct Link of Biochemical Networks to Dynamic Models.** *Genome Res* 2005, **15:**590-600.

26. Kurata H, Matoba N, Shimizu N: **CADLIVE for constructing a large-scale biochemical network based on a simulation-directed notation and its application to yeast cell cycle.** *Nucleic Acids Res* 2003, **31:**4071-4084.