Journal of Bioscience and Bioengineering
VOL. 110 No. 2, 254–261, 2010

www.elsevier.com/locate/jbiosc

Use of maximum entropy principle with Lagrange multipliers extends the feasibility of elementary mode analysis

Quanyu Zhao, and Hiroyuki Kurata*

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

Received 8 December 2009; accepted 12 January 2010
Available online 6 February 2010

Elementary mode (EM) analysis is potentially effective in integrating transcriptome or proteome data into metabolic network analyses and in exploring the mechanism of how phenotypic or metabolic flux distribution is changed with respect to environmental and genetic perturbations. The EM coefficients (EMCs) indicate the quantitative contribution of their associated EMs and can be estimated by maximizing Shannon's entropy as a general objective function in our previous study, but the use of EMCs is still restricted to a relatively small-scale networks. We propose a fast and universal method that optimizes hundreds of thousands of EMCs under the constraint of the Maximum entropy principle (MEP). Lagrange multipliers (LMs) are applied to maximize the Shannon's entropy-based objective function, analytically solving each EMC as the function of LMs. Consequently, the number of such search variables, the EMC number, is dramatically reduced to the reaction number. To demonstrate the feasibility of the MEP with Lagrange multipliers (MEPLM), it is coupled with enzyme control flux (ECF) to predict the flux distributions of *Escherichia coli* and *Saccharomyces cerevisiae* for different conditions (gene deletion, adaptive evolution, temperature, and dilution rate) and to provide a quantitative understanding of how metabolic or physiological states are changed in response to these genetic or environmental perturbations at the elementary mode level. It is shown that the ECF-based method is a feasible framework for the prediction of metabolic flux distribution by integrating enzyme activity data into EMs to genetic and environmental perturbations.

[**Key words:** Elementary mode; Maximum entropy principle; Lagrange multiplier; Genetic and environmental perturbation; Enzyme control flux; Systems biology]

Metabolic engineering is successfully applied to the synthesis and analysis for productions of biofuels, pharmaceuticals, and other products (1). In silico simulation is an important tool. Flux Balance Analysis (FBA) (2) could predict physiological behaviors in genome scale for *Escherichia coli*, while the employed objective function including maximum molar yield (biomass or other) was not always suitable (3). A new modeling framework for metabolic networks is required for integration of the experimental data from genomics, transcriptomics, proteomics, metabolomics, and fluxomics, which are determined by high-throughput technologies.

A few methods have been proposed for integration of omics data into constraint-based flux analysis (4–6). A problem common to those studies is that transcriptional regulations or gene expressions are given as the Boolean logic. It shows only two states for genes, expressed or not, while gene expression profiles are changed in a large range with respect to genetic or environmental perturbations.

Elementary mode (EM) analysis is potentially effective in correlating transcriptome or proteome data to their associated metabolic network architecture or flux distributions (7–10). To establish a linkage between gene expression profile and metabolic network structure, several approaches were proposed based on elementary modes (EMs), which are all of the possible and non-decomposable pathways in a steady-state biochemical network (11, 12). Control Effective Flux (CEF) was developed to predict the change in transcriptional regulations (7). A modified CEF (mCEF) was presented to predict how gene expression profiles are changed with respect to various types of genetic modifications (10). Enzyme control flux (ECF) was proposed to integrate enzyme activity data into EMs in a multiplication formula for estimating the flux distributions of genetic mutants of *E. coli* and *Bacillus subtilis* (8). ECF is currently a promising algorithm that quantitatively correlates gene or protein expression profiles to their associated metabolic flux distributions.

In ECF, maximum entropy principle (MEP) was employed to optimize the EMCs (9). MEP is a universal principle established based on Shannon entropy (13) when insufficient information is available. A problem for EM-based analyses is calculation complexity, which makes it difficult to estimate EMCs for a moderate or large-scale metabolic model (9, 14). To obtain reliable EMCs, we propose the MEP algorithm coupled with Lagrange Multipliers (LMs), which is named MEPLM. MEPLM readily optimizes hundreds of thousands of the EMCs in large-scale networks under different types of environmental and genetic perturbations, such as temperature (15), dilution rates (16),

All of additional files including the programs are freely available on our homepage: http://www.cadlive.jp/JBB/suppl.htm.

* Corresponding author. Tel./fax: +81 948 29 7828.
E-mail address: kurata@bio.kyutech.ac.jp (H. Kurata).

1389-1723/$ - see front matter © 2010, The Society for Biotechnology, Japan. All rights reserved.
doi:10.1016/j.jbiosc.2010.01.015

oxygen supply, osmotic pressure, substrate concentrations, gene deletion or addition, and partially deficiency or overexpression of enzymes. The feasibility of MEPLM is demonstrated by applying it to ECF-based flux estimation and to EM-based analysis of physiological states.

MATERIALS AND METHODS

Metabolic network models The metabolic network models in detail are shown in Table 1. The metabolic network models of *E. coli* (Figure S1) were revised from the model registered in CellNetAnalyzer (Table S1). Model I includes reaction 1–6, 8, 11, 13–35, 38–50, 52–100, 103–159 (the reaction number is shown in Table S1); model II includes reaction 1–100, 103, 104, and 107–159; model III includes reaction 1–104 and 107–159; model IV includes reaction 1–5, 7– 104, and 107–159. There are 106 reactions and 136,086 EMs in the metabolic network of *Saccharomyces cerevisiae*, including central carbon metabolism with amino acid syntheses, as shown in Table S2 and Figure S2. The flux distributions and enzyme activities data for *S. cerevisiae* were obtained from Tai et al. (15). The experimental flux distributions for *E. coli* were determined by ^{13}C tracer experiments (16–19).

Maximum entropy principle algorithm for evaluation of EMCs Generally, the flux distribution at steady state can be decomposed onto EMs (12):

$$\mathbf{P}_d \cdot \lambda = \mathbf{v}_d. \tag{1}$$

$\mathbf{P}_d$ is the sub-matrix of EM matrix $\mathbf{P}$ in which the rows represent the reactions with the determined fluxes and the columns correspond to the elementary modes. λ is the EMC vector and $\mathbf{v}_d$ is the flux vector for the determined reactions.

In our previous studies (9), the probability of EM was presented as.

$$\rho_i = \frac{1}{v_{\text{substrate uptake}}} p_{\text{substrate uptake},\,i} \cdot \lambda_i \left(\sum_{i=1}^{n} \rho_i = 1 \right), \tag{2}$$

where $v_{\text{substrateuptake}}$ is the flux for substrate uptake, $p_{\text{substrateuptake},i}$ is the element of the ith EM, n is the number of EMs. It assumes that contribution of the internal loops ($p_{\text{substrateuptake},i} = 0$) is neglected based on loop law thermodynamic constraints (20). In the employed metabolic network model of *E. coli*, the internal loop has two reactions, *sdh* and *frd*. For *S. cerevisiae*, there are two internal loops: one includes *mdh*, *mdh2*, and shuttlex, and another one is composed of *osm* and *sdh*. ρ_i is provided by solving the following optimization problem:

$$\text{Maximize} \; -\sum_{i=1}^{n} \rho_i \log \rho_i, \tag{3}$$

$$s.t. \sum_{i=1}^{n} \rho_i = 1 \tag{4}$$

$$\sum_{i=1}^{n} \rho_i x_{r,\,i} = v_r (r = 1, 2, \ldots, m), \tag{5}$$

where v_r is the rth determined flux; m is the number of the determined fluxes. $\mathbf{X}(x_{ij})$ is the new matrix converted from EM matrix $\mathbf{P}_d$, given by:

$$x_{r,\,i} = \begin{cases} \dfrac{v_{\text{substrate uptake}}}{p_{\text{substrate uptake},\,i}} p_{r,\,i} & (if\ p_{\text{substrate uptake},\,i} \neq 0) \\ 0 & (if\ p_{\text{substrate uptake},\,i} = 0) \end{cases} \tag{6}$$

The Shannon's entropy (Eq. (3)) should be maximized to provide a most probable distribution of ρ_i under the constraints (Eqs. (4) and (5)). A problem of this optimization is that it is hard to estimate ρ_i at a large-scale network model due to a huge number of n.

MEP coupled with the Lagrange multipliers In this study, to overcome the above limitation, the method of Lagrange multipliers is proposed for converting the optimization problem (Eqs. (3)–(6)), whereby the number of the search variables (EMCs) (n) is greatly reduced to that of the determined fluxes (m). The Lagrange function is provided by:

$$F(\rho_1, .., \rho_n, \varphi_0, .., \varphi_m) = -\sum_{i=1}^{n} \rho_i \log \rho_i - \varphi_0 \left(\sum_{i=1}^{n} \rho_i - 1 \right) - \sum_{r=1}^{m} \varphi_r \left(\sum_{i=1}^{n} \rho_i x_{r,\,i} - v_r \right) \tag{7}$$

$$\frac{\partial F(\rho_1, .., \rho_n, \varphi_0, .., \varphi_n)}{\partial \rho_i} = -1 - \log \rho_i - \varphi_0 - \sum_{r=1}^{m} \varphi_r x_{r,\,i} = 0 (i = 1, .., n), \tag{8}$$

where φ_i ($i = 0,1,2\ldots,m$) is the Lagrange multiplier for constraints. If we let $\varphi_0 = \log Z - 1$, then

$$\rho_i = \frac{\exp\left(-\sum_{r=1}^{m} \varphi_r x_{r,i} \right)}{Z} \quad (i = 1, 2, \ldots, n) \tag{9}$$

$\sum_{i=1}^{n} \rho_i = 1$ so

$$Z(\varphi) = \sum_{i=1}^{n} \exp\left(-\sum_{r=1}^{m} \varphi_r x_{r,i} \right) \tag{10}$$

$$\text{Then}, \frac{\sum_{i=1}^{n} x_{r,\,i} \exp\left(-\sum_{r=1}^{m} \varphi_r x_{r,\,i} \right)}{\sum_{i=1}^{n} \exp\left(-\sum_{r=1}^{m} \varphi_r x_{r,\,i} \right)} - v_r = 0 \quad (r = 1, 2, \ldots, m) \tag{11}$$

The nonlinear equation (11) for φ could be solved by mmfsolve in Matlab (21). The probabilities (ρ_i) and EMCs (λ_i) of the ith EMs are calculated as follows:

$$\rho_i = \frac{\exp\left(-\sum_{r=1}^{m} \varphi_r x_{r,\,i} \right)}{\sum_{i=1}^{n} \exp\left(-\sum_{r=1}^{m} \varphi_r x_{r,\,i} \right)} \quad (i = 1, 2, \ldots, n) \tag{12}$$

$$\lambda_i = \begin{cases} \dfrac{v_{\text{substrate uptake}}}{p_{\text{substrate uptake},\,i}} \cdot \rho_i & (\text{if } p_{\text{substrate uptake},\,i} \neq 0) \\ 0 & (\text{if } p_{\text{substrate uptake},\,i} = 0) \end{cases}. \tag{13}$$

Enzyme control flux coupled with maximum entropy principle with Lagrange multiplier method (ECF-MEPLM) In enzyme control flux (ECF), the steady-state flux distribution of a reference type is defined by:

$$\mathbf{v}^{\text{ref}} = \mathbf{P} \cdot \lambda^{\text{ref}}, \tag{14}$$

where the element of λ^{ref} are the EMCs of a reference model, optimized under the MEP constraint. Here, ECF calculates the EMCs of a target model. The EMC of the ith EM for a target model $\lambda_i^{\text{target}}$ is presented by:

$$\lambda_i^{\text{target}} = \gamma \cdot \lambda_i^{\text{ref}} \prod_{j=1}^{mr} a_{j,i}, \tag{15}$$

where λ_i^{ref} is the EMC of the ith EM for the reference model; $a_{j,i}$ is the parameter for the enzyme activity of the jth reaction; mr is the number of reactions in metabolic models; γ is a parameter to adjust the flux of the substrate uptake reaction of the target to the determined value for the reference type. The enzyme activity parameter is defined as:

$$a_{j,\,i} = \begin{cases} a_j & (\text{if } p_{j,\,i} \neq 0) \\ 1 & (\text{if } p_{j,\,i} = 0) \end{cases} \tag{16}$$

a_j is the relative enzyme activity of the target to the reference model for the jth reaction, which is the experimental data. $p_{j,i}$ is the element for the jth reaction and ith EM in EM matrix $\mathbf{P}$. The flux distribution of the target model is predicted by:

$$v^{\text{target}} = P \cdot \lambda^{\text{target}}. \tag{17}$$

Detailed explanation is described elsewhere (8).

TABLE 1. Details for metabolic network models for *E. coli* (I, II, III, and IV) and *S. cerevisiae* (V).

Model	I	II	III	IV	V
O_2	Aerobic	Anaerobic	Anaerobic	Anaerobic	Aerobic
Substrates	Glucose	Glucose	Glucose	Glucose	Glucose
Products	Acetate, CO_2	Acetate, ethanol, succinate, formate, lactate, CO_2	Acetate, ethanol, succinate, glycerol, formate, lactate, CO_2	Acetate, ethanol, succinate, glycerol, formate, lactate, CO_2	Acetate, ethanol, glycerol, CO_2
Total number of reactions	149	155	157	156	106
EMs	30579	98338	321416	122126	136086
Calculation time (s)	48	264	1674	417	32[a]

[a] The determined fluxes are limited, so the optimization is performed quickly.

TABLE 2. Speed and accuracy for the calculation of flux distributions for *S. cerevisiae*.

Specific growth rate	Running time (s)		Calculation accuracy	
	MEP	MEPLM	MEP	MEPLM
$\mu = 0.15\ h^{-1}$	20.41	0.28	5.75	5.75
$\mu = 0.30\ h^{-1}$	19.07	0.13	7.44	7.40
$\mu = 0.40\ h^{-1}$	24.50	0.16	2.46	2.46

Calculation accuracy is defined by Eq. (18). The metabolic model is shown elsewhere (9). There are 62 EMs.

Prediction accuracy The prediction errors in the flux distributions estimated by ECF, ECF-MEPLM, or ECF-B are calculated by:

$$\text{Prediction error} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(v_{i,\text{prediction}} - v_{i,\text{exp}}\right)^2} \quad (18)$$

where $v_{i,\text{prediction}}$ is the predicted flux for the ith reaction; $v_{i,\text{exp}}$ is the experimental data of the ith reaction; m is the number of the determined fluxes.

Control method ECF-B is a control method for ECF. In ECF-B, the determined enzyme activity parameters were replaced by the binary Boolean states: the activity of a deleted enzyme is set to zero and the others are set to one. The EMCs of a target model are calculated in the same manner as ECF.

Implementation EMs were calculated by CellNetAnalyzer (22). The optimization of EMCs and the prediction of the flux distribution were implemented in MATLAB (Mathworks Inc., Natick, MA). The employed computer for the simulation is Dell-Optiplex 755 (Intel-R Core-TM 2 Duo; CPU 2.33 GHz; Memory-RAM, 2.00 GB).

RESULTS

Validation of maximum entropy principle with Lagrange Multipliers (MEPLM) Shannon's entropy is a physical measure of the average information content of random events occurring in a system. Use of MEP presented a most probable distribution of EMCs in the absence of any biological hypotheses describing the physiological state within cells (Eqs. (1)–(6)) (9). MEP is different from typical objective functions. Typically, the prediction accuracy under such biologically specific objective functions depends on environmental and genetic conditions (3). On the other hand, optimization under the MEP constraint predicts the EMCs with relatively high accuracy under a variety of environmental and genetic conditions (9, 23). Particularly, MEP is a reasonable choice in the cases where biological objective functions are not specifically defined.

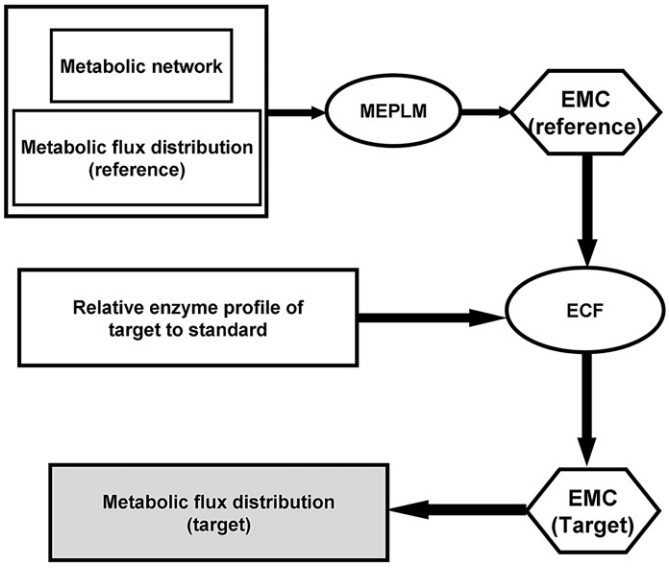

FIG. 1. A flow chart of ECFLM. The white square boxes are given; the grey square box is predicted. The ovals are algorithms.

The distributions of EMCs optimized by linear programming with maximum biomass formation or by minimization of the squared sum of EMCs have much more zero values than those by MEP (9). It suggests that MEP explores a most probable distribution without any biases deriving from the specific objective functions, while use of those specific functions would make the distribution of EMCs narrow.

A problem of EMC optimization under the MEP constraint is that the number of EMCs exponentially increases with an increase in network size (24). To circumvent this problem, we propose the MEPLM algorithm. Use of it remarkably decreases the number of the variables to that of metabolic fluxes. The feasibility of MEPLM

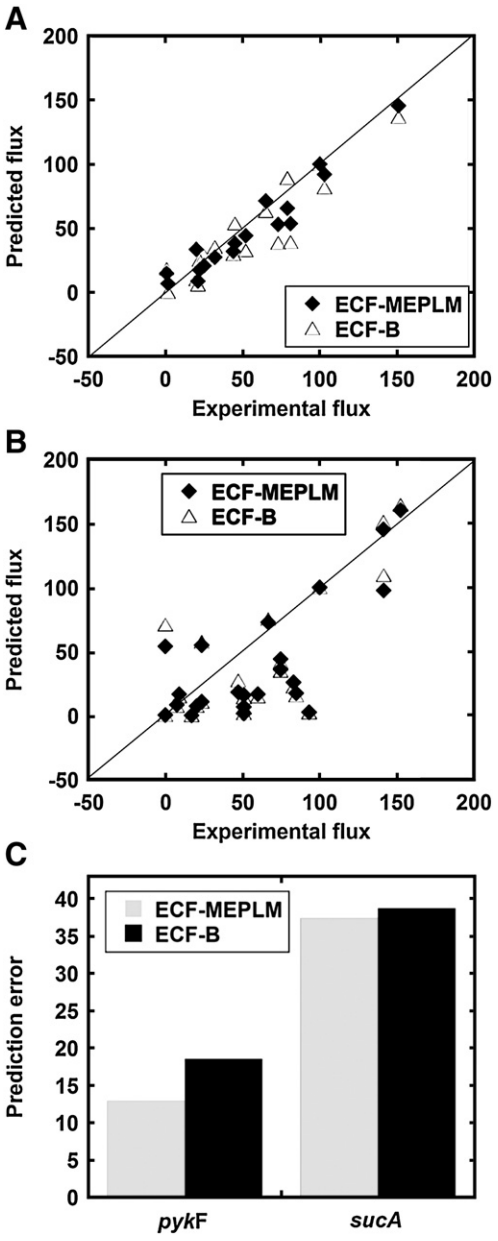

FIG. 2. Flux prediction for *E. coli* gene deletion mutants by ECF-MEPLM. Upper panel (A, B): the flux distributions are predicted by ECF-MEPLM and ECF-B under aerobic conditions for two gene deletion mutants in *E. coli*: (A) *pykF*; (B) *sucA*. Model I shown in Table 1 was used and the predicted flux distributions were compared with 20 (A) and 24 (B) experimental fluxes. Thirteen (A) and eighteen (B) relative enzyme activity data of mutants to wild type were used for the calculation by ECF-MEPLM. Lower panel (C): the prediction errors by ECF-MEPLM (grey) and ECF-B (black) are calculated for the above mutants (A, B).

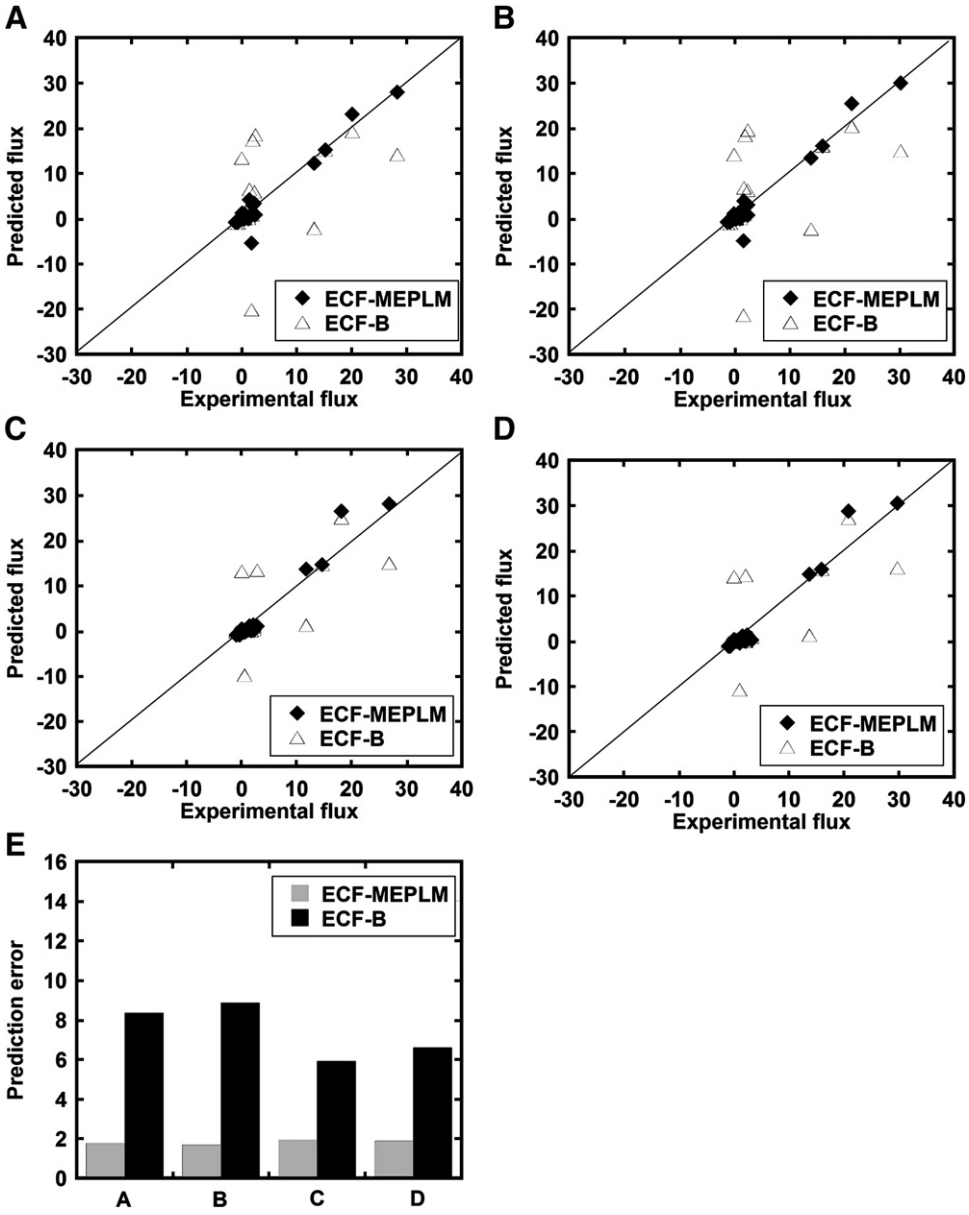

FIG. 3. Flux prediction for *E. coli* mutants undergoing adaptive evolution. Upper panel (A–D): the flux distributions are predicted by ECF-MEPLM and ECF-B for the *E. coli* gene deletion mutants undergoing adaptive evolution under anaerobic conditions: (A) *pta-pfkA* gene knockout mutant, evolved 30 days; (B) *pta-pfkA* gene knockout mutant, evolved 60 days; (C) *pta-adhE-pfkA-glk* gene knockout mutant, evolved 30 days; (D) *pta-adhE-pfkA-glk* gene knockout mutant, evolved 60 days. Model III shown in Table 1 was used for the *pta-pfkA* gene knockout mutant and model IV was used for the *pta-adhE-pfkA-glk* gene knockout mutant. The predicted flux distributions were compared with 26 (A and B) and 25 (C and D) experimental fluxes. Five relative enzyme activity data of adapted cells to the starring (reference) cells were used for the calculation by ECF-MEPLM. Lowe panel (E): the prediction errors are calculated for ECF-MEPLM (grey) and ECF-B (black) for the above *E. coli* gene deletion mutants (A–D).

is validated using a plain metabolic model for *S. cerevisiae*, as shown in Figs. S3–S4 and Table 2, where the experimental flux distribution was determined by ^{13}C trace technology (25). When seven fluxes were given, the EMCs and flux distribution were readily estimated by MEPLM. Use of LMs enhanced the calculation speed by more than 100-fold but never deteriorated the prediction accuracy of the flux distributions (Eq. (18)). These results demonstrate that MEPLM is very effective in the fast and accurate optimization of EMCs.

A link between enzyme profiles and flux profiles by ECF-MEPLM in genetic mutants ECF correlates the relationship between an enzyme activity profile and its associated flux distribution (Eqs. (14)–(17)) (8). Assuming that the flux passing through each EM is synergistically affected by all enzyme activities that belong to the EM, a multiplication formula is used to integrate all enzyme activities into the EMs. The ECF model neither considers any allosteric kinetics nor reduces large-scale network analysis to local one. Despite such plain ideas, ECF predicts how the change in enzyme profiles affects the flux distribution. This indicates that the total change in an enzyme activity profile plays a major role in determining flux distributions or the flux distribution would be determined by the effects of multiple enzyme activities rather than by a few rate-limiting reactions. A bottleneck of ECF is that it is available only to a small-scale metabolic network (8, 9, 23).

To extend the coverage of ECF, ECF is coupled with MEPLM to predict how the change in an enzyme activity profile alters the flux distributions in large scale metabolic networks of gene deletion mutants, as shown in Figure S1 (*E. coli*) and Figure S2 (*S. cerevisiae*).

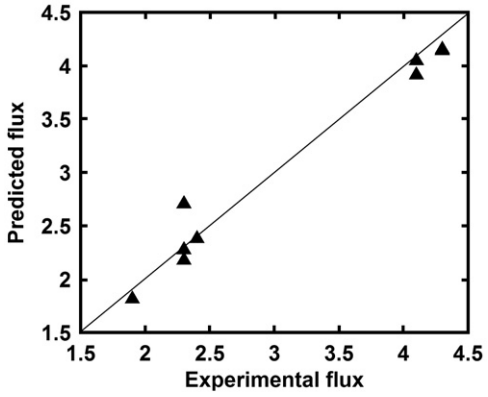

FIG. 4. Predicted flux distribution for temperature perturbation. The flux distribution is predicted by ECF-MEPLM in *S. cerevisiae* when the cultivation temperature decreases from 30 to 12 °C. Model V was used and the predicted flux distribution were compared with 12 experimental fluxes. Eleven relative enzyme activity data at 12 to at 30 °C were used for the calculation by ECF-MEPLM.

This coupling method is named ECF-MEPLM and its algorithm chart is shown in Fig. 1. The resultant predicted fluxes in two gene deletion mutants of *E. coli* (17, 18) are shown in Fig. 2. A reference method (ECF-B) uses the Boolean logic instead of the real enzyme activity profiles. ECF-MEPLM provided a high correlation between an enzyme activity profile and its associated flux distribution, compared with ECF-B. ECF-MEPLM takes an advantage in the use of the overall changes in enzyme activities.

Here, we compare the performance of ECF-MEPLM with that of typically used methods: FBA (2) and Minimization of Metabolic Adjustment (MOMA) (26). In FBA and MOMA, the flux related to gene deletion is set to zero and the flux distributions of gene deletion mutants are optimized by a specific biological function and by quadratic programming, respectively. The hypothesis in MOMA is the minimal flux adjustments between a mutant and wild type, assuming the robust property of a metabolic flux distribution. In principle, they do not use any enzyme activity profile caused by genetic modification but only reflect the change in the stoichiometric matrix. ECF-MEPLM is confirmed to take an advantage in the use of an enzyme activity profile and to accurately estimate the flux distributions (Figure S5). Note that we have no intention that ECF-MEPLM indicates a higher capability to predict flux distributions than MOMA and FBA, because ECF-MEPLM uses an enzyme activity profile whereas they do not. ECF-MEPLM presents a mathematical model that precisely correlates an enzyme activity profile to its associated flux distribution.

ECF-MEPLM in environmental changes The metabolic phenotype for genetically modified mutants can be improved after adaptive evolution process. The productive capabilities of lactate were enhanced by adaptive evolution in *E. coli* mutants after the cultivation of 1000 generations (27). The ECF-MEPLM-predicted flux distributions of two mutants of *E. coli* evolved in 30 days and 60 days are shown in Fig. 3, where the ratios of the enzyme activity profile of adapted cells to that of the starting cells (reference cells) were introduced to the optimized EMCs of the starting cells. The predicted flux distributions by ECF-MEPLM were consistent with the experimental data (19) and the prediction errors were low. These results show that ECF-MEPLM is effective in estimating a quantitative correlation between an enzyme activity profile and its associated flux distribution in adaptation stages. Note that neither MOMA nor FBA is available for this experiment because the stoichiometric matrix does not change during the adaptive evolution process.

Next, ECF-MEPLM was applied to estimating how the change in an enzyme activity profile alters its associated flux distribution with respect to the temperature perturbation in *S. cerevisiae*. The cultivation temperature decreased from 30 to 12 °C (15). The reference EMCs were evaluated by MEPLM with the flux distribution at 30 °C. Then ECF predicted the flux distribution at 12 °C using the relative change of the enzyme activity profile at 12 °C to that at 30 °C. The *in vivo* enzyme activities are changed greatly in response to the decrease in temperature. The predicted result coincides with the experimental fluxes, as shown in Fig. 4. The prediction error is almost the same as that by ECF-B (data not shown).

Finally, the flux distribution of *E. coli* was predicted by ECF-MEPLM when the dilution rate was changed from 0.10 to 0.32 and 0.55 h^{-1}

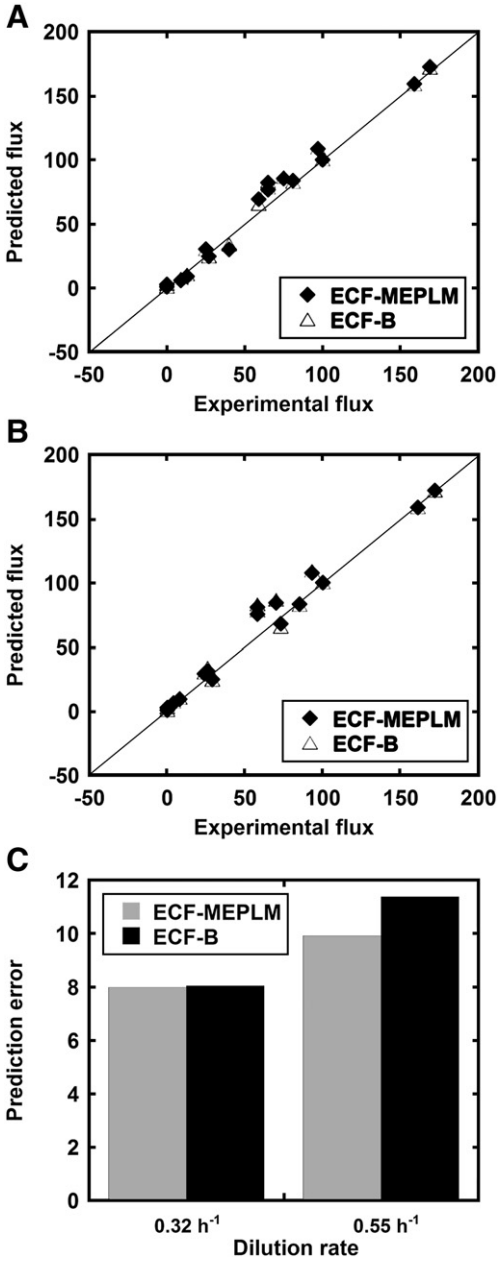

FIG. 5. Flux prediction for dilution rate change. Upper panel (A, B): the flux distribution is predicted by ECF-MEPLM and ECF-B for *E. coli* when the dilution rate changes from 0.1 h^{-1} to (A) 0.32 h^{-1} or (B) 0.55 h^{-1}. Model I shown in Table 1 was used and the predicted flux distribution was compared with 19 experimental fluxes. Three relative enzyme activity data of at 0.32 h^{-1} or 0.55 h^{-1} to at 0.1 h^{-1} were used for the calculation by ECF-MEPLM. Lower panel (C): the prediction errors are calculated for ECF-MEPLM (grey) and ECF-B (black) for the above conditions (A, B).

(16). MEPLM evaluated the reference EMCs from the flux distribution at the dilution rate 0.10 h^{-1}. Then ECF-MEPLM predicted the flux distribution at a dilution rate of 0.32 or 0.55 h^{-1} using the relative change of the enzyme activity profile at 0.32 or 0.55 h^{-1} to that at 0.10 h^{-1}. The predicted flux distributions and prediction errors for different dilution rates are shown in Fig. 5. The predicted flux distributions are consistent with the experimental ones. Since the prediction error decreases by incorporating an enzyme activity profile, ECF-MEPLM successfully incorporates the overall changes in enzyme activities into EMCs.

In summary, use of MEPLM greatly extends the coverage of ECF, although the originally developed ECF is restricted to small-scale networks (8–9). Consequently, ECF-MEPLM is demonstrated to correlates an enzyme activity profile to its associated metabolic flux distribution under different types of genetic and environmental perturbations.

Analysis of physiological states by ECF-MEPLM Generally, the macroscopic behavior of complex systems can be attributed to a collection of microscopic states (28). A metabolic flux distribution is one of the macroscopic properties for a biological system; EMs could be regarded as intracellular microscopic structures. The macroscopic behavior, a flux distribution, could be evaluated by the sum of each microscopic state. In this section, the changes in physiological states are characterized by using the most probable distribution of the EMCs optimized ECF-MEPLM (Eq. (3)).

We demonstrate that the EMC profile is effective in the analysis of the physiological sates of metabolic networks. The physiological states in *E. coli* were analyzed under aerobic and anaerobic conditions, as shown in Table 1 (29). The pathway length distributions for two conditions, calculated by CellNetAnalyzer (22), were

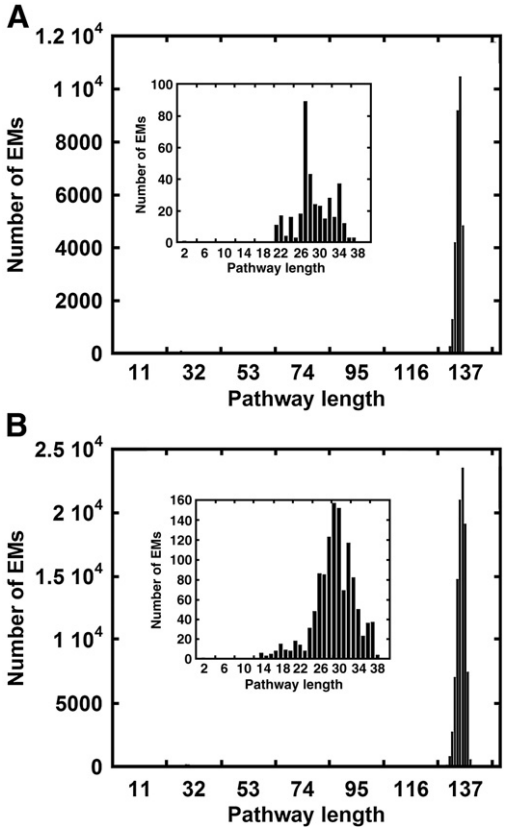

FIG. 6. Pathway length distribution. Model I shown in Table 1 was used for the pathway length distribution of *E. coli* under aerobic conditions (A) and model II was used for the analysis under anaerobic conditions (B).

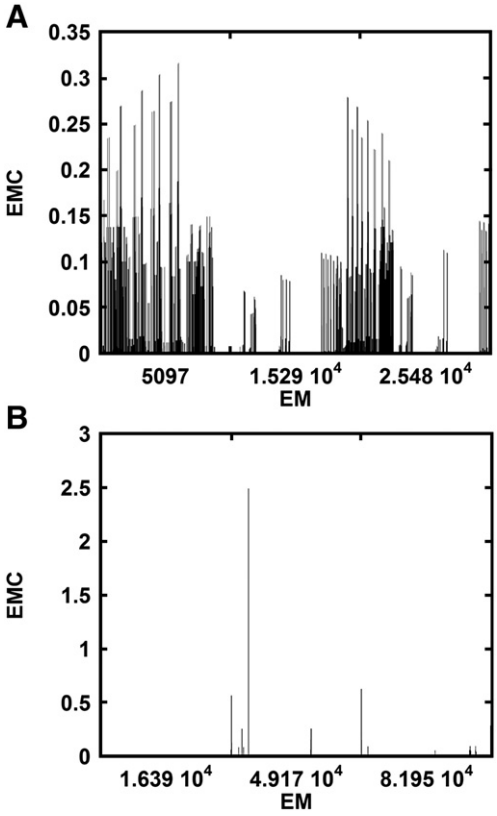

FIG. 7. EMC distribution of elementary modes. The EMC spectrums evaluated by MEPLM for *E. coli* under aerobic conditions (A) and anaerobic conditions (B). Model I and model II were used for aerobic and anaerobic conditions, respectively.

shown in Fig. 6. There are two groups of EMs denoted as groups I and II. Group I, in which the EMs have short pathway length, is mainly related to the ATP and product formations, while there is no EM related to growth. Group II consists of the EMs with long pathway length. All the EMs related to biomass formation are involved in group II. This classification was the same as discussed by Gagneur and Klamt (30). Under aerobic conditions, the pathway lengths are from 20 to 36 in group I; they are from 130 to 135 in group II. Under anaerobic conditions, the pathway lengths vary from 13 to 37 in group I and they range from 128 to 138 in group II. Most of the EMs are involved in group II under both the aerobic and anaerobic conditions.

Next, the EMC distributions under aerobic and anaerobic conditions are evaluated by MEPLM, as shown in Fig. 7. The EMC distributions are rather different between both the conditions, although the pathway length distributions are relatively similar (Fig. 6). Under anaerobic conditions, the number of the dominant EMs is limited and their overall reactions are shown in Table 3. There are four EMs in each subgroup of the dominant EMs, which are related to the formation of ethanol and lactate; some of them are coupled with ATP drain. EMs that have the same overall reaction stoichiometry can be grouped into one EM family (31). The EMC

TABLE 3. Overall reactions for the dominant EMs for *E. coli* under anaerobic conditions.

No.	Over all reaction	No. of EMs	EMCs
1	Glucose → 2 ethanol + 2 CO_2	37194, 37208, 37210, 37212[a]	2.4890
2	Glucose → 2 lactate	37225, 37237, 37239, 37241[a]	1.3272
3	Glucose → 2 ethanol + 2 CO_2	37193, 37207, 37209, 37211[a]	0.9738

[a] There are 98338 EMs for the metabolic network model under anaerobic conditions. The three EMs, 37212, 37241, and 37211, are also related to ATP drain, while the mass balance remains for the overall reactions.

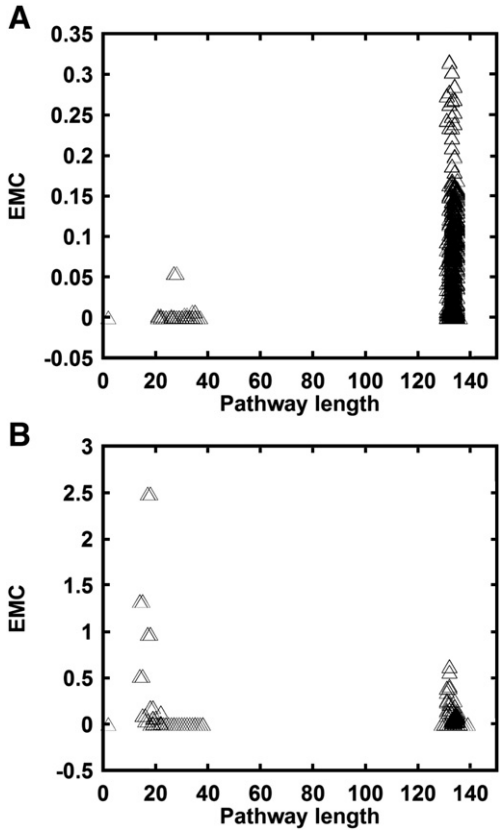

FIG. 8. EMC distribution with respect to pathway length. EMC evaluated by MEPLM versus pathway length for *E. coli* under aerobic conditions (A) and anaerobic conditions (B). Model I and model II were used for aerobic and anaerobic conditions, respectively, as shown in Table 1.

subgroup belonging to the same EM family (1, 3 in Table 3) are not consistent: the EMCs of one subgroup are 2.489 and those of the other 0.9738. There are amounts of alternative pathways in metabolic networks with same overall reactions. It ensures that metabolic networks are not destroyed by genetic and environmental perturbations. On the other hand, there are hundreds of dominant EMs under aerobic conditions as shown in Fig. 7. The largest EMC is approximately 0.32 under aerobic conditions, whereas it is 2.5 under anaerobic conditions. Many reactions are involved in cell metabolism under aerobic conditions, because the EMC profile is much scattered.

The relationship between the EMCs and pathway lengths are shown in Fig. 8. In general, the pathway length of biomass formation is long, because it includes many reactions for amino acids, DNA, and RNA syntheses. The pathway length of ethanol and lactate productions is short, as they belong to glycolysis and pyruvate metabolism. The pathway length of the largest EMC is 132 under aerobic conditions, while it is less than 20 under anaerobic conditions. This result shows that physiological states are clearly different between under both the conditions. Biomass formation is active under aerobic conditions, while ethanol/lactate formation and ATP drain are majorly active under anaerobic conditions. The changes in physiological states with respect to oxygen supply are characterized by those in the EMC profiles optimized by MEPLM.

DISCUSSION

ECF with MEP presents a quantitative and phenomenological relationship between enzyme activity data and the associated flux distribution without any biologically specific objective functions (8, 9). MEP can be a reasonable or standard choice in cases where biological objective functions are not specified. However, ECF-MEP (9) is not suitable for large-scale metabolic networks. To overcome this problem, MEPLM is proposed, where the number of the search variables is greatly reduced from the number of EMs to that of the determined fluxes. MEPLM enables estimating hundreds of thousands of EMCs in a large-scale metabolic network, under different types of environmental and genetic changes. Since MEPLM greatly extends the coverage of EM analysis, it can be a standard objective function, especially in cases where specific objective functions are not available.

While ECF-MEPLM is a non-mechanistic model that considers neither detailed enzyme kinetics, such as allosteric binding of inhibitors and activators, nor the concentrations of substrates (8,9), it effectively makes use of an enzyme activity profile for estimating its associated flux distributions in large-scale metabolic networks. A high correlation between an enzyme activity profile and its associated flux distribution is shown under different environmental and genetic perturbations. In addition, the calculation of the EMC distributions identifies the physiological states under aerobic and anaerobic conditions. Use of MEPLM greatly enhances the feasibility of EM-based analyses: integration of enzyme profile data into metabolic flux distributions under different types of genetic and environmental perturbations and identification of the physiological states of metabolic networks.

Generally, metabolic flux distributions are very informative to analyze the physiological state of microorganisms, e.g., cell growth and biosynthesis, while flux distribution data are not abundant in human cells, compared with proteome and transcriptome data, due to experimental complexity. It is critically important to predict flux distributions from available transcriptome and proteome data and to characterize the physiological state of disease cells. ECF-MEPLM coupled with such omics data is expected to classify the steady-state flux space, resulting in a characterization of all feasible steady-state flux distributions: normal physiological condition, diabetic condition, ischemic condition, diet condition, providing a basis for the quantitative analysis or diagnosis of metabolic diseases and prediction of effects of potential disease treatments.

The size of networks analyzed by ECF-MEPLM may be more enlarged to cope with a genome-scale model. Since the number of EMs increases exponentially with the network size, it may be still hard to explore EMCs at genome-scale networks with more than several hundreds of reactions due to combinatorial explosion. It is the bottleneck problem for the application of EMs while FBA is available for genome-scale networks. A few methods, which divide the network into subsystems by redefining internal and external metabolites or improve the algorithm deriving EMs from a stoichiometric matrix (32,33), have been proposed to reduce calculation complexity, but they have not been fully established yet.

ACKNOWLEDGMENTS

This work was supported by KAKENHI (Grant-in-Aid for Scientific Research) on Priority Areas “Systems Genomics” from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

APPENDIX A. SUPPLEMENTARY DATA

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jbiosc.2010.01.015.

References

1. **Stephanopoulos, G. and Vallino, J. J.:** Network rigidity and metabolic engineering in metabolite overproduction, Science, **252**, 1675–1681 (1991).
2. **Kauffman, K. J., Prakash, P., and Edwards, J. S.:** Advances in flux balance analysis, Curr. Opin. Biotechnol., **14**, 491–496 (2003).

3. **Schuster, S., Pfeiffer, T., and Fell, D. A.:** Is maximization of molar yield in metabolic networks favoured by evolution? J. Theor. Biol., **252**, 497–504 (2008).
4. **Åkesson, M., Förster, J., and Nielsen, J.:** Integration of gene expression data into genome-scale metabolic models, Metab. Eng., **6**, 285–293 (2004).
5. **Covert, M. W., Knight, E. M., Reed, J. L., Herrgard, M. J., and Palsson, B.Ø.:** Integrating high-throughput and computational data elucidates bacterial networks, Nature, **429**, 92–96 (2004).
6. **Shlomi, T., Eisenberg, Y., Sharan, R., and Ruppin, E:** A genome-scale computational study of the interplay between transcriptional regulation and metabolism, Mol. Syst. Biol., **3**, 101 (2007).
7. **Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., and Gilles, E. D.:** Metabolic network structure determines key aspects of functionality and regulation, Nature, **420**, 190–193 (2002).
8. **Kurata, H., Zhao, Q. Y., Okuda, R., and Shimizu, K.:** Integration of enzyme activities into metabolic flux distributions by elementary mode analysis, BMC Syst. Biol., **1**, 31 (2007).
9. **Zhao, Q. Y. and Kurata, H.:** Maximum entropy decomposition of flux distribution at steady state to elementary modes, J. Biosci. Bioeng., **107**, 84–89 (2009).
10. **Zhao, Q. Y. and Kurata, H.:** Genetic Modification of Flux (GMF) for flux prediction of mutants, Bioinformatics, **25**, 1702–1708 (2009).
11. **Papin, J. A., Stelling, J., Price, N. D., Klamt, S., Schuster, S., and Palsson, B.Ø.:** Comparison of network-based pathway analysis methods, Trends Biotechnol., **22**, 400–405 (2004).
12. **Schuster, S., Fell, D. A., and Dandekar, T.:** A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks, Nat. Biotechnol., **18**, 326–332 (2000).
13. **Shannon, C.E.:** A mathematical theory of communication. Bell Syst. Technol. J., 27, 379–423, 623–656 (1948).
14. **Acuña, V., Chierichetti, F., Lacroix, V., Marchetti-Spaccamela, A., Sagot, M., and Stougie, L.:** Modes and cuts in metabolic networks: complexity and algorithms, Biosystems, **95**, 51–60 (2009).
15. **Tai, S. L., Daran-Lapujade, P., Luttik, M. A., Walsh, M. C., Diderich, J. A., Krijger, G.C., van Gulik, W. M., Pronk, J. T., and Daran, J. M.:** Control of the glycolytic flux in *Saccharomyces cerevisiae* grown at low temperature: a multi-level analysis in anaerobic chemostat cultures, J. Biol. Chem., **282**, 10243–10251 (2007).
16. **Yang, C., Hua, Q., Baba, T., Mori, H., and Shimizu, K.:** Analysis of *Escherichia coli* anaplerotic metabolism and its regulation mechanisms from the metabolic responses to altered dilution rates and phosphoenolpyruvate carboxykinase knockout, Biotechnol. Bioeng., **84**, 129–144 (2003).
17. **Siddiquee, K. A., Arauzo-Bravo, M. J., and Shimizu, K:** Effect of a pyruvate kinase (*pykF*-gene) knockout mutation on the control of gene expression and metabolic fluxes in *Escherichia coli*, FEBS Lett., **235**, 25–33 (2004).
18. **Li, M., Ho, P. Y., Yao, S. J., and Shimizu, K:** Effect of *suc*A or *suc*C gene knockout on the metabolism in Escherichia coli based on gene expressions, enzyme activities, intracellular metabolite concentrations and metabolic fluxes by C-13-labeling experiments, Biochem. Eng. J., **30**, 286–296 (2009).
19. **Hua, Q., Joyce, A. R., Fong, S. S., and Palsson, B.Ø.:** Metabolic analysis of adaptive evolution for *in silico*-designed lactate-producing strains, Biotechnol. Bioeng., **95**, 992–1002 (2006).
20. **Price, N. D., Thiele, I., and Palsson, B.Ø.:** Candidate states of *Helicobacter pylori*'s genome-scale metabolic network upon application of "loop law" thermodynamic constraints, Biophys. J., **90**, 3919–3928 (2006).
21. **Hasbun, J. E.:** Classical Mechanics with MATLAB Applications, Jones and Bartlett, Sudbury, 2008.
22. **Klamt, S., Saez-Rodriguez, J., and Gilles, E. D.:** Structural and functional analysis of cellular networks with CellNetAnalyzer, BMC Syst. Biol., **1**, 2 (2007).
23. **Zhao, Q. Y. and Kurata, H.:** Estimation of intracellular flux distribution under underdetermined and uncertain conditions by maximum entropy principle, Chin. J. Biotechnol., **24**, 2135–2136 (2008).
24. **Trinh, C. T., Wlaschin, A., and Srienc, F.:** Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism, Appl. Microbiol. Biotechnol., **81**, 813–826 (2009).
25. **Frick, O. and Wittmann, C.:** Characterization of the metabolic shift between oxidative and fermentative growth in *Saccharomyces cerevisiae* by comparative ^{13}C flux analysis, Microb. Cell Fact., **4**, 30 (2005).
26. **Segrè, D., Vitkup, D., and Church, G. M.:** Analysis of optimality in natural and perturbed metabolic networks, Proc. Natl. Acad. Sci. USA, **99**, 15112–15117 (2002).
27. **Ibarra, R. U., Edwards, J. S., and Palsson, B.Ø.:** *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth, Nature, **420**, 186–189 (2002).
28. **Demetrius, L.:** Directionality principles in thermodynamics and evolution, Proc. Natl. Acad. Sci. USA, **94**, 3491–3498 (1997).
29. **Schmidt, K., Nielsen, J., and Villadsen, J.:** Quantitative analysis of metabolic fluxes in *Escherichia coli*, using two-dimensional NMR spectroscopy and complete isotopomer models, J. Biotechnol., **71**, 175–189 (1999).
30. **Gagneur, J. and Klamt, S.:** Computation of elementary modes: a unifying framework and the new binary approach, BMC Bioinformatics, **5**, 175 (2004).
31. **Wlaschin, A. P., Trinh, C. T., Carlson, R., and Srienc, F.:** The fractional contributions of elementary modes to the metabolism of *Escherichia coli* and their estimation from reaction entropies, Metab. Eng., **8**, 338–352 (2006).
32. **Urbanczik, R. and Wagner, C.:** An improved algorithm for stoichiometric network analysis: theory and applications, Bioinformatics, **21**, 1203–1210 (2005).
33. **Terzer, M. and Stelling, J.:** Large-scale computation of elementary flux modes with bit pattern trees, Bioinformatics, **24**, 2229–2235 (2008).