Accurate Metabolic Flux Analysis through Data Reconciliation of Isotope Balance-Based Data

KIM, TAE YONG[1] AND SANG YUP LEE[1,2]*

[1]Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical & Biomolecular Engineering and BioProcess Engineering Research Center,
[2]Department of BioSystems and Bioinformatics Research Center, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea

Abstract Various techniques and strategies have been developed for the identification of intracellular metabolic conditions, and among them, isotope balance-based flux analysis with gas chromatography/mass spectrometry (GC/MS) has recently become popular. Even though isotope balance-based flux analysis allows a more accurate estimation of intracellular fluxes, its application has been restricted to relatively small metabolic systems because of the limited number of measurable metabolites. In this paper, a strategy for incorporating isotope balance-based flux data obtained for a small network into metabolic flux analysis was examined as a feasible alternative allowing more accurate quantification of intracellular flux distribution in a large metabolic system. To impose GC/MS based data into a large metabolic network and obtain optimum flux distribution profile, data reconciliation procedure was applied. As a result, metabolic flux values of 308 intracellular reactions could be estimated from 29 GC/MS based fluxes with higher accuracy.

Key words: Data reconciliation, metabolic flux analysis, isotopic flux data, GC/MS, *Escherichia coli*

Metabolic engineering has become an essential discipline in successful development of biotechnological processes. Metabolic engineering can be defined as directed modification of cellular metabolism and properties through the introduction, deletion and modification of metabolic pathways by recombinant DNA technology and other molecular biological tools [1, 8]. Considering the many possible combinatorial manipulations that can be made on a given system, it is almost impossible to perform all these experiments. Thus,

it is more desirable to perform *in silico* metabolic analysis first, which is followed by much reduced number of actual experiments. For this reason, the quantification of intracellular metabolism is a key element of metabolic engineering, and has been realized using various mathematical and experimental techniques such as metabolic flux balance analysis (MFA), pathway analysis and gas chromatography/mass spectrometry (GC/MS)-based isotope analysis [10, 11, 13].

Generally, the MFA techniques are based on mass balance equations, which can be represented by a linear combination of stoichiometric reactions: in matrix form $Sv=0$, where S is the stoichiometric matrix describing all the reactions in the network and v is a vector describing the fluxes through each of the reactions. To obtain an exact flux distribution profile, the number of constraints (substrate uptake rates, product formation rates and mass balances around intracellular metabolites) should be equal to or greater than the number of variables (unmeasured intracellular fluxes). In other words, there should be no degrees of freedom [12]. In general, however, the number of measurable extracellular metabolite formation rates is significantly smaller than the number of intracellular fluxes. Therefore, linear optimization algorithms satisfying the objective function have been widely employed to obtain the optimum flux distribution under various metabolic conditions [13].

Also, tracer experiments using ^{13}C labeled compounds and the isotope balance-based flux analysis on GC/MS measurements has recently been intensively studied due to its high accuracy resulting from the actual measurement of intracellular carbon flux distribution. In GC/MS measurements, ^{13}C labeled carbon fragments are ionized by fragmentation of the molecular ions. The ionized carbon fragments are then seperated by GC, and the labeling patterns of the carbon fragments are analyzed by MS as they elute [4, 11]. Several isotope balance-based flux analyses have been

---
*Corresponding author
Phone: 82-42-869-3930; Fax: 82-42-869-8800;
E-mail: leesy@kaist.ac.kr

successfully demonstrated. The intracellular flux distribution of penicillin producing *Penicillium chrisogenum* was estimated by GC/MS based analysis [3]. In another study, GC/MS was used in combination with ^{13}C NMR for the identification of key metabolic network in lysine producing *Corynebacterium glutamicum* [9]. The response of intracellular metabolism of *Synechocystis* on the disturbance of extracellular environments was also evaluated through isotope balance-based analysis [14].

In spite of these successful applications of the isotope balance-based flux analysis towards understanding of intracellular metabolism, much improvement is needed since it has so far been applied to relatively small metabolic systems. Obviously, one can expand the metabolic network if more metabolites can be analyzed. However, this is currently difficult because of the limited number of measurable metabolites and the increasing number of nonlinear equations. Also, when the isotope balance-based fluxes were just provided as the constraints during the analysis of a large scale metabolic network, infeasible solutions were generated. Therefore, a new mathematical approach needs to be developed to use the isotope balance-based data in an expanded metabolic network, and consequently, to achieve more accurate and comprehensive understanding of intracellular metabolism.

In this paper, we describe a new strategy that combines the isotope balance-based data with MFA, which can handle large metabolic networks. The combined analysis allows more accurate quantification of intracellular flux distributions in large metabolic networks through integrating the benefits of both methods, e.g., accuracy of the isotope balance-based method and capability of handling large metabolic networks by MFA. Here, we describe how isotope balance-based flux data were introduced as constraints during the MFA. We also describe data reconciliation procedure in order to obtain optimal intracellular flux distribution.

MATERIALS AND METHODS

System Classification

The general strategy for determining metabolic fluxes is presented in Fig. 1. In a metabolic network, the relationships among all metabolites and reactions are balanced stoichiometrically as follows [12]:

$$Sv=S_m v_m+S_c v_c \qquad (1)$$

where S is stoichiometric matrix and v is flux vector, and the subscripts "m" and "c" indicate "measured" and "calculated", respectively. The system is, then, classified by one of four possible cases according to *Determinacy* and *Redundancy* [6]. In the case of the determined system, a unique solution or a least-squares solution is obtained by matrix operations if the system is observable. Otherwise,
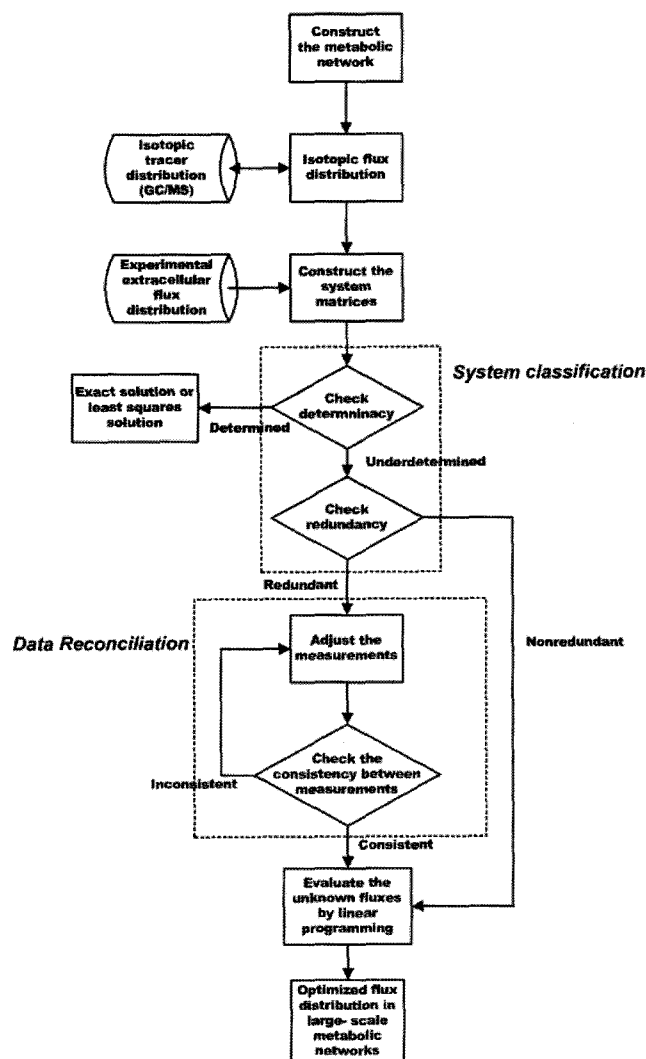
Fig. 1. Flow chart for the flux estimation by data reconciliation of the isotope balance-based flux data during the metabolic flux analysis of a large metabolic network. Reconciliation takes advantages of the accuracy of isotope balance-based flux analysis and applicability of metabolic flux analysis to a large metabolic network.

measured fluxes are reconciled to remove the inconsistency in the case of the redundant system, followed by inspecting calculable fluxes which can be uniquely determined by the least-squares solution using the pseudo-inverse (Fig. 1) [12]. If any row of S_c is expressed as a linear combination of other rows, the metabolic system is classified as a redundant one, in which the stoichiometric balances are often not satisfied due to unavoidable measurement errors and/or modeling inaccuracy. This makes the system inconsistent.

For the metabolic system to be consistent, the following condition should be satisfied:

$$Rv_m=0 \qquad (2)$$

where the reduced redundancy matrix R is defined as

$$R=S_m-S_c S_c^{\#} S_m \qquad (3)$$

$S_c^{\#}$ is the Penrose pseudo inverse matrix of S_c [2, 12]. If the matrix R is null, every measured rate is linearly independent and hence the adjustment of v_m is not required because there is no redundancy in the measurements. However, if any nonzero columns remain in the matrix R, the measured flux vector v_m should be adjusted to satisfy the conservation law and any other constraints imposed on it through data reconciliation.

Data Reconciliation

Experimental data generally contain noise, which generates residuals in Eq. (2). This would result in flux deviations between the actual rate vector v_m and $\bar{v}_m$ the measured rate vector:

$$\delta=\bar{v}_m - v_m \qquad (4)$$

where v_m, $\bar{v}_m$ and δ are vectors of the actual rate, measured rate and error, respectively.

Combining equation (2) and (4) gives the following equation for the vector of residual ε:
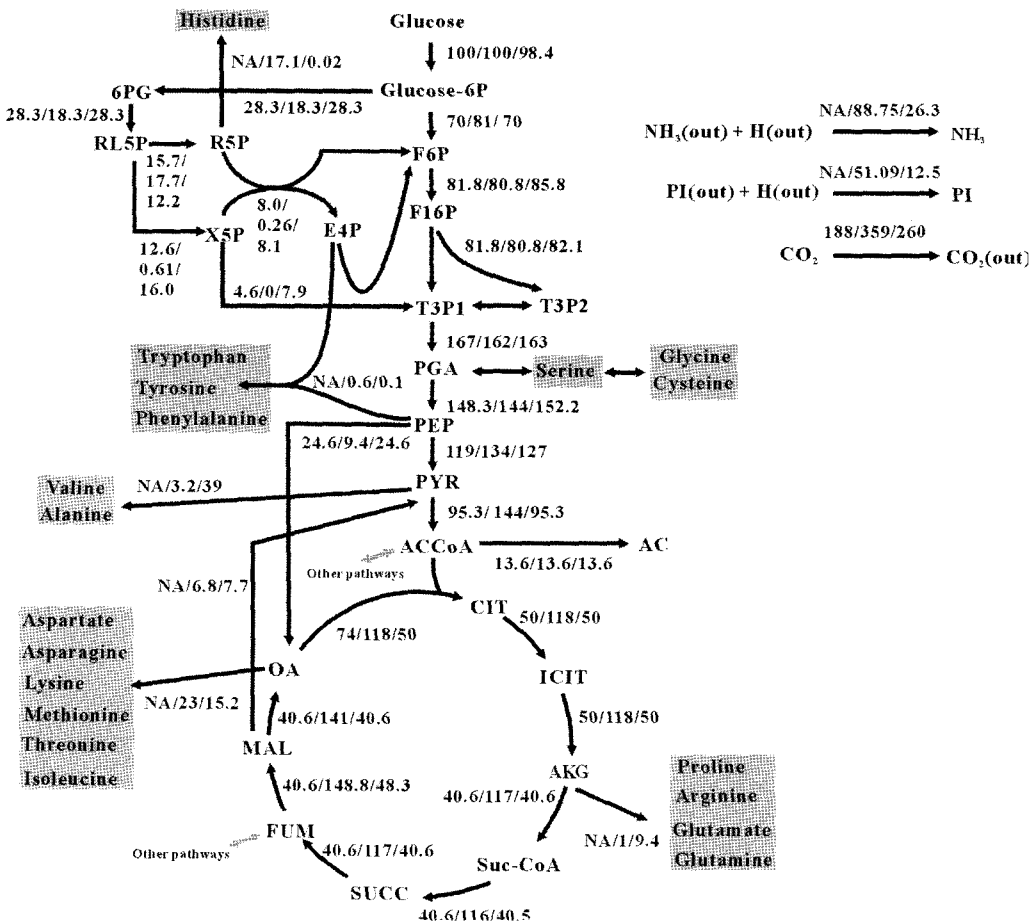
$$\varepsilon=R\bar{v}_m=R(\delta+v_m)=R\delta \qquad (5)$$

The minimum value of the error vector δ can be estimated by minimizing the sum of squared errors scaled according to their variance-covariance matrix F:

$$\min_{\delta}=(\delta^T F^{-1}\delta) \qquad (6)$$

It is assumed that the error vector is distributed normally with a mean value of zero.

A solution strategy based on the data reconciliation concepts yields the following solution:

Fig. 2. Comparison of the normalized metabolic fluxes obtained by (A) isotope balance-based flux analysis, (B) flux balance analysis based on linear optimization, and (C) combination of the two through data reconciliation in central metabolic network. Isotope balance-based flux data were taken from Zhao and Shimizu [15]. Fluxes were normalized with respect to the glucose uptake rate, and are presented in the order of (A)/(B)/(C). Abbreviations are: 6PG, 6-phospho-gluconate; ACCoA, acetyl-CoA; CIT, citrate; E4P, erythrose-4-phosphate; F6P, fructose-6-phosphate; F16P, fructose-1,6-phosphate; FUM, fumarate; Glucose-6P, glucose-6-phosphate; ICIT, isocitrate; MAL, malate; OA, oxaloacetate; AKG, -ketoglutarate; PYR, pyruvate; PEP, phosphoenolpyruvate; PGA, 3-phosphoglycerate; R5P, ribose-5-phosphate; RL5P, ribulose-5-phosphate; S7P, sedoheptulose-7-phosphate; SUC, succinate; Suc-CoA, succinyl-CoA; T3P1, glyceraldehyde-3-phosphate; T3P2, dihydroxyacetone phosphate; X5P, xylulose-5-phosphate; NA, not available.

$$\mathbf{v}_m = \bar{\mathbf{v}}_m - \delta = (\mathbf{I} - \mathbf{FR}^T(\mathbf{RFR}^T)^{-1}\mathbf{R})\bar{\mathbf{v}}_m \qquad (7)$$

Finally, MFA with constraints-based linear programming approach can be employed to quantify optimal flux distribution by optimizing a desired physiological property (objective function) such as growth rate, substrate uptake rate and product formation rate [13].

In this study, the biochemical reaction network of *E. coli* was constructed with 308 reactions and 275 metabolites, which contains all metabolic reaction pathways required for growth on glucose based on the previous work [5]. MFA was carried out using the MetaFluxNet package [7] using the maximum growth as the objective function.

RESULTS AND DISCUSSION

System Classification

The proposed new strategy of combining the isotope balance-based data and MFA was applied for the elucidation of the intracellular metabolism of *E. coli* under aerobic glucose-limited condition as a model system. Because the GC/MS-based flux distribution results are balanced in a relatively small metabolic network, the characteristics of the metabolic network should be examined to avoid possible self-discrepancies between the metabolic systems, which could arise during direct application of isotope balance-based flux distribution to a large metabolic network.

First, a large-scale metabolic network was constructed with 308 reactions and 275 metabolites that included all the reactions for glycolysis, the TCA cycle, the pentose phosphate pathway and respiration as well as biomass synthesis pathways [5]. The isotope balance-based fluxes obtained by Zhao *et al.* [15] for *E. coli* cultured under an aerobic glucose-limited condition were then applied as constraints. The metabolic network used for isotope balance-based analysis is composed of 29 reactions and 25 metabolites, and only covers the central metabolic network of *E. coli*. This intracellular flux distribution is presented in Fig. 2. A redundancy analysis indicated that there existed three redundant measurements when the GC/MS results were applied directly to the large metabolic network, indicating that the redundancy matrix **R** contains three nonzero columns corresponding to measured fluxes. Therefore, these fluxes should be adjusted to satisfy constraints imposed on them through the data reconciliation procedure.

We then carried out MFA on 308 reaction network by linear optimization using MetaFluxNet, and compared the results with the flux profiles obtained by isotope balance-based flux analysis. Due to the limited number of fluxes available for isotope balance-based flux analysis, direct comparison of flux values was limited to the central metabolic pathways as shown in Fig. 2. The calculated fluxes were normalized with respect to the glucose uptake rate. For MFA, achieving maximum growth was used as an objective

function, and glucose uptake and acetate excretion rates were applied as constraints. The general tendency of flux distribution obtained by MFA was in good agreement with that obtained by GC/MS-based flux analysis. Glycolysis functioned as a major catabolic metabolism, and up to 20% of carbon flux entered the pentose phosphate pathway. However, it can be seen from Fig. 2 that some of the fluxes calculated by MFA deviate considerably from those determined by GC/MS-based flux analysis. In particular, the fluxes of pentose phosphate pathway and TCA cycle show such deviations. This is due to the less accuracy of fluxes optimized for the underdetermined system. Therefore, to enhance the accuracy of MFA and to extract more biological information from intracellular metabolic network, we combined MFA with GC/MS-based flux analysis as described below.

Data Reconciliation

Because the large metabolic network was inconsistent with the isotope balance-based flux distribution, data reconciliation was performed to obtain optimized flux distribution. Again, the maximum growth rate was used as an objective function, and the results are presented in Fig. 2. The fluxes calculated by data reconciliation were not normalized with respect to the glucose up take rate. If the flux from glucose to glucose-6-phosphate is normalized, it will have a value of 100. However, we show the direct results from data reconciliation that consequently resulted in a value different from 100, based on the assumption that the measurement of glucose uptake is also vulnerable to measurement errors and so it also needs to be reconciled.

Comparison of the isotope balance-based flux distribution with the fluxes obtained after data reconciliation procedure suggested that the latter could yield more comprehensive information on intracellular metabolic conditions. In particular, the isotope balance-based flux analysis provided little information about anaplerotic fluxes in spite of its importance in aerobic metabolism (Fig. 2). As the intermediates of the TCA cycle are removed to serve as biosynthetic precursors, they are replenished by anaplerotic reactions from other pathways including glycolysis and pyruvate conversion pathways. Through data reconciliation, it could be predicted that PEP carboxylase and malic enzyme fluxes in anaplerotic reactions were 54.22 and 7.7 mmol/gDCW/h, respectively.

As shown above, data reconciliation led to the more accurate quantification of intracellular fluxes in a large metabolic network. However, the accuracy of data reconciliation is in question because is the results are still based on the optimization by linear programming. To evaluate the validity of data reconciliation results, the predicted CO_2 excretion flux was compared with the measured CO_2 excretion flux. This was possible because the measured CO_2 excretion rate was not applied as a constraint throughout the data reconciliation, which means that the CO_2 excretion rate could vary freely. The data reconciliation predicted that 220 mmol/

Table 1. Normalized sensitivity of carbon dioxide excretion rate with respect to changes in fluxes.

Pathway	$\dfrac{v_x}{q_{CO_2}}\dfrac{\partial v_{CO_2}}{\partial v_x}$
EMP pathway	0.47–1.06
Pentose phosphate pathway	0.22–0.39
TCA cycle	–0.2–0.2

gDCW/h of CO_2 was excreted, which is in good agreement with the measured CO_2 excretion rate (188 mmol/gDCW/h). This result suggests that data reconciliation can provide accurate and reliable flux values. Then, the data reconciliation was applied for the calculation of nonmeasured extracellular fluxes, e.g., the production/consumption rates of ammonia, phosphate and oxygen. The data reconciliation results predicted that the excretion and consumption rates of ammonia, phosphate, and oxygen were 65.7, 27.5, and 139 mmol/gDCW/h, respectively (Fig. 2).

For nonmeasured fluxes, it is important to estimate the sensitivities of the calculated fluxes with respect to small perturbations in the measured fluxes. The normalized sensitivity of v_c to the changes of v_m can be determined as $(v_m/v_c)\cdot(\partial v_c/\partial v_m)$, where v_m is the measured flux and v_c is the calculated flux, respectively. Thus, the sensitivities of the calculated fluxes with respect to the variations in the measured fluxes can provide information on the general sensitivity of the large metabolic system. The effects of altering each independent flux on the CO_2 excretion flux were thus investigated (Table 1). The obtained sensitivity range indicated that the TCA cycle fluxes have low sensitivity values or exert the least effect on the CO_2 excretion flux.

Metabolic flux analysis allows determination of fluxes in a large metabolic network, but the results may not be accurate due to the characteristics of linear optimization with limited number of constraints. On the other hand, isotope balance-based flux analysis allows a more accurate flux information, but is restricted to a relatively small metabolic network due to the limited number of measurable metabolites. Therefore, a new strategy described in this paper aims at providing the combined advantages of the two methods: accuracy of isotope balance-based flux analysis and applicability of metabolic flux analysis to a large metabolic network. As described in this paper, data reconciliation of isotope balance-based flux data during the MFA of a large metabolic system allows a more accurate determination of fluxes in a large metabolic network, which allows for a better understanding of the metabolic characteristics.

Acknowledgments

REFERENCES

1. Bailey, J. E. 1991. Towards a science of metabolic engineering. *Science* **252:** 1668–1674.
2. Ben-Israel, A. and T. N. E. Greville. 1974. Generalized inverses: Theory and applications. John Wiley & Sons, New York.
3. Christensen, B. and J. Nielsen. 2000. Metabolic network analysis of *P. chrisogenum* using 13C-labeled glucose. *Biotechnol. Bioeng.* **68:** 652–659.
4. Fischer, E. and U. Sauer. 2003. Metabolic flux profiling of *Escherichia coli* mutants in central carbon metabolism by GC-MS. *Eur. J. Biochem.* **270:** 880–891.
5. Hong, S. H. and S. Y. Lee. 2001. Metabolic flux analysis for succinic acid production by recombinant *Escherichia coli* with amplified malic enzyme activity. *Biotechnol. Bioeng.* **74:** 89–95.
6. Klamt, S., S. Schuster, and E. D. Gilles. 2002. Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism in purple nonsulfur bacteria. *Biotechnol. Bioeng.* **77:** 734–751.
7. Lee, D.-Y., H. S. Yun, S. W. Park, and S. Y. Lee. 2003. MetaFluxNet: The management of metabolic reaction information and quantitative metabolic flux analysis. *Bioinformatics* **19:** 2144–2146
8. Lee, S. Y. and E. T. Papoutsakis. 1999. Metabolic Engineering. Marcel Dekker, New York.
9. Park, S. M., C. Shaw-Reid, A. J. Sinskey, and G. Stephanopoulos. 1997. Elucidation of anapletoric pathway in *Corynebacterium glutamicum* via 13C-NMR spectroscopy and GC-MS. *Appl. Microbiol. Biotechnol.* **47:** 430–440.
10. Schuster, S., T. Dandekar, and D. A. Fell. 1999. Detection of elementary flux modes in biochemical networks: A promising tool for pathway analysis and metabolic engineering. *Trends in Biotechnology* **17:** 53–60.
11. Shimizu, K. 2002. A review on metabolic pathway analysis with emphasis on isotope labeling approach. *Biotechnol. Bioprocess Eng.* **7:** 237–251.
12. Stephanopoulos, G. A. A. Aristidou, and J. Nielsen. 1998. Metabolic engineering. Academic Press, San Diego.
13. Varma, A. and B. Ø. Palsson. 1994. Metabolic flux balancing: Basic concepts, scientific and practical use. *Bio/Technology* **12:** 994–998.
14. Yang, C., Q. Hua, and K. Shimizu. 2002. Integration of the information from gene expression and metabolic fluxes for the analysis of the regulatory mechanisms in *Synechocystis. Appl. Microbiol. Bioeng.* **58:** 813–822.
15. Zhao, J. and K. Shimizu. 2003. Metabolic flux analysis of *Escherichia coli* K12 grown on 13C labeled acetate and glucose using GC-MS and powerful flux calculation method. *J. Biotechnol.* **101:** 101–117.