Available online at www.sciencedirect.com
SCIENCE DIRECT®
Metabolic Engineering 7 (2005) 182–200
METABOLIC ENGINEERING
www.elsevier.com/locate/ymben

Complementary identification of multiple flux distributions and multiple metabolic pathways

Dong-Yup Lee[a,b,c], L.T. Fan[c,*], Sunwon Park[b], Sang Yup Lee[a,b,*], Shahram Shafie[c], Botond Bertók[d], Ferenc Friedler[d]

[a]*Metabolic and Biomolecular Engineering National Research Laboratory, Korea Advanced Institute of Science and Technology, 373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Korea*
[b]*Department of Chemical and Biomolecular Engineering, and Bioinformatics Research Center, Korea Advanced Institute of Science and Technology, 373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Korea*
[c]*Department of Chemical Engineering, Kansas State University, Manhattan, KS 66506, USA*
[d]*Department of Computer Science, University of Veszprém, Veszprém, Egyetem u. 10., H-8200 Hungary*

Received 17 September 2004; accepted 8 February 2005
Available online 30 March 2005

Abstract

Cell robustness and complexity have been recognized as unique features of biological systems. Such robustness and complexity of metabolic-reaction systems can be explored by discovering, or identifying, the multiple flux distributions (MFD) and redundant pathways that lead to a given external state; however, this is exceedingly cumbersome to accomplish. It is, therefore, highly desirable to establish an effective computational method for their identification, which, in turn, gives rise to a novel insight into the cellular function. An effective approach is proposed for complementarily identifying MFD in metabolic flux analysis and multiple metabolic pathways (MMP) in structural pathway analysis. This approach judiciously integrates flux balance analysis (FBA) based on linear programming and the graph-theoretic method for determining reaction pathways. A single metabolic pathway, with the concomitant flux distribution and the overall reaction manifesting itself as the desired phenotype under some environmental conditions, is determined by FBA from the initial candidate sequence of metabolic reactions. Subsequently, the graph-theoretic method recovers all feasible MMP and the corresponding MFD. The approach's efficacy is demonstrated by applying it to the *in silico Escherichia coli* model under various culture conditions. The resultant MMP and MFD attaining a unique external state reveal the surprising adaptability and robustness of the intricate cellular network as a key to cell survival against environmental or genetic changes. These results indicate that the proposed approach would be useful in facilitating drug discovery.

Keywords: Multiple flux distributions; Flux balance analysis; Multiple metabolic pathways; P-graphs; Systems biology

1. Introduction

Today, biotechnology manifests itself in one form or another in some sectors of wide-ranging industries having global economic impact. These industries include healthcare, pharmaceutical, chemical, food and agricultural industries. These industries are destined to be benefited from the development of inexpensive and higher-yield processes fueled by biotechnology. In this connection, it is highly likely that the rapidly growing field of quantitative analysis and modeling in metabolic engineering would become a promising tool in elucidating the functions and characteristics of complex biological systems essential for biotechnology. Moreover, the quantitative analysis and modeling will certainly facilitate the prediction of cellular behavior

*Corresponding author. Department of Chemical and Biomolecular Engineering, and Bioinformatics Research Center, Korea Advanced Institute of Science and Technology, 373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Korea.
E-mail addresses: fan@cheme.ksu.edu (L.T. Fan), leesy@kaist.ac.kr (S.Y. Lee).

1096-7176/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymben.2005.02.002

of microorganisms under various perturbations, e.g., the genetic modifications and/or environmental changes.

Several available approaches for such analysis and modeling include structural (topological) pathway analysis (Clarke, 1988; Seressiotis and Bailey, 1988; Mavrovouniotis et al., 1990; Liao et al., 1996; Simpson et al., 1999; Schilling et al., 2000; Schuster et al., 2000; Seo et al., 2001), metabolic flux analysis (MFA) (Stephanopoulos et al., 1998), metabolic control analysis (Kacser and Burns, 1973; Heinrich and Rapoport, 1974; Fell, 1996), and dynamic simulation (Tomita et al., 1999). Among them, MFA is most widely adopted for rational design and *in silico* engineering of metabolic pathways: the only information required is the stoichiometry of metabolic reactions and mass balances around the metabolites under pseudo-steady state, or stationary, assumption (Lee and Papoutsakis, 1999). Nevertheless, the number of reactions almost always exceeds the number of metabolites in any of the metabolic reaction systems; as such, the algorithmic identification of all metabolic fluxes is underdetermined (Klamt et al., 2002; Stephanopoulos et al., 1998). The flux distribution, therefore, has often been determined by means of flux balance analysis (FBA) based on linear programming (LP), thereby resulting in the flux distribution.

FBA has been firmly established theoretically (Varma and Palsson, 1994a; Bonarius et al., 1997; Edwards and Palsson, 1998; Sauer et al., 1998; Edwards et al., 1999; Schilling et al., 1999). It has effectively dealt with the metabolic networks of various kinds; however, several critical issues remain unresolved (Edwards et al., 2002). One such issue pertains to the uniqueness of the flux distribution. In general, FBA provides one desired physiological endpoint, e.g., the maximum growth rate, and its corresponding flux distribution under some culture conditions. It is uncertain, however, if this solution is unique. The implementation of LP in FBA frequently leads to multiple (or alternate) optima, thereby signifying the existence of multiple solutions corresponding to multiple flux distributions (MFD) (Lee et al., 2000). In the biological sense, these MFD imply the existence of redundant pathways in a metabolic network; the difference among the multiple solutions is attributable to the alternate equivalent sets of reactions, which have been investigated in detail by Mahadevan and Schilling (2003). This redundancy renders the network robust against the breakdown of the components, such as genes and enzymes, that disrupt some, but not all, of the pathways capable of achieving the same external state (Papin et al., 2002). To establish a rational metabolic engineering strategy requires that identification of the cellular state and prediction of the cellular behavior be sufficiently precise. Nevertheless, such identification and prediction are severely hindered by the existence of MFD. This entails the establishment of a method for identifying these MFD, which has been the focus of only a handful of works to date (Lee et al., 2000; Papin et al., 2002; Phalakornkule et al., 2001).

Lee et al. (2000) and Phalakornkule et al. (2001) have opened a new avenue for exploring MFD. They have resorted to the approach based on the recursive MILP to enumerate all the multiple optimal solutions for the given objective function in the MFA model. Recently, the modified MILP approach has been applied to the genome-size *Escherichia coli* model to generate the limited number of multiple equivalent phenotype states (Reed and Palsson, 2004). In fact, Schuster et al. (1999) have indicated that the elementary mode analysis gives rise to a systematic overview of the multitude of flux distributions realizable in the metabolic system. In this regard, Papin et al. (2002) have shown that the extreme pathway analysis (Schilling et al., 2000) and elementary flux modes (Schuster et al., 2000), both based on the iterative algebraic algorithms, can serve the same purpose.

The current work proposes a unified approach for identifying MFD in MFA and multiple metabolic pathways (MMP) in structural pathway analysis. This approach combines the flux balance model based on LP and the graph-theoretic method for reaction-pathway identification based on P-graphs. At the outset, the stoichiometric expression of an overall reaction, which is the manifestation of the relationship among extracellular metabolites, is obtained from a series of candidate metabolic reactions as well as the objective function specified by resorting to FBA. This is followed by the determination of the MMP satisfying the resultant overall reaction through the graph-theoretic identification of reaction pathways. Eventually, the corresponding MFD are recovered through FBA of each of the resultant MMP. The proposed approach is applied to *E. coli* to demonstrate its profound efficacy.

2. Flux balance analysis

In FBA, a metabolic reaction model is derived under the stationary hypothesis on the basis of measured fluxes. In deriving such a model, the relationships among all metabolites and reactions are balanced in terms of stoichiometry. Nevertheless, the resultant balanced reaction model is almost always underdetermined in calculating the flux distribution due to insufficient measurements or to constraints (Klamt et al., 2002). Thus, the unknown fluxes within the metabolic reaction network are evaluated by LP, subject to the constraints pertaining to mass conservation, reaction thermodynamics, and capacity as described elsewhere (Bonarius et al., 1997; Edwards et al., 1999; Varma and Palsson, 1994a). In this work, FBA has been implemented by program MetaFluxNet (version 1.6) developed for quantitatively analyzing metabolic fluxes

(Lee et al., 2003). The software is available from http://mbel.kaist.ac.kr.

Note that the stoichiometric coefficient, in conjunction with the flux distribution determined by FBA, leads to the net reaction balance equation; consequently, only the relationship among the fluxes of extracellular metabolites gives rise to the stoichiometric expression of the overall reaction.

3. Identification of MMP

With the stoichiometric expression of the overall reaction determined by FBA in hand, the proposed unified approach identifies all the feasible MMP from the initially proposed series of candidate metabolic reactions by the graph-theoretic method for reaction-pathway identification based on P-graphs originally developed for catalytic reactions (Fan et al., 1999, 2001, 2002). The method comprises the unique graph-representation of networks in terms of process graphs (P-graphs), two sets of axioms, and a group of 3 combinatorial algorithms. It has been unambiguously demonstrated that the method is applicable to the identification of feasible biochemical, or metabolic, pathways (Seo et al., 2001). This is not unexpected: metabolic reactions are catalytic reactions with enzymes serving as catalysts (Voet and Voet, 1995).

3.1. P-graph representation

An unambiguous network representation is required in the biochemical pathway determination through the synthesis of elementary reactions if the resultant networks are to be mathematically exact so that they can be analyzed formally. The elementary-reaction steps or metabolic reactions are directed; thus, every network representing a reaction pathway including these steps can be represented by directed graphs. In contrast, conventional graphs are incapable of uniquely representing such networks (Fan et al., 1999, 2001, 2002; Seo et al., 2001).

P-graphs are directed bipartite graphs (Friedler et al., 1992, 1993, 1995). Let O be the set of metabolic-reaction steps and M be the set of metabolites under consideration; then, $O \subseteq \wp(M) \times \wp(M)$, where $O \cap M = \varnothing$. If (s, p) is a reaction step, i.e., $(s,p) \in O$, then s is called the set of substrates, and p, the set of products of this reaction step. Pair (M, O) is termed a P-graph with the set of vertices $M \cup O$, and the set of arcs $\{(x,y)$: $y = (s,p) \in O$ and $x \in s\} \cup \{(y,x) : y = (s,p) \in O$ and $x \in p\}$. In P-graph representation, metabolic reactions of which the pathways are composed are symbolized by horizontal bars; and metabolites, by circles. If a metabolite is an input to or output from a metabolic reaction, the vertex representing this metabolite is linked by an arc to or from the vertex representing this metabolic reaction, respectively, as illustrated in Fig. 1.

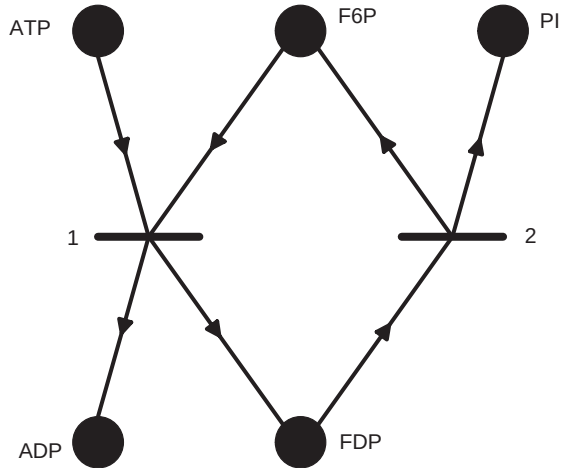

Fig. 1. P-graph representation of two metabolic reactions, F6P + ATP → ADP + FDP and FDP → F6P + PI in the glycolytic pathway. m = {F6P, FDP, ATP, ADP, PI}; o = {1 →, 2 →} = {({F6P, ATP},{ADP,FDP}), ({FDP},{F6P,PI})}.

3.2. Formal graph-theoretic description of the problem

Here, a formal description is given of the problem of the identification of MMP in the parlance of graph theory in general and that of P-graph in particular (Roberts, 1984; Friedler et al., 1992, 1993; Imreh et al., 1996). Let a reaction-pathway-identification problem be defined by triplet (E, O, M), where E is the overall reaction; $O = \{e_1, e_2, \ldots, e_n\}$, the finite ordered set of metabolic reactions; and $M = \{a_1, a_2, \ldots, a_l\}$, the finite ordered set of metabolites. For any reversible reaction step e_i defined, its reverse step, denoted by $-e_i$, is also included in set O. It is assumed that

$$M \cap O = \varnothing \quad \text{and} \quad E \notin O \cup M \tag{1}$$

For the overall reaction, E, let $\omega^-(E)$ and $\omega^+(E)$ denote the set of starting substrates and final products, respectively. Notice that a stoichiometric expression for the overall reaction must be known a priori to graph-theoretically identify any metabolic pathway. Frequently, such an expression is unavailable for complex or realistic metabolic pathways. In the proposed unified approach, it is determined through FBA at the outset. If $\omega(E)$ is the set of extracellular metabolites consumed or produced by the overall reaction, E, we have

$$\omega(E) = \omega^-(E) \cup \omega^+(E). \tag{2}$$

Similarly, for any metabolic reaction step $e_i \in O$, let $\omega^-(e_i)$ and $\omega^+(e_i)$ denote the set of reactants and products of e_i, respectively. If $\omega(e_i)$ denotes the set of metabolites consumed or produced by the reaction step e_i, we have

$$\omega(e_i) = \omega^-(e_i) \cup \omega^+(e_i). \tag{3}$$

For any metabolites $a_j \in M$, let $\upsilon^-(a_j)$ and $\upsilon^-(a_j)$ denote the set of metabolic reaction steps consuming and producing a_j, respectively. If $\upsilon(a_j)$ denotes the set of metabolic reaction steps consuming or producing a_j, we have obviously

$$\upsilon(a_j) = \upsilon^-(a_j) \cup \upsilon^-(a_j). \tag{4}$$

For any set of the reaction steps, $o \subseteq O$, let $\Psi^-(o)$ and $\Psi^+(o)$ denote the set of active metabolites consumed and produced by any element of o, respectively; it follows that

$$\Psi^-(o) = \bigcup_{e_i \in o} \omega^-(e_i) \tag{5}$$

and

$$\Psi^+(o) = \bigcup_{e_i \in o} \omega^+(e_i). \tag{6}$$

If $\Psi(o)$ is the set of metabolites consumed or produced by any element of o, we have

$$\Psi(o) = \Psi^-(o) \cup \Psi^+(o). \tag{7}$$

For any set of metabolites $m \subseteq M$, let $\varphi^-(m)$ and $\varphi^+(m)$ denote the set of reaction steps producing and consuming any element of m, respectively; it follows that

$$\varphi^-(m) = \bigcup_{a_j \in m} \upsilon^-(a_j) \tag{8}$$

and

$$\varphi^+(m) = \bigcup_{a_j \in m} \upsilon^+(a_j). \tag{9}$$

If $\varphi(m)$ is the set of reaction steps producing or consuming any element of m, we have

$$\varphi(m) = \varphi^-(m) \cup \varphi^+(m). \tag{10}$$

For any set of reaction steps $o \subseteq O$, let $\mathrm{X}(o)$ denote the set of opposite steps of the elementary-reaction steps included in set o; then,

$$\mathrm{X}(o) = \{e_i : -e_i \in o\}. \tag{11}$$

Any P-graph representing a set of metabolites and metabolic reaction steps is given by pair (m, o), where $o \subseteq O$ is the set of the reaction steps, and $m \subseteq M$ is the set of metabolites, where

$$\Psi(o) \subseteq m. \tag{12}$$

The set of vertices of the graph is

$$V = o \cup m, \tag{13}$$

where any vertex corresponding to set m is termed M-type, and any vertex corresponding to set o is termed O-type. The set of arcs is

$$A = A_1 \cup A_2, \tag{14}$$

where

$$A_1 = \{(a_j, e_i) : a_j \in m,\ \ e_i \in o,\ \ a_j \in \omega^-(e_i)\} \tag{15}$$

and

$$A_2 = \{(e_i, a_j) : e_i \in o,\ \ a_j \in m,\ \ a_j \in \omega^+(e_i)\}. \tag{16}$$

In graphical representation, vertices of the O-type are denoted by horizontal bars, and vertices of the M-type are denoted by solid circles. It is worth noting that from the standpoint of formal representation, P-graphs are isomorphic to Petri nets which have been extensively adopted for the analysis of metabolic pathways (Reddy et al., 1996; Küffner et al., 2000; Oliveira et al., 2001). Petri nets are bipartite directed graphs $G = (V, E)$ composed of the two kinds of vertices, V_1 and V_2; naturally $V_1 \cup V_2 = V$. The former is termed places ($V_1 = P$), and the latter is termed transitions ($V_2 = T$); edges $ed \in Ed \subseteq (V_1 \times V_2) \cup (V_2 \times V_1)$ link the places with the transitions, and vice versa (Murata, 1989).

As far as the static structure, the M-type and O-type vertices in a P-graph can be regarded, respectively, as disjoint sets of the places, P, and the transitions, T, in a proper Petri net while edges Ed in Petri nets correspond to the set of arc, A, in P-graphs. Obviously, a strong correspondence exists between the P-graphs and Petri nets; nevertheless, each plays individually a unique role in theoretical discourse or in application.

Petri nets have been developed for representing and exploring a system of concurrent events. An additional term, i.e., token, is introduced to take into account the dynamic behavior of the system of events. On the other hand, P-graphs have originally been conceived to describe the structure of a process system involving the transformation of material species, especially for its synthesis based on the algorithmic framework. The procedure for implementing the algorithms based on the P-graphs can be readily modified. Thus, the P-graphs can be extended to a variety of network systems to analyze these structural properties through syntheses.

3.3. Axioms

For any given overall reaction representing the starting substrates (e.g., carbon sources) and the final metabolic products, P-graph (M, O) composed of the metabolic reactions in set O and the concomitant metabolites in set M is combinatorially feasible if it satisfies the set of 7 axioms of combinatorially feasible metabolic reaction networks. Moreover, P-graph (M, O) is a feasible metabolic pathway if it satisfies the set of 6 axioms of feasible metabolic pathways for the given overall reaction. In the parlance of metabolic reaction networks or pathways, these two sets of axioms can be stated as follows:

(a) *Seven axioms of combinatorially feasible metabolic networks*

(T1) Every final product (target metabolite) is represented in the network.

(T2 Every starting substrate (precursor metabolite) is represented in the network.
(T3) Each metabolic reaction represented in the network is defined a priori.
(T4) Every metabolite represented in the network has at least one path leading to a final product (target metabolite) of the overall reaction.
(T5) Every metabolite represented in the network must be a substrate for or a product from at least one metabolic reaction represented in the network.
(T6) A substrate of any metabolic reaction represented in the network is a starting substrate (precursor metabolite), if it is not produced by any metabolic reaction represented in the network.
(T7) The network includes at most either the forward or reverse step of each metabolic reaction represented in the network.
(b) *Six axioms of feasible metabolic pathways*
(R1) Every final product (target metabolite) is totally produced by the metabolic reactions represented in the pathway.
(R2) Every starting substrate (precursor metabolite) is totally consumed by the metabolic reactions represented in the pathway.
(R3) Every intermediate metabolite produced by any metabolic reaction represented in the pathway is totally consumed by one or more metabolic reactions in the pathway, and every intermediate metabolite consumed by any metabolic reaction represented in the pathway is totally produced by one of more metabolic reactions in the pathway.
(R4) All metabolic reactions represented in the pathway are defined a priori.
(R5) The metabolic network representing the pathway is acyclic.
(R6) At least one metabolic reaction represented in the pathway activates a starting substrate (precursor metabolite).

It is noteworthy that all metabolites except extracellular metabolites are active intermediates or simply intermediates.

3.4. Algorithms

The two sets of axioms given above naturally give rise to three efficient algorithms. The first is algorithm Reaction Pathway Identification for Maximal Structure Generation (RPIMSG) for generating the maximal metabolic reaction network. The second is algorithm Reaction Pathway Identification for Solution Structure Generation (RPISSG) for generating the combinatorially feasible metabolic pathways. The third is algorithm Pathway BackTracking (PBT) for the final determination of all the feasible metabolic pathways directly from the maximal metabolic reaction network (Fan et al., 2002).

The maximal structure contains all combinatorially feasible structures, i.e., reaction networks or pathways, each leading from the starting substrates to the final metabolic products, without violating axioms (T1)–(T7); note that not every combinatorially feasible structure constitutes a feasible pathway. Moreover, such a structure must satisfy the mass conservation, as expressed by axioms (R1)–(R3); must not contain a cycle satisfying the principle of microscopic reversibility, as expressed by axiom (R5); and must contain at least one metabolic reaction step activating a starting substrate, as expressed by axiom (R6). Fig. 2 contains the computer program for implementing algorithm RPIMSG in terms of the formal graph-theoretic description of the metabolic-pathway-identification problem. The algorithm consists of two major parts, reduction and composition. In the former, the metabolites, i.e., starting substrates, final products, or intermediates, and the metabolic reaction steps that must not belong to the maximal structure are excluded from the initial structure to the maximum extent possible on the basis of Axioms (T1)–(T7). Obviously, all dead ends and isolated reactions disconnected from the network can be identified to preclude their participation in the formation of the network structure. Note that a dead end is an intermediate metabolite either only produced or consumed in the network (Reed et al., 2003). To initiate the latter, i.e., composition, every step of each reaction, which has survived the elimination and is deemed plausible for inclusion, is properly identified on the basis of Axiom (T3), and each final product is correctly specified on the basis of Axiom (T1). Hereafter, the maximal structure is constructed stepwisely by collecting the reaction steps so as to satisfy Axioms (T4) and (T5).

The algorithm for the solution structure generation, algorithm RPISSG, yields the set of all combinatorially feasible reaction networks from the maximal structure of reaction networks. To drastically reduce the computational time necessary to ascertain if each combinatorially feasible reaction network or pathway is indeed a feasible pathway in the light of Axioms (R1)–(R5), a branch-and-bound-like algorithm termed Pathway-Back-Tracking algorithm (algorithm PBT) has been developed. The procedure for implementing algorithm PBT, or equivalently the search through the enumeration tree, is initiated at the maximal structure of reaction networks obtained by virtue of algorithm RPIMSG; this structure is at the root of the tree. Algorithm PBT, facilitated by the subsidiary algorithms, eventually generates the complete set of feasible pathways and the multipliers of the resultant reaction steps for a given reaction-pathway-identification problem. Herein, each feasible pathway attains the same overall reaction by the linear combination of reaction steps with multipliers or weights which are

inputs: RPI problem (E, M, O);
output: maximal structure (m, o);

begin
  comment: reduction part of the algorithm;
  repeat
    $M := \Psi(O)$;
    $exc := \varnothing$;
    for all $x \in M$
    begin
      comment: Case 1
      if $x \notin \omega(E)$ **and** $v^-(x) = \varnothing$ **then** $exc := exc \cup v^+(x)$;
      comment: Case 2
      if $x \notin \omega(E)$ **and** $v^+(x) = \varnothing$ **then** $exc := exc \cup v^-(x)$;
      comment: Case 3
      if $x \notin \omega(E)$ **and** $|v^-(x)| = 1$ **and** $v^+(x) = X(v^-(x))$ **then** $exc := exc \cup v(x)$;
      comment: Case 4
      if $x \notin \omega^-(E)$ **and** $|v^-(x)| = 1$ **then** $exc := exc \cup X(v^-(x))$;
      comment: Case 5
      if $x \notin \omega^+(E)$ **and** $|v^+(x)| = 1$ **then** $exc := exc \cup X(v^+(x))$;
    end;
    $O := O \backslash exc$;
  until $exc = \varnothing$;

  comment: composition part of the algorithm;
  $m := \omega^+(E)$;
  $o := \varnothing$;
  repeat
    $add := \varphi^-(m) \backslash o$;
    $o := o \cup add$;
    $m := m \cup \Psi^-(o)$;
  until $add = \varnothing$;
  if $\omega^-(E) \backslash m \neq \varnothing$ **or** $\omega^+(E) \backslash m \neq \varnothing$ **then stop**; **comment:** There is no maximal structure.
  write (m, o);
end.

Fig. 2. Algorithm RPIMSG for the identification of multiple metabolic pathways.

stoichiometric numbers. Note that the stoichiometric numbers are defined for the elementary reactions constituting the mechanism of a catalytic reaction (Boudart and Djega-Mariadassou, 1984; Fan et al., 1999, 2001, 2002). The metabolic reactions are nothing but the elementary reactions in an enzymatically catalyzed reaction system. Supporting material detailing the detailed procedure for implementing algorithms RPISSG and PBT is available at http://mbel.kaist.ac.kr/publication/MEsuppl/.

3.5. Identification of MFD

FBA of MMP is straightforward; MMP is determined by the aforementioned graph-theoretic method. Specifically, the objective function as given in FBA is maximized or minimized by LP for each of the MMP subject to the mass balance constraints in terms of all the metabolites in the pathway and the substrate uptake rate, thus yielding MFD.

4. Simple example for the identification of MFD and MMP

A simple example is presented herein to illustrate the current methodology and procedure (Fig. 3). The example also provides numerical proofs for the validity of the methodology and procedure

4.1. Flux balance analysis

At the outset of performing metabolic balance analysis, it is visualized that a metabolic pathway exists, which comprises a set of 7 metabolic reactions listed below.

$$R_1: \quad A \overset{v_1}{\leftrightarrow} B,$$

$$R_2: \quad B \overset{v_2}{\longrightarrow} 2E,$$

$$R_3: \quad E \overset{v_3}{\leftrightarrow} F,$$

$$R_4: \quad B \overset{v_4}{\leftrightarrow} C,$$

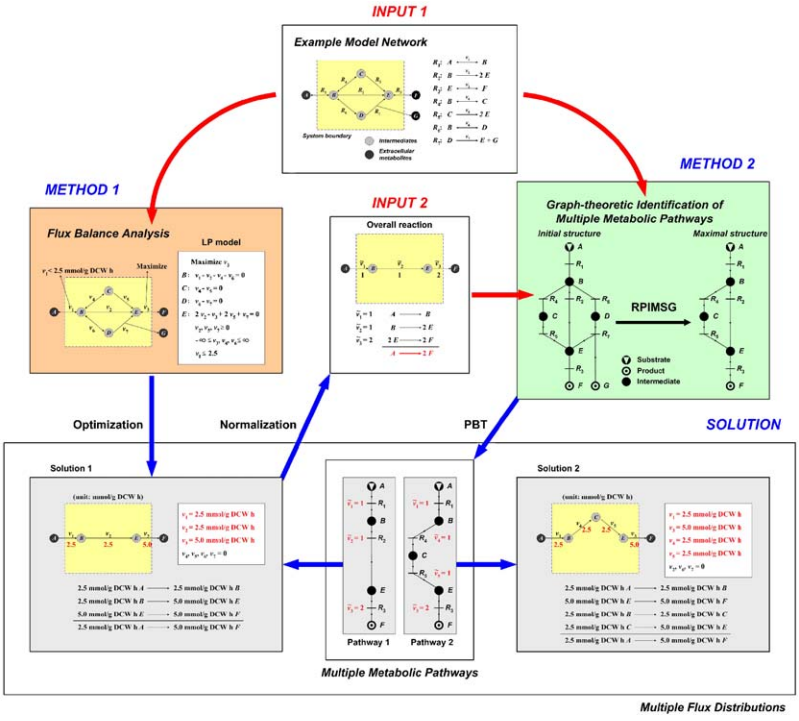

Fig. 3. Schematic procedure of the identification of multiple metabolic pathways and multiple flux distributions for a simple example. INPUT 1: an example model network comprising 7 metabolic reactions; INPUT 2: an overall reaction determined by means of flux balance analysis (FBA); METHOD 1: FBA; METHOD 2: graph-theoretic approach; SOLUTION: multiple metabolic pathways and concomitant flux distributions. See the text for the detailed procedure.

$$R_5: \quad C \xrightarrow{v_5} 2E,$$

$$R_6: \quad B \overset{v_6}{\leftrightarrow} D,$$

$$R_7: \quad D \xrightarrow{v_7} E + G.$$

For the sake of demonstration, the overall reaction of this pathway is regarded as unknown; for many of the complex or realistic metabolic pathways, this is indeed the case. Note that B, C, D, and E are intermediate metabolites, or simply, intermediates; A, F and G are extracellular metabolites, metabolite A being the precursor, or substrate and F and G, the product; and v_i, $i = 1, 2, \ldots, 7$, is the flux of metabolic reaction R_i, $i = 1, 2, \ldots, 7$. An example model network (INPUT 1) in Fig. 3 depicts the initially hypothesized metabolic pathway. For the pathway under consideration, the mass balance around the system boundary in terms of moles of intermediates gives rise to

$$\begin{array}{c} B \\ C \\ D \\ E \end{array} \begin{bmatrix} 1 & -1 & 0 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 2 & -1 & 0 & 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \end{bmatrix} = 0 \quad (17)$$

which, in turn, yields

$$\begin{aligned} B: &\quad v_1 - v_2 - v_4 - v_6 = 0 \\ C: &\quad v_4 - v_5 = 0 \\ D: &\quad v_6 - v_7 = 0 \\ E: &\quad 2v_2 - v_3 + 2v_5 + v_7 = 0. \end{aligned} \quad (18)$$

These expressions are imposed as the constraints in implementing LP for FBA of the hypothesized metabolic pathway. Moreover, the production of final product F is defined as the objective function to be maximized for a given steady uptake rate of substrate A. As indicated in Fig. 3 (METHOD 1), the maximization of the product generation is equivalent to the maximization of flux v_3 and that the numerical value of the rate of the substrate uptake is specified to be no more than 2.5 mmol/g DCW h. The LP problem to be solved can be stated as follows:

$$\begin{aligned} &\text{Maximize}: \quad v_3 \\ &\text{subject to}: \\ &B: \quad v_1 - v_2 - v_4 - v_6 = 0 \\ &C: \quad v_4 - v_5 = 0 \\ &D: \quad v_6 - v_7 = 0 \end{aligned}$$

$$E: \quad 2v_2 - v_3 + 2v_5 + v_7 = 0.$$

$$v_2, v_3, v_6 \geqslant 0$$

$$-\infty \leqslant v_4, v_5, v_7 \leqslant \infty$$

$$v_1 \leqslant 2.5. \tag{19}$$

Upon solution, we obtain

$$v_1 = 2.5 \text{ mmol/g DCW h}$$

$$v_2 = 2.5 \text{ mmol/g DCW h}$$

$$v_3 = 5.0 \text{ mmol/g DCW h}$$

$$v_4 = 0.0$$

$$v_5 = 0.0$$

$$v_6 = 0.0$$

$$v_7 = 0.0$$

On the basis of one g DCW and 1 h, these results imply that

2.5 mmol/g DCW h $A \rightarrow$ 2.5 mmol/g DCW h B,
2.5 mmol/g DCW h $B \rightarrow$ 5.0 mmol/g DCW h E,
5.0 mmol/g DCW h $E \rightarrow$ 5.0 mmol/g DCW h F.

Normalizing with the substrate uptake rate of 2.5 mmol/g DCW h gives

$$A \xrightarrow[1]{\tilde{v}_1} B,$$

$$B \xrightarrow[1]{\tilde{v}_2} 2E,$$

$$2E \xrightarrow[2]{\tilde{v}_3} 2F.$$

Naturally, the corresponding overall reaction is

$$A \rightarrow 2F.$$

When some $\tilde{v}_i$'s are non-integer rational numbers, all $\tilde{v}_i$'s are multiplied by their least common denominator (LCD) to transform them into integers. This is often the case for realistic metabolic networks. It is worth noting that the computational time to perform exact rational arithmetic can be exceedingly large when the magnitude of the constituent integers is huge in the large metabolic networks. When inaccurate numerical coefficients are acceptable, however, non-integer rational numbers can be approximated by floating-point numbers, and numerically rationalized, thus resulting in the relevant integer numerator and denominator. INPUT 2 in Fig. 3 presents the resultant pathway and the normalized flux distribution.

4.2. Identification of MMP

With the overall reaction (INPUT 2) in hand, all feasible metabolic pathways can be recovered from the initially hypothesized pathway (INPUT 1) comprising the aforementioned metabolic reactions, R_1–R_7. This is accomplished by resorting to algorithms RPIMSG and PBT based on P-graphs. Initial structure in METHOD 2 of Fig. 3 shows the P-graph representation of the initially hypothesized pathway corresponding to the pathway's diagram illustrated in INPUT 1. This P-graph serves as the input to algorithm RPIMSG. In the initial reaction pathway that appears on the left-hand side of METHOD 2, notice that no arrows are indicated on all the arcs connecting to reversible reactions, R_1, R_3, R_4 and R_6, and thus, no selection can be made a priori as to which step, forward or reverse, contributes to the pathway proceeding from the substrate to the product. Algorithm RPIMSG eventually generates the maximal reaction network on the right-hand side of METHOD 2. In this maximal reaction network, arrows are placed in all the arcs, thereby indicating that the forward step participates in the pathway. Moreover, R_6 and R_7 have been eliminated in the reduction part of algorithm RPIMSG. The resultant maximal reaction network becomes the input to algorithm PBT, thereby generating two feasible pathways, labeled as pathway 1 and pathway 2; the former is identical to that obtained by FBA. These two pathways are given in Table 1 and their P-graphs are illustrated in SOLUTION of Fig. 3.

4.3. Identification of MFD

Naturally, the FBA of pathway 2 is straightforward. Specifically, v_3 is maximized by LP subject to the mass-balance constraints in terms of moles of metabolites B,

Table 1
Feasible reaction pathways of a simple example with the overall reaction, $A \rightarrow 2F$ (note that figures are stoichiometric numbers for the corresponding feasible pathways)

Reactions	Pathway 1	Pathway 2
R_1: $A \rightarrow B$	1	1
R_2: $B \rightarrow 2E$	1	
R_3: $E \rightarrow F$	2	2
R_4: $B \rightarrow C$		1
R_5: $C \rightarrow 2E$		1
Overall reaction[a]	$A \rightarrow 2F$	$A \rightarrow 2F$

[a]Each of the metabolic reactions, R_i's, $i = 1, 2, \ldots, 7$, is to be multiplied by the stoichiometric number to reach the overall reaction.

C, and E, and the substrate uptake rate of no more than 2.5 mmol/g DCW h. This results in

$v_1 = 2.5$ mmol/g DCW h,

$v_3 = 5.0$ mmol/g DCW h,

$v_4 = 2.5$ mmol/g DCW h,

$v_5 = 2.5$ mmol/g DCW h

or equivalently, upon normalization,

$$R_1: \quad A \xrightarrow[1]{\tilde{v}_1} B,$$

$$R_3: \quad E \xrightarrow[2]{\tilde{v}_3} 2F,$$

$$R_4: \quad B \xrightarrow[1]{\tilde{v}_4} C,$$

$$R_5: \quad C \xrightarrow[1]{\tilde{v}_5} 2E.$$

In summary, we have obtained 2 flux distributions corresponding to 2 metabolic pathways. In other words, we have MFD and MMP. Moreover, this illustrative example unequivocally demonstrates numerically that when normalized, the metabolic flux is identical to the corresponding stoichiometric number for each metabolic reaction.

5. Application to a model of *E. coli* metabolism

5.1. In silico representation of the E. coli metabolic network

The *in silico* model of *E. coli* metabolism (Schilling et al., 2001) is reconstructed to illustrate the proposed unified approach for identifying MFD and MMP. Fig. 4 depicts an overview of the metabolic network of the model. This network incorporates 52 metabolites (6 extracellular metabolites and 46 intermediates) and 48 metabolic reactions (see Tables 2 and 3). Embedded in the metabolic network are the glycolytic pathway, the pentose phosphate pathway (PPP), the tricarboxylic acid (TCA) cycle, and the energy and redox metabolisms, along with the necessary transport reactions for extracellular metabolites. In addition, growth is quantified by a biomass equation derived from the drain of biosynthetic precursors (11 intermediates) into *E. coli* biomass with their appropriate ratios (see Table 2).

Of extracellular metabolites, glucose is regarded as the only carbon source consumed through the system while metabolic products, i.e., ethanol and acetate, and biomass are allowed to be secreted or accumulated; however, all the intermediates are equally constrained. Fueling of the metabolic network is rendered possible by a constrained amount of glucose (<10 mmol/g DCW h) representing limited substrate availability, along with unconstrained uptake/secretion routes for inorganic

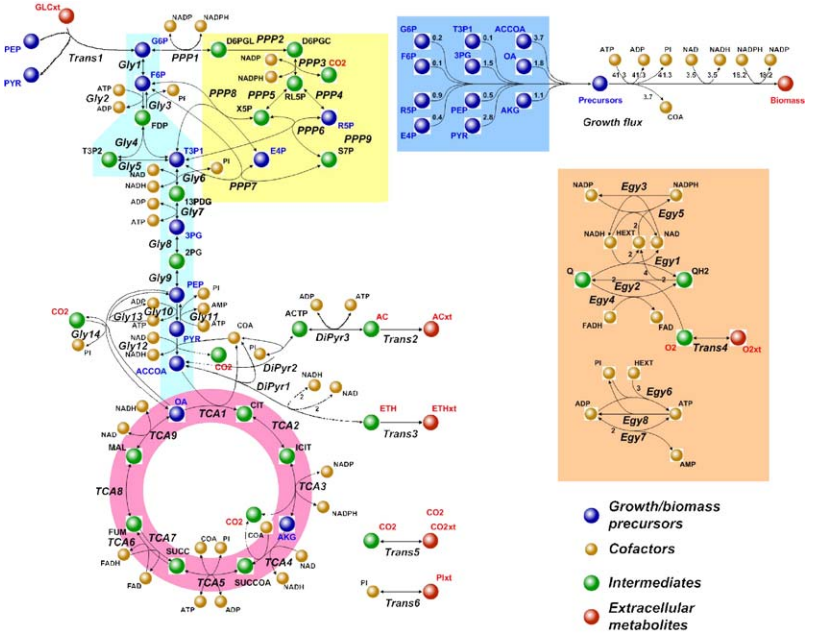

Fig. 4. Overview of the metabolic network of *E. coli* model (Schilling et al., 2001): The network consists of 52 metabolites (6 extracellular metabolites and 46 intermediates) and 48 reactions (24 reversible and 24 irreversible reactions); when growth is quantified by the biomass equation derived from the drain of precursors (11 intermediates) into *E. coli* biomass with their appropriate ratios, biomass is regarded as an additional extracellular metabolite within the model; and 11 precursors for the production of biomass and cofactors should be equally constrained in FBA and be regarded as intermediates to satisfy axiom R3 (see Tables 2 and 3).

Table 2
Metabolic reactions for the *E. coli* model

Enzyme	Gene	Rxn no.	Reaction	EC no.
Membrane transport (*6*)		*Trans*		
Phosphotransferase system	*Pts*	1	GLCxt + PEP → G6P + PYR	
Acetate transport		2	AC → ACxt	
Ethanol transport		3	ETH → ETHxt	
Oxygen transport		4	O2 ↔ O2xt	
Carbon dioxide transport		5	CO2 ↔ CO2xt	
Phosphate transport		6	PI ↔ PIxt	
Glycolysis (*14*)		*Gly*		
Phosphoglucose isomerase	*pgi*	1	G6P ↔ F6P	5.3.1.9
Phosphafructokinase	*pfkA*	2	F6P + ATP → ADP + F16P	2.7.1.11
Fructose-1,6-bisphosphatase	*fbp*	3	FDP → F6P + PI	3.1.3.11
Fructose-1,6-bisphophate aldolase	*fba*	4	FDP ↔ T3P1 + T3P2	4.1.2.13
Triosphosphate isomerase	*tpiA*	5	T3P1 ↔ T3P2	5.3.1.1
Glyceraldehyde-3-phosphate dehydrogenase	*gapA*	6	T3P1 + PI + NAD ↔ NADH + 13PDG	1.2.1.12
Phosphoglycerate kinase	*pgk*	7	13PDG + ADP ↔ ATP + 3PG	2.7.2.3
Phosphoglycerate mutase	*gpmA*	8	3PG ↔ 2PG	5.4.2.1
Enolase	*eno*	9	2PG ↔ PEP	4.2.1.11
Pyruvate kinase	*pyk*	10	PEP + ADP → ATP + PYR	2.7.1.40
PEP synthase	*ppsA*	11	PYR + ATP → AMP + PI + PEP	2.7.9.2
Pyruvate dehydrogenase	*lpdA, aceEF*	12	PYR + COA + NAD → NADH + CO2 + ACCOA	1.2.4.1, 2.3.1.12
PEP carboxykinase	*pckA*	13	OA + ATP → ADP + PEP + CO2	4.1.1.49
PEP carboxylase	*ppc*	14	PEP + CO2 → PI + OA	4.1.1.31
Pentose phosphate pathway (*8*)		*PPP*		
Glucose-6-phosphate dehydrogenase	*zwf*	1	G6P + NADP ↔ NADPH + D6PGL	1.1.1.49
6-Phophogluconolactonase	*pgl*	2	D6PGL → D6PGC	3.1.1.31
6-Phosphogluconate dehydrogenase	*gnd*	3	D6PGC + NADP → NADPH + CO2 + RL5P	1.1.1.44
Ribose-5-phosphate isomerase	*rpiA*	4	RL5P ↔ R5P	5.3.1.6
Ribulose phosphate 3-epimerase	*rpe*	5	RL5P ↔ X5P	5.1.3.1
Transketolase 1	*tktAB*	6	R5P + X5P ↔ T3P1 + S7P	2.2.1.1
Transaldolase	*talB*	7	T3P1 + S7P ↔ E4P + F6P	2.2.1.2
Transketolase 2	*tktAB*	8	X5P + E4P ↔ F6P + T3P1	2.2.1.1
TCA cycle (*9*)		*TCA*		
Citrate synthase	*gltA*	1	ACCOA + OA → COA + CIT	4.1.3.7
Aconitase	*acnA*	2	CIT ↔ ICIT	4.2.1.3
Isocitrate dehydrogenase	*icdA*	3	ICIT + NADP ↔ CO2 + NADPH + AKG	1.1.1.42
2-Ketoglutarate dehydrogenase	*sucAB lpdA*	4	AKG + NAD + COA → CO2 + NADH + SUCCOA	1.2.4.2, 2.3.1.61, 1.8.1.4
Succinate thiokinase	*sucCD*	5	SUCCOA + ADP + PI ↔ ATP + COA + SUCC	6.2.1.5
Succinate dehydrogenase	*sdhABCD*	6	SUCC + FAD → FADH + FUM	1.3.99.1
Fumurate reductase	*frdABCD*	7	FUM + FADH → FAD + SUCC	1.3.99.1
Fumarase	*fumAB*	8	FUM ↔ MAL	4.2.1.2
Malate dehydrogenase	*mdh*	9	MAL + NAD ↔ NADH + OA	1.1.1.37
Dissimilation of pyruvate (*3*)		*DiPyr*		
Acetaldehyde dehydrogenase	*adhE*	1	ACCOA + 2 NADH ↔ 2 NAD + COA + ETH	1.2.1.10
Phosphotransacetylase	*pta*	2	ACCOA + PI ↔ COA + ACTP	2.3.1.8
Acetate kinase	*ackA*	3	ACTP + ADP ↔ ATP + AC	2.7.2.1
Energy/redox metabolism (*8*)		*Egy*		
NADH dehydrogenase I	*nuoA*	1	NADH + Q → NAD + QH2 + 2 HEXT	1.6.5.3
Cytochrome oxidase bo3	*cyoA*	2	2 QH2 + O2 → 2 Q + 4 HEXT	1.10.2.2, 1.9.3.1
Pyridine nucleotide transhydrogenase	*pntAB*	3	NADPH + NAD → NADP + NADH	1.6.1.1
Succinate dehydrogenase complex	*sdhABCD*	4	FADH + Q → FAD + QH2	1.3.5.1
Pyridine nucleotide transhydrogenase	*pntAB*	5	NADP + NADH + 2 HEXT → NADPH + NAD	1.6.1.1

Table 2 (*continued*)

Enzyme	Gene	Rxn no.	Reaction	EC no.
F0F1-ATPase	*atpABCD*	6	ADP + PI + 3 HEXT → ATP	3.6.1.34
Adenylate kinase	*adk*	7	ATP + AMP ↔ 2 ADP	2.7.4.3
ATP drain		8	ATP → ADP + PI	
Growth flux (*1*)			41.3 ATP + 3.5 NAD + 18.2 NADPH + 0.2 G6P + 0.1 F6P + 0.9 R5P + 0.4 E4P + 0.1 T3P1 + 1.5 3PG + 0.5 PEP + 2.8 PYR + 3.7 ACCOA + 1.8 OA + 1.1 AKG → 41.3 ADP + 41.3 PI + 3.5 NADH + 18.2 NADP + 3.7 COA + BIOMASS	

Table 3
Abbreviations of metabolites in the reactions of the *E. coli* model

Abbreviation	Compound
Extracellular metabolites (*7*)	
GLCxt	Glucose (external)
ACxt	Acetate (external)
ETHxt	Ethanol (external)
O2xt	Oxygen (external)
CO2xt	Carbon dioxide (external)
PIxt	Phosphate (external)
BIOMASS	Biomass
Intracellular metabolites (*46*)	
13P2DG	1,3-*P*- *d* glycerate
2K3D6PG	2-Dehydro-3-deoxy-6-*P*-gluconate
2PG	2-*P*- *d* glycerate
3PG	3-*P*- *d* glycerate
AC	Acetate
ACCOA	Acetyl-CoA
ACTP	Acetyl-phosphate
ADP	Adenosine diphosphate
AKG	A-Ketoglutarate
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
CIT	Citrate
CO2	Carbon dioxide
COA	Coenzyme A-SH
D6PGC	*d*-6-Phosphoglucono-*δ*-lactone
D6PGL	*d*-6-Phosphogluconate
E4P	Erythrose 4-phosphate
ETH	Ethanol
FDP	Fructose 1,6-diphosphate
F6P	Fructose 6-phosphate
FAD	Flavin adenine dinucleotide
FADH	
FUM	Fumarate
G6P	Glucose 6-phosphate
HEXT	External H^+
ICIT	Isocitrate
MAL	Malate
NAD	Nicotinamide adenine dinucleotide
NADH	
NADP	Nicotinamide adenine dinucleotide phosphate
NADPH	
O2	Oxygen
OA	Oxaloacetate
PEP	Phosphoenolpyruvate
PI	Phosphate (inorganic)
PYR	Pyruvate
Q	Ubiquinone
QH2	Ubiquinol
R5P	Ribose 5-phosphate
RL5P	*d*-Ribulose 5-phosphate
S7P	*d*-Sedoheptulose-7-*P*
SUCC	Succinate
SUCCOA	Succinyl-CoA
T3P1	Glyceraldehyde-3-phosphate
T3P2	Dihydroxyacetone phosphate
X5P	Xylulose-5-phosphate

phosphate, oxygen and carbon dioxide. (Note: g DCW stands for grams dry cell weight.)

Based on the hypothesis that biological functions in the cell evolve optimally under the given environmental conditions (Bialy, 2001; Edwards et al., 2002), the cell behavior can be predicted under various culture conditions by means of FBA. Herein, three scenarios or cases representing different culture conditions are considered for physiologically meaningful results by setting some of the fluxes as the desired targets (objective functions) within the defined system. In the first, the growth flux generating biomass is maximized under nutritionally rich growth conditions (Varma and Palsson, 1994b). In the second, the generation of acetate is maximized at a limited level of oxygen (slightly anaerobic conditions) (Majewski and Domach, 1990; Varma et al., 1993). In the third, the ethanol production is set as the desired target to be maximized within the defined system under highly oxygen-limited or anaerobic conditions (Wong et al., 1999; Varma et al., 1993). In the latter two cases, the use of non-growing cells is assumed for the production of metabolic products to obtain the maximum possible yield.

5.2. Maximization of biomass production

For this first case (maximization of biomass production), the objective function to be maximized in FBA is the growth flux of the biosynthetic routes involving the

aforementioned 11 precursors (intermediates) for the cell growth in terms of biomass production as indicated in the upper right-hand corner of Fig. 4; the availability of glucose is specified to be less than 10 mmol/g DCW h. The maximization has yielded the theoretical maximum growth rate of 1.05 g biomass/g DCW h, which is in accord with the available result (Schilling et al., 2001). Moreover, the net reaction balance in Table 4 is obtained from the net transport flux vector that is calculated from the resultant flux distribution given in Table 5 in conjunction with the stoichiometric coefficients. Fig. 5a exhibits the flux distribution normalized by the glucose uptake rate of 10 mmol/g DCW h, thus resulting in the normalized net reaction balance

$$\mathrm{GLCxt} + 0.390\mathrm{PIxt} + 1.376\mathrm{O2xt}$$

$$\rightarrow 0.105\mathrm{Biomass} + 1.498\mathrm{CO2xt}.$$

The overall reaction is obtained by rendering the stoichiometric coefficients in the above expression to be integers by multiplying all the coefficients with their LCD, thereby giving rise to

$$58610\mathrm{GLCxt} + 80667\mathrm{O2xt} + 22866\mathrm{PIxt}$$

$$\rightarrow 87774\mathrm{CO2xt} + 6180\mathrm{BIOMASS}.$$

Fig. 6 illustrates the P-graph representation of the pathways corresponding to the glycolytic pathway, PPP and TCA cycle of the *E. coli* model. This P-graph, combined with the pathways pertaining to the energy and redox metabolisms exhibited in the middle of the right-hand side of Fig. 4, serves as the input to algorithm RPIMSG for generating the maximal metabolic reaction network. In general, all feasible metabolic pathways, i.e., MMP are generated via algorithm PBT with the maximal metabolic reaction network as its input. Nevertheless, only a single metabolic pathway is recovered for the case under consideration; in fact, this pathway is identical to that recovered by FBA at the outset. Naturally, FBA of this pathway gives rise to the same flux distribution as that obtained at the outset.

5.3. Maximization of acetate production

For this second case (maximization of acetate production), the objective function to be maximized in FBA is the acetate production which can be discerned in the middle portion of Fig. 4. In addition to the limitation of glucose availability, the oxygen supply level is specified at 10 mmol/g DCW h (see Table 4), which signifies slightly anaerobic conditions. Under these conditions, the theoretical maximum acetate production rate of 10 mmol/g DCW h has been obtained in the absence of cell growth. The resultant flux distribution is given in Table 5, which gives rise to the net reaction balance given in Table 5. Fig. 5b exhibits the flux distribution normalized by the glucose uptake rate of 10 mmol/g DCW h. This, in turn, leads to the overall reaction,

$$\mathrm{GLCxt} + \mathrm{O2xt} \rightarrow \mathrm{ACxt} + \mathrm{ETHxt} + 2\mathrm{CO2xt},$$

which signifies the external state.

Based on the above overall reaction, eight feasible metabolic pathways have been recovered via algorithms RPIMSG and PBT for the graph-theoretic pathway identification (MPI) from the aforementioned *E. coli* network model and concomitant metabolic reactions. Subsequently, the corresponding 8 flux distributions are obtained through the FBA of each of the 8 metabolic pathways (see Table 5).

Table 4
Net reaction balance equations and the resultant overall reactions of the *E. coli* metabolic model under various culture conditions

Cases	Conditions for FBA	Net reaction balance and corresponding overall reaction (OR)	MMP & MFD	
			Numbers identified	Computational time (s)[a]
Case 1	Max. Biomass production Limited glucose uptake (< 10 mmol/g DCW h)	10.0GLCxt + 3.90PIxt + 13.76O2xt → 1.05Biomass + 14.98CO2xt OR: 58610GLCxt + 80667O2xt + 22866PIxt → 87774CO2xt + 6180BIOMASS	1	0.04
Case 2	Max. ACxt production Limited glucose uptake (< 10 mmol/g DCW h) Limited oxygen uptake (< 10 mmol/g DCW h)	10.0GLCxt + 10.0O2xt → 10.0ACxt + 10.0ETHxt + 20.0CO2xt OR: GLCxt + O2xt → ACxt + ETHxt + 2CO2xt	8	0.241
Case 3	Max. ETHxt production Limited glucose uptake (< 10 mmol/g DCW h) Disallowed oxygen uptake (= 0 mmol/g DCW h)	10.0 GLCxt → 20.0ETHxt + 20.0CO2xt OR: GLCxt → 2ETHxt + 2CO2xt	4	0.111

Abbreviation: ACxt, acetate (external); CO2xt, carbon dioxide (external); ETHxt, ethanol (external); GLCxt, glucose (external); O2xt, Oxygen (external); PIxt, phosphate (external); FBA, flux balance analysis; MMP, multiple metabolic pathways; MFD, multiple flux distributions.

[a]PC (Pentium-IV 1.8 GHz, 768 MB RAM).

Table 5
Resultant flux distributions for three cases and the corresponding multiple flux distributions (unit: mmol/g DCW h)

Flux	Case 1[a]	Case 2[b]								Case 3[c]			
		Solution 1	Solution 2	Solution 3	Solution 4	Solution 5	Solution 6	Solution 7	Solution 8	Solution 1	Solution 2	Solution 3	Solution 4
Trans1	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
Trans2		10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00				
Trans3		10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	20.00	20.00	20.00	20.00
Trans4	−13.76	−10.00	−10.00	−10.00	−10.00	−10.00	−10.00	−10.00	−10.00				
Trans5	14.98	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00
Trans6	−3.90												
Gly1	1.33	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
Gly2	6.09	10.00	40.00	10.00	10.00	66.67	10.00	10.00	10.00	10.00	30.00	10.00	10.00
Gly3			30.00			56.67					20.00		
Gly4	6.09	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
Gly5	−6.09	−10.00	−10.00	−10.00	−10.00	−10.00	−10.00	−10.00	−10.00	−10.00	−10.00	−10.00	−10.00
Gly6	14.30	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00
Gly7	14.30	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00
Gly8	12.72	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00
Gly9	12.72	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00
Gly10		40.00	10.00	10.00	10.00	10.00	10.00	10.00	66.67	10.00	10.00	10.00	30.00
Gly11	0.87	30.00							56.67				20.00
Gly12	6.18	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00
Gly13					30.00			56.67		20.00			
Gly14	3.06				30.00			56.67		20.00			
PPP1	8.46												
PPP2	8.46												
PPP3	8.46												
PPP4	3.59												
PPP5	4.86												
PPP6	2.64												
PPP7	2.64												
PPP8	2.22												
TCA1	2.28												
TCA2	2.28												
TCA3	2.28												
TCA4	1.12												
TCA5	1.12												
TCA6	1.12												
TCA7													
TCA8	1.12												
TCA9	1.12												
DiPyr1		10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	20.00	20.00	20.00	20.00
DiPyr2		10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00				
DiPyr3		10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00				
Egy1	26.41	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00				
Egy2	13.76	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00				
Egy3		40.00	40.00	40.00	40.00								
Egy4	1.12												
Egy5		40.00	40.00	40.00	40.00								
Egy6	35.96					26.67	26.67	26.67	26.67				
Egy7	0.87	30.00							56.67				20.00
Egy8				30.00			56.67					20.00	
Growth	1.05												

Note: Solution 1 for cases 2 and 3 are obtained by means of FBA; for the resultant net reaction balance equation of each case, see Table 4.
[a]Case 1: maximization of biomass production.
[b]Case 2: maximization of acetate production.
[c]Case 3: maximization of ethanol production.

5.4. Maximization of ethanol production

For this third case (maximization of ethanol production), in addition to the constraint that the glucose availability is limited, the anaerobic condition is imposed (see Table 4). The objective function to be maximized in FBA is the ethanol production which can be achieved through the dissimilation of pyruvate

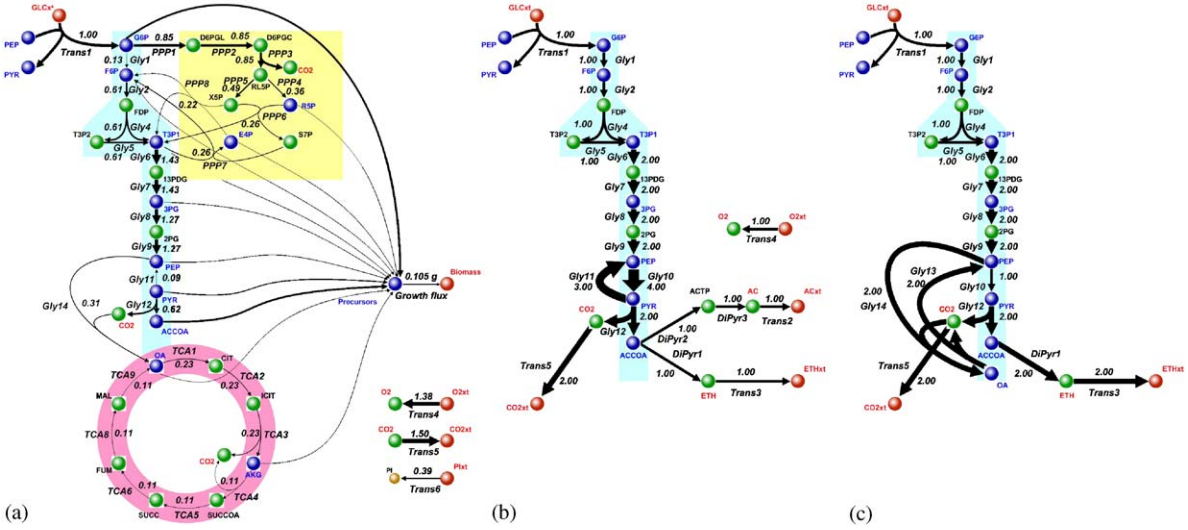

Fig. 5. Normalized flux distributions for (a) the maximum biomass production, (b) maximum acetate production, and (c) maximum ethanol production of *E. coli* on glucose as the substrate: The thickness of arrows is proportional to the value of normalized flux; in the case of (a) the maximum biomass production, the theoretic maximum biomass yield of 0.105 g DCW/mmol glucose is achieved during the growth on glucose; the resultant flux distribution gives rise to the net reaction balance, GLCxt + 0.39PIxt + 1.376O2xt → 0.105Biomass + 1.498CO2xt, which is transformed into overall reaction, 58610GLCxt + 80667O2xt + 22866PIxt → 87774CO2xt + 6180BIOMASS, by multiplying the coefficients by their LCD; and for detail, see the text, Tables 4 and 5.

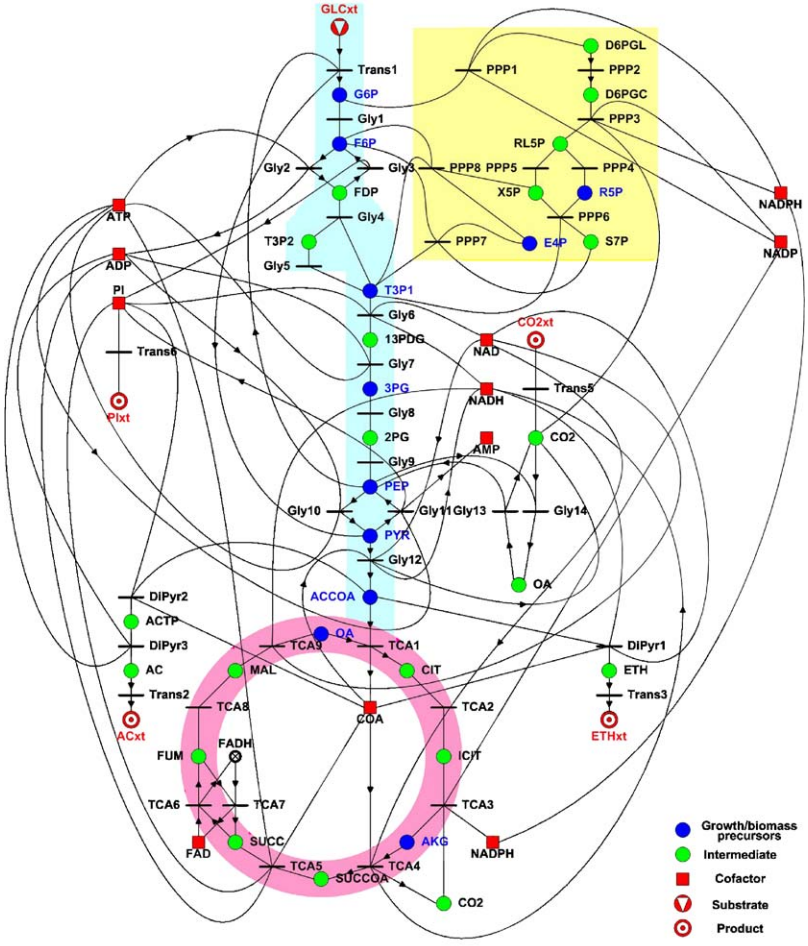

Fig. 6. P-graph representation of the glycolytic pathway, pentose phosphate pathway, TCA cycle and dissimilation of pyruvate in the *E. coli* model: this P-graph can be combined with other pathways of the model, which, in turn, serves as the input to algorithm RPIMSG for generating the maximal metabolic reaction network.

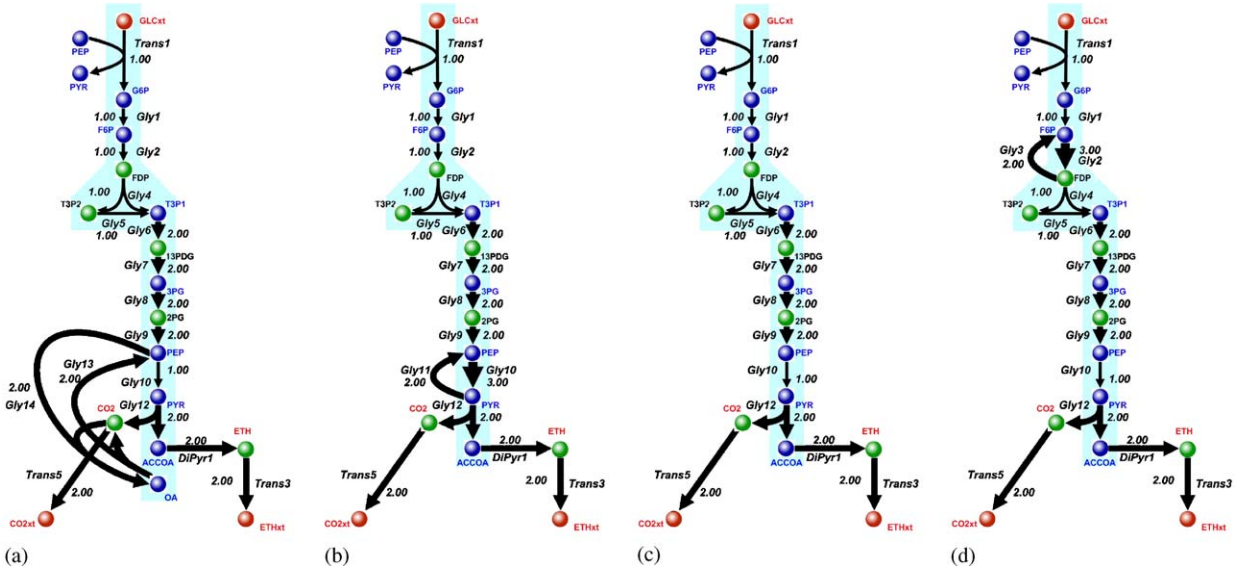

Fig. 7. Normalized multiple flux distributions for the maximum ethanol production: The thickness of the arrows is proportional to the value of normalized flux; there exist the four different flux distributions, (a), (b), (c) and (d), each leading to the same external state, i.e., the net reaction balance, GLCxt → 2 ETHxt + 2 CO2xt; and the values of fluxes in each distributions are given in Table 5.

initiated by its conversion into acetyl-CoA through reaction Gly12 (see Fig. 4). The theoretical maximum ethanol production rate of 20 mmol/g DCW h has been obtained in the absence of cell growth. The resultant flux distribution given in Table 5 leads to the net reaction balance listed in Table 4. Fig. 5c exhibits the flux distribution normalized by the glucose uptake rate of 10 mmol/g DCW h, thus resulting in the overall reaction,

$$\mathrm{GLCxt} \rightarrow 2\mathrm{ETHxt} + 2\mathrm{CO2xt},$$

which represents the external state.

Based on the above overall reaction, four feasible metabolic pathways have been recovered via algorithms RPIMSG and PBT for MPI from the metabolic reactions in the *E. coli* network model. Subsequently, the corresponding 4 flux distributions, illustrated in Fig. 7, are obtained through the FBA of each of the 4 feasible metabolic pathways. In the figure, note that the fluxes are distributed mainly through the glycolytic pathway in all the metabolic pathways; however, they involve distinctly different bypasses through the pathway.

6. Discussion

6.1. Stoichiometric numbers and parameter-independent metabolic fluxes

For the three conditions, i.e., the nutritionally rich growth, slightly anaerobic and anaerobic conditions, considered, the third phase of the proposed approach has yielded the fluxes of the metabolic reactions belonging to every feasible metabolic pathway in the MFD; and the second phase, the stoichiometric numbers of these metabolic reactions. Moreover, it has been demonstrated unequivocally that if normalized with the smallest substrate intake and rendered integers by resorting to their LCD, the fluxes are identical to the stoichiometric numbers provided that each stoichiometric number is regarded as positive for each forward direction and negative for each reverse reaction (see the simple example). This obviously implies that the third phase of the proposed approach, i.e., the identification of MFD corresponding to the MMP identified by MPI in the second phase, is indeed redundant: The stoichiometric numbers of metabolic reactions in each pathway recovered in the second phase can be taken as the parameter-independent metabolic fluxes through the pathway. Naturally, the normalized fluxes can be recovered from the stoichiometric numbers that, in turn, give rises to the corresponding dimensional fluxes for any specific substrate intake rate. This renders it possible and meaningful to systematically generate the stoichiometric numbers of MMP as the multiple MFD in the parameter independent form from the available metabolic reaction databases for various microorganisms under wide ranging environmental conditions. As clearly articulated by several researchers (Lee et al., 2000; Papin et al., 2002; Phalakornkule et al., 2001; Mahadevan and Schilling, 2003), these MMP and the associated MFD have practical implication or utilities besides their obvious theoretical significance. Eventually, they can be organized as a library for broad dissemination. Some of the available metabolic reaction databases are: BioSilico (http://biosilico.kaist.ac.kr/), BRENDA (http://www.brenda.uni-koeln.de/), ENZYME (http://us.expasy.org/

enzyme/), LIGAND (http://www.genome.ad.jp/ligand/), EcoCyc (http://www.ecocyc.org/) and MetaCyc (http://www.metacyc.org/).

6.2. Computational efficacy

The proposed approach appears to be an alternative to the two well-known approaches (Schilling et al., 2000; Schuster et al., 2000) for identifying a complete set of feasible metabolic pathways (MMP) and the concomitant flux distributions (MFD), which give rise to a unique overall reaction representing an external state. These two approaches are the extreme pathways analysis (Schilling et al., 2000) and the elementary mode analysis (Schuster et al., 2000), both of which are based on the convex analysis involving iterative operations of linear algebra. The former identifies all extreme pathways; and the latter, all elementary flux modes. Thus, each generates a complete set of feasible metabolic pathways together with the associated flux distributions for a multitude of the overall reactions representing various external states.

Any algorithm involving the enumeration of all feasible and independent pathways of a network tends to encounter difficulty due to the combinatorial explosion referred to as an NP-hard problem where the complexity is exponential in nature (Mavrovouniotis, 1995; Schilling et al., 2001; Klamt and Stelling, 2002; Schuster et al., 2002). In this regard, Schilling et al. (2001) state: "… as the size of the network increases in a linear fashion, the time to calculate the extreme pathways as well as the number of pathways increases in an exponential fashion." Moreover, according to Schuster et al. (2002), "… an elementary flux mode is a minimal set of enzymes that can operate at steady state with all irreversible reactions proceeding in the appropriate direction. In complex and dense networks, the computation of elementary modes often meets with the problem of combinatorial explosion." To cope with the difficulty due to this combinatorial explosion, the decomposition algorithm for metabolic networks based on the local connectivity of metabolites (Schuster et al., 2002) and the optimized algorithm in terms of speed, memory requirement and numerical stability (Klamt et al., 2003) have been developed. In our proposed approach, the difficulty in solving the combinatorial explosion problem can also be overcome by the profound reduction in the space in which at most the combinatorially feasible pathways are searched even in the worst case. This is partly accomplished by eliminating metabolic reactions which cannot be contained in any of such pathways through the construction of the maximal metabolic reaction network with the aid of algorithm RPIMSG at the outset of the graph-theoretic pathway identification. This algorithm is an exact polynomial algorithm (Fan et al., 1999, 2001, 2002; Friedler et al., 1993).

The proposed approach's computational efficacy is also partly attributable to the fact that the final selection of the feasible metabolic pathways directly from the maximal reaction network by algorithm PBT is immeasurably accelerated by means of a unique branch-and-bound scheme. A traditional implicit enumeration procedure highly likely generates an enormous number of redundant combinations. The number of combinations, however, can be significantly reduced by resorting to an axiom system expressing the obvious combinatorial properties inherent in feasible metabolic pathways. This scheme exploits the structural features of feasible metabolic reaction networks as expressed by axioms T1–T7.

In the simple example, only 3 pathways are combinatorially feasible among $(3^4 \cdot 2^3 - 1)$, or 647, possible pathways from the 4 reversible and 3 irreversible reactions. For the third case (maximization of ethanol production) of the *E. coli* model, only 236 pathways are combinatorially feasible among $(3^{25} \cdot 2^{23} - 1)$ possible combinations from the 25 reversible and 23 irreversible reactions; this number is merely 0.0001% of the total number of possible networks. Note that eliminating the majority, e.g., 99%, of infeasible or redundant pathways to accelerate the search by many orders of magnitude can hardly be achieved by any conventional Mixed Integer Programming (MIP) method. The computational efficacy of our approach is amply reflected in the computational time of less than 1 s with a PC (Intel Pentium IV, 1.8 GHz, 768 MB RAM) to obtain the results for each of the 3 cases (see Table 1).

6.3. Large-size application and cell robustness

The unified approach proposed is applied to problems of maximal production of various metabolic products by *E. coli* based on its large-scale model involving as many as 300 metabolic reactions. The preliminary results obtained unequivocally indicate that problems of even such complexity are amenable to the proposed approach when implemented on a PC of modest capacity, e.g., the one adopted in the current work. In fact, the approach has given rise to hundreds or thousands of MMP under various culture conditions. This pathway redundancy implies cell robustness which is a unique feature of complex systems (Carlson and Doyle, 2002; Stelling et al., 2002): the network function can be sustained towards some internal disturbances, e.g., gene mutations by possibly redistributing fluxes through one of the MMP. Details will be presented in our forthcoming contribution.

The computing efficiency of the proposed approach can be further enhanced by means of parallel or grid computing. Algorithm PBT preserves the useful property of conventional branch-and-bound of being suitable for parallel implementation. Specifically, all

subproblems in the branches of an enumeration tree in algorithm PBT can be divided to generate a master-slave architecture (Varga et al., 1995); subsequently, they can be solved separately in the multi-processor or parallel computing environment. The resulting computing scheme or system can be promising for the analysis of the genome-wide networks with more than 1000 metabolic reactions and 700 metabolites. This would facilitate the discovery of antimicrobial drug targets that are essential for the growth or survival of a pathogen satisfying the principle of selective toxicity.

Conventionally, drug targets are identified by comparative analysis of a host (i.e., human cell) and the microbial pathogen on the metabolic pathway maps, where enzymes unique to the pathogen are regarded as potential drug targets: inactivating those enzymes may selectively threaten the pathogen without harm to the host (Fairlamb, 2002; Karp et al., 1999; Lindroos and Andersson, 2002). The existence of multiple pathways or flux distributions, however, implies that their "back-up" pathways may possibly be activated to perform the same function even though some potential drug-target enzymes identified are disrupted under certain environmental conditions in the cell. Such fault-tolerance or robustness may be a key to cell survival against environmental or genetic change. Thus, inactivating common enzymes among the drug targets involved in MMP may attenuate the cellular function, thereby killing the pathogen. In this regard, the proposed approach would provide novel insight into drug discovery.

7. Concluding remarks

The current work proposes a computationally efficient approach for identifying MMP and the concomitant MFD. These MMP or MFD attain a unique phenotypic state whose manifestation is the overall reaction. Nevertheless, the final identification of a valid flux distribution among the different distributions must await experimental verification by means of various methods, e.g., an isotopic tracer method with nuclear magnetic resonance (NMR) spectroscopy and/or gas chromatography-mass spectrometry (GC-MS). The approach would serve as an alternative to the methods based on convex analysis adopted in most, if not all, of the available approaches. The unified approach proposed complementarily identifies MMP and MFD by judiciously integrating FBA based on LP and the graph-theoretic method for reaction-pathway identification. The results from the application of the unified approach proposed to the *E. coli* model demonstrate its profound efficiency and efficacy and reveal that the stoichiometric numbers of metabolic reactions are identical to the parameter-independent metabolic fluxes. The computational efficacy of the proposed unified approach is mainly attributable to the fact that it resorts to the P-graph. The P-graph, a unique bipartite graph, together with the representation of the structure of any material or molecular transformation network obeying the laws of mass conservation and/or stiochiometry, e.g., metabolic pathway, renders it possible to capture the syntactic and semantic contents of the network and to craft a set of axiomatic statements depicting the network's structure. Such axiomatic statements have naturally given rise to a set of three highly compact and efficient algorithms. The computational efficacy of the proposed unified approach has been ascertained by successfully extending its application to the *E. coli* models involving 300 and 700 metabolic reactions, the results which will be in our forthcoming contributions. The resultant MMP and MFD attaining a unique external state imply the surprising adaptability and robustness of the intricate cellular network as a key to cell survival against environmental or genetic change. It is highly plausible that the approach proposed herein is applicable to drug discovery.

Acknowledgments

This work was supported by the Korean Systems Biology Research Program (M10309020000-03B5002-00000) of the Ministry of Science and Technology and by the BK21 project. Further supports by the LG Chemicals Chair Professorship, IBM-SUR program, and the Center for Ultramicrochemical Process Systems sponsored by KOSEF are appreciated. The participation of the second, fifth, sixth and seventh authors in the project was financially supported by their respective institutions.

Appendix A. Supplementary Materials

The online version of this article contains additional supplementary data. Please visit doi:10.1016/j.ymben.2005.02.002.

References

Bialy, H., 2001. Living on the edges. Nat. Biotechnol. 19, 111–112.
Bonarius, H.P.J., Schmid, G., Tramper, J., 1997. Flux analysis of underdetermined metabolic networks: the quest for the missing constraints. Trends Biotechnol. 15, 308–314.
Boudart, M., Djega-Mariadassou, G., 1984. Kinetics of Heterogeneous Catalytic Reactions. Princeton University Press, Princeton.
Carlson, J.M., Doyle, J., 2002. Complexity and robustness. Proc. Natl Acad. Sci. USA 99, 2538–2545.
Clarke, B.L., 1988. Stoichiometric network analysis. Cell Biophys. 12, 237–253.
Edwards, J.S., Palsson, B.Ø., 1998. How will bioinformatics influence metabolic engineering? Biotechnol. Bioeng. 58, 162–169.

Edwards, J.S., Ramakrishna, R., Schilling, C.S., Palsson, B.Ø., 1999. Metabolic flux balance analysis. In: Lee, S.Y., Papoutsakis, E.T. (Eds.), Metabolic Engineering. Marcel Dekker, New York, pp. 13–57.

Edwards, J.S., Covert, M., Palsson, B., 2002. Metabolic modelling of microbes: the flux-balance approach. Environ. Microbiol. 4, 133–140.

Fairlamb, A.H., 2002. Metabolic pathway analysis in trypanosomes and malaria parasites. Phil. Trans. R. Soc. Lond. B. 357, 101–107.

Fan, L.T., Bertók, B., Friedler, F., 1999. Combinatorial framework for the systematic generation of reaction pathways. Paper Presented at the AIChE Annual Meeting, Dallas, TX, ISBN 0-8169-0805-2.

Fan, L.T., Bertók, B., Friedler, F., Shafie, S., 2001. Mechanisms of ammonia-synthesis reaction revisited with the aid of a novel graph-theoretic method for determining candidate mechanisms in deriving the rate law of a catalytic reaction. Hungarian J. Ind. Chem. 29, 71–80.

Fan, L.T., Bertók, B., Friedler, F., 2002. A graph-theoretic method to identify candidate mechanisms for deriving the rate law of a catalytic reaction. Comput. Chem. 26, 265–292.

Fell, D., 1996. Understanding the Control of Metabolism. Portland Press, London.

Friedler, F., Tarjan, K., Huang, Y.W., Fan, L.T., 1992. Graph-theoretic approach to process synthesis: axioms and theorems. Chem. Eng. Sci. 47, 1973–1988.

Friedler, F., Tarjan, K., Huang, Y.W., Fan, L.T., 1993. Graph-theoretic approach to process synthesis: polynomial algorithm for maximum structure generation. Comput. Chem. Eng. 17, 929–942.

Friedler, F., Varga, J.B., Fan, L.T., 1995. Decision mapping: a tool for consistent and complete decisions in process synthesis. Chem. Eng. Sci. 50, 1755–1768.

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

Imreh, B., Friedler, F., Fan, L.T., 1996. An algorithm for improving the bounding procedure in solving process network synthesis by a branch-and-bound method. In: Bomze, I., Csendes, T., Horst, R., Pardalos, P. (Eds.), Nonconvex Optimization and its Applications: Developments in Global Optimization. Kluwer Academic Publishers, Dordrecht, pp. 315–348.

Kacser, H., Burns, J., 1973. The control of flux. Symp. Soc. Exp. Biol. 27, 65–104.

Karp, P.D., Krummenacker, M., Paley, S., Wagg, J., 1999. Integrated pathway-genome databases and their role in drug discovery. Trends Biotechnol. 17, 275–281.

Klamt, S., Stelling, J., 2002. Combinatorial complexity of pathway analysis in metabolic networks. Mol. Biol. Rep. 29, 222–236.

Klamt, S., Schuster, S., Gilles, E.D., 2002. Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism in purple nonsulfur bacteria. Biotechnol. Bioeng. 77, 734–751.

Klamt, S., Stelling, J., Ginkel, M., Gilles, E.D., 2003. FluxAnalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps. Bioinformatics 19, 261–269.

Küffner, R., Zimmer, R., Lengauer, T., 2000. Pathway analysis in metabolic databases via differential metabolic display (DMD). Bioinformatics 16, 825–836.

Lee, D.-Y., Yun, H., Lee, S.Y., Park, S., 2003. MetaFluxNet: the management of metabolic reaction information and quantitative metabolic flux analysis. Bioinformatics 19, 2144–2146.

Lee, S., Phalakornkule, C., Domuch, M.M., Grossman, I.E., 2000. Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. Comput. Chem. Eng. 24, 711–716.

Lee, S.Y., Papoutsakis, E.T. (Eds.), 1999. Metabolic Engineering. Marcel Dekker, New York.

Liao, J.C., Hou, S.-Y., Chao, Y.-P., 1996. Pathway analysis, engineering, and physiological considerations for redirecting central metabolism. Biotechnol. Bioeng. 52, 129–140.

Lindroos, H., Andersson, S.G.E., 2002. Visualizing metabolic pathways: comparative genomics and expression analysis. P. IEEE. 90, 1793–1802.

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

Majewski, R.A., Domach, M.M., 1990. Simple constrained optimization view of acetate overflow in *E. coli*. Biotechnol. Bioeng. 35, 732–738.

Mavrovouniotis, M.L., 1995. Symbolic and quantitative reasoning: design of reaction pathways through recursive satisfaction of constraints. In: Stephanopoulos, G. (Ed.), Advanced in Chemical Engineering, vol. 21. Academic Press, New York, pp. 147–186.

Mavrovouniotis, M.L., Stephanopoulos, G., Stephanopoulos, G., 1990. Computer aided synthesis of biochemical pathways. Biotechnol. Bioeng. 36, 1119–1132.

Murata, T., 1989. Petri nets: properties, analysis and applications. P. IEEE. 77, 541–580.

Oliveira, J.S., Bailey, C.G., Jones-Oliveira, J.B., Dixon, D.A., 2001. An algebraic-combinatorial model for the identification and mapping of biochemical pathways. Bull. Math. Biol. 63, 1163–1196.

Papin, J.A., Price, N.D., Edwards, J.S., Palsson, B.Ø., 2002. The genome-scale metabolic extreme pathway structure in *Haemophilus influenzae* shows significant network redundancy. J. Theor. Biol. 215, 67–82.

Phalakornkule, C., Lee, S., Zhu, T., Koepsel, R., Ataai, M.M., Grossmann, I.E., Domuch, M.M., 2001. A MILP-based flux alternative generation and NMR experimental design strategy for metabolic engineering. Metab. Eng. 3, 124–137.

Reddy, V.N., Liebman, M.N., Mavrovouniotis, M.L., 1996. Qualitative analysis of biochemical reaction systems. Comput. Biol. Med. 26, 9–24.

Reed, J.L., Palsson, B.Ø., 2004. Genome-scale in silico models of *E. coli* have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states. Genome Res. 14, 1797–1805.

Reed, J.L., Vo, T.D., Schilling, C.H., Palsson, B.Ø., 2003. An expanded genome-scale model of *Escherichia coli* K-12 (iJR903 GSM/GPR). Genome Biol. 4, R54.1–R54.12.

Roberts, F.S., 1984. Applied Combinatorics. Prentice-Hall, Englewood Cliffs, NJ.

Sauer, U., Cameron, D.C., Bailey, J.E., 1998. Metabolic capacity of *Bacillus subtilis* for the production of purine nucleosides, riboflavin, and folic acid. Biotechnol. Bioeng. 59, 227–238.

Schilling, C.H., Edwards, J.S., Palsson, B.Ø., 1999. Towards metabolic phenomics: analysis of genomic data using flux balances. Biotechnol. Prog. 15, 288–295.

Schilling, C.H., Letscher, D., Palsson, B.Ø., 2000. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. J. Theor. Biol. 203, 229–248.

Schilling, C.H., Edwards, J.S., Letscher, D., Palsson, B.Ø., 2001. Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. Biotechnol. Bioeng. 71, 286–306.

Schuster, S., Dandekar, T., Fell, D.A., 1999. Detection of elementary flux modes in biochemical networks: a promising tool for analysis and metabolic engineering. Trends Biotechnol. 17, 53–60.

Schuster, S., Fell, D.A., Dandekar, T., 2000. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat. Biotechnol. 18, 326–332.

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

Seo, H., Lee, D.-Y., Park, S., Fan, L.T., Shafie, S., Bertók, B., Friedler, F., 2001. Graph-theoretical identification of pathways for biochemical reactions. Biotechnol. Lett. 23, 1551–1557.

Seressiotis, A., Bailey, J.E., 1988. MPS: an artificially intelligent software system for the analysis and synthesis of metabolic pathways. Biotechnol. Bioeng. 31, 587–602.

Simpson, T.W., Follstad, B.D., Stephanopoulos, G., 1999. Analysis of the pathway structure of metabolic networks. J. Biotechnol. 71, 207–223.

Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., Gilles, E.D., 2002. Metabolic network structure determines key aspects of functionality and regulation. Nature 420, 190–193.

Stephanopoulos, G.N., Aristidou, A.A., Nielsen, J., 1998. Metabolic Engineering. Academic Press, New York.

Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T.S., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J.C., et al., 1999. E-CELL: software environment for whole-cell simulation. Bioinformatics 15, 72–84.

Varga, J.B., Friedler, F., Fan, L.T., 1995. Parallelization of the accelerated branch-and-bound algorithm of process synthesis: application in total flowsheet synthesis. Acta Chim. Slovenica 42, 15–20.

Varma, A., Palsson, B.Ø., 1994a. Metabolic flux balancing: basic concepts, scientific and practical use. Bio-Technology 12, 994–998.

Varma, A., Palsson, B.Ø., 1994b. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wildtype *Escherichia coli* W3110. Appl. Environ. Microbiol. 60, 3724–3731.

Varma, A., Boesch, B.W., Palsson, B.Ø., 1993. Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. Appl. Environ. Microbiol. 59, 2465–2473.

Voet, D., Voet, J.G., 1995. Biochemistry. Wiley, New York.

Wong, H.H., van Wegen, R.J., Choi, J.-I., Lee, S.Y., Middelberg, A.P.J., 1999. Metabolic analysis of poly(3-hydroxybutyrate) production by recombinant *Escherichia coli*. J. Microbiol. Biotechnol. 9, 593–603.