Bioprocess Biosyst Eng (2013) 36:1199–1203
DOI 10.1007/s00449-012-0847-5

ORIGINAL PAPER

Relationship between extreme pathways and structurally minimal pathways

Mate Barany · Botond Bertok · L. T. Fan · Ferenc Friedler

Received: 6 August 2012 / Accepted: 2 October 2012 / Published online: 10 November 2012

Abstract The determination of reaction pathways is one of the most important functions that should be performed in exploring the kinetics of catalyzed chemical reactions or biochemical reactions, the latter being generally catalyzed by enzymes. It is proven that the terms, "type-I extreme pathway" and "structurally minimal pathway", both introduced to characterize the kinetics of a catalyzed reaction are equivalent. These two terms are based on two distinct methodologies, one mainly rooted in convex analysis and the other in graph theory. The equivalence promises further even more effective methods for reaction-pathway identification by synergistic integration of existing ones.

Keywords Reaction-pathway analysis · Convex analysis · Graph theory · Metabolism · Catalysis · P-graph

Introduction

Metabolic pathway analysis is of great interest and has been extensively studied in the past decade. Two popular concepts for analyzing metabolic pathways are found in the literature; one is based on extreme pathways [13], and the other on elementary flux modes [15]. Both concepts deploy the toolbox of convex analysis and they are closely related, i.e., extreme pathways are a subset of elementary flux modes [7].

In contrast to extreme pathways and elementary flux modes, the concept of structurally minimal pathways [2] is rooted in graph theory. Moreover, it has been proven in a recent paper [1] that the concept of structurally minimal pathway is equivalent to the concept of direct mechanism [6] in catalytic pathway identification.

The current contribution is aimed at clarifying the relationships between structurally minimal pathways and extreme pathways or elementary flux modes. The relationships to be discovered herein may open novel avenues for integrating steps of different reaction pathway identification methods toward exhaustively analyzing mechanisms of never reached complexity.

M. Barany · B. Bertok (✉) · F. Friedler
Department of Computer Science and Systems Technology, University of Pannonia, Egyetem u. 10, 8200 Veszprém, Hungary
e-mail: bertok@dcs.uni-pannon.hu

L. T. Fan
Department of Chemical Engineering, Kansas State University, Manhattan, Kansas 66506, USA

Extreme pathways

Extreme pathways were introduced by Schilling et al. in [13] and have been extensively deployed for metabolic pathway analysis (see, e.g., [10, 11, 14, 16]). For generating extreme pathways, the metabolic network is represented by an $m \times n$ stoichiometric matrix $\mathbf{S}$ where m is the number of metabolites and n is the number of reactions. These reactions are classified as internal or exchange reactions, based on whether they cross the system boundary, or not.

Exchange fluxes render it possible for metabolites to enter or exit through the theoretical system boundary, and as such they can be visualized as the inputs to and outputs from the system. Reversible internal reactions are considered as two reactions in opposite directions. Exchange reactions can be reversible but each metabolite can participate in at most one exchange reaction.

Under the assumption that the system is in steady state, material balances can be described as

$$\mathbf{Sv} = \mathbf{0} \tag{1}$$

where $\mathbf{v} = [v_1, v_2, \ldots, v_n]^{\mathrm{T}}$ is an n-dimensional vector representing the flux through each reaction. This is the (right) null space of **S** and it contains the steady-state flux distributions. Each feasible steady state is represented by a flux vector **v** and the null space of **S** corresponds to the set of all solutions of Eq. (1). To completely describe the system, the following mass balance constraints have to be taken account. Internal fluxes must be non-negative yielding

$$v_i \geq 0, \quad \forall i. \tag{2}$$

The constraint on exchange flux b_j depends on the input and output status of the metabolite. The lower bound (α_j) and the upper bound (β_j) can be either 0, $-\infty$, or ∞ based on the direction of the exchange flux. If the exchange flux is bidirectional, i.e., both a source and a sink is present for the metabolite, then α_j is set to $-\infty$ and β_j to ∞ leaving the exchange flux unconstrained. Thus, the constraint can be written as

$$\alpha_j \leq b_j \leq \beta_j. \tag{3}$$

The stoichiometric matrix is usually structured in the following manner. The columns representing the internal reactions are moved to the first part of the matrix followed by the columns representing the exchange reactions. This is illustrated by an example taken from [8]. Figure 1 depicts a simple biochemical network consisting of 7 metabolites, 11 internal and 3 exchange fluxes. All three exchange fluxes are unconstrained. This network can be formulated according to Eqs. (1)–(3) as

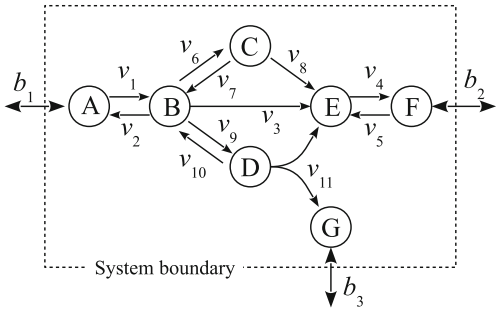

Fig. 1 Biochemical network consisting of 7 metabolites, 11 internal and 3 exchange fluxes

where

$$v_i \geq 0; \quad i = 1, 2, \ldots, 11, \quad -\infty \leq b_j \leq \infty, \quad j = 1, 2, 3. \tag{5}$$

Thus, the metabolic network in steady state is represented by a system of linear equalities and inequalities, thereby giving rise to convex analysis. The solution set for this system can be described geometrically as convex polyhedral cone emanating from the origin in the n-dimensional space. Within this cone, all the possible steady-state solutions lie; it is called as the steady-state flux cone.

In convex analysis, the edges of the cone are half-lines emanating from the origin and are called extreme rays. These rays are said to generate the cone and are systematically independent, since they cannot be decomposed into a non-trivial convex combination of any other extremal rays or vectors residing in the cone. In contrast to the basis concept of linear algebra, this minimal generating set is unique.

In the context of metabolic systems, the edges of the cone are termed as extreme pathways [13] as each edge

$$\begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 1 & -1 & -1 & 0 & 0 & -1 & 1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 2 & -1 & 1 & 0 & 0 & 2 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \\ v_9 \\ v_{10} \\ v_{11} \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{4}$$

corresponds to a particular pathway which satisfies Eqs. (1)–(3). Every point within the cone (C) can be written as a convex combination of extreme pathways. By denoting the extreme pathways by $\mathbf{p}_i$ and the total number of extreme pathways to generate C for a system by k, we have

$$C = \left\{ \mathbf{v} : \mathbf{v} = \sum_{i=1}^{k} w_i \mathbf{p}_i, \quad w_i \geq 0, \quad \forall i \right\}. \tag{6}$$

Extreme pathways can be grouped into three classes according to their exchange fluxes. The most important pathways are the type I extreme pathways, which involve the conversion of primary inputs into primary outputs and thus contain exchange fluxes with the environment. These pathways are the major contributors to the decomposition of almost any steady-state flux distribution. For additional details, see [9, 13].

Structurally minimal pathways

The P-graph framework [4, 5] is an effective method for solving realistic and exceedingly complex industrial process synthesis problems. The method has been adapted by Fan et. al. [2] for reaction-pathway identification and has been applied for metabolic pathway analysis in [3, 8, 12].

The P-graph is a unique bipartite graph which unambiguously represents the structure of chemical or biochemical networks. In the graphical representation of a P-graph, elementary biochemical reactions are represented by horizontal bars while metabolites by circles. Figure 2 depicts the P-graph representation of the simple example taken from [8].

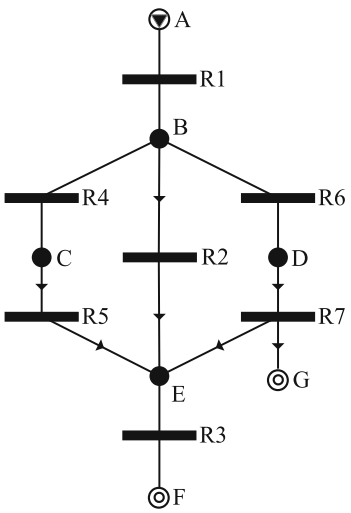

Fig. 2 P-graph representing a biochemical network consisting of seven metabolites, four reversible and three irreversible reactions

For formal definition, let $\mathcal{M} = \{m_1, m_2, \ldots, m_m\}$ be the finite ordered set of metabolites and $\mathcal{O} = \{\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_n\}$ be the finite ordered set of reactions. Every reaction $\mathbf{o}_i$ is represented by an m-dimensional vector of real numbers, $\mathbf{o}_i = [o_{1,i}, o_{2,i}, \ldots, o_{m,i}]^{\mathrm{T}} \in \mathbb{R}^m$ where $o_{j,i}$ indicates the difference of the rates of consumption and production of the metabolite, $m_j (j = 1, 2, \ldots, m)$ by reaction i. Similarly to extreme pathway analysis, reversible elementary reactions are considered as two reactions operating in opposite directions. The overall reaction denoted by $\mathbf{E}$ is represented by an m-dimensional vector of real numbers $\mathbf{E} = [E_1, E_2, \ldots, E_m]^{\mathrm{T}} \in \mathbb{R}^m$ where E_j signifies the difference of the production and consumption rates of the metabolite $m_j (j = 1, 2, \ldots, m)$ by the overall reaction.

Formally, a P-graph is defined as pair (m, o), where $m \subseteq \mathcal{M}$ is the set of metabolites while $o \subseteq \mathcal{O}$ is the set of reactions. P-graph (m', o') is a subgraph of P-graph (m'', o''), i.e., $(m', o') \subseteq (m'', o'')$, if $m' \subseteq m''$ and $o' \subseteq o''$. The union of P-graphs (m', o') and (m'', o'') is defined to be the P-graph $(m' \cup m'', o' \cup o'')$.

A P-graph is termed a feasible pathway if it satisfies a set of axioms [2]. These axioms can be formulated as Eqs. (7)–(9). Equation (7) implies that P-graph (m, o) represents a feasible pathway if the overall reaction $\mathbf{E}$ can be written as a linear combination of the elementary reactions in set o with positive coefficients. Thus,

$$\exists \boldsymbol{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_n]^{\mathrm{T}} : \sum_{o_i \in o} \lambda_i o_i = \mathbf{E}, \quad o_i \in o \Longleftrightarrow \lambda_i > 0. \tag{7}$$

Note that the above equation implies that the system must be at steady state. The following equation states that all elementary reactions represented in the pathway must be defined a priori, i.e.,

$$o \subseteq \mathcal{O}. \tag{8}$$

Finally, the following equation states that the network representing a feasible pathway must be acyclic;

$$\nexists o' : o' \subseteq o, o' \neq \emptyset, \exists \boldsymbol{\lambda}' = [\lambda'_1, \lambda'_2, \ldots, \lambda'_n]^{\mathrm{T}} : \sum_{o_i \in o} o_i \lambda'_i = \mathbf{0},$$
$$o_i \in o' \Longleftrightarrow \lambda'_i > 0. \tag{9}$$

A P-graph is termed a structurally minimal pathway or "independent pathway" if it represents a feasible pathway and none of its proper subgraphs can represent a feasible pathway. P-graph (m, o) is termed structurally minimal pathway, if it satisfies the following: there exist such positive coefficients for the reaction steps included in o, that the overall reaction can be written as a linear combination of the elements of o with the given

coefficients. Moreover (m, o) is minimal in the sense that it has no proper subgraphs satisfying this criteria.

Equivalence of type-I extreme pathways and structurally minimal pathways

One of the two different approaches introduced so far is based on convex analysis, and the other is based on graph theory. Unlike the approach based on convex analysis, the overall reaction is defined a priori in the approach based on graph theory. On the other hand, unlike the approach based on graph theory, the fluxes assigned to the input and output metabolites may vary in the approach based on convex analysis. The P-graph framework identifies the structurally minimal pathways leading from the input (starting) metabolites to the output (product) metabolites, which are defined by the vector of the overall reaction. Similarly, type-I extreme pathways involve the conversion of primary inputs into primary outputs. In what follows, it will be proven that for a given overall reaction, the type-I extreme pathways and structurally minimal pathways are equivalent. The following is a theorem resultant from [13]:

Theorem 1 *A convex flux cone determined by Eqs.* (1)–(3) *has a set of systematically independent generating vectors. Furthermore, these generating vectors (extremal rays) are unique up to a multiplication by a positive scalar. The generating vectors are called extreme pathways.*

First, it must be shown that for a given overall reaction, the convex cone determined by Eqs. (1)–(3) is identical to the cone determined by Eqs. (7)–(9). Note that the reaction vectors $\mathbf{o}_i = [o_{1,i}, o_{2,i}, \ldots, o_{m,i}]^{\mathrm{T}}$ are the columns of the stoichiometric matrix; thus, Eq. (7) can be rewritten as

$$\mathbf{S}'\boldsymbol{\lambda} = \mathbf{E}, \quad \lambda_i \geq 0 \quad i = 1, 2, \ldots, n \tag{10}$$

where $\mathbf{S}' \in \mathbb{R}^{m \times n}$, $\boldsymbol{\lambda} \in \mathbb{R}^n$ and $\mathbf{E} \in \mathbb{R}^m$. The difference between the matrices $\mathbf{S}$ and $\mathbf{S}'$ and the vectors $\mathbf{v}$ and $\boldsymbol{\lambda}$ is the absence of the exchange flux components from $\mathbf{S}'$ and $\boldsymbol{\lambda}$. Thus, the column dimension of $\mathbf{S}$ and the dimension of $\mathbf{v}$ is always greater than the dimension of $\mathbf{S}'$ and $\boldsymbol{\lambda}$, respectively.

There can be only one exchange flux per metabolite [13]. The model defined by Eqs. (1)–(3), therefore, can be "generalized" as follows: append an $m \times m$ diagonal matrix denoted by $\mathbf{I}'$ to the first n columns of $\mathbf{S}$. If a metabolite has an exchange flux assigned to it, then $I'_{m,m}$ is -1 and 0 otherwise; $\mathbf{v}$ is expanded accordingly. Thus, $\mathbf{S} \in \mathbb{R}^{m \times (n+m)}$ and $\mathbf{v} \in \mathbb{R}^{(n+m)}$, thereby resulting in the "generalized" form of Eq. (4) as given below.

$$\begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 0 & 0 & -1 & 1 & 0 & -1 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 2 & -1 & 1 & 0 & 0 & 2 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \\ v_9 \\ v_{10} \\ v_{11} \\ b_1 \\ b_e \\ b_e \\ b_e \\ b_e \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{11}$$

The $v_i \geq 0$ $(i = 1, 2, \ldots, n)$ inequalities are still valid, but for a given overall reaction, the b_j $(j = 1, 2, \ldots, m)$ values will be known. Furthermore, note that the b_j components of $\mathbf{v}$ constitute $\mathbf{E}$, the vector of the overall reaction from Eq. (10), while the v_i components of $\mathbf{v}$, the fluxes through the reactions are the same as the λ_i components of the vector $\boldsymbol{\lambda}$ from Eq. (10). Thus, Eq. (11) has the following structure;

$$[\mathbf{S}'|\mathbf{I}']\begin{bmatrix}\boldsymbol{\lambda}\\ \mathbf{E}\end{bmatrix}=\mathbf{0}. \quad (12)$$

The result of this matrix–vector multiplication can be written as

$$\mathbf{S}'\boldsymbol{\lambda}+\mathbf{I}'\mathbf{E}=\mathbf{0} \quad (13)$$

Since $\mathbf{I}'$ is diagonal and has only -1 as nonzero elements, Eq. (13) becomes

$$\mathbf{S}'\boldsymbol{\lambda}-\mathbf{E}=\mathbf{0} \quad (14)$$

which is the same as Eq. (10). This proves that the convex cone determined by Eqs. (1)–(3) are the same as the cone determined by Eqs. (7)–(9).

Now it has to be shown that like the extreme pathways, the structurally minimal pathways are the edges of the cone, i.e., first, they have to be systematically independent, and second, they have to generate the cone. The property of systematic independence follows from the definition of structurally minimal pathways, i.e., a vector corresponding to a structurally minimal pathway cannot be written as a positive linear combination of other structurally minimal pathways. The second property is proven indirectly.

It has been proven recently that structurally minimal pathways are equivalent to direct mechanisms [1]. The set of all direct mechanisms in a system containing a basis for the vector space of all mechanisms, and unlike a linear algebraic basis, the set of direct mechanisms is a unique property of the system [6]. It follows that the set of structurally minimal pathways is unique and every reaction pathway can be expressed in terms of structurally minimal pathways (i.e., they generate the cone). Thus, the following has been proven.

Theorem 2 *For a given overall reaction, the type-I extreme pathways and structurally minimal pathways are equivalent.*

Corollary 1 *For a given overall reaction, the type-I extreme pathways and direct mechanisms are equivalent.*

Practically, for the illustrative example, the software for generating type-I extreme pathways [17] results vectors [1 0 1 2 0 0 0 0 0 0 0 −1 2 0], [1 0 0 2 0 1 0 1 0 0 0 −1 2 0], and [1 0 0 1 0 0 0 0 1 0 1 −1 1 1] representing reaction networks $R1 + R2 + R3 = (A \Leftrightarrow 2\ F)$, $R1 + R3 + R4 + R5 = (A \Leftrightarrow 2\ F)$, and $R1 + R3 + R6 + R7 = (A \Leftrightarrow F + G)$, respectively. For the overall reaction $(A \Leftrightarrow 2\ F)$, the P-graph software [18] provides structurally minimal pathways $(\{A, B, E, F\}, \{R1 \rightarrow, R2 \rightarrow, R3 \rightarrow\})$ and $(\{A, B, C, E, F\}, \{R1 \rightarrow, R3 \rightarrow, R4 \rightarrow, R5 \rightarrow\})$; while for the overall reaction $(A \Leftrightarrow F + G)$ structurally minimal pathway $(\{A, B, D, F, G\}, \{R1 \rightarrow, R3 \rightarrow, R6 \rightarrow, R7 \rightarrow\})$.

Concluding remarks

For a fixed overall reaction, the equivalence of the terms, "type-I extreme pathway and structurally minimal pathway," has been proven. As a corollary, the term, "type-I extreme pathway," is equivalent to the term, "direct mechanism." The equivalence revealed herein opens promising novel avenues for reaction or metabolic pathway identification by integrating different methods, thus resulting in equivalent outcomes.

Acknowledgments Authors acknowledge the support of the Hungarian Research Fund under project OTKA 81493K.

References

1. Barany M, Bertok B, Imreh C, Fan LT, Friedler F (2012) J Math Chem 50:1347–1361
2. Fan LT, Bertok B, Friedler F (2002) Comput Chem 26:265–292
3. Fan LT, Shafie S, Bertok B, Friedler F, Lee DY, Seo H, Park SW, Lee SY (2005) J Chin Inst Eng 28:1021–1037
4. Friedler F, Tarjan K, Huang YW, Fan LT (1992) Comput Chem Eng 16:S313–S320
5. Friedler F, Tarjan K, Huang YW, Fan LT (1992) Chem Eng Sci 47:1973–1988
6. Happel J, Sellers PH (1983) Adv Catal 32:273–323
7. Klamt S, Stelling J (2003) Trends Biotechnol 21:64–69
8. Lee DY, Fan LT, Park S, Lee SY, Shafie S, Bertok B, Friedler F (2005) Metab Eng 7:182–200
9. Palsson BO (2006) Systems biology: properties of reconstructed networks. Cambridge University Press,
10. Papin JA, Price ND, Edwards JS, Palsson BO (2002) J Theor Biol 215:67–82
11. Price ND, Papin JA, Palsson BO (2002) Genome Res 12:760–769
12. Seo H, Lee DY, Park S, Fan LT, Shafie S, Bertok B, Friedler F (2001) Biotechnol Lett 23:1551–1557
13. Schilling CH, Letscher D, Palsson BO (2000) J Theor Biol 203:229–248
14. Schilling CH, Palsson BO (2000) J Theor Biol 203:249–283
15. Schuster S, Dandekar T, Fell DA (1999) Trends Biotechnol 17:53–60
16. Wiback SJ, Palsson BO (2002) Biophys J 83:808–818
17. http://gcrg.ucsd.edu/Downloads/ExtremePathwayAnalysis. Accessed 27 Sept 2012
18. http://www.p-graph.com/demo/rpi/comp_biol_chem/rpi.html. Accessed 27 Sept 2012