Biotechnology Letters **23:** 1551–1557, 2001.

Graph-theoretical identification of pathways for biochemical reactions

H. Seo[1], D.-Y. Lee[1], S. Park[1,*], L.T. Fan[2], S. Shafie[2], B. Bertók[3] & F. Friedler[2,3]
[1]*Department of Chemical Engineering, KAIST, 373-1 Kusong-dong, Yusong-ku, Taejon 305-701, South Korea*
[2]*Department of Chemical Engineering, Kansas State University, Manhattan, KS 66506, USA*
[3]*Department of Computer Science, University of Veszprém, Veszprém, Egyetem u. 10., H-8200 Hungary*
[*]*Author for correspondence (Fax: +82-42-869-3910; E-mail: swpark@mail.kaist.ac.kr)*

Received 6 June 2001; Revisions requested 4 July 2001; Revisions received 20 July 2001; Accepted 20 July 2001

Key words: algorithm, identification, metabolic pathways, *P*-graph

Abstract

A rigorous method for identifying biochemical reaction or metabolic pathways through its systematic synthesis has been established. The current method for synthesizing networks of metabolic pathways follows the general framework of a highly exacting combinatorial method. The method is capable of generating not only all combinatorially independent, feasible reaction networks only once, but also those combinations of independent pathways. A case study involving the conversion of glucose to pyruvate with 14 elementary reactions illustrates the efficiency and efficacy of the method. All the results have been obtained with a PC (Pentium-III 550 MHz, 256 MB RAM) within 1 s.

Introduction

A logical approach for circumventing the enormous combinatorial complexities involved in identifying a biochemical or metabolic pathway is through its systematic synthesis. A major effort is required to establish a method to perform this systematic synthesis; moreover, it is highly desirable that the method would eventually give rise to a computer-implementable algorithm; such an effort was initiated by Seressiotis & Bailey (1988). Since then, appreciable progress has been made through the participation of a number of other researchers (Happel *et al.* 1990, Mavrovouniotis *et al.* 1990, Hatzimanikatis *et al.* 1996a,b). Much remains to be done to establish an efficient algorithmic method for identifying biochemical reaction pathways. The current work explores if a novel algorithmic method devised by Fan *et al.* (1999, 2001) for the identification of pathways and mechanism of catalytic and other complex reactions could contribute towards achieving such a goal.

The method of Fan *et al.* (1999, 2001) is capable of rapidly yielding a complete metabolic pathway network, the maximal structure with minimal complexity, for a given set of candidate elementary reactions. A complete set of feasible sub-networks corresponding to feasible pathways, in turn, can be extracted from the maximal structure. The feasible pathways or

Table 1. Overall and elementary reactions for the conversion of glucose to pyruvate.

Overall reactions		Elementary reactions	
d_1:	$G \rightleftharpoons 2P$	(1)	$R + X \rightleftharpoons S + Y$
d_2:	$P \rightleftharpoons 3C$	(2)	$L \rightleftharpoons R$
d_3:	$G \rightleftharpoons 6C$	(3)	$N \rightleftharpoons L + C$
r_0:	$3G \rightleftharpoons 5P + 3C$	(4)	$G \rightleftharpoons N$
		(5)	$F \rightleftharpoons D + Y$
		(6)	$G \rightleftharpoons F$
		(7)	$D \rightleftharpoons Y$
		(8)	$N \rightleftharpoons K$
		(9)	$L \rightleftharpoons X$
		(10)	$E + X \rightleftharpoons Y + F$
		(11)	$Y \rightleftharpoons P$
		(12)	$D \rightleftharpoons P$
		(13)	$K \rightleftharpoons Y + P$
		(14)	$S + Y \rightleftharpoons E + F$

mechanisms are to be examined experimentally, computationally and/or theoretically for the final selection of bioreaction pathways, i.e., mechanism identification, which is outside the scope of the present work. The efficacy of the method is demonstrated through application to a well-known example, the conversion of glucose to pyruvate (Seressiotis & Bailey 1988, Happel *et al.* 1990, see Figure 1).

Given in Table 1 are the overall reactions and the 14 elementary reactions in the pathway with letter abbreviations (Seressiotis & Bailey 1988, Happel *et al.* 1990). Note that Happel *et al.* (1990) have generated the 3 direct overall reactions, d_1, d_2, and d_3, by recognizing that all elementary reactions must be reversible. To the contrary, Seressiotis & Bailey (1988) have chosen r_0 as an overall reaction which is not direct. Happel *et al.* (1990) have unequivocally shown that r_0 can be produced by summing d_1 and d_2, i.e., $3d_1 + d_2 = r_0$. The elementary reactions in Table 1 are the same as those considered by Seressiotis & Bailey (1988) except elementary reaction (3) where CO_2 (C) is added to satisfy a carbon balance for various reasons (Happel *et al.* 1990). It is worth noting that elementary reaction (4) or (6) is an initiation step activating the precursor, G; and elementary reactions (11), (12), and (13) are termination steps forming the target, P; one metabolic byproduct, CO_2, from elementary reaction (3) is released to the environment; and another metabolic product, H_2O, is released to the environment from elementary reactions (4) and (8). All the elementary reactions are written in a simple form of pseudoequations. Strictly speaking from the thermodynamic point of view, every elementary reaction is reversible. Nevertheless, elementary reactions (3), (4), (5), (8), (11), (12), and (13) are regarded as irreversible in Seressiotis & Bailey (1988): the rate of the reverse step ($\leftarrow$) in each of these reactions must be far smaller than that of the forward step ($\rightarrow$).

Graphical 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 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. The P-graphs, which are bipartite graphs, serve this purpose (Friedler *et al.* 1992, 1993, 1995). In a P-graph, elementary-reaction steps are represented by horizontal bars, biochemicals and active species by circles.

Results and discussion

The present results have been obtained by means of the 3 algorithms based on the 2 sets of axioms; one is the 6 axioms of *feasible reaction pathways for any given overall reaction,* and the other is the 7 axioms of *combinatorially feasible reaction networks* (Fan *et al.* 1999, 2001; also see Table 2). The 3 algorithms are RPIMSG for maximal structure generation, RPISSG for solution structure generation, and PBT for feasible pathway generation. Figure 2 shows the results obtained by applying algorithm RPIMSG to a set of biochemical reactions of interest involving multiple overall reactions.

When applied to the conversion of glucose to pyruvate, algorithm RPISSG, together with the subsidiary algorithms, has generated the combinatorially feasible biochemical reaction networks. Note that the order of the generation of the combinatorially feasible networks may be effected by the implementation of the algorithms; however, the resultant set of networks is obviously invariant. Moreover, it is remarkable that algorithm RPISSG reduces the search space of $(3^{14} - 1) = 4\,782\,968$ combinations of the 14 elementary reactions with overall reaction d_1 for the conversion of glucose to pyruvate to only 46 feasible combinations, i.e., networks, of elementary-reaction steps. In each case of overall reactions, d_2, d_3, and r_0, 115, 104, and 174 feasible combinations are generated, respectively.

If the number of feasible pathways is large, it is more convenient to first examine the independent pathways only, as proposed by Seressiotis & Bailey (1988), and Happel *et al.* (1990). Algorithm PBT, together with the subsidiary algorithms and functions (Fan *et al.* 1999, 2001), has yielded from the maximal structure, i.e., networks, 6 feasible independent pathways and 11 feasible acyclic combined pathways for the overall reaction, d_1: $G \rightleftharpoons 2P$; 5 feasible independent pathways and 8 feasible acyclic combined pathways for the overall reaction, d_2: $P \rightleftharpoons 3C$; 10 feasible independent pathways and 21 feasible acyclic combined pathways for the overall reaction, d_3: $G \rightleftharpoons 6C$; and 5 feasible independent pathways and 8 feasible

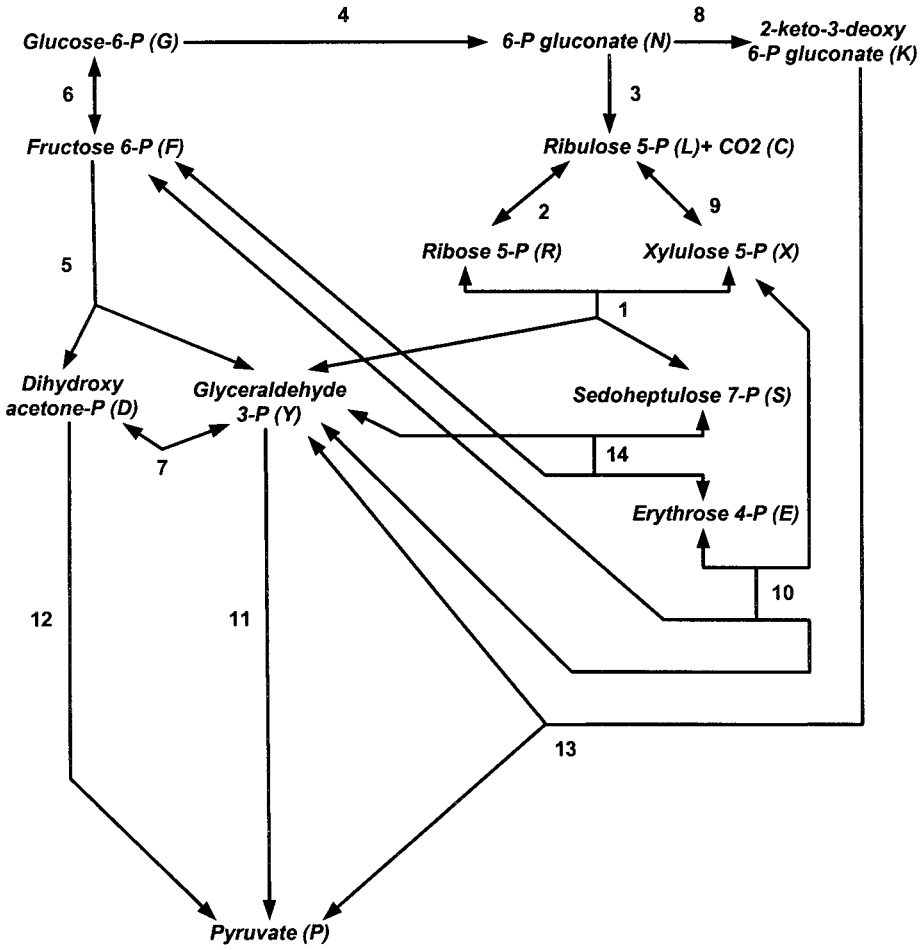

Fig. 1. Metabolic chart for the conversion of glucose to pyruvate by the Embden–Meyerhof–Parnas pathway, the Pentose Phosphate pathway, and the Entner–Doudoroff pathway (Seressiotis & Bailey 1988).

Table 2. Sets of axioms for feasible reaction pathways and combinatorially feasible reaction networks (Fan *et al.* 1999, 2001).

Six axioms of feasible reaction pathways	Seven axioms of combinatorially reaction networks
(R1) Every final product (target) is totally produced by reaction steps represented in the pathway. (R2) Every starting reactant (precursor) is totally consumed by reaction steps represented in the pathway. (R3) Every active intermediate produced by any reaction step represented in the pathway is totally consumed by one or more reaction steps in the pathway; and every active intermediate consumed by any reaction step represented in the pathway is totally produced by one or more reaction steps in the pathway. (R4) All reaction steps represented in the pathway are defined *a priori*. (R5) The network representing the pathway is acyclic. (R6) At least one elementary-reaction step represented in the pathway effects the activation of a starting reactant (precursor).	(T1) Every final product (target) is represented in the network. (T2) Every starting reactant (precursor) is represented in the network. (T3) Each reaction step represented in the network is defined *a priori*. (T4) Every active species represented in the network has at least one path leading to a final product (target) of the overall reaction. (T5) Every chemical or active species represented in the network must be a reactant for or a product from at least one reaction step represented in the network. (T6) A reactant of any elementary reaction represented in the reaction network is a starting reactant (precursor) if it is not produced by any reaction step represented in the network. (T7) The network includes at most either the forward or reverse step of each elementary reaction represented in the network.

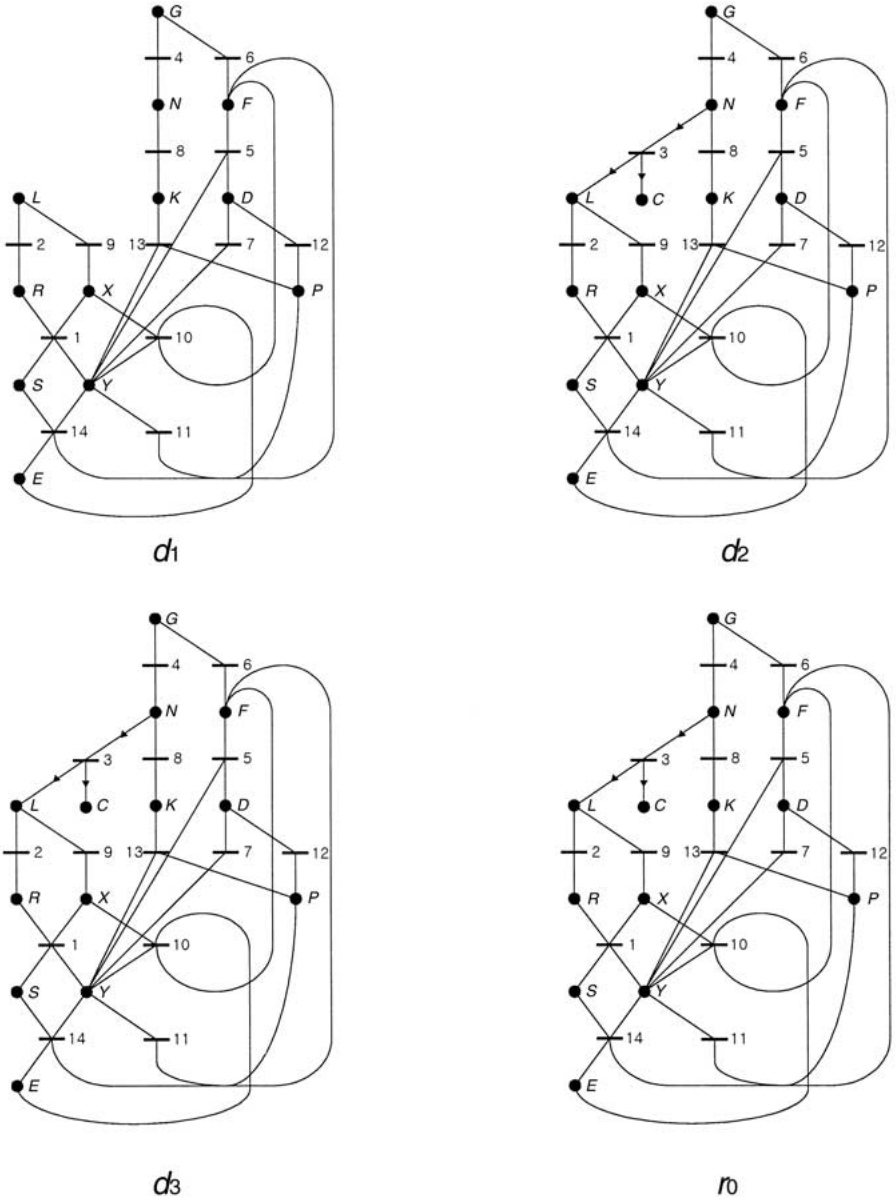

Fig. 2. P-graph representation of maximal structures for each overall reaction in converting glucose to pyruvate (Seressiotis & Bailey 1988).

acyclic combined pathways for the overall reaction, r_0: $3G \rightleftharpoons 5P + 3C$.

All available results from the earlier works have been obtained under the premise that elementary reactions (3)–(5), (8), and (11)–(13) are irreversible; therefore, such results are not comparable to the present results. The irreversibility of the 7 elementary reactions appears to be widely accepted. With the revised inputs comprising the overall reactions, d_1, d_2, d_3 and r_0, and the 14 elementary reactions from which the reverse steps of the 7 elementary reactions are removed, algorithms RPIMSG, RPISSG and PBT and the subsidiary algorithms (Fan *et al.* 1999, 2001) have eventually given rise to 5 feasible independent pathways and 8 feasible acyclic combined pathways for the overall reaction, d_1: $G \rightleftharpoons 2P$ as given in Table 3; no feasible pathways for the overall reactions, d_2: $P \rightleftharpoons 3C$, and d_3: $G \rightleftharpoons 6C$; and 5 feasible independent pathways and 8 feasible acyclic combined pathways for the overall reaction, r_0: $3G \rightleftharpoons 5P + 3C$.

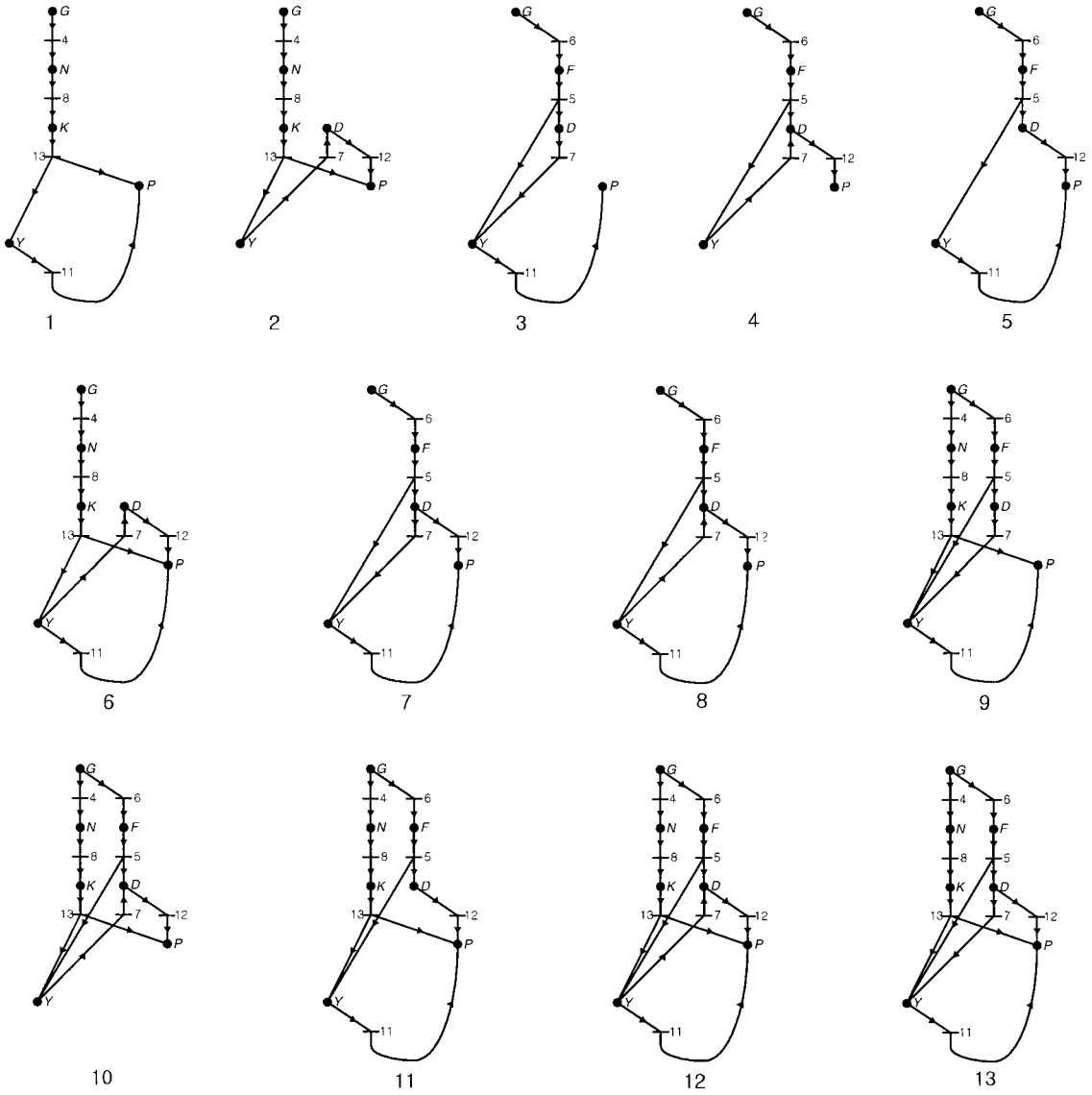

Fig. 3. P-graph representation of 13 feasible acyclic pathways for the overall reaction, d_1, $G \rightleftharpoons 2P$, in converting glucose to pyruvate under the assumption of irreversibility of elementary reactions (3), (4), (5), (8), (11), (12) and (13).

Figure 3 shows the P-graph representation of feasible pathways in Table 3. All the results have been obtained with a PC (Pentium-III, 550 MHz, 256 MB RAM). The computational time required on this particular PC is less than 1 s.

With the assumption of the irreversibility of elementary reactions (3)–(5), (8), and (11)–(13), Seressiotis & Bailey (1988) have listed only 3 feasible independent pathways, while Happel *et al.* (1990) as well as the current work have yielded 5 feasible independent pathways for r_0. The difference arises from the fact that Seressiotis & Bailey (1988) have assumed that if reaction r_0 occurs, no other reaction, e.g., d_1, takes place simultaneously in the system. In reality, r_0 can be generated as the sum of d_1 and d_2, i.e., $3d_1 + d_2 = r_0$. Apparently, no experimental evidence exists to ascertain the assumption made by Seressiotis & Bailey (1988).

As mentioned above, the pathways for the conversion of glucose to pyruvate identified by the method of Fan *et al.* (1999, 2001) in the current work are deemed valid or feasible in the light of previous works. It is, however, worth cautioning that some of the pathways might have been identified fortuitously because the

Table 3. Feasible biochemical reaction pathways of the conversion of glucose to pyruvate with the overall reaction, d_1: $G \rightleftharpoons 2P$ under the assumption of irreversibility of elementary reactions (3), (4), (5), (8), (11), (12), and (13) (note that figures are stoichiometric numbers for the corresponding feasible pathways; positive sign indicates the forward step of the corresponding elementary reaction; negative sign, reverse step).

Elementary reactions		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Independent	Pathway 1				1				1			1		1	
pathways	Pathway 2				1			−1	1				1	1	
	Pathway 3					1	1	1				2			
	Pathway 4					1	1	−1					2		
	Pathway 5					1	1					1	1		
Acyclic	Pathway 6				2			−1	2			1	1	2	
combined	Pathway 7					2	2	1				3	1		
pathways	Pathway 8					2	2	−1				1	3		
	Pathway 9				1	1	1	1	1			3		1	
	Pathway 10				1	1	1	−2	3				1	1	
	Pathway 11				1	1	1		1			2	1	1	
	Pathway 12				1	1	1	−1	1			1	2	1	
	Pathway 13				2	2	2	1	2			5	1	2	

method has been applied in somewhat of an ad hoc manner as delineated below.

(a) The active components of enzymes, i.e., the active sites of catalysts, are not explicitly expressed in the formulas, or equivalently P-graph representations, of elementary reactions, as required by the method.

(b) All the participating species, including the precursors, intermediates and targets, are expressed only symbolically in the overall as well as elementary reactions. It is, therefore, nearly impossible to verify with absolute certainty that the elementary balances, the fundamental constraints of the method, are exactly satisfied.

(c) The method has been rigorously established strictly for the identification of the pathways of an individual catalytic or complex overall reaction. Nevertheless, it has been demonstrated that the pathways of a set of multiple overall reactions, e.g., of reactions occurring in series, can be identified as long as the method is applied separately to individual overall reactions in the set.

For any given overall reaction and a set of plausible elementary reactions, the current method exactly gives all feasible mechanisms, each containing a set of elementary reactions with varied multipliers, i.e., stoichiometric numbers. Nevertheless, the final selection of valid mechanisms from the set of feasible mechanisms, i.e., correct identification, must await the comparison of the rate expressions derived from them with the experimental data; these rate expressions or laws are almost always obtained under some assumptions, e.g., the existence of rate-controlling and/or equilibrium steps. It should be emphasized that the final rate law of any biochemical reaction would emerge from one of the stoichiometrically exact mechanisms identified by the current method.

Concluding remarks

We have demonstrated that a mathematically exact, graph-theoretic method proposed by Fan *et al.* (1999, 2001) for the identification, i.e., determination, of the mechanisms of complex chemical reactions is applicable to biochemical reactions.

Acknowledgement

This work was partially supported by the Brain Korea 21 project.

References

Fan LT, 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 LT, Bertók B, Friedler F (2001) A graph-theoretical method to identify candidate mechanisms for deriving the rate law of a catalytic reaction. *Comput. Chem.*, in press.

Friedler F, Tarjan K, Huang YW, Fan LT (1992) Graph-theoretic approach to process synthesis: axioms and theorems. *Chem. Eng. Sci.* **47**: 1973–1988.

Friedler F, Tarjan K, Huang YW, Fan LT (1993) Graph-theoretic approach to process synthesis: polynomial algorithm for maximum structure generation. *Comput. Chem. Eng.* **17**: 929–942.

Friedler F, Varga JB, Fan LT (1995) Decision mapping: a tool for consistent and complete decisions in process synthesis. *Chem. Eng. Sci.* **50**: 1755–1768.

Happel J, Sellers PH, Otarod M (1990) Mechanistic study of chemical reaction systems. *Ind. Eng. Chem. Res*. **29**: 1057–1064.

Hatzimanikatis V, Floudas CA, Bailey JE (1996a) Optimization of regulatory architectures in metabolic reaction networks. *Biotechnol. Bioeng.* **52**: 485–500.

Hatzimanikatis V, Floudas CA, Bailey JE (1996b) Analysis and design of metabolic reaction networks via mixed-integer linear optimization. *AIChE J.* **42**: 1277–1292.

Mavrovouniotis ML, Stephanopoulos G, Stephanopoulos G (1990) Computer aided synthesis of biochemical pathways. *Biotechnol. Bioeng*. **36**: 1119–1132.

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