I&EC research
Industrial & Engineering Chemistry Research

Article

pubs.acs.org/IECR

Determination of the Thermodynamically Dominant Metabolic Pathways

Choamun Yun,[†,▽] Tae Yong Kim,[†] Tengyan Zhang,[‡,⊥] Young Kim,[†,○] Sang Yup Lee,[†] Sunwon Park,*[,†] Ferenc Friedler,[§] and Botond Bertok[§]

[†]Department of Chemical and Biomolecular Engineering, KAIST, Daejeon 305-701, Korea

[‡]Department of Chemical Engineering, Kansas State University, Manhattan, Kansas 66506, United States

[§]Department of Computer Science, University of Pannonia, Egyetem u. 10, Veszprem, H-8200, Hungary

[○]Department of Thermal Systems, Korea Institute of Machinery and Materials, Daejeon 305-343, Korea

Supporting Information

ABSTRACT: An effective strategy comprising two phases is proposed to determine the thermodynamically dominant pathways in a metabolic network of a given phenotype, involving several metabolic reactions. In the first phase, stoichiometrically feasible metabolic pathways are exhaustively identified through the flux balance analysis and the graph-theoretic method based on P-graphs. In the second phase, thermodynamically dominant pathways are selected from these stoichiometrically feasible metabolic pathways on the basis of the Gibbs free energy change of reaction. The proposed strategy's efficacy is demonstrated by applying it to two *E. coli* models: one is for maximal acetate and ethanol production, and the other is for maximal poly(3-hydroxybutyrate) production.

INTRODUCTION

Bio-based products are increasingly gaining worldwide interest as substitutes for petrochemical products to reduce the dependency on fossil fuels and to exploit their environmentally benign characteristics. For the efficient production of these bio-based products from the microorganisms, metabolic engineering is indispensable to render it possible for microorganisms to become suitable for such production.[1] Metabolic networks are adapted for the altered objectives in metabolic engineering with a variety of approaches to achieve the high-yield processes at lower costs.[2,3] It is, however, rather convoluted to identify the engineering targets, because of the complex interactions of various components in metabolic networks. With the ever-growing information on the functions and phenotypes in metabolic networks, because of the technological advances, metabolic engineers are searching for increasingly effective tools that will facilitate the metabolic engineering of microorganisms through the systematic analysis and prediction of biological behavior. Metabolic flux analysis (MFA), among others, has contributed significantly to advancing metabolic engineering, based on the pseudo-steady-state assumption and linear programming (LP). By resorting to MFA, the overall reaction, or the overall mass balance of consumed nutrients, secreted metabolites, and byproduct, and the intracellular flux distribution can be observed for a given objective function (e.g., maximum target production).

For simplicity, metabolic flux analysis often neglects the fact that an overall reaction may be generated not from a unique pathway but from one of multiple possible pathways. Multiple pathways, however, are attainable through elementary flux modes or extreme pathway analysis,[4,5] and are referred to as alternate optimal or equivalent pathways.[6] Although these methods are well-known to be efficient and informative in handling small-scale metabolic networks, they are ineffective for generating the equivalent pathways of large-scale metabolic networks, because of the exponentially increasing combinatorial complexity of the networks.[7] In the current contribution, the equivalent pathways of large-scale metabolic networks are searched by resorting to a graph-theoretic approach based on P-graphs to overcome the combinatorial explosion issue.[8−10] Herein, the equivalent pathways are also referred to stoichiometrically feasible pathways in accordance with the definitions in the P-graph-based approach. Moreover, the resultant solution pathways are prioritized based on thermodynamic principles.

In principle, all chemical reactions, including metabolic reactions, are reversible: a reaction favors either the forward or backward direction, depending on the Gibbs free energy change of reaction (ΔG_r).[11] Extensive research has been carried out to estimate ΔG_r for identifying the favored direction of every reaction to determine the thermodynamically feasible pathways in a metabolic network.[12−15] Even though the thermodynamic criterion is invaluable in analyzing the metabolic network, this criterion requires an exception: The available experimental results imply that some of the thermodynamically unfavorable reactions are essential to the cell.[16] This inconsistency arises from the uncertainties contained in the data for calculating ΔG_r and the incomplete understanding of intracellular reactions in the living cells, such as the effect of the energy produced by the common-intermediate strategy of ATP (adenosine triphosphate) and NADH (nicontinamide adenine dinucleotide). Thus, unless a reaction step is assured to be absent from the

Special Issue: L. T. Fan Festschrift

Received: March 12, 2012
Revised: July 5, 2012
Accepted: July 9, 2012
Published: July 12, 2012

metabolic pathway, it should not be eliminated in the search space solely based on the thermodynamic criterion.

An effective strategy comprising two phases has been proposed in the current work to determine the thermodynamically dominant pathways in a metabolic network. The method is applied to the pathway analysis of *E. coli*, the most widely deployed microorganism for the synthesis of biochemicals. In the first example, the overall procedure is illustrated with a small model for maximal acetate and ethanol production. Then, the efficacy of the proposed method is demonstrated in the second example by identifying the important reaction steps for maximal poly(3-hydroxybutyrate) [P(3HB)] production on the basis of the thermodynamically dominant pathways. P(3HB) has the superior characteristics as the raw material for the biodegradable plastics.

Table 1. Numbers of Essential, Substitutable, And Blocked Reactions in Examples 1 and 2

types of reactions	Example 1		Example 2
	maximal acetate production	maximal ethanol production	maximal P(3HB) production
essential	17	12	6
substitutable	8	7	89
blocked	23	29	215

Table 2. Values of $(\Delta G_r)_{min}$ and $(\Delta G_r)_{max}$ for the Substitutable Reactions in the Stoichiometrically Feasible Metabolic Pathways for Maximal Acetate and Ethanol Production[a]

reaction name	reaction	$(\Delta G_r)_{min}$ (kcal/mol)	$(\Delta G_r)_{max}$ (kcal/mol)
Gly3	FDP + H_2O → F6P + Pi	−2220	2480
Gly11	PYR + ATP + H_2O → AMP + Pi + PEP + $2H^+$	−20 600	−4160
Gly13	OA + ATP → CO_2 + ADP + PEP	−1700	8720
Gly14	PEP + CO_2 + H_2O → OA + Pi + H^+	−14 500	−3390
Egy3	NADPH + NAD^+ → NADP + NADH	−3810	3810
Egy5	2H(e) + NADP + NADH → $2H^+$ + NADPH + NAD^+	−5460	7630
Egy6	ADP + Pi + H^+ → ATP + H_2O	1270	9590
Egy7	ATP + AMP ↔ 2ADP	−3810	3810
Egy8	ATP + H_2O → ADP + Pi	−1950	3640

[a]The values are estimated for T = 25°C, pH 7.6, and the ionic strength of 0.15 M (data taken from Kummel et al.[13]

This strategy is akin to that of the flowsheet synthesis for any chemical or biochemical process: The process flowsheet is first composed on the basis of the mass balances prior to performing any thermodynamic analysis, including energy and exergy balances.

METHODS

The methods include two phases. The first executes the identification of stoichiometrically feasible metabolic pathways; and the second involves the selection of thermodynamically dominant pathways from such feasible pathways. Presumably, it would be most logical to select the dominant pathways on the basis of an energetic or thermodynamic criterion.

Identification of Stoichiometrically Feasible Metabolic Pathways. Stoichiometrically feasible metabolic pathways and their flux distributions can be exhaustively identified through the flux balance analysis (FBA)[17] and the graph-theoretic method based on process graphs (P-graphs)[8,18] executed sequentially.[9,10] The overall reaction is obtained via FBA from a series of candidate metabolic reactions and the objective metabolites to be maximized or minimized. This is followed by the identification of the stoichiometrically feasible pathways and the reaction fluxes in them, which satisfy the overall reaction, via algorithm PBT. At the outset, the maximal structure, which is the maximally connected network of the metabolites and the reactions, is generated to exclude the combinatorially infeasible pathways. Subsequently, the stoichiometrically feasible (i.e., balanced) pathways are exhaustively recovered from this maximal structure via algorithm PBT. The details are available elsewhere.[8−10,18]

For a large-scale metabolic pathway, knowledge of essential, substitutable, and blocked reactions will facilitate the identification of stoichiometrically feasible metabolic pathways. The resultant flux distributions of a set of stoichiometrically feasible pathways naturally reveal the flux variability of each reaction step in the pathways.[6] In generating the overall reaction, the reaction steps with nonzero fluxes in every solution pathway are essential; those with zero fluxes are blocked; and the remaining ones are substitutable.[15] While the flux variability analysis measures only the extent of variability, the graph-theoretic method provides the exact values of reaction fluxes in every stoichiometrically feasible pathway. Accordingly, the latter renders it possible to distinguish those pathways by further investigating the differentiated properties of substitutable reactions.

Table 3. Values of ΔG_r^d for Substitutable Reactions in the Stoichiometrically Feasible Pathways for Maximal Acetate Production in Example 1

reaction name	ΔG_r^d (kcal/mol)							
	1	2	3	4	5	6[a]	7	8
Gly3		2480			2480			
Gly11	−4160							−4160
Gly13				8720			8720	
Gly14				−3390			−3390	
Egy3	3810	3810	3810	3810				
Egy5	7630	7630	7630	7630				
Egy6					9590	9590	9590	9590
Egy7	3810							3810
Egy8			3640			640		
rank	3	1	2	4	5	6	8	7

[a]Here, the optimal pathway has been identified by linear programming.

Selection of Thermodynamically Dominant Pathways. Thermodynamically dominant pathways are selected from the stoichiometrically feasible metabolic pathways on the basis of ΔG_r of every metabolic reaction step in the pathways. Unfortunately, however, the value of ΔG_r for a metabolic reaction under specific conditions is usually unknown; instead, only its minimum value, $(\Delta G_r)_{min}$, and maximum value, $(\Delta G_r)_{max}$, are known, because of the variations of concentrations of metabolites involved in the reaction.

Reiterating, any reaction with a more-negative ΔG_r value is favored over a reaction with a less-negative (or more-positive) ΔG_r value.[8] It is logical to envision that the tendency or extent of any feasible pathway to proceed is controlled by the metabolic reaction step in the pathway with the largest possible $(\Delta G_r)_{max}$. This, in turn, renders it possible to select the pathway for which its largest possible $(\Delta G_r)_{max}$ value is the lowest among all the competing feasible pathways as the thermodynamically dominant one.

Naturally, for any reaction step occurring in the opposite direction as indicated by its negative flux, the change of any extensive thermodynamic property, for instance, ΔG_r, reverses its sign. Thus, $-(\Delta G_r)_{min}$ takes the place of $(\Delta G_r)_{max}$. For convenience, therefore, the quantity ΔG_r^d is defined such that if J_r of a reaction step is positive, it is ΔG_r; and if J_r of a reaction step is negative, it is $-\Delta G_r$. With this definition of ΔG_r^d, the pathway with the smallest value of $[\Delta G_r^d]_{max}$ can be selected as the dominant one. If the values of $[\Delta G_r^d]_{max}$ among all the pathways in two or more pathways are equal, these pathways are similarly ranked for their thermodynamic dominance, in terms of the second largest ΔG_r^d, and so on.

Table 4. Values of ΔG_r^d for Substitutable Reactions in the Stoichiometrically Feasible Pathways for Maximal Ethanol Production in Example 1

	ΔG_r^d (kcal/mol)			
reaction name	1	2	3[a]	4
Gly3				2480
Gly11		−4160		
Gly13	8720			
Gly14	−3390			
Egy7		3810		
Egy8			3640	
rank	4	3	2	1

[a]Here, the optimal pathway has been identified by linear programming.

Application to the Metabolic Network Models of *E. coli*. The proposed method is illustrated with the two examples of the metabolic network models of *E. coli*. One model contains 52 metabolites in 48 reactions consisting of the glycolytic pathway, the tricarboxylic acid (TCA) cycle, the pentose phosphate pathway, and the transport reactions.[8] The overall procedure of determining the thermodynamically dominant pathways among the optimal equivalent pathways is described with this relatively simple model. The other contains 295 metabolites in 310 reactions including the P(3HB) biosynthesis pathway. This model was previously investigated for maximal P(3HB) production employing FBA and validated through experiments.[19] Accordingly, the efficacy of the method proposed herein could be elucidated by comparing it with the results of the previous work.

The ΔG_r values of the substitutable reaction steps, under the specific conditions, have been estimated by resorting to the

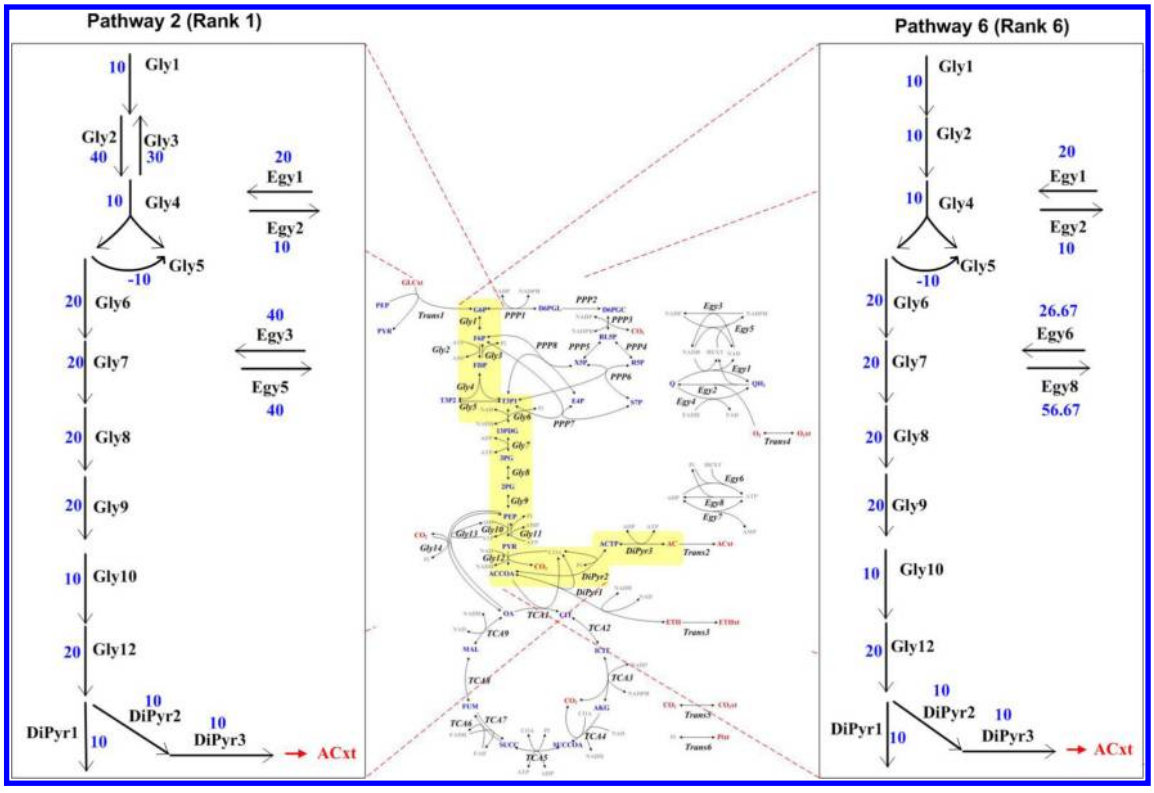

Figure 1. Pathways 2 and 6 for maximum acetate production: The former being the most thermodynamically dominant pathway, and the latter being the optimal pathway identified solely based on linear programming.

Table 5. Number of Pathways, Containing the Essential or Substitutable Reactions among All the Stoichiometrically Feasible Pathways and the Thermodynamically Dominant 100 Pathways, For Maximal P(3HB) Production in Example 2

reaction name	enzyme	reaction	No. of Pathways[a]	
			all 11 455	Top 100
ptsIH	phosphotransferase system	GLC + PEP → PYR + G6P	11 455	100
pgi	phosphoglucose isomerase	G6P ↔ F6P	4840 (3996)	92
pfkAB	phosphofructokinase	F6P + ATP ↔ F16P + ADP	6328 (1580)	49
fbp	fructose-1,6-bisphosphate aldolase	F16P + PI → F6P	999	2
fba	fructose-1,6-bisphosphatase	F16P ↔ T3P1 + T3P2	6227 (2064)	49 (24)
tpiA	triosphosphate isomerase	T3P1 ↔ T3P2	5580 (2526)	100
gapA	glyceraldehyde-3-phosphate dehydrogenase	T3P1 + PI + NAD ↔ a13P2DG + NADH	11,335 (19)	100
pgk	phosphoglycerate kinase	a13P2DG + ADP ↔ a3PDGL + ATP	11 335 (19)	100
gpmAB	phosphoglycerate mutase	a3PDGL ↔ a2PDGL	10 298 (231)	95 (1)
eno	enolase	a2PDGL ↔ PEP	10 298 (231)	95 (1)
pykAF	pyruvate kinase	PEP + ADP → PYR + ATP	4303	29
pckA	PEP carboxykinase	OA + ATP ↔ PEP + ADP + CO_2	5202 (2087)	23 (19)
ppc	PEP carboxylase	PEP + CO_2 → PI + OA	2671	20
lpdA	pyruvate dehydrogenase	PYR + COA + NAD → ACCOA + CO_2 + NADH	6484	8
ppsA	PEP synthase	PYR + ATP → PEP +AMP + PI	2652	33
zwf	glucose-6-phosphate dehydrogenase	G6P + NADP ↔ D6PGL + NADPH	6615	8
pgl	6-phosphogluconolactonase	D6PGL → D6PGC	6615	8
gnd	6-phosphogluconate dehydrogenase	D6PGC + NADP ↔ RL5P + CO_2 + NADPH	4533 (3073)	(44)
rpiAB	ribose-5-phosphate isomerase	RL5P ↔ R5P	4533 (3073)	(44)
rpe	ribose-5-phosphate epimerase	RL5P ↔ X5P	4533 (3073)	(44)
tktAB1	transketolase 1	R5P + X5P ↔ T3P1 + S7P	4533 (3073)	(44)
talB	transaldolase	T3P1 + S7P ↔ E4P + F6P	4533 (3073)	(44)
tktAB2	transketolase 2	X5P + E4P ↔ F6P + T3P1	4533 (3073)	(44)
edd	6-phosphogluconate dehydrase	D6PGC → a2K3D6PG	6569	52
eda	2-keto-3-deoxy-6-phosphogluconate aldolase	a2K3D6PG → T3P1 + PYR	6569	52
Glycogen02	glycogen synthase	G1P + ATP → GLYCOGEN + ADP + PPI	1035	0
Glycogen03	glycogen phosphorylase	GLYCOGEN + PI → G1P	1035	0
Diss_Pyruvate04	pyruvate formate lyase	PYR + COA → FORMATE + ACCOA	6112	93
gltA	citrate synthase	ACCOA + OA ↔ CIT+ COA	(11 356)	(76)
acnAB	aconitase	CIT ↔ ICIT	(11 356)	(76)
icdA	isocitrate dehydrogenase	ICIT + NADP ↔ AKG + CO_2 + NADPH	(11 369)	(76)
sucAB	2-ketoglutarate dehydrogenase	AKG + COA + NAD ↔ SUCCOA + CO_2 + NADH	(11 369)	(76)
sucCD	succinate thiokinase	SUCCOA + GDP + PI ↔ SUCC + COA + GTP	(11 369)	(76)
sdhABCD	succinate dehydrogenase	SUCC + FAD → FUM + $FADH_2$	1682	11
frdABCD	fumurate reductase	FUM + $FADH_2$ → SUCC + FAD	11 358	76
fumABC	fumarase	FUM ↔ MAL	(11 356)	(76)
mdh	malate dehydrogenase	MAL + NAD ↔ OA + NADH	1024 (6738)	26 (63)
mez1	malic enzyme	MAL + NADP → PYR + CO_2 + NADPH	4967	0
mez2	malic enzyme	MAL + NAD ↔ PYR + CO_2 + NADH	1375 (6302)	30 (48)
TCA12	isocitrate lyase	ICIT → SUCC + GLX	2809	12
TCA13	malate synthase	ACCOA + GLX → MAL + COA	2809	12
Respiration01	NADH dehydrogenase II	NADH + Q → NAD + QH_2	2917	40
Respiration02	NADH dehydrogenase I	NADH + Q → NAD + QH_2 + 4 Hext	4663	0
Respiration03	formate dehydrogenase	FORMATE + Q → QH_2 + CO_2 + 2 Hext	4801	0
Respiration06	succinate dehydrogenase complex	FADH2 + Q ↔ FAD + QH_2	(10 041)	(40)
ATP_synthesis	F0F1-ATPase	ATP ↔ ADP + PI + 3 Hext	(8256)	0
Asp01	aspartate transaminase	GLU + OA↔ ASP + AKG	4576	42
Glu_Gln01	glutamate dehydrogenase	AKG + NH_3 + NADPH → NADP + GLU	5390	56
Glu_Gln02	glutamine synthatase	GLU + NH_3 + ATP → GLN + ADP + PI	3829	6
Glu_Gln03	glutamate synthase	AKG + GLN + NADPH → 2 GLU + NADP	3829	6
Ser_Gly01	3-phosphoglycerate dehydrogenase	a3PDGL + NAD → PHP + NADH	4784	21
Ser_Gly02	phosphoserine transaminase	PHP + GLU → AKG + a3PSER	4784	21
Ser_Gly03	phosphoserine phosphatase	a3PSER → SER + PI	4784	21
Ser_Gly04	serine hydroxymethyltransferase	GLY + METTHF ↔ SER + THF	(4784)	(21)
Ser_Gly05	glycine cleavage system	GLY + THF + NAD → METTHF + CO_2 + NH_3 + NADH	8445	59
Ser_Gly06	threonine dehydrogenase	THR + NAD ↔ AABK + NADH	4576	42
Ser_Gly07	amino-b-ketobutyrase	AABK + COA ↔ GLY + ACCOA	4576	42
Ser_Gly08	formate dehydrogenase	FORMATE + NAD → CO_2 +NADH	7469	95

Table 5. continued

reaction name	enzyme	reaction	No. of Pathways[a] all 11 455	No. of Pathways[a] Top 100
Ser_Gly09	formate THF ligase	THF + FORMATE + ATP → FTHF + ADP + PI	1218	3
Ser_Gly10	formyl THF deformylase	FTHF → FORMATE + THF	8732	61
Cys02	APS kinase	APS + ATP → PAPS + ADP	122	0
Cys08	adenylyl sulfate kinase	PAPS + ADP → APS + ATP	122	0
Thr_Lys01	aspartate kinase	ASP + ATP ↔ BASP + ADP	4576	42
Thr_Lys02	aspartate semialdehyde dehydrogenase	BASP + NADPH ↔ ASPSA + NADP + PI	4576	42
Thr_Lys03	Homoserine dehydrogenase	ASPSA + NADPH ↔ HSER + NADP	4576	42
Thr_Lys04	homoserine kinase	HSER + ATP → PHSER + ADP	4576	42
Thr_Lys05	threonine synthase	PHSER → THR + PI	4576	42
Nucleotides13	AMP phosphatase	AMP → ADN + PI	780	16
Nucleotides14	adenylate kinase	ATP + ADN → AMP + ADP	780	16
Nucleotides15	adenylate kinase	ATP + AMP → 2 ADP	2652	33
Nucleotides19	GDP kinase	GDP + ATP ↔ GTP + ADP	11 369	76
Pyrimidines18	dUDP kinase	DUDP + ATP ↔ DUTP + ADP	526	0
Pyrimidines19	dUTP pyrophosphatase	DUTP → DUMP + PPI	526	0
Pyrimidines20	dUMP kinase	DUMP + ATP ↔ DUDP + ADP	526	0
THF02	methylene THF dehydrogenase	METTHF + NADP ↔ METHF + NADPH	8445	59
THF03	methenyl tetrahydrofolate cyclehydrolase	METHF ↔FTHF	8445	59
Lipids01	acetyl-CoA carboxylase	ACCOA + ATP + CO_2 ↔ MALCOA + ADP + PI	1243	8
Lipids02	malonyl-CoA:ACP transacylase	MALCOA + ACP ↔ MALACP + COA	1243	8
Lipids03	b-ketoacyl-ACP synthase	MALACP → ACACP + CO_2	1243	8
Lipids04	acetyl-CoA:ACP transacylase	ACACP + COA ↔ ACP + ACCOA	1243	8
Lipids10	glycerol-3-phosphate dehydrogenase	T3P2 + NADH ↔ GL3P + NAD	8163	100
Isoprenoids01	aldose reductase	GL + NADP ↔ GLAL + NADPH	7278	100
Isoprenoids02	glyceraldehyde kinase	GLAL + ATP → T3P1 + ADP	7278	100
NAD05	NAD kinase	NAD + ATP → NADP + ADP	1174	0
NAD06	NADP phosphatase	NADP → NAD + PI	1174	0
PolyPI01	pyrophosphatase	PPI → 2 PI	1561	0
PolyPI02	polyphosphate kinase	1000 ATP ↔ 1000 ADP + POLYP	1144	13
PolyPI03	polyphosphatase	POLYP → 1000 PI	1144	13
Glycerol01	glycerol kinase	GL + ATP ↔ GL3P + ADP	(7278)	(100)
Glycerol02	glycerol-3-phosphate dehydrogenase	GL3P + FAD → T3P2 + $FADH_2$	2905	39
Transport10	glucose transport	GLCext ↔ GLC	11 455	100
Transport11	carbon dioxide transport	CO2ext ↔ CO_2	(11 455)	(100)
P3HB_syn_1	*β*-ketothiolase	2 ACCOA → COA + ACETOCOA	11 455	100
P3HB_syn_2	acetoacetyl-CoA reductase	ACETOCOA + NADPH → C4COA + NADP	11 455	100
P3HB_syn_3	P(3HB) synthase	C4COA → P(3HB) + COA	11 455	100

[a]Parentheses () indicates reactions in the negative direction.

transformation of the experimental data[20,21] and the group contribution method.[16] Subsequently, the ranges of ΔG_r have been calculated for the approximate ranges of metabolite concentrations reported in recent works.[13,22]

■ RESULTS

Analyzed herein are the results from applying the proposed strategy to the two *E. coli* models.

Example 1: A Simple *E. coli* Model for Maximal Acetate and Ethanol Production. The model reported by Schilling and colleagues[4] is optimized for maximal acetate and maximal ethanol production. This *E. coli* model gives rise to eight stoichiometrically feasible metabolic pathways for the maximal acetate production and four stoichiometrically feasible metabolic pathways for the maximal ethanol production.[10] Table 1 presents the numbers of essential, substitutable, and blocked reactions for the model. The values of $(\Delta G_r)_{min}$, the minimum value of ΔG_r, and $(\Delta G_r)_{max}$, and the maximum value of ΔG_r, for each substitutable reaction, are listed in Table 2. By multiplying these values with the sign of concomitant flux, the ΔG_r^d values have been obtained for the maximal acetate production and the maximal ethanol production. This has rendered it possible to rank the pathways according to $[\Delta G_r^d]_{max}$, as illustrated in Table 3 for the maximal acetate production and Table 4 for the maximal ethanol production. For maximal acetate production, Figure 1 depicts pathway 2, which has been identified as the most thermodynamically dominant pathway, and pathway 6, which has been identified to correspond with the result obtained by FBA. Similarly, for maximal ethanol production, pathway 4 has been determined to be the most thermodynamically dominant while pathway 3 has been identified by FBA.

Example 2: An *E. coli* Model Modified for Maximal P(3HB) Production. This *E. coli* model gives rise to 11 455 stoichiometrically feasible metabolic pathways for the maximal production of poly(3-hydroxybutyrate) [P(3HB)].[19] The numbers of essential, substitutable, and blocked reactions for the model are also presented in Table 1. The computational time for identifying all the stoichiometrically feasible metabolic pathways is ∼34 h on a 2.93 GHz Core 2 PC for this example.

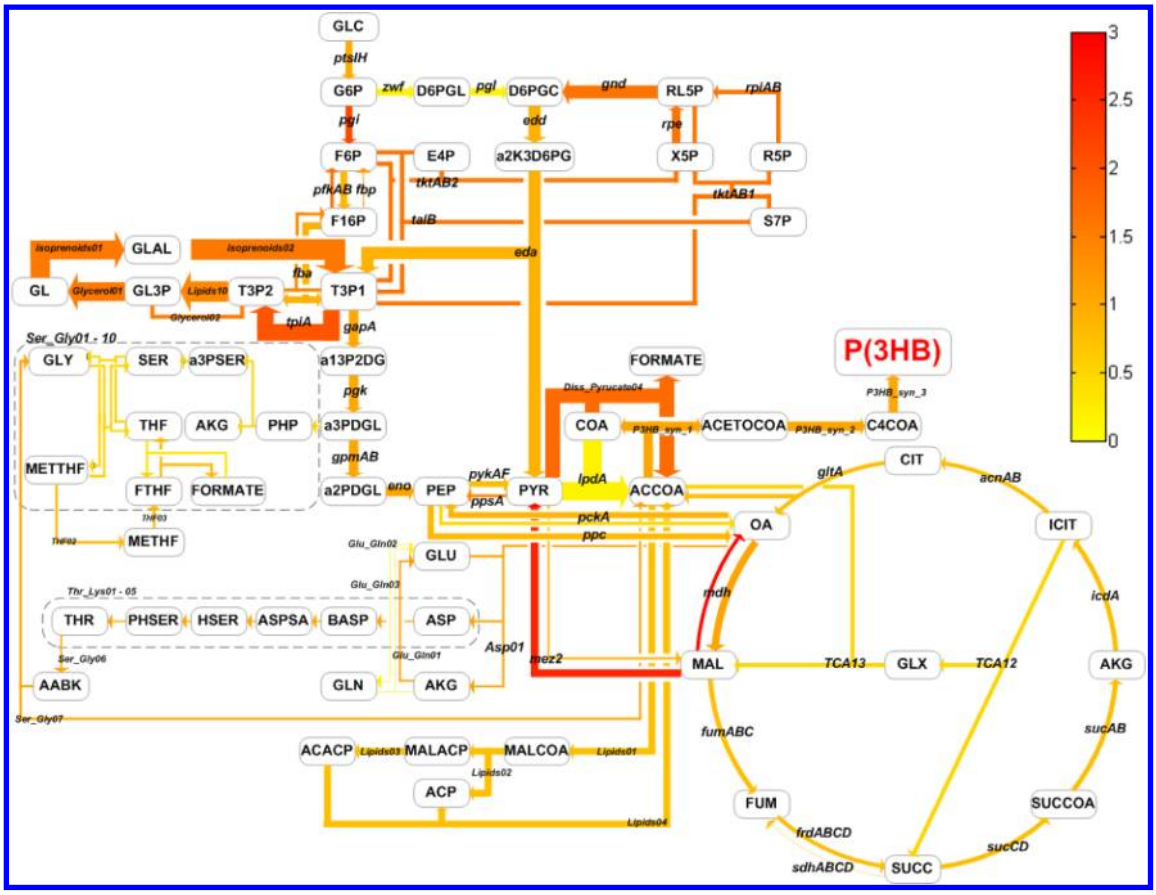

Figure 2. Metabolic reactions in the thermodynamically dominant 100 pathways for maximal P(3HB) production in *E. coli*. The thickness of each arrow indicates the average flux, and the color indicates the presence ratio of the reaction in the 100 pathways to that in all of the stoichiometrically feasible pathways. The yellow to red colors in the color bar indicate the increase in the ratio from 0 to 3. The full names of metabolites are listed in the Supporting Information.

The ranges of ΔG_r and the concomitant ΔG_r^d have been calculated for the 89 substitutable reactions.

Three reactions catalyzed by malic enzyme, pyrophosphatase, and NADP phosphatase appearing in 6483 pathways have been determined to be thermodynamically unfavorable: Both of their $(\Delta G_r)_{\min}$ and $(\Delta G_r)_{\max}$ values are positive. Rationally, the pathways with those reaction steps are regarded thermodynamically less dominant, and therefore, ranked lower than the remaining ones.

The current work indicates that the number of stoichiometrically feasible pathways leading to the maximal theoretical yield of P(3HB) production is on the order of 10^4 for the metabolic network comprising 310 reactions. Table 5 lists the number of pathways, containing the essential and the substitutable reactions in generating the optimal overall reactions. The reactions that are essential for maximal P(3HB) production are those included in all 11 455 identified pathways. As depicted in Figure 2, the reactions contained in the thermodynamically dominant 100 pathways, which are ~1% of the total pathways, are examined to investigate the important reaction steps and their fluxes for the maximal production of P(3HB). This result may also contribute to the reduction of the errors of the study that arise from the uncertainties involved in the estimation of thermodynamic properties. The frequency of reactions in the thermodynamically dominant pathways is revealed through the color of arrows changing from yellow to red in Figure 2. The frequency number represents the presence ratio of the reaction in the thermodynamically dominant 100 pathways to that in all the stoichiometrically feasible pathways.

DISCUSSION

The graph-theoretic approach based on *P*-graphs not only reduces the combinatorial complexity, but also provides a deeper understanding of the metabolic networks. As can be observed in Figure 1, *Gly 2* and *Gly 3* are represented as two separate reactions by different enzymes in the opposite directions. The *P*-graph approach may present exchange fluxes in this futile cycle, in addition to their net flux, while only the net flux can be identified in MFA. The physiological meaning of this futile cycle can be further consulted elsewhere.[23]

The complete set of ranked pathways has been identified to facilitate the experimental design of gene regulation. The results depicted in Figure 2 reveal the thermodynamically important reactions for this goal. Most noticeably in the middle part of the figure, some of the reactions producing pyruvate and acetyl-CoA, which are the precursors of P(3HB), are found to appear frequently in the thermodynamically dominant pathways. The *mez2* and *Diss_Pyruvate 04* reactions appear, respectively, at 2.5 and 1.7 times of the average frequency. In the previous work, Hong and his co-workers[19] presented the significance of the *eda* reaction, along with that of the aforementioned two precursors. They have attributed this to the requirement of NADPH in the P(3HB) biosynthesis flux. The *eda* reaction by 2-Keto-3-deoxy-6-phosphogluconate aldolase, however, is not exceptionally favored in the thermodynamically dominant pathways. Note on the upper left side of Figure 2, that the cyclic reactions containing the *Isoprenoids01* reaction by Aldose reductase,

appear ~1.5−2.1 times more often in ~1% of thermodynamically dominant pathways than the average. Summation over these reactions reveals that this cyclic pathway contributes to the production of the NADPH, performing the similar role as the *eda* reaction in the P(3HB) biosynthesis. It should be noted that this study does not include the other metabolic reactions in *E. coli* (e.g., the production of biomass) that are competing with the P(3HB) biosynthesis pathway. Thus, the results can be further reinforced by enhancing the overall reactions. Nevertheless, even neglecting the accurate prediction of metabolic behaviors, the important reactions for the P(3HB) biosynthesis could be thoroughly identified based on the mass balances and the thermodynamic principles.

■ CONCLUSION

The efficacy of the proposed strategy has been ascertained through the exploration of two *E. coli* models. For both models, the sets of stoichiometrically feasible pathways leading to their respective optimal overall reactions have been exhaustively identified on a PC of moderate size by resorting to the graph-theoretic method, based on process graphs (*P*-graphs). Subsequently, these pathways have been differentiated by determining the thermodynamically dominant pathways among the feasible pathways on the basis of the Gibbs free-energy changes of reactions in the pathways. The proposed approach will contribute to understanding biochemical networks and facilitating the experimental design of gene regulation for enhanced production of the desired product.

■ APPENDIX A. *P*-GRAPH APPROACH FOR METABOLIC NETWORK

P-Graph

P-graph was developed for the algorithmic network synthesis of chemical processes.[24] It is comprised of the nodes of materials and operating units shown in Figure A.1. For metabolic networks, the role of operating nodes is played by metabolic reactions, and the materials are enzymes and nutrients. The nodes are connected with directed arcs, which indicate the direction of reaction pathways, and connection rules are expressed with the following axioms:[18]

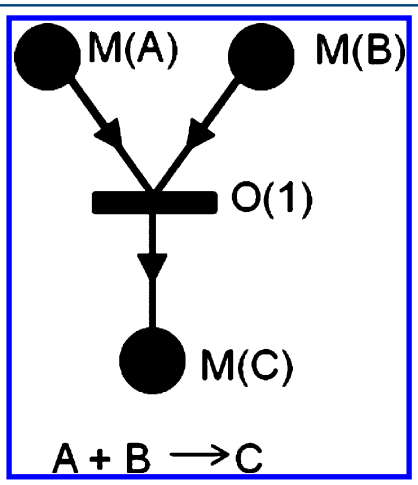

Figure A.1. *P*-graph representation for reaction A+B → C. O(1) is a reaction node, and M(A), M(B), and M(C) are material nodes.

(a) Six axioms of feasible reaction pathways

(R1) Every final product (target) is totally produced by the reaction steps represented in the pathway.

(R2) Every starting reactant (precursor) is totally consumed by the 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 activates a starting reactant (precursor).

(b) Seven axioms of the combinatorially feasible reaction networks

(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) 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.

In this way, syntactic and semantic contents of biochemical networks may be described and implemented in algorithms.

Maximal Structure

The maximal structure of a network is generated by identifying input and output materials of operating units, and then merging all the common material nodes. Accordingly, it contains all the combinatorially feasible pathways.

Stoichiometrically Feasible Pathways

Combinatorially feasible pathways of a network, which are subsets of the maximal structure, are generated based on the names of input and output materials. By further imposing stoichiometric ratios of metabolic reactions and mass balance constraints, stoichiometrically feasible pathways are obtained as the candidate solutions of network synthesis. In process networks, the most economical structure among them is selected as a solution network.

Algorithm PBT

Identification of metabolic pathways is different from designing a chemical process in that the former do not have measurable economic costs of reactions, unlike the latter, where each operating unit has operating and investment costs. This attribute is responsible for the generation of cyclic or dependent pathways during the calculation of solution pathways, which all produce a predefined overall reaction. It is worth noting that the desired

overall reaction should be calculated by linear programming. While the cyclic pathways frustrate the successful calculation of feasible pathways, generating linear combinations of independent pathways is unnecessary and raises the computational burden. Therefore, a Pathway Back-Tracking algorithm (algorithm PBT) is developed to generate the complete set of acyclic and independent feasible pathways by discriminating cyclic and dependent pathways during computation.

■ ASSOCIATED CONTENT

Supporting Information

(A) Estimation of the standard Gibbs free energy with the group contribution method; (B) calculation of the Gibbs free energy change of reaction; (C) the flux distributions of the thermodynamically dominant 100 pathways for maximal P (3HB) production and the abbreviations for metabolite names. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: +82-42-350-3920. Fax: +82-42-350-3910. E-mail: sunwon@kaist.ac.kr.

Present Addresses

[∇]Doosan Heavy Industries & Construction, Daejeon 305-348, Korea.

[⊥]Western Research Institute, Laramie, WY 82072.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Project for Developing Systems Metabolic Engineering Platform Technologies for Biorefineries, Technology Development Program to Solve Climate Changes (NRF-2012-C1AAA001-2012M1A2A2026556) and World Class University program (R322009000101420) of the Ministry of Education, Science and Technology (MEST) through the National Research Foundation of Korea; the Institute for Systems Design and Optimization at Kansas State University; and the Hungarian Scientific Research Fund (Project F61227). The authors deeply appreciate the advice and guidance received from Prof. L. T. Fan throughout the course of this work.

■ REFERENCES

(1) Kim, H. U.; Kim, T. Y.; Lee, S. Y. Metabolic flux analysis and metabolic engineering of microorganisms. *Mol. BioSyst.* **2008**, *4*, 113−120.

(2) Lee., S. Y.; Lee, D.-Y.; Kim, T. Y. Systems biotechnology for strain improvement. *Trends Biotechnol.* **2005**, *23*, 349−358.

(3) Lee, D.-Y.; Zimmer, R.; Lee, S. Y.; Park, S. Colored Petri net modeling and simulation of signal transduction pathways. *Metab. Eng.* **2006**, *8*, 112−122.

(4) Schilling, C. H.; Edwards, H. S.; Letscher, D.; Palsson, B. Ø. Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnol. Bioeng.* **2001**, *71*, 286−306.

(5) Schwartz, J. M.; Kanehisa, M. Quantitative elementary mode analysis of metabolic pathways: The example of yeast glycolysis. *BMC Bioinf.* **2006**, *7*, 186−205.

(6) Mahadevan, R.; Schilling, C. H. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* **2003**, *5*, 264−276.

(7) Klamt, S.; Stelling, J. Combinatorial complexity of pathway analysis in metabolic networks. *Mol. Biol. Rep.* **2002**, *29*, 233−236.

(8) Fan, L. T.; Bertok, B.; Friedler, F. Combinatorial framework for the systematic generation of reaction pathways. Presented at the *AIChE Annual Meeting*, Dallas, TX, 1999 (ISBNO-8169-0805-2).

(9) Seo, H.; Lee, D.-Y.; Park, S.; Fan, L. T.; Shafie, S.; Bertok, B.; Friedler, F. Graph-theoretical identification of pathways for biochemical reactions. *Biotechnol. Lett.* **2001**, *23*, 1551−1557.

(10) Lee, D.-Y.; Fan, L. T.; Park, S.; Lee, S. Y.; Shafie, S.; Bertok, B.; Friedler, F. Complementary identification of multiple flux distributions and multiple metabolic pathways. *Metab. Eng.* **2005**, *7*, 182−200.

(11) Stanley, I. S. *Chemical and Engineering Thermodynamics*, 3rd ed.; John Wiley & Sons, Inc.: New York, 1999.

(12) Yang, F.; Qian, H.; Beard, D. A. Ab initio prediction of thermodynamically feasible reaction directions from biochemical network stoichiometry. *Metab. Eng.* **2005**, *7*, 251−259.

(13) Kummel, A.; Panke, S.; Heinemann, M. Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Mol. Syst. Biol.* **2006**, *2*, Art. No. 2006.0034.

(14) Nolan, R. P.; Fenley, A. P.; Lee, K. Identification of distributed metabolic objectives in the hypermetabolic liver by flux and energy balance analysis. *Metab. Eng.* **2006**, *8*, 30−45.

(15) Henry, C. S.; Broadbelt, L. J.; Hatzimanikatis, V. Thermodynamics-based metabolic flux analysis. *Biophys. J.* **2007**, *92*, 1792−1805.

(16) Henry, C. S.; Jankowski, M. D.; Broadbelt, L. D.; Hatzimanikatis, V. Genome-scale thermodynamic analysis of *Escherichia coli* metabolism. *Biophys. J.* **2006**, *90*, 1453−1461.

(17) Varma, A.; Palsson, B. Ø. Metabolic flux balancing: Basic concepts, scientific and practical use. *Nat. Biotechnol.* **1994**, *12*, 994−998.

(18) Fan, L. T.; Bertok, B.; Friedler, F. A graph-theoretical method to identify candidate mechanisms for deriving the rate law of a catalytic reaction. *Comput. Chem.* **2002**, *26*, 265−292.

(19) Hong, S. H.; Park, S. J.; Moon, S. Y.; Park, J. P.; Lee, S. Y. In silico prediction and validation of the importance of the Entner-Doudoroff pathway in poly(3-hydroxybutyrate) production by metabolically engineered *Escherichia coli*. *Biotechnol. Bioeng.* **2003**, *83*, 854−863.

(20) Alberty, R. A. *Biochemical Thermodynamics: Applications of Mathematica*; John Wiley & Sons, Inc.: Hoboken, NJ, 2006.

(21) NIST database, http://xpdb.nist.gov/enzyme_thermodynamics/.

(22) Ishii, N.; Nakahigashi, K.; Bata, T.; Robert, M.; Soga, T.; et al. Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science* **2007**, *316*, 593−597.

(23) Qian, H.; Beard, D. A. Metabolic futile cycles and their functions: A systems analysis of energy and control. *Syst. Biol.* **2006**, *153*, 192−200.

(24) Friedler, F.; Tarjan, K.; Huang, Y. W.; Fan, L. T. Combinatorial algorithms for process synthesis. *Comput. Chem.* **1992**, *16*, S313−320.