Prediction of Maximum Yields of Metabolites and Optimal Pathways for Their Production by Metabolic Flux Analysis

HONG, SOON HO, SOO YUN MOON, AND SANG YUP LEE*

Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical and Biomolecular Engineering, BioProcess Engineering Research Center and Bioinformatics Research Center, Korea Advanced Institute of Science and Technology, 373-I Guseong-dong, Yuseong-gu, Daejeon 305-701, Korea

Received: January 28, 2003
Accepted: April 17, 2003

Abstract The intracellular metabolic fluxes can be calculated by metabolic flux analysis, which uses a stoichiometric model for the intracellular reactions along with mass balances around the intracellular metabolites. In this study, metabolic flux analyses were carried out to estimate flux distributions for the maximum in silico yields of various metabolites in Escherichia coli. The maximum in silico yields of acetic acid and lactic acid were identical to their theoretical yields. On the other hand, the in silico yields of succinic acid and ethanol were only 83% and 6.5% of their theoretical yields, respectively. The lower in silico yield of succinic acid was found to be due to the insufficient reducing power, but this lower yield could be increased to its theoretical yield by supplying more reducing power. The maximum theoretical yield of ethanol could be achieved, when a reaction catalyzed by pyruvate decarboxylase was added in the metabolic network. Furthermore, optimal metabolic pathways for the production of various metabolites could be proposed, based on the results of metabolic flux analyses. In the case of succinic acid production, it was found that the pyruvate carboxylation pathway should be used for its optimal production in E. coli rather than the phosphoenolpyruvate carboxylation pathway.

Key words: Metabolic engineering, metabolic flux analysis, in silico yield, optimal pathway

Metabolic engineering can be defined as the directed modification of cellular metabolism and properties through the introduction, deletion, and modification of metabolic pathways by using recombinant DNA and other molecular biological tools [1, 13]. Various metabolic engineering strategies have been applied for the more efficient production of desired metabolites and biomolecules [13]. Even though enhanced production of some biomolecules has been successful, many other attempts have failed due to the lack of rational strategies based on a predictable technique. Therefore, metabolic flux analysis, which allows the calculation of intracellular metabolic fluxes based on the stoichiometry of intracellular reactions and mass balances around the intracellular metabolites, has become an essential tool for metabolic engineering [4, 16].

Metabolic flux analysis is based on the pseudo-steady state assumption, which means no net intracellular accumulation of intermediates, considering the high turnover of intracellular metabolite pools [19, 21]. Metabolic flux analysis has been applied to calculate the maximum theoretical yield of a desired metabolite to be produced, and to identify the rigidity of branch points in the metabolic pathways [11, 20]. Another possible application is the identification of alternative metabolic pathways that lead to the desired product [14, 22]. Detailed theories and applications of the flux analysis can be found in recent reviews [4, 19].

Determination of the achievable maximum yield and optimal metabolic pathways is essential for the engineering of the metabolic pathways and for the redirection of metabolic fluxes towards the desired bioproducts. In this study, metabolic flux analyses were carried out for the estimation of the maximum in silico yields of four metabolites in Escherichia coli and for the determination of the optimal flux distribution.

Theoretical Backgrounds

Construction of In Silico Metabolic Network. From the results of E. coli genome sequencing project, about 600 to 700 metabolic reactions were identified [2]. Most of the reactions such as amino acid transport and utilization are, however, not essential for cell growth, and consequently, 224 metabolic reactions can support growth of E. coli on glucose [3, 17]. Therefore, in this study, a metabolic network was constructed consisting of 127 reversible and
174 irreversible reactions and 294 metabolites, containing all metabolic pathways required for the growth of E. coli on glucose. The list of reactions and metabolites can be found from our website (http://mbel.kaist.ac.kr).

**Metabolic Flux Analysis.** Metabolic flux analyses were carried out for the calculation of the volumetric rates of formation of intracellular metabolites [13, 16]. For the calculation of intracellular metabolic fluxes, the composition of E. coli biomass was assumed to be 55% protein, 20.5% RNA, 3.1% DNA, 9.1% lipids, 3.4% lipopolysaccharides, 2.5% peptidoglycan, 2.5% glycogen, 0.4% polyamines, and 3.5% other metabolites, cofactors, and ions [15]. The types and amounts of precursors required to synthesize these macromolecules were determined from the average composition of each of the macromolecules in E. coli: Protein: Ala 9.6%, Arg 5.53%, Asn 4.51%, Asp 4.51%, Cys 1.71%, Gln 4.92%, Gly 4.92%, His 1.77%, Ile 5.43%, Leu 8.42%, Lys 6.42%, Met 2.87%, Phe 3.46%, Pro 4.13%, Ser 4.03%, Thr 4.74%, Trp 1.06%, Tyr 2.58%, Val 7.91%; RNA: ATP 26.2%, GTP 32.2%, CTP 20.0%, UTP 21.6%; DNA: dATP 24.7%, dGTP 25.4%, dCTP 25.4%, dTTP 24.7%; phospholipids: phosphatidylethanolamine 72.8%, phosphatidylglycerol 19.8%, cardiolipin 3.4% [15].

The flux balance model was set up as follows:

$$ S \cdot v = b $$

where $S$ is the stoichiometric matrix, $v$ is a vector for reaction rates or fluxes, and $b$ is the vector for the rates of metabolites consumption and excretion [4]. Since this study's system is underdetermined, i.e. the number of reactions is greater than the number of metabolites or constraints, linear optimization was carried out with the following objective function:

Maximize: $Z = \sum c_i v_i$

where $c_i$ are the weights and $v_i$ are the elements of the flux vector [21]. Linear optimization was carried out using the program LP_solve (State University of New York, NY, U.S.A.). To obtain feasible solutions, constraints on the rates of metabolites uptake and excretion were applied (Table 1). As seen in Table 1, cells are allowed to operate both aerobic and anaerobic pathways with no limit of nutrient.

**Definition of In Silico Yield and Theoretical Yield.** The in silico yield is defined as the yield calculated by metabolic flux analysis, which represents the actual amount of metabolite that can be formed through the metabolic reactions in operation from one mole of substrate. The theoretical yield is defined as the yield which can be achieved by the complete conversion of substrate to metabolite.

### RESULTS

#### Maximum Metabolite Yields

The maximum capacities of metabolites production were estimated by metabolic flux analysis. The maximum theoretical and in silico yields of acetic acid, lactic acid, succinic acid, and ethanol on glucose are shown in Table 2. As expected, the maximum in silico yields of acetic acid and lactic acid were 2 mole/mole glucose, which are identical to their theoretical yields.

However, the maximum in silico yield of succinic acid and ethanol were only 1.65 mole/mole glucose (83% of the maximum theoretical yield) and 0.13 mole/mole glucose (6.5% of the maximum theoretical yield), respectively. It was found that the CO$_2$ availability for maximum production of succinic acid was 0.82 mole/mole glucose/h even though the CO$_2$ consumption rate was not restricted (Table 1). CO$_2$ is required for the carboxylation of C$_2$-compounds.

<table>
<thead>
<tr>
<th>Metabolites</th>
<th>Acetic acid</th>
<th>Lactic acid</th>
<th>Succinic acid</th>
<th>Ethanol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum theoretical yield</td>
<td>0.67* (2***)</td>
<td>1 (2)</td>
<td>1.31 (2)</td>
<td>0.51 (2)</td>
</tr>
<tr>
<td>Maximum in silico yield</td>
<td>0.67 (2)</td>
<td>1 (2)</td>
<td>1.08 (1.647)</td>
<td>0.51*** (2****)</td>
</tr>
</tbody>
</table>

*aMass yield (g metabolites/g glucose).

**Molar yield (mole metabolites/mole glucose).

***Mass yield when an additional reaction catalyzed by pyruvate decarboxylase was introduced.

****Molar yield when an additional reaction catalyzed by pyruvate decarboxylase was introduced.
Fig. 1. External supply of reducing power. Ubiquinol or menaquinol (QH$_2$) is oxidized to ubiquinone or menaquinone (Q), during which the electron acceptor (shown here is FAD) is reduced.

(phosphoenolpyruvate and pyruvate) to C$_7$-compounds (oxaloacetate and malic acid), which are further converted to succinic acid [10]. Therefore, 2 moles of CO$_2$ are theoretically required for each mole of glucose to achieve the maximum yield of succinic acid. These results indicate that there exists a limiting factor. As stated earlier, however, cells are operating metabolic pathways without any nutrient limitation. Examination of the metabolic reaction network that was constructed suggested that there was no reaction which incorporated external reducing power. It was therefore reasoned that the limiting factor for succinic acid production might be reducing power. To examine this hypothesis, the following reaction was added to the metabolic reaction network:

$$QH_2 \rightarrow Q + 2H^+ + 2e^-$$  \hspace{1cm} (3)

where QH$_2$ is ubiquinol or menaquinol and Q is ubiquinone or menaquinone. The electrons generated can be transferred to several different acceptors including FAD (Fig. 1) and fumaric acid. After incorporating this reaction, intracellular flux distribution was estimated again for all four cases. The maximum in silico yield of acetic acid, lactic acid, and ethanol were the same as before. However, the maximum in silico yield of 2 moles of succinic acid/mole glucose was achieved, instead of 1.65 mole/mole glucose, by supplying more reducing power. This result is consistent with our previous report showing that succinic acid flux was controlled by reducing power in *E. coli*, and that the succinic acid production could be enhanced by using a more reduced carbon substrate such as sorbitol [7].

Ethanol production is rather inefficient in *E. coli*. Good ethanol producers such as yeast and *Zymomonas* possess pyruvate decarboxylase for the conversion of pyruvate to acetaldehyde, which in turn is converted to ethanol. Interestingly, when this pyruvate decarboxylase reaction was added to the metabolic network of *E. coli*, the maximum ethanol yield of 2 moles/mole glucose could be achieved. It has already been reported that the introduction of the *Zymomonas mobilis* pyruvate decarboxylase gene in *E. coli* allowed the efficient production of ethanol [9].

**Optimal Flux Distributions for Maximum Metabolite Production**

The optimal metabolic flux distributions for the production of acetic acid, lactic acid, succinic acid, and ethanol are shown in Fig. 2. The optimal flux distribution for the production of lactic acid is quite simple. All glucose is converted to pyruvate through glycolysis and further converted

---

Fig. 2. Optimized metabolic flux distributions during the production of (A) acetic acid, (B) lactic acid, (C) succinic acid, and (D) ethanol. Abbreviations: ACCOA, acetyl-CoA; AKG, α-ketoglutarate; CIT, citrate; CO$_2$, carbon dioxide; DHAP, dihydroxyacetone phosphate; E4P, erythrose-4-phosphate; F16P, fructose-1,6-bisphosphate; F6P, fructose-6-phosphate; FUM, fumaric acid; G3P, glyceraldehyde-3-phosphate; G6P, glucose-6-phosphate; ICT, isocitrate; MAL, malic acid; O2, oxygen; OA, oxaloacetate; PEP, phosphoenolpyruvate; PYR, pyruvate; R5P, ribose-5-phosphate; SUCCOA, succinyl-CoA; Q, ubiquinone or menaquinone; QH$_2$, ubiquinol or menaquinol; ext, external.
to lactic acid. No fluxes to the pentose phosphate pathway nor to the macromolecule biosynthesis, which included amino acids and lipid, were detected from flux analysis.

For acetic acid production, all glucose is converted to phosphoenolpyruvate (PEP) in the same manner as in lactic acid production. However, half of PEP is carboxylated to oxaloacetate by PEP carboxylase and further converted to pyruvate by malate dehydrogenase and malic enzyme:

\[
\text{Phosphoenolpyruvate} + \text{CO}_2 \rightarrow \text{Oxaloacetate} + \text{Phosphate} \\
\text{Oxaloacetate} + \text{NADH} \rightarrow \text{Malic acid} + \text{NAD} \\
\text{Malic acid} + \text{NAD} \rightarrow \text{CO}_2 + \text{NADH} + \text{Pyruvate}
\]

Through this pathway, free energy of PEP is wasted as inorganic phosphate while ATP can be generated by pyruvate kinase.

\[
\text{PEP} + \text{ADP} \rightarrow \text{Pyruvate} + \text{ATP}
\]

Even though half of the PEP's free energy was wasted, the ATP produced through acetic acid production was three times higher than that obtained through other metabolite production pathways (Table 3):

<table>
<thead>
<tr>
<th>Metabolites</th>
<th>Acetyl-CoA</th>
<th>Pyruvate</th>
<th>ATP</th>
<th>NADH</th>
<th>FADH₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetic acid</td>
<td>2</td>
<td>2</td>
<td>6.3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Lactic acid</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Succinic acid</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Ethanol</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

*Formation rates: sum of metabolites producing fluxes (mM/g DCW/h).

decarboxylase was added to the metabolic network (Fig. 2D). As in lactic acid production, all glucose was converted to pyruvate through glycolysis and further converted to ethanol.

**DISCUSSION**

Through metabolic flux analysis, the maximum in silico yields and the optimal metabolic pathways for the production of four metabolites could be determined. We used well-known single primary metabolites as model systems to demonstrate the usefulness of this analysis. Based on the yields and optimal pathways predicted in silico, rational metabolic engineering strategies can be established. For example, resting cell could be applied for the efficient production of lactic acid, succinic acid, and ethanol, since any macromolecule biosynthesis fluxes are not detected from the optimal pathways. However, growing cell might be more suitable for acetic acid production, since large amount of ATP which may exceed the capacity of the E. coli metabolic network is produced. It can be utilized through macromolecule biosynthesis, and thus ATP balance could be achieved in the growing cell. Furthermore, metabolic flux analysis can also suggest suitable culture conditions for metabolites production. As shown in Fig. 2, aerobic culture condition is suitable for acetic acid and ethanol production, while anaerobic culture condition is more appropriate for the production of lactic acid and succinic acid.

As mentioned above, the proposed optimal pathway for succinic acid production is different from the typical succinic acid production pathway. Usually, succinic acid is produced from PEP by PEP carboxylation. However, considering the low energy efficiency of the PEP carboxylation pathway (loss of high energy phosphate bond), pyruvate is a more appropriate substrate for succinic acid production. These results suggest that the metabolism of E. coli has not been optimized for succinic acid production. This was supported by previous reports that efficient production of succinic acid could be achieved by the amplification of the malic enzyme in the ldhA pfl double-mutant E. coli strain NZN111 [6, 7]. Moreover, it was found that succinic acid productivity could be enhanced if redox potential was balanced by the supply of additional reducing power [7].

The maximum theoretical yield can be defined as the yield which can be achieved by complete conversion of substrate to metabolite. However, the maximum theoretical yield is different from the real maximum yield achievable under physiological conditions. To estimate a reasonable in silico maximum yield, ATP requirements for maintenance can be incorporated. Maintenance energy can be classified into growth associated and non-growth associated fractions. For the maximum metabolites production, the growth
Table 4. The maximum in silico yield of various metabolites when maintenance energy was considered.

<table>
<thead>
<tr>
<th>Metabolites</th>
<th>Acetic acid</th>
<th>Lactic acid</th>
<th>Succinic acid</th>
<th>Ethanol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum in silico yield</td>
<td>2°</td>
<td>2</td>
<td>1.647 (2**)</td>
<td>0.073 (1.16***)</td>
</tr>
</tbody>
</table>

°Molar yield (mole metabolites/mole glucose).
**Molar yield with sufficient supply of external reducing power.
***Molar yield when an additional reaction catalyzed by pyruvate decarboxylase was introduced.

associated fraction can be ignored. It was reported that the non-growth associated maintenance requirement is 7.6 mmol of ATP/g DCW/h in E. coli under normal growth condition [21]. Under anaerobic condition, however, the non-growth associated ATP requirement is only 0.69 mmol of ATP/g DCW/h [6]. As mentioned above, optimal culture conditions for metabolites production can be formulated based on flux analysis. Using these proposed culture conditions, metabolic flux analyses were carried out again to predict the optimal metabolic pathway and the maximum

Fig. 3. Optimized metabolic flux distributions re-estimated by considering ATP requirement during the production of (A) acetic acid, (B) lactic acid, (C) succinic acid, and ethanol (D) without and (E) with additional reaction catalyzed by pyruvate decarboxylase. Abbreviations: 3PDGlu, 3-phosphoglycerate; AC, acetate; ACCOA, acetyl-CoA; AKG, α-ketoglutarate; ALA, alanine; ASP, O-acetylseryine; ASP, aspartate; CIT, citrate; CO2, carbon dioxide; CYS, cysteine; DHAP, dihydroxyacetone phosphate; E4P, erythrose-4-phosphate; F6P, fructose-6-phosphate; F6P, fructose-6-phosphate; FORMATE, formate; FTHF, 10-formyl-tetrahydrofolate; FUM, fumaric acid; G3P, glyceroldehyde-3-phosphate; G6P, glucose-6-phosphate; GLN, glutamine; GLU, glutamate; GLY, glycine; H2S, hydrogen sulfide; H2SO3, hydrogen sulfite; H2SO4, hydrogen sulfate; ICT, isocitrate; MAL, malic acid; METTHF, 5,10-methylene tetrahydrofolate; O2, oxygen; OA, oxaloacetate; PAP, adenosine-3,5-diphosphate; PEP, phosphoenolpyruvate; PRP, phosphoribosyl pyrophosphate; PVR, pyruvate; Q, ubiquinone or menaquinone; QH2, ubiquinol or menaquinol; RSP, ribose-5-phosphate; RLS, ribulose 5-phosphate; S7P, sedo-heptulose; SER, serine; SUCCOA, succinyl-CoA; THR, threonine; XSP, xylulose-5-phosphate; ext, external.
in silico yield with consideration of maintenance requirements. The results of metabolic flux analyses are presented in Table 4 and Fig. 3. The optimal metabolic pathway for acetic acid production was the same as that obtained when not considering the ATP requirement. However, the oxygen uptake rate increased 2-fold. The CO₂ excretion rate was 2 mol/e DCW/h when considering ATP requirement, while there was no excretion of CO₂ when the maintenance energy was not considered (Fig. 3A). These results indicate that respiration plays an important role in ATP generation. The optimal metabolic pathway for lactic acid production was also the same as that obtained when not considering the maintenance energy. Similarly, metabolic flux distribution for succinic acid production under sufficient supply of external reducing power was equal to the flux distribution without considering the maintenance energy. The optimal metabolic pathway for ethanol production was somewhat different when the ATP requirement was considered. A major difference was early found was the production of acetic acid. With consideration of ATP requirement, 0.47 mole of acetic acid/mole glucose was produced while only 0.073 mole of ethanol/mole glucose was produced. The molar ratio of acetic acid to ethanol is 6.5 mole acetic acid/mole ethanol. This result again indicates that wild-type E. coli is not suitable for the production of ethanol. The effect of introducing pyruvate decarboxylase into E. coli was estimated by metabolic flux analysis with maintenance requirement (Fig. 3E). When the reaction catalyzed by pyruvate decarboxylase was added to the metabolic reaction network, the maximum yield increased to 1.16 mole ethanol/mole glucose, which is 16 times higher than that obtained without introducing this reaction. Furthermore, acetic acid production decreased to 0.28 mole ethanol/mole glucose from 0.47 mole ethanol/mole glucose by introducing this reaction into the metabolic network. From these results, it can be concluded that pyruvate decarboxylase is an attractive enzyme as it provides two beneficial effects: enhanced metabolite (ethanol) production and decreased byproduct (acetic acid) formation.

In this study, we applied metabolic flux analysis technique to estimate the maximum metabolic capacities of the metabolic network of E. coli and to evaluate the achievable in silico maximum theoretical yields. In some cases, the maximum theoretical yield cannot be achieved even under ideal conditions due to the lack of metabolic limitation inherent in the metabolic network. It was found that this limiting factor could be identified by flux analysis, and thus, the optimal metabolite production fluxes could be proposed. Additionally, the effects of introducing new reaction pathways into metabolites production can be predicted by metabolic flux analysis. Based on the optimal metabolic pathways predicted in silico, new efficient metabolite production systems can be suggested. Therefore, this type of in silico metabolic flux analysis will reduce the time, money, and effort required to develop metabolically engineered strains for the enhanced production of various bioproducts.

Acknowledgments

This work was supported by the IMT2000 project (IMT2000-C3-1) of the Ministry of Information and Communication (MIC) and Ministry of Science and Technology (MOST), and by the National Research Laboratory Program (2000-N-NL-01-C-237) of MOST. Hardware for computational analysis was supported by the IBM-SUR program.

REFERENCES