Estimating Optimal Profiles of Genetic Alterations Using Constraint-Based Models

**Kapil G. Gadkar,[1] Francis J. Doyle III,[1]
Jeremy S. Edwards,[2] Radhakrishnan Mahadevan[3]**

[1]Department of Chemical Engineering, University of California, Santa Barbara, California
[2]Department of Chemical Engineering, University of Delaware, Newark, Delaware
[3]Department of Bioengineering, University of California, San Diego, La Jolla, California

Received 21 January 2004; accepted 29 September 2004

Published online 7 December 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/bit.20349

Abstract: Metabolic engineering involves application of recombinant DNA methods to manipulate metabolic networks to improve cellular properties. It is critical that the genetic alterations be performed in an optimal manner to maximize profit. In addition to the product yield, productivity consideration is also critical, especially for the production of bulk chemicals such as 1,3-propanediol. In this work, we demonstrate that it is suboptimal from the standpoint of productivity to induce genetic alteration at the start of the production process. A bilevel optimization scheme is formulated to determine the optimal temporal flux profile for the manipulated reaction. In the first case study, an optimal flux in the reaction catalyzed by glycerol kinase is determined to maximize the glycerol production at the end of a 6-h batch cultivation of *Escherichia coli* under aerobic conditions. The final glycerol concentration is 30% higher for the optimal flux profile compared with having an active flux during the entire batch. The effect of the mass transfer coefficient on the optimal profile and the glycerol concentration is also determined. In the second case study, the anaerobic batch fermentation of the *ldh*$^-$ strain of *Escherichia coli* is considered. The optimal flux in the acetate pathway is determined to maximize the final ethanol concentration. The optimal flux results in higher ethanol concentration (11.92 mmol L^{-1}) compared to strains with no acetate flux (8.36 mmol L^{-1}) and fully active acetate flux (6.22 mmol L^{-1}). We also examine the effects of growth inhibition due to high ethanol concentrations and variations in final batch time on ethanol production.

Keywords: metabolic engineering; bilevel optimization; constraint-based modeling; ethanol overproduction; flux balance analysis

INTRODUCTION

Metabolic engineering involves application of recombinant DNA methods to manipulate metabolic networks for improvement of metabolite and protein production capabilities (Bailey, 1991). The successful application of metabolic engineering involves redirection of metabolic flux from the production of biomass to the production of metabolite products; however, metabolic pathways have evolved to exhibit control architectures that resist flux alterations (Stephanopoulos and Vallino, 1991). The demands of metabolite production placed upon the cell interfere with the cell's built-in objective to grow, and gene knockouts and high levels of gene induction lead to growth retardation (Kurland and Dong, 1996). With advances in metabolic manipulation techniques (Herring et al., 2003; Kim and Cha, 2003), there are two broad classes of challenges. The first is to identify the optimal genetic changes to achieve the optimal product yield (Nielsen, 1998). The second is to design a dynamic controller for gene expression that allows for optimal productivity (Farmer and Liao, 2000). This involves re-engineering of the gene-regulatory hierarchy that controls gene expression to achieve enhanced production, minimized growth retardation, and reduced toxic by-product formation.

A critical issue for engineering microbial strains for metabolic production is identification of the optimal production pathway. Varma and Palsson (1994) utilized linear programming to identify optimal production of various metabolites in *Escherichia coli.* Burgard and Maranas (2001) developed a modeling tool to identify optimal selection of genes to introduce into *E. coli* for optimal production. Burgard and co-workers (2003) also developed a bilevel optimization strategy, OptKnock, to identify gene knockout strategies for biochemical overproduction of the desired product. Their strategy incorporates the knowledge that the cell will redistribute the metabolic fluxes to achieve its intrinsic objective to maximize growth rate, as has been demonstrated experimentally for the *E. coli* MG1655 strain under certain growth conditions (Edwards et al., 2001; Ibarra et al., 2002). The OptKnock strategy can only identify the key reactions that effect product formation and, hence, need to be manipulated. However, it may

Correspondence to: R. Mahadevan, Genomatica, Inc., 5405 Morehouse Drive, Suite 210, San Diego, CA 92121; telephone: 858-362-8562; fax: 858-824-1772; e-mail: rmahadevan@genomatica.com

Contract grant sponsor: National Science Foundation
Contract grant number: BES-0000961

not be optimal to implement the suggested gene alterations at the start of the process. Dynamic regulation of the corresponding gene expressions is required to maximize the total product formed at the end of the process. In this work, a bilevel optimization scheme is formulated in which the optimal flux profiles are determined, corresponding to the key reactions identified using an approach such as OptKnock.

The next important issue for obtaining optimal productivity is the controlled synthesis of key metabolic enzymes. Kim and Cha (2003) employed the antisense RNA strategy (Desai and Papoutsakis, 1999) as a sophisticated metabolic engineering tool. The antisense strategy was employed to regulate synthesis of acetate pathway enzymes to enhance production of foreign proteins in *E. coli*. Regulation of the acetate pathway was such that there was sufficient ATP production due to acetate secretion, thereby preventing inhibition of growth and recombinant protein production. It was shown that 10% to 20% transcriptional downregulation of the acetate production pathway genes resulted in a 1.6-fold increase in foreign protein production. Furthermore, the timing of antisense expression was controlled using the intrinsic *ackA* promoter, which resulted in efficient utilization of the metabolic pathway. Optimal control of expression of the acetate production pathway was critical to increase protein production, and a simple knockout mutation of the target gene was not an appropriate method to increase foreign protein production. In this study, a modeling tool is developed to identify the optimal control strategy to maximize productivity using bilevel optimization. The method is demonstrated for glycerol and ethanol production in *E. coli*.

CASE STUDY

In the first case study, glycerol production at the end of a batch in an aerobic cultivation of *E. coli* was maximized by determining the optimal reaction flux catalyzed by glycerol kinase. This could be achieved by manipulating the gene expression of *glpK*, which codes for the glycerol kinase enzyme. Flux from the glycolysis pathway intermediate diacetone phosphate is diverted toward glycerol via two reactions. The first reaction catalyzed by the enzyme glycerol-3-phosphate dehydrogenase converts diacetone phosphate to glycerol-3-phosphate, which is further converted to glycerol by the enzyme glycerol kinase. Without metabolic engineering, all the carbon flux is used to produce growth precursors and no glycerol is produced. A flux diverted toward glycerol production affects growth rate as illustrated in Figure 1. The glycerol kinase–catalyzed reaction flux was optimized such that a balance was achieved between glycerol production rate and growth rate to maximize productivity given a defined batch time. The effect of variations in oxygen availability on the optimal flux profile for glycerol production was examined. Although glycerol was considered in this case study, the approach can easily be extended to analyze other industrially relevant bulk chemicals that are synthesized by a biological route, such as 1,3-propanediol (Nakamura and Whited, 2003).

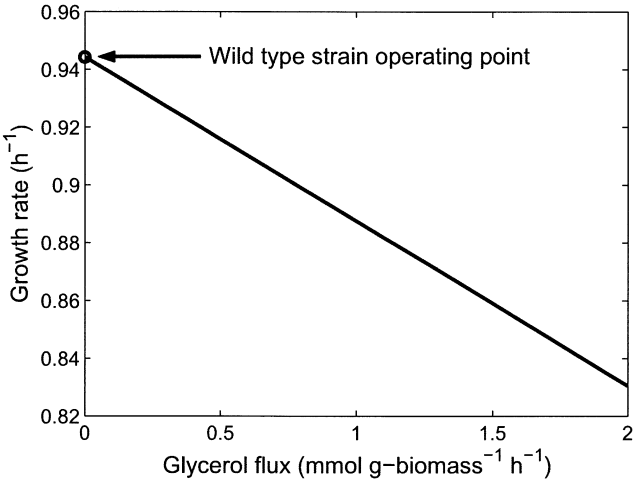

Figure 1. Effect of glycerol flux on growth rate of *Escherichia coli* under aerobic conditions.

The second case study considered ethanol production in an anaerobic batch fermentation of *E. coli*. The use of genetic engineering for ethanol production has been well established (Ingram et al., 1987). Under anaerobic growth with glucose as the substrate, *E. coli* produces a number of reduced byproducts including acetate, ethanol, lactate, and succinate. Acetate is formed mainly to regenerate ATP levels in the cells (Atkinson, 1977), whereas ethanol and lactate production is associated with oxidation of NADH (Leonardo et al., 1996). The lactate-producing pathway competes with the ethanol-producing pathway for NADH oxidation inside the cell. An *ldh* knockout strain (gene coding for lactate dehydrogenase) cannot produce lactate, which results in an increased ethanol flux (Yang et al., 1999). For increased ethanol production rates, it is also beneficial to knock out the *ackA* gene that codes for acetate kinase enzyme, which shuts off the acetate flux. This results in an increased carbon flux available for ethanol production; however, as a result of the knockout, ATP generation in the mutant is decreased and the growth rate is severely affected. Figure 2 shows the variations in growth rate and ethanol flux with changes in acetate flux. In maximizing the end-ethanol production, an optimal temporal acetate flux must be determined over the batch time so there is optimal distribution of fluxes favoring growth and ethanol production over the period of the batch. Underwood et al. (2002) demonstrated that changes in carbon partitioning between fermentation reactions and biosynthesis is critical in reducing time of fermentation during ethanol production. As an extension of the ethanol production case study, the effect of growth inhibition at high ethanol concentrations and varying batch times on

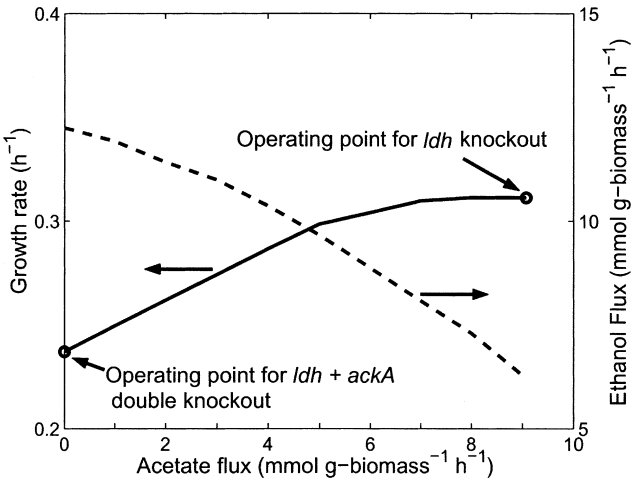

Figure 2. Effect of acetate flux on growth rate and ethanol flux in *Escherichia coli* under anaerobic conditions.

optimal acetate flux profiles and ethanol production was also examined.

METHODS

In this work, an *in silico* representation of *E. coli* (Covert and Palsson, 2002; Schilling et al., 2000) was employed. The model included all reactions in the central metabolism of the cell, consisting of a total of 57 internal reactions, 13 exchange fluxes, and 54 internal metabolites. The flux balance analysis (FBA) approach was used to determine the metabolic flux distribution in the network (Edwards et al., 2002). The FBA solution is based on the assumption that the metabolic fluxes are distributed optimally such that a built-in cellular objective is achieved. It has been demonstrated, under various conditions, that the objective is maximization of cell growth rate (Ibarra et al., 2002); therefore, growth rate was used as the objective in this work. Glucose was considered the only carbon source. The maximum glucose uptake rate was prespecified at 10 mmol g biomass^{-1} h^{-1} and uptake was enabled via both the phosphotransferase system and the glucokinase enzyme. Unconstrained uptake of inorganic phosphate was considered and secretion fluxes for acetate, ethanol, formate, lactate, glycerol, succinate, pyruvate, and carbon dioxide were enabled.

A bilevel optimization scheme was formulated to obtain the optimal profile for the manipulated flux to maximize the desired product concentration at the end of the batch. The optimal manipulated flux profile was achieved through temporal genetic modifications of the corresponding gene. However, the desired objective of maximum metabolite production was always in conflict with the cellular objective to maximize growth rate. The unregulated fluxes in the network were optimally distributed by the cell to maximize growth. Using this approach, the trade-off between growth and product maximization was addressed in a systematic fashion. Bilevel optimization can be formulated as follows:

$$\max_{v_c(t_0),\cdots,v_c(t_{f-1})} P(t_f)$$

$$\text{subject to :} \left(\begin{array}{lll} \max\limits_{v_j(t_i)} & v^{\text{growth}}(t_i) & \forall \quad j \in Q \\ \text{subject to :} & \sum\limits_{j=1}^{M} S_{kj} v_j(t_i) = 0 & \forall \quad k \in N \\ & \alpha_j \le v_j(t_i) \le \beta_j & \forall \quad j \in M \\ & v_j(t_i) = v_c(t_i) & \forall \quad j \in U \end{array} \right)$$

$$\frac{dP}{dt} = v^{\text{product}}(t_i)X$$

$$\frac{dX}{dt} = v^{\text{growth}}(t_i)X$$

$$\frac{dO_2}{dt} = v^{O_2}(t_i)X + k_L a(O_2^* - O_2)$$

$$\frac{dCO_2}{dt} = v^{CO_2}(t_i)X + k_L a(CO_2^* - CO_2)$$

$$t_i = t_0 + iT_s \quad i = 1, \cdots, f \tag{1}$$

In the outer optimization the optimal profile of the manipulated flux, v_c, was determined such that the desired product concentration at the end of the batch, $P(t_f)$, was maximized. This outer optimization was subject to an inner optimization at each sampling time over the period of the batch. This inner optimization forced the unregulated fluxes to be distributed such that growth rate was maximized at each sample time. The manipulated flux profile determined in the outer optimization was forced in the inner optimization. All fluxes were assumed to be constant between the two sampling times.

The optimal glycerol kinase flux for maximum production of glycerol at the end of the batch is considered. Determination of optimal flux in the acetate pathway for the maximum ethanol production with and without growth inhibition by high ethanol concentrations is also considered. Growth inhibition was modeled by reducing the maximum glucose uptake rate with an increase in ethanol concentration:

$$v_{\text{glucose}}^{\text{uptake}} = v_{\text{glucose}}^{\text{uptake}^*} \frac{1}{1 + \frac{[E]}{K_i}} \tag{2}$$

For these cases, the optimization shown in Eq. (1) was used. Further, the effect of oxygen availability on optimal glycerol kinase flux for glycerol production and the effect of end-batch times on optimal flux in the acetate pathway for ethanol production are considered. In these two cases, the bilevel optimization was modified slightly. Instead of determining the optimal flux profiles over the entire batch, the strategy involves determination of the optimal time of regulation of the flux from an initial value

to a final value. The resulting bilevel optimization was as follows:

$$\max_{t_{\text{reg}}} \; P(t_f)$$

$$\text{subject to :} \left(\begin{array}{lll} \max\limits_{v_j} & v^{\text{growth}} & \forall \quad j \in Q \\ \text{subject to :} & \sum\limits_{j=1}^{M} S_{kj} v_j = 0 & \forall \quad k \in N \\ & \alpha_j \le v_j \le \beta_j & \forall \quad j \in M \\ & v_j = v_c^- & \forall \quad j \in U, t < t_{\text{reg}} \\ & v_j = v_c^+ & \forall \quad j \in U, t \ge t_{\text{reg}} \end{array} \right)$$

$$\frac{dP}{dt} = v^{\text{product}} X$$

$$\frac{dX}{dt} = v^{\text{growth}} X$$

$$\frac{dO_2}{dt} = v^{O_2} X + k_L a (O_2^* - O_2)$$

$$\frac{dCO_2}{dt} = v^{CO_2} X + k_L a (CO_2^* - CO_2) \qquad (3)$$

In this case, in the outer optimization, the optimal time of regulation, t_{reg}, was determined to maximize the desired product concentration at the end of the batch. The manipulated flux was maintained at v_c^- for times $<t_{reg}$ and v_c^+ for times $\ge t_{reg}$. The quantities v_c^- and v_c^+ are fixed quantities and they represent the values of the manipulated flux before and after induction/repression, respectively. These values were predetermined and used as fixed parameters in the bilevel optimization. The outer optimization was subject to the inner optimization, similar to the previous case.

The bilevel optimization problem was solved in MATLAB. The outer nonlinear optimization was evaluated using *fmincon*, whereas the inner linear optimization was evaluated using *linprog*. In the case of nonlinear optimization, evaluation was performed from multiple initial conditions to approximate the globally optimal result. All simulations were performed on a 2.4-GHz Intel Pentium processor with 512-MB memory.

RESULTS

Glycerol Production

Batch growth over 6 h with an initial glucose concentration of 12 mmol L^{-1} was considered. For the bilevel formulation, the sample time (T_s) over which the manipulated flux was kept constant was 0.25 h. A high k_La value of 80 h^{-1} was used in this simulation to ensure that oxygen uptake was not limiting. Figure 3 shows the optimal flux catalyzed by the glycerol kinase enzyme, which had levels controlled by modulating expression of the corresponding gene, *glpK*. This could be achieved using a promoter system with an inducer such as arabinose, although this does not guarantee that flux through glycerol is at the maximum allowable value (Khlebnikov et al., 2002). The optimal flux was zero for the first 3.25 h, subsequently reaching the maximum allowable value. This can be explained by the fact that, initially, the entire carbon flux was directed toward growth to build up the biomass. During this time, there was no accumulation of glycerol. When the biomass was increased to an optimal level, glycerol production flux was increased to its upper bound of 2 mmol g biomass^{-1} h^{-1}. During this phase, a portion of the carbon flux was diverted to glycerol production at the expense of a suboptimal growth flux. The time at which the glycerol pathway was induced was optimized in such a way that all glucose was utilized by the cell in the defined batch time. The final glycerol concentration was 2.28 mmol L^{-1}. Figure 3 also shows the profiles of glycerol, glucose, and biomass concentrations during the batch. The average computation time for this bilevel optimization using Eq. (1) was 1 h (Table I). If the glycerol pathway was induced at the start of the batch, the growth rate was affected and biomass was not built up in an optimal manner. Consequently, not all available glucose was used in the batch time of 6 h and, more importantly, the final glycerol concentration was suboptimal (1.75 mmol L^{-1}).

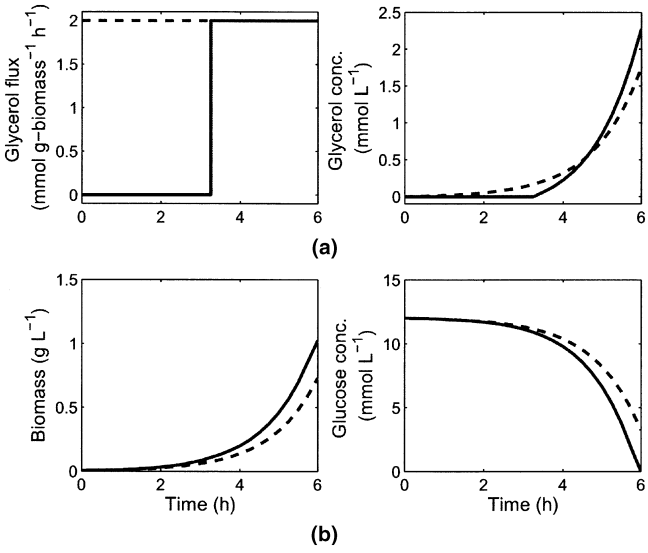

Figure 3. (a) Glycerol flux and glycerol concentration profiles, (b) biomass and glycerol concentration profiles. Solid line, strain with optimal glycerol flux; dashed line, strain with maximum glycerol flux for the entire batch time.

Effect of k_La on Glycerol Production

The effect of changes in oxygen availability to the cell on the optimal glycerol kinase flux for maximizing glycerol production at the end of the batch was investigated. Variations in the optimal flux profile for different values of the mass transfer coefficient, k_La, were studied. Earlier in this study we noted that the optimal flux was zero for the initial part of the batch and then increased to a maximum value of 2 mmol g biomass^{-1} h^{-1}. The trend of the profile did not change for different k_La values, but the time for

Table I. Average computation time for the two bilevel optimization schemes.

	Computation time[a]	
Case study	Bilevel scheme 1	Bilevel scheme 2
Glycerol	1.00 h	035 s
Ethanol	2.87 h	300 s

[a]All simulations performed on a 2.4-GHz Intel Pentium processor and 512-MB memory.

upregulating the flux varied (data not shown). This motivated the use of the bilevel optimization scheme shown in Eq. (3). The time of regulation in this case was the optimal time for induction of glycerol kinase flux, which was determined by the optimization. The flux prior to the optimal induction time was maintained at zero, and was upregulated to 2 mmol g biomass^{-1} h^{-1} beyond the induction time. The time of the batch was taken to be 6 h. The average computation time for the simulation using different k_La values was 35 s (Table I).

Figure 4 shows the optimal time of induction and the final concentrations of the glycerol, biomass, and residual glucose at the end of the batch. It was observed that, for high k_La values (>40 h^{-1}), there was no oxygen limitation to the growth. In this case, a high growth rate can be maintained. The glycerol flux was induced at an optimal time (approximately 3.35 h) such that all the glucose was consumed by the end of the batch. For lower k_La values, growth was limited by oxygen availability during the batch. The growth rate was affected by oxygen limitation. Under these conditions, the glycerol flux was activated at the onset of oxygen-limiting conditions, which occurs when the dissolved oxygen concentration falls below 0.08 mmol L^{-1}. It was not optimal to divert the glucose toward biomass build-up because of the reduced growth rates due to limited oxygen availability. Smaller k_La values led to an early onset of oxygen-limiting conditions, which resulted in a smaller optimal induction time. Under these conditions, not all glucose was consumed by the end of the batch, as seen in Figure 4. As expected, the glycerol and biomass concentrations at the end of the batch increased with increasing k_La values.

Figure 5 shows the temporal profiles of the glycerol flux and the profiles of glycerol, biomass, and dissolved oxygen for two different k_La values (5 h^{-1} and 100 h^{-1}). For the higher k_La value, there was no oxygen limitation and the glycerol flux was activated optimally to utilize all glucose. For the k_La value of 5 h^{-1}, the oxygen concentration was <0.08 mmol L^{-1} at approximately 2.2 h, and the glycerol flux was activated at this time. The end glycerol concentration was much lower than in the previous case, although the flux was active for a longer period. This follows from the fact that the growth rates in the oxygen-limiting case were lower and there was insufficient build-up of biomass as seen in Figure 5.

Ethanol Production

Ethanol production in a batch culture under anaerobic growth conditions was considered. The acetate-producing flux was manipulated by modulating expression of the *ackA* gene in the *E. coli* strain with the *ldh* gene deletion. The growth rate and ethanol flux were obtained in the inner optimization problem and depended on the acetate flux determined in the outer optimization problem. The batch time was defined as 12 h with an initial glucose concentration of 10 mmol L^{-1}. The sampling time was taken to be 0.5 h. This optimization was performed using the bilevel scheme shown in Eq. (1) and required a computation time of 2.87 h (Table I). Figure 6A shows the optimal

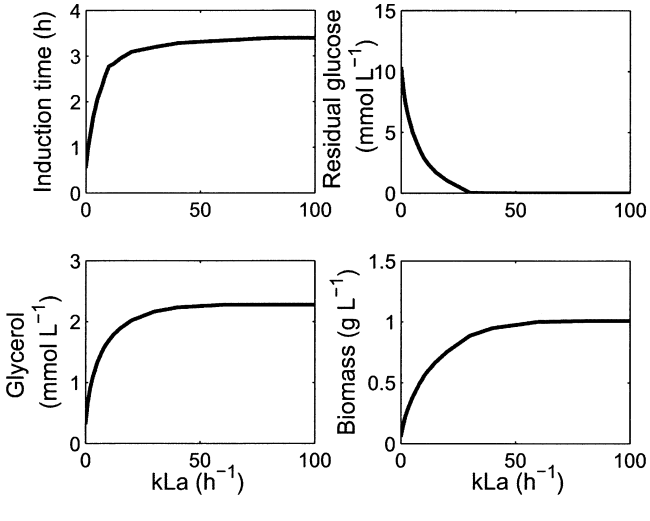

Figure 4. Optimal induction time of glycerol flux and concentrations of glycerol, biomass, and residual glucose at the end of the batch as a function of k_La.

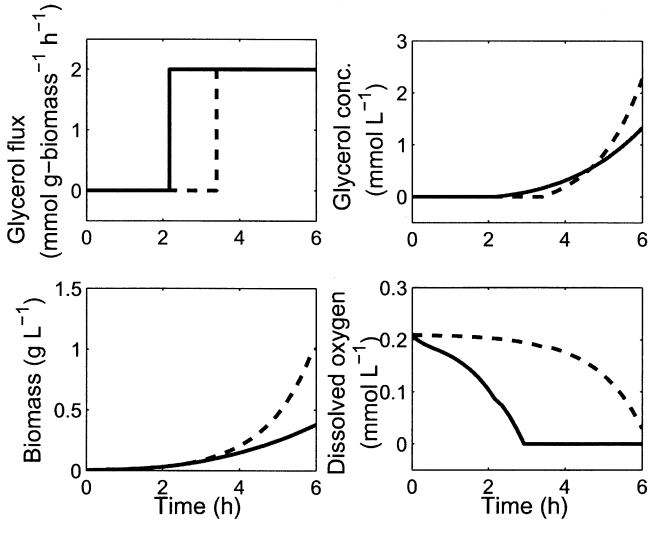

Figure 5. Profiles of optimal glycerol flux and concentrations of glycerol, biomass, and dissolved oxygen for two k_La values: k_La = 5 h^{-1} (solid line); and k_La = 100 h^{-1} (dashed line).

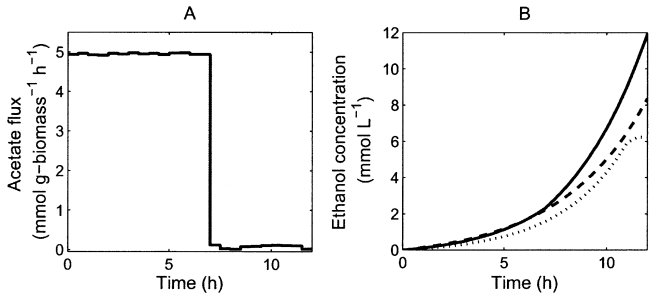

Figure 6. (A) Optimal acetate flux for maximum ethanol production at the end of the batch. (B) Ethanol concentration profiles: optimal strain (solid line); *ldh* + *ackA* knockout strain (dashed line); and *ldh* knockout strain (dotted line).

acetate flux for maximizing the ethanol concentration at the end of the batch. The flux was maintained at approximately 5 mmol g biomass^{-1} h^{-1} up to 7 h. During this time, the carbon flux was diverted toward the acetate pathway to maintain a growth rate of 0.3 h^{-1}, resulting in build-up of biomass. Beyond this time, the flux was reduced to zero so that an increased carbon flux was diverted toward the ethanol pathway and the growth rate was reduced to 0.23 h^{-1}. However, there was sufficient biomass that consumed all the glucose at the end of the batch and the flux to the ethanol was maximized during this period. The ethanol concentration at the end of the batch was 11.92 mmol L^{-1}. The ethanol profile is shown in Figure 6B. Figure 6B also shows the ethanol profiles for the strains in which the *ldh* gene was deleted and the one in which both the *ldh* and the *ackA* genes were deleted at the start of the batch. For the *ldh* knockout, the acetate flux was unconstrained. In this case, the cell maintained high growth rates of 0.31 h^{-1} by having an active flux through the acetate pathway of around 9 mmol g biomass^{-1} h^{-1}. The ethanol produced at the end of the batch was only 6.22 mmol L^{-1}. In the second case, where both the lactate and the acetate pathways were inactive, the flux to ethanol was maximal, but in this case the growth rate was reduced to 0.23 h^{-1}. As a result, although the flux was maximal, there was insufficient biomass to consume all the glucose in the 12-h batch time and only 8.36 mmol L^{-1} of ethanol was formed. It is important to note that if the batch was run for a longer time, then all the glucose would be consumed and this strain would yield a maximum ethanol concentration.

Figure 7 shows the concentration profiles for the glucose and biomass for the aforementioned cases. For the *ldh* knockout, there was an active flux toward acetate to support growth. In this case, the growth rate was maximal. All glucose was consumed before the batch time of 12 h. The biomass concentrations increased to >0.3 g L^{-1}. For *ldh* + *ackA* double knockout, there was no flux going to the acetate branch, and hence growth rate was affected. All glucose was not consumed in the 12-h batch time and only 0.17 g L^{-1} of biomass was formed. In this case, the flux toward ethanol was maximal, but there was not enough biomass for high productivity. For an optimal profile, a balance between ethanol flux and growth flux must be obtained.

Ethanol Production With Growth Inhibition by Product

The optimal acetate kinase flux for maximum ethanol production at the end of the batch in the presence of growth inhibition by ethanol was studied. Growth inhibition was modeled as a reduction in the maximum glucose uptake rate, as shown in Eq. (2). A batch time of 12 h and an initial glucose concentration of 10 mmol L^{-1} was considered. Figure 8A shows the optimal acetate flux that maximizes the end-ethanol concentration. The trends in this profile were similar to the previous case in which there was no growth inhibition. The acetate flux was high in the earlier period of the batch, during which biomass was built up after which the flux was reduced to zero to provide a higher ethanol production flux. However, in this case, the acetate flux was approximately 7 mmol g biomass^{-1} h^{-1} compared with 5 mmol g biomass^{-1} h^{-1} in the case of uninhibited growth. Also, this high flux was maintained for a longer period (9 h). Higher acetate flux was required to maintain growth rates and overcome the inhibition due to high ethanol concentrations. The ethanol concentration profile is shown in Figure 8B. The concentration increased to 9.4 mmol L^{-1} at the end of the batch. The figure also

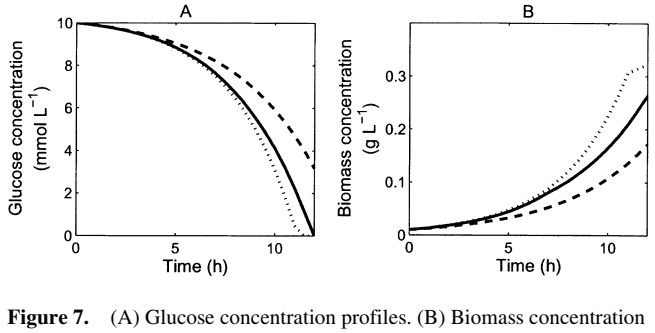

Figure 7. (A) Glucose concentration profiles. (B) Biomass concentration profiles: optimal strain (solid line); *ldh* + *ackA* knockout strain (dashed line); and *ldh* knockout strain (dotted line).

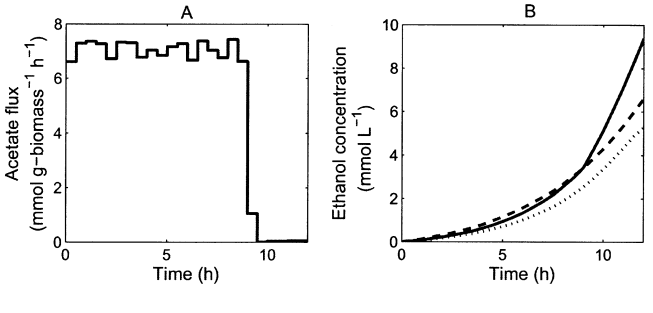

Figure 8. (A) Optimal acetate flux for maximum ethanol production in the presence of growth inhibition. (B) Ethanol concentration profiles: optimal strain (solid line); *ldh* + *ackA* knockout strain (dashed line); and *ldh* knockout strain (dotted line).

shows the ethanol concentration profiles for the *ldh* knockout strain and the *ldh* + *ackA* double-knockout strain. In all these cases, the ethanol concentration was considerably lower than the concentrations achieved in the absence of growth inhibition. This was expected because the requirement for a higher acetate flux converts a higher proportion of glucose to acetate. The final concentrations of acetate, ethanol, and biomass obtained for all the strains with and without growth inhibition are summarized in Table II.

Effect of Batch Time on Ethanol Production

The effect of batch time on optimal acetate flux profile and ethanol production was considered in the *E. coli* strain with the *ldh* knockout. The bilevel optimization scheme of Eq. (3) was used. In this case, the acetate flux was maintained at 5 mmol g biomass^{-1} h^{-1} until the optimal regulation time, after which it was reduced to zero. This was a case of repressing the active flux. The initial glucose concentration was 10 mmol L^{-1} and growth inhibition by build-up of ethanol was not considered. The average computation time for these simulations with different batch times was 300 s (Table I).

Figure 9 shows the optimal repression time for acetate flux and the final ethanol concentration for different batch times. For batch times of $\geq$13.75 h, it was optimal to use a strain with the *ldh* + *ackA* knockout. This case would lead to no acetate production. As a result, the ethanol flux was high, but the growth rates were low (0.23 h^{-1}). This was optimal because the batch time was sufficiently long to utilize all glucose, even when growing at a reduced growth rate. The ethanol was produced at a maximum flux of 12.25 mmol g biomass^{-1} h^{-1}, which results in a final ethanol concentration of 12.37 mmol L^{-1}. For batch times <13.75 h, the low growth rates of the *ldh* + *ackA* knockout strain made full consumption of glucose impossible. Thus, for lower batch times, it was optimal to have an acetate flux held at 5 mmol g biomass^{-1} h^{-1} during the initial period of the batch up to an optimal time after which it was repressed. The growth rate of the cell during this period was 0.2985 h^{-1}. The optimal time of repression increased with a decrease in final batch time. This was expected because shorter batch times require that higher growth rates be maintained for longer times to utilize all glucose by the end of the batch. Once the acetate flux was repressed, a greater flux was diverted toward ethanol production. Here, it was observed that, for maximum ethanol production, it was critical to utilize all the glucose. For a batch time of 11.65 h, the optimal repression time was as late as 9 h. If the batch time was reduced to <11.65 h, then, to utilize all the glucose, it would be necessary to have an active acetate flux for >9 h. This would leave insufficient time to produce ethanol after the repression. In these cases it was not optimal to use all the glucose. In fact, the repression time

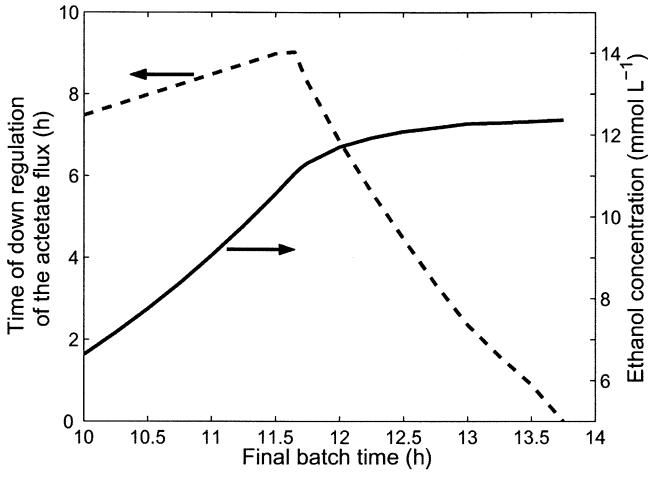

Figure 9. Optimal time for downregulation of acetate flux for different batch times and the corresponding ethanol concentrations at the end of the batch.

Table II. Species concentrations at the end of batch in anaerobic growth of *Escherichia coli* with and without growth inhibition by ethanol (initial glucose concentration 10 mmol L^{-1}).

	With growth inhibition			Without growth inhibition		
	Strain A	Strain B	Strain C	Strain A	Strain B	Strain C
Ethanol (mmol L^{-1})	6.22	11.92	8.36	5.28	9.38	6.60
Acetate (mmol L^{-1})	9.43	1.23	0	10.21	3.37	0
Residual glucose (mmol L^{-1})	0	0	3.17	0	0.87	4.63
Biomass (g L^{-1})	0.32	0.26	0.17	0.31	0.25	0.13

Strain A, *ldh* knockout; strain B, *ldh* knockout with optimal acetate flux; strain C; *ldh* + *ackA* double knockout.

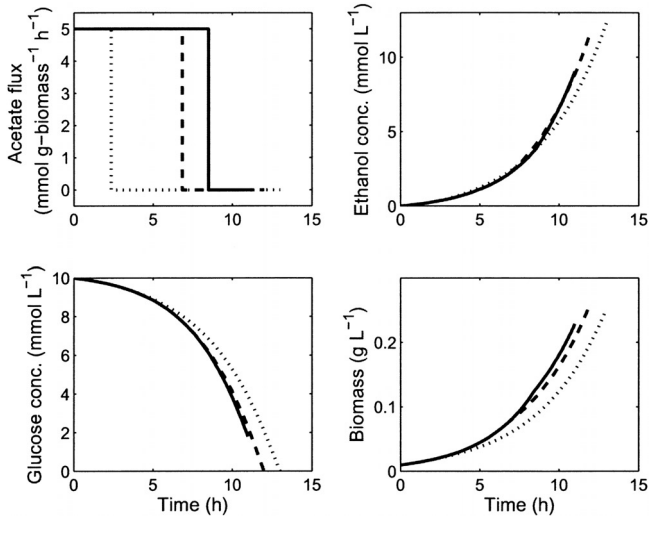

Figure 10. The optimal acetate flux profile and the profiles of ethanol, glucose, and biomass concentrations for three different batch times: 13 h (dotted line); 12 h (dashed line); and 11 h (solid line).

decreased as the batch time decreased below 11.65 h, and for a batch time of 10 h it was reduced to 7.48 h. The repression in this case was done at an optimal time such that a required amount of biomass was built up initially and a high ethanol flux was maintained later so as to maximize the ethanol concentration at the end of the batch. Figure 10 shows the profiles of the acetate flux and the concentrations of ethanol, glucose, and biomass for three different batch times of 11, 12, and 13 h. The ethanol flux and the growth rates corresponding to a particular value of the acetate flux are shown in Figure 2. As discussed earlier, for the batch times of 12 h and 13 h, the glucose was optimally consumed at the end of the batch. For the batch of 11 h, a residual glucose concentration of 1.79 mmol L^{-1} remained in the batch. The biomass growth rate was much less for the 13-h batch and the biomass concentration at the end of the batch was 0.25 g L^{-1}, which is less than that for the 12-h batch. The ethanol concentration increased with an increase in batch time, as seen in Figure 9. The increase was sharp until approximately hour 12, after which the increase was rather gradual. This suggests that it would be feasible to run a 12-h batch. Marginal improvements in ethanol concentration beyond hour 12 may not be justified and it may be advisable to reduce batch time due to productivity considerations.

DISCUSSION

In this study, a bilevel optimization scheme was developed to determine the optimal temporal flux profile in a specific reaction to maximize the desired product concentrations at the end of the batch culture. The particular reaction with flux manipulation was identified a priori, and is known to have an impact on the desired product formation. The optimal flux profile can be achieved by controlling the expression profile of the corresponding gene. Maximization of glycerol production by manipulating the flux through the reaction catalyzed by glycerol kinase enzyme and ethanol production by manipulating the acetate flux was considered. In either case, we found that to maximize the desired product concentrations it was critical that the induction or repression of the flux be introduced at an optimal time rather than beginning with a strain with gene alteration. Furthermore, it was also shown that the optimal profiles varied with operating conditions such as growth inhibition, oxygen availability, and batch time.

In this work, manipulation of only one flux in the bilevel optimization scheme was considered. The number of genetic alterations predicted for maximization of a particular product flux was usually two or three (Burgard et al., 2003). The bilevel optimization scheme can be expanded easily to determine the optimal flux profiles through multiple reactions. In the case of multiple fluxes, a combined effect of the fluxes would be crucial for maximization of the desired product. In these cases, it was even more important that the flux profiles were optimal.

The bilevel optimization scheme uses the flux balance analysis (FBA) approach to determine flux distributions in the network based on the cellular objective to maximize growth rate. The scheme is based on a steady-state assumption, however, and the model lacks detailed kinetic or regulatory information. Thus, there is a possibility that the flux distributions predicted by the FBA may not be accurate. The bilevel scheme is not restricted to using the FBA and it is flexible enough to incorporate any arbitrary metabolic network model. However, more complex models may lead to a greater computational burden when solving the bilevel optimization. Transcriptional regulation information has been incorporated into the FBA framework by Covert and Palsson (2002), and these regulatory flux balance analysis (rFBA) models can be used in the bilevel scheme. Regulatory information will not only improve the predictions of the flux distributions but the regulatory modifications could be optimized in a temporal fashion by including them as manipulated variables in the outer optimization of the bilevel scheme. Alternative modeling approaches, such as cybernetic models (Varner and Ramkrishna, 1999) and minimization of metabolic adjustment models (Segre et al., 2002), can also be incorporated in the bilevel framework. These models may provide better predictions of the flux distribution through the metabolic network under specific conditions. However, the primary aim of this work has been to develop the bilevel optimization framework and demonstrate the advantages of temporal genetic manipulations in metabolic engineering. Thus, for simplicity, the inner optimization has been restricted to the FBA framework.

The computation burden incurred in the bilevel optimization is moderate. The CPU time required for the bilevel scheme of Eq. (1) is considerably higher than that of Eq. (3). The reason for this is that, in the first case, the optimal flux is determined at each sampling time increasing the variables in the optimization, whereas, in the second case, only the time of regulation is optimized. As expected, the required CPU time for Eq. (1) depends on the batch time and the sampling time, whereas, in the case of Eq. (3), it is independent of these parameters. In both optimization schemes, the inner optimization is performed using FBA, which has minimal computation burden. The overall CPU time would increase in cases where alternative modeling approaches that have a higher computational requirement are used. For the case of a larger metabolic network, the computational cost would not be expected to rise exponentially, because the cost of solving a larger linear programming problem in the inner optimization is not significant for a larger metabolic network (a linear programming problem with ~ 250 variables takes on the order of 0.4 s compared with 0.15 s for a problem with 70 variables on a 2.4-GHz system).

The bilevel optimization scheme developed in this work shows utility in the field of metabolic engineering to design strains and provide optimal conditions for maximizing the desired product. The scheme could be expanded easily to incorporate a broad class of metabolic models as

well as various operating constraints. The tool thus shows tremendous promise as a framework for providing a dynamic strategy in metabolic engineering.

NOMENCLATURE

CO_2^*	equilibrium interfacial concentration of carbon dioxide (0.5 mmol L^{-1})
k_La	mass transfer coefficient (h^{-1})
K_i	inhibition constant (25 mmol L^{-1})
M	total number of fluxes
N	total number of metabolites
O_2^*	equilibrium interfacial concentration of oxygen (0.21 mmol L^{-1})
P	product concentration (mmol L^{-1})
Q	set of all unregulated fluxes
S	stoichiometric matrix
t_0	initial time (h)
t_f	final time (h)
t_{reg}	optimal time of induction/repression of manipulated flux (h)
T_s	sampling time (h)
U	set of regulated fluxes
v	flux (mmol g biomass^{-1} h^{-1})
v_c	manipulated flux (mmol g biomass^{-1} h^{-1})
v_c^-	manipulated flux value for $t < t_{reg}$ (mmol g biomass^{-1} h^{-1})
v_c^+	manipulated flux value for $t \geq t_{reg}$ (mmol g biomass^{-1} h^{-1})
v^{CO_2}	carbon dioxide flux (mmol g biomass^{-1} h^{-1})
$v_{glucose}^{uptake}$	maximum glucose uptake flux in the presence of growth inhibition (mmol g biomass^{-1} h^{-1})
$v_{glucose}^{uptake*}$	maximum glucose uptake flux in the absence of growth inhibition (mmol g biomass^{-1} h^{-1})
v^{growth}	growth rate (h^{-1})
v^{O_2}	oxygen flux (mmol g biomass^{-1} h^{-1})
$v^{product}$	product flux (mmol g biomass^{-1} h^{-1})
X	biomass concentration (g L^{-1})
α_j	lower bound of flux j (mmol g biomass^{-1} h^{-1})
β_j	upper bound of flux j (mmol g biomass^{-1} h^{-1})

References

Atkinson DE. 1977. Cellular energy metabolism and its regulation. New York: Academic Press.

Bailey JE. 1991. Towards a science of metabolic engineering. Science 252:1668–1675.

Burgard AP, Maranas CD. 2001. Probing the performance limits of the *Escherichia coli* metabolic network subject to gene additions or deletions. Biotechnol Bioeng 74:364–375.

Burgard AP, Pharkya P, Maranas CD. 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng 84:647–657.

Covert MW, Palsson BO. 2002. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. J Biol Chem 277: 28058–28064.

Desai RP, Papoutsakis ET. 1999. Antisense RNA strategies for metabolic engineering of *Clostridium acetobutylicum*. Appl Environ Microbiol 65:936–945.

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

Edwards JS, Ibarra RU, Palsson BO. 2001. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. Nat Biotechnol 19:125–130.

Farmer WR, Liao JC. 2000. Improving lycopene production in *Escherichia coli* by engineering metabolic control. Nat Biotechnol 18:533–537.

Herring CD, Glasner JD, Blattner FR. 2003. Gene replacement without selection: regulated suppression of amber mutations in *Escherichia coli*. Gene 311:153–163.

Ibarra RU, Edwards JS, Palsson BO. 2002. *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. Nature 420:186–189.

Ingram LO, Conway T, Clark DP, Sewell GW, Preston JF. 1987. Genetic engineering of ethanol production in *Escherichia coli*. Appl Environ Microbiol 53:2420–2425.

Khlebnikov A, Skaug T, Keasling JD. 2002. Modulation of gene expression from the arabinose-inducible *araBAD* promoter. J Indust Microbiol Biotechnol 29:34–37.

Kim JY, Cha HJ. 2003. Down-regulation of acetate pathway through antisense strategy in *Escherichia coli*: improved foreign protein production. Biotechnol Bioeng 83:841–853.

Kurland CG, Dong H. 1996. Bacterial growth inhibition by overproduction of protein. Mol Microbiol 21:1–4.

Leonardo MR, Dailly Y, Clark DP. 1996. Role of NAD in regulating the *adhE* gene of *Escherichia coli*. J Bacteriol 178:6013–6018.

Nakamura CE, Whited GM. 2003. Metabolic engineering for the microbial production of 1,3-propanediol. Curr Opin Biotechnol 14:454–459.

Nielsen J. 1998. Metabolic engineering: techniques for analysis of targets for genetic manipulations. Biotechnol Bioeng 58:125–132.

Schilling CH, Edwards JS, Letscher D, Palsson BO. 2000. Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. Biotechnol Bioeng 71:286–306.

Segre D, Vitkup D, Church GM. 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci USA 99: 15112–15117.

Stephanopoulos G, Vallino JJ. 1991. Network rigidity and metabolic engineering in metabolite overproduction. Science 252:1675–1681.

Underwood SA, Zhou S, Causey TB, Yomano LP, Shanmugam KT, Ingram LO. 2002. Genetic changes to optimize carbon partitioning between ethanol and biosynthesis in ethanologenic *Escherichia coli*. Appl Environ Microbiol 68:6263–6272.

Varma A, Palsson BO. 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. Appl Environ Microbiol 60: 3724–3731.

Varner J, Ramkrishna D. 1999. Metabolic engineering from a cybernetic perspective. 1. Theoretical preliminaries. Biotechnol Progr 15: 407–425.

Yang YT, San KY, Bennett GN. 1999. Redistribution of metabolic fluxes in *Escherichia coli* with fermentative lactate dehydrogenase overexpression and deletion. Metab Eng 1:141–152.