

# Exploring the Overproduction of Amino Acids Using the Bilevel Optimization Framework OptKnock

Priti Pharkya, Anthony P. Burgard, Costas D. Maranas

*Department of Chemical Engineering, The Pennsylvania State University, University Park, Pennsylvania 16802; telephone: 814-863-9958; fax: 814-865-7846; e-mail: costas@psu.edu*

*Published online 24 November 2003 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/bit.10857*

**Abstract:** In this study, we modify and extend the bilevel optimization framework OptKnock for identifying gene knockout strategies in the *Escherichia coli* metabolic network, leading to the overproduction of representative amino acids and key precursors for all five families. These strategies span not only the central metabolic network genes but also the amino acid biosynthetic and degradation pathways. In addition to gene deletions, the transport rates of carbon dioxide, ammonia, and oxygen, as well as the secretion pathways for key metabolites, are introduced as optimization variables in the framework. Computational results demonstrate the importance of manipulating energy-producing/consuming pathways, controlling the uptake of nitrogen and oxygen, and blocking the secretion pathways of key competing metabolites. The identified pathway modifications include not only straightforward elimination of competing reactions but also a number of nonintuitive knockouts quite distant from the amino acid-producing pathways. Specifically, OptKnock suggests three reactions (i.e., pyruvate kinase, phosphotransacetylase, and ATPase) for deletion, in addition to the straightforward elimination of 2-ketoglutarate dehydrogenase, to generate a glutamate-overproducing mutant. Similarly, phosphofructokinase and ATPase are identified as promising knockout targets to complement the removal of pyruvate formate lyase and pyruvate dehydrogenase for enhancing the yield of alanine. Although OptKnock in its present form does not consider regulatory constraints, it does provide useful suggestions largely in agreement with existing practices and, more importantly, introduces a framework for incorporating additional modeling refinements as they become available. © 2003 Wiley Periodicals, Inc.

**Keywords:** amino acid overproduction; bilevel optimization; transport rates; energy consuming/producing pathways

## INTRODUCTION

The significance of amino acid production in the biotechnology industry is underlined by the fact that the annual

production of amino acids worldwide is approximately two million metric tons, the majority of which is accomplished by microbial methods (Eggeling and Sahm, 1999). Initial attempts to produce strains with improved yields of amino acids relied on the relatively straightforward approach of eliminating the competing pathways and on removing feedback inhibitions in the biosynthetic pathways (Stephanopoulos et al., 1998). These approaches did not always result in the anticipated yields because of the difficulty in achieving complete deregulation of the regulatory steps by mutagenic methods. Recombinant DNA technology methods initially focused on the terminal steps of the amino acid biosynthetic pathways and emphasized primarily on (1) overexpression of the rate-limiting enzyme; (2) introduction of an enzyme that could bypass the regulatory steps in the biosynthetic pathway; and (3) enhancement of the activity of the first enzyme in the terminal pathway (Ikeda, 2003). The yield improvements accomplished by such techniques were, however, in many cases quite limited because they relied on the manipulation of only terminal pathways. Efforts have been made in the past decade to address this problem of redirecting additional carbon flux from the central metabolism into the amino acid-producing pathways (Flores et al., 1996; Patnaik and Liao, 1994) by adopting a more global view of metabolism. However, so far, it has been difficult to predict the synergistic effect of multiple changes introduced in the central metabolism, or even to reproduce the same behavior under different experimental conditions (Chandran et al., 2003; Chen et al., 1997; Gubler et al., 1994; Miller et al., 1987; Patnaik et al., 1995). This motivates the need for a computational approach that can systematically consider the effect of genetic modifications on the entire metabolic network of a production system.

The recent availability of genome-scale models of microbial organisms (Edwards et al., 2001; Schilling et al., 2002) has provided the pathway reconstructions necessary for developing novel computational methods aimed at identifying strain-engineering strategies. These methods enable the investigation of the effect of a genetic manipulation, such as a gene addition or deletion, in an organism on a systemic

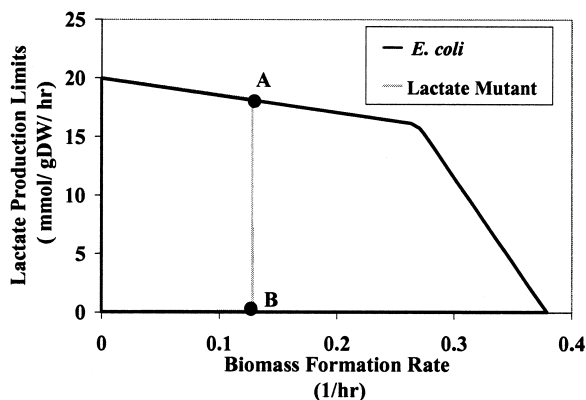
*Correspondence to:* C. Maranas

Contract grant sponsors: National Science Foundation; Department of Energy

Contract grant number: BES 0120277

level and thus help to accelerate the construction of mutant strains with the desired characteristics. In an earlier study, we introduced the bilevel optimization framework, OptKnock, aimed at proposing reaction eliminations from the *Escherichia coli* central metabolic network for maximizing the production of simple compounds such as succinate, lactate, and 1,3-propanediol (Burgard et al., 2003). This was achieved by relying on the maximization of biomass formation assumption for identifying plausible flux distribution in the network, while the maximization of the desired product flux was achieved by eliminating key reaction steps. This multiobjective model aimed at ensuring that, due to the stoichiometric constraints and pathway connectivity, the desired biochemical had to be produced as an obligatory byproduct of growth.

In this work, we modify and customize this bilevel framework for the enhancement of amino acid overproduction. We focus on the five families into which the amino acids are categorized: the pyruvate family; the 3-phosphoglycerate family; the  $\alpha$ -ketoglutarate family; the aspartate family; and the family of aromatic amino acids. For cases in which a single precursor exists for all the amino acids in the family, we concentrate on improving its yield over that in the wild-type *E. coli* network. For example, chorismate is the common intermediate to the three amino acids in the family of aromatic amino acids. Alternatively, when increasing the formation of the immediate reactant does not map one-to-one with the enhancement of the desired amino acids, we focus on the individual amino acids. For example, increasing the flux toward pyruvate does not necessarily lead to more alanine production, due to the large number of reactions in which the former is involved. Subsequently, we use OptKnock to detect the reactions whose removal leads to a stronger coupling between biomass formation and alanine overproduction. The *E. coli* K-12 stoichiometric model (Edwards and Palsson, 2000), comprised of 720 reactions, is used to describe the metabolic network.

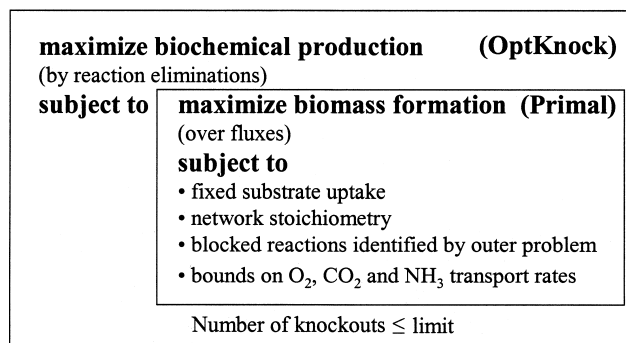


**Figure 1.** The limits of lactate production of the network obtained by maximizing and then minimizing lactate production at different biomass levels available to the network. Point A represents the maximal yield suggested by OptKnock at maximum biomass production. Point B is associated with zero lactate production. A basis glucose uptake of 10 mmol/gDW-hr is assumed.

## MODIFICATIONS TO THE OPTKNOCK FRAMEWORK

Here, we explore reaction deletions for the enhanced production of the desired amino acids not only in the central metabolic network of *E. coli*, but also in the amino acid biosynthetic and degradation pathways. The resulting size of the explored network, in conjunction with its redundancy, provides a large number of alternative routes for channeling its carbon flux. This may not ensure the secretion of the desired biochemical, even after the eliminations identified by OptKnock are imposed. In fact, we first observed this type of behavior when we investigated the mutants for lactate formation (Burgard et al., 2003). Specifically, after eliminating two reactions (i.e., acetate kinase and phosphofructokinase) identified by OptKnock, we maximized and then minimized the lactate secretion as a function of different biomass production levels. As shown in Figure 1, this mutant could still avoid lactate formation (see point B) at maximum biomass formation and secrete other compounds because of a plethora of other options to redirect its carbon flux. Point A, identified by OptKnock, corresponds to maximum overproduction of lactate; however, point B, involving no lactate production, is also an equivalent solution. This is because the rightmost boundary of the range of feasible phenotypes is a vertical line and not a single point. In optimization terminology, there is not a single optimum, but a family of alternative solutions defined by the convex combinations of points A and B (i.e., the straight line joining points A and B).

Therefore, we decided to introduce the cellular transport rates of ammonia, carbon dioxide, and oxygen as well as key secretion pathways as optimization variables in the OptKnock framework to reduce the degrees of freedom available to the network and to force it to assume flux distributions as close as possible to the desired point A. Specifically, upon fixing the values of these variables equal to the ones for point A, we observe a strong coupling between the biochemical overproduction and the cellular objective of maximizing biomass formation. Note that ammonia, carbon dioxide, and oxygen uptake rates are inputs that can be effectively



**Figure 2.** The bilevel formulation of OptKnock. The inner problem allocates the fluxes for maximum biomass production. The outer problem maximizes the biochemical production by blocking fluxes through the reactions necessary for attaining the highest yields.

controlled in production systems. The rate of carbon dioxide evolution can be varied by controlling pressure in industrial reactors. Nitrogen is often provided through ammonia whose concentration that can be effectively maintained by monitoring the pH of the medium. The aeration rate and the oxygen transport coefficient ( $k_L a$ ) are used to maintain the requisite oxygen supply to the medium, while continuous stirring helps to ensure uniform oxygen concentration in the bioreactor (Ikeda, 2003; Kumagai, 2000). Equivalently, the secretion of key compounds can be blocked by deleting the corresponding genes coding for their transferases.

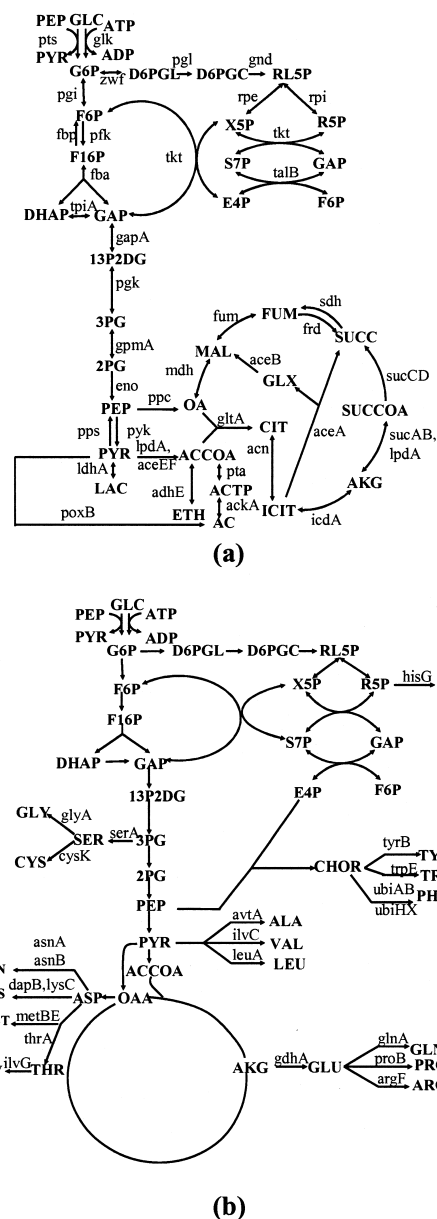
A graphic depiction of the OptKnock bilevel optimization model is presented in Figure 2. The following stepwise procedure summarizes the main computational steps undertaken for all amino acid studies. The technical details for each of these steps are described in the Appendix.

### Modified OptKnock Procedure

- Step 1.* Solve the OptKnock optimization problem (see Appendix) to identify reaction deletions (i.e.,  $y_j = 0$ ) for enhanced production of an amino acid (or a key precursor).
- Step 2.* Solve the max/min problem (see Appendix) for finding the production limits of the biochemical of interest after fixing the fluxes through the deleted reactions to zero.
- Step 3.* If alternative optimal solutions exist for the inner biomass maximization problem, fix transport rates of carbon dioxide, ammonia, and/or oxygen to their values identified from the OptKnock solution and return to step 2.
- Step 4.* If alternative solutions still exist, prevent the network from secreting metabolites associated with alternative solutions. Return to step 2.

## RESULTS

In this section, we describe the specific strategies identified by OptKnock for enhancing the production of amino acids by employing the *E. coli* K-12 stoichiometric model (Edwards and Palsson, 2000). These approaches are based on deletions of not only the reactions present in the central metabolic network but also in the amino acid biosynthetic and degradation pathways, thus increasing the complexity of the model. Figure 3 depicts the central metabolic and the amino acid biosynthetic pathways considered here to achieve amino acid overproduction. The fluxes through the metabolic network are scaled by allowing the uptake of a fixed amount of glucose (10 millimoles per gram dry weight of cell per unit hour [or 10 mmol/g DW·hr]) through the phosphotransferase system, the glucokinase pathway, or both. The OptKnock framework allows for aerobic or anaerobic metabolism by enabling or eliminating the oxygen transport reaction. The network is allowed to secrete all metabolites for which it possesses the mechanism of export. In all five cases

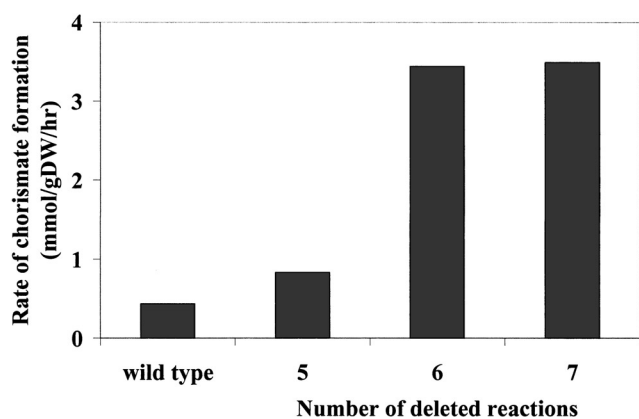


**Figure 3.** (a) The central metabolism in *E. coli*. (b) The amino acid biosynthetic pathways in *E. coli*.

tested, an aerobic mode of growth is found necessary for overproduction of amino acids.

### Chorismate Formation (Precursor to Aromatic Amino Acids)

In the production of aromatic amino acids (tyrosine, tryptophan, and phenylalanine), chorismate is the key intermediate. The aromatic amino acid biosynthesis pathway divides at the chorismate branching point of the metabolic network (see Fig. 3b) to ensure the formation of all three aromatic amino acids. Consequently, we use chorismate formation as the objective function in OptKnock for overproduction of aromatic amino acids. Figure 4 compares the chorismate formation rates obtained for different total numbers of reactions deleted from the network. These rates



**Figure 4.** Comparison of chorismate formation rates for different numbers of reaction eliminations in the network.

are also compared with the chorismate production limits of the wild-type *E. coli* network. As shown, while the flux toward chorismate is very small when only five reactions are eliminated, the six-reaction strategy offers a substantial improvement with negligible changes found for seven or more deletions. Table I describes the six-deletion mutant, and Table II summarizes the current strategies for aromatic amino acid overproduction.

The precursors for chorismate are erythrose-4-phosphate (E4P) and phosphoenolpyruvate (PEP). Not surprisingly, the reactions catalyzed by transaldolase B (*talB*) and phosphoenolpyruvate carboxylase (*ppc*) (Fig. 3a), which lead to the consumption of these compounds, are selected by OptKnock to be eliminated. Interestingly, the removal of the *ppc* gene, which converts PEP to oxaloacetate (OAA), was proposed originally by (Miller et al., 1987) and later patented (Backman, 1992) for increasing carbon flux toward aromatic amino acids. Subsequently, it was suggested (Patnaik et al., 1995) that the exclusion of this gene may not always have the desired effect of increasing the flux toward chorismate under all experimental set-ups. Therefore, we decided to investigate the effect of the *ppc* deletion by OptKnock with this perspective in mind. The model predicts that a *ppc*-deficient *E. coli* network by itself does not lead to increased flux towards aromatic amino acids at maximum biomass yield because PEP is preferably converted to pyruvate instead of OAA, which is utilized for biomass formation in the TCA cycle. Specifically, it appears that the deletion of *ppc* reaction in conjunction with five others is needed to redirect flux

toward chorismate. The calculated flux distributions for this mutant are shown in Figure 5a. Three of these reactions, pyruvate oxidase (*pox*), pyruvate dehydrogenase (*lpdA*), and pyruvate formate lyase reactions (*pfl*) (see Fig. 3a) involve pyruvate as a reactant. PEP is converted to pyruvate during glycolysis and this carbon flux may be lost for chorismate production because recycling of pyruvate to PEP is energetically expensive (Patnaik and Liao, 1994). These eliminations are thus aimed at preventing the conversion of PEP to pyruvate and help redirect more carbon flux toward chorismate production. This strategy, although quite non-intuitive, is in agreement with the observation that increasing the fluxes from PEP towards a specific pathway by the elimination of competing enzymatic activities is not very straightforward (Stephanopoulos and Vallino, 1991). The removal of these three pyruvate-consuming reactions is similar to increasing the expression of pyruvate synthase (*pps*), which recycles pyruvate to PEP (Patnaik and Liao, 1994). The deletion of 6-phosphogluconolactonase (*pgl*) suggested by OptKnock is also essential for attaining the predicted yield because the removal of *pgl* and *talB* has the synergistic effect of making more E4P available. This is comparable to the overexpression of transketolase (*tktA*) (Draths et al., 1992), transaldolase (*talB*), or both (Lu and Liao, 1997; Sprenger et al., 1998). It should be noted that in the strategy identified by OptKnock, no flux is observed through the oxidative part of the pentose phosphate pathway. Accordingly, the availability of E4P can be increased by eliminating *talB*. Interestingly, the identified knockout strategy utilizes a complex pathway of multiple reactions for linking flux from ribose-5-phosphate to glyceraldehyde-3-phosphate and acetate, thus fueling the TCA cycle.

Figure 5b explores the performance limits of the wild-type *E. coli* network with respect to chorismate formation and compares them with those of the suggested mutant network. These trade-off curves have been obtained by solving the (max/min) formulation for chorismate production subject to fixed rates of growth to identify its allowable production range. It is found that approximately 65% of the maximum theoretical yield of chorismate can be reached at maximum biomass formation. In turn, most of this chorismate (about 88%) can be converted to the required aromatic amino acid by blocking secretion of the other two aromatic amino acids. The remainder is converted to di- and trihydrofolates and to the two other amino acids in

**Table I.** Deleted reactions in the mutant for the maximum formation of chorismate.<sup>a</sup>

No. of knockouts	ID	Reaction	Enzyme	Growth rate (1/hr)	Chorismate (mmol/g DW.hr)	
6	A	1	PYR + COA + NAD → NADH + CO <sub>2</sub> + ACCOA	Pyruvate dehydrogenase	0.32	3.44
		2	D6PGL → D6PGC	6-Phosphogluconolactonase		
		3	T3P1 + S7P ↔ E4P + F6P	Transaldolase B		
		4	PYR + COA → ACCOA + FOR	Pyruvate formate lyase 1		
		5	PEP + CO <sub>2</sub> → OA + PI	Phosphoenolpyruvate carboxylase		
		6	PYR + Q → AC + CO <sub>2</sub> + QH <sub>2</sub>	Pyruvate oxidase		

<sup>a</sup>The maximum theoretical yield of chorismate is 5.33 mmol/g DW.hr

**Table II.** Different approaches for maximizing flux toward aromatic acid production from the central metabolic pathway.

No.	Approach	Effect	Reference
1	Overexpression of transketolase gene <i>tktA</i>	Increases the availability of erythrose-4-phosphate (E4P)	Draths et al. (1992)
2	Inactivation of phosphoenolpyruvate carboxylase ( <i>ppc</i> )	Prevents the consumption of phosphoenolpyruvate (PEP)	Miller et al. (1987), Backman (1992)
3	Amplification of phosphoenolpyruvate synthase	Recycles pyruvate to PEP	Patnaik and Liao (1994)
4	Using non-PTS sugar	Prevents the conversion of PEP to pyruvate	Flores et al. (1996)
5	Overexpression of transaldolase	Increases the availability of E4P	Sprenger et al. (1998), Lu and Liao (1997)

proportion to their utilization in the formation of biomass. In the following study, we observe that the energy balance of the network and the fixing of the transport rates of carbon

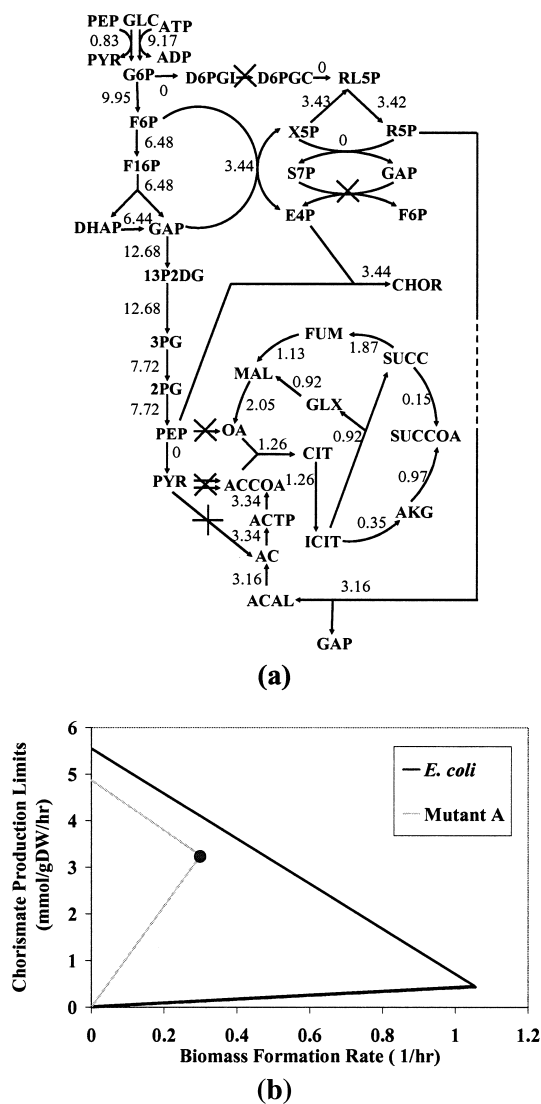
dioxide and ammonia have a key role in the production of the pyruvate family of amino acids.

### Alanine Overproduction (Pyruvate Family)

The pyruvate family of amino acids is comprised of three amino acids: alanine, valine, and leucine. Alanine is formed by the transamination of pyruvate. The identified reaction eliminations for alanine secretion are listed in Table III, and the flux distributions are illustrated in Figure 6a and b.

Mutant A involves the removal of three reactions that decouple alanine production and biomass formation. These deletions have a cumulative effect of reducing the carbon flux from pyruvate to other compounds and of transaminating more pyruvate toward alanine. The removal of pyruvate dehydrogenase (*lpdA*, *aceEF*) inhibits the conversion of pyruvate to acetyl-CoA (ACCOA) and that of pyruvate formate lyase (*pfl*) prevents pyruvate from converting to formate and ACCOA. The elimination of the ATPase reaction (which interconverts ATP into ADP with exchange of a proton) is essential for attaining the predicted theoretical yield, because it augments the amount of glucose through the glycolysis pathway (Jensen et al., 1993), increasing the availability of the precursors for alanine overproduction. In mutant B, the phosphofructokinase reaction (*pfk*), which converts fructose 6-phosphate (F6P) to fructose 1,6-diphosphate (FDP) with the utilization of one ATP molecule, is also deleted, relying on the Entner–Doudoroff pathway for the formation of pyruvate. The removal of this extra reaction alters the cofactor balance in such a way that the flux through the pyruvate oxidase reaction (*pox*) is reduced. This implies that less pyruvate is converted to acetate and subsequently to ACCOA for production of biomass through the TCA cycle. Interestingly, the predicted alanine yield of mutant B is approximately 24% higher than that of mutant A and is equal to 94% of the maximum theoretical yield of alanine. This amounts to 91.5% of the yield based on weight of the raw material (glucose), which is considerably higher than the range of 45% to 55% yield presently being obtained in industry (Ikeda, 2003).

We then investigated the degree of coupling between biomass production and alanine secretion by solving the max/min formulation for varying levels of biomass production. Figure 6c shows that even at maximum biomass, a wide range of phenotypic behaviors (the collection of points between the desired point, D, and the undesired point of zero



**Figure 5.** (a) Flux distribution for mutant A (six reaction deletions) with enhanced chorismate formation rate. (b) Chorismate formation limits of the mutant A (gray line). The filled circle is the desired point identified by OptKnock where maximum chorismate can be formed at the maximum biomass yields. The black lines depict the production limits of the wild-type *E. coli* network.

**Table III.** The deletion mutants for overproducing alanine.<sup>a</sup>

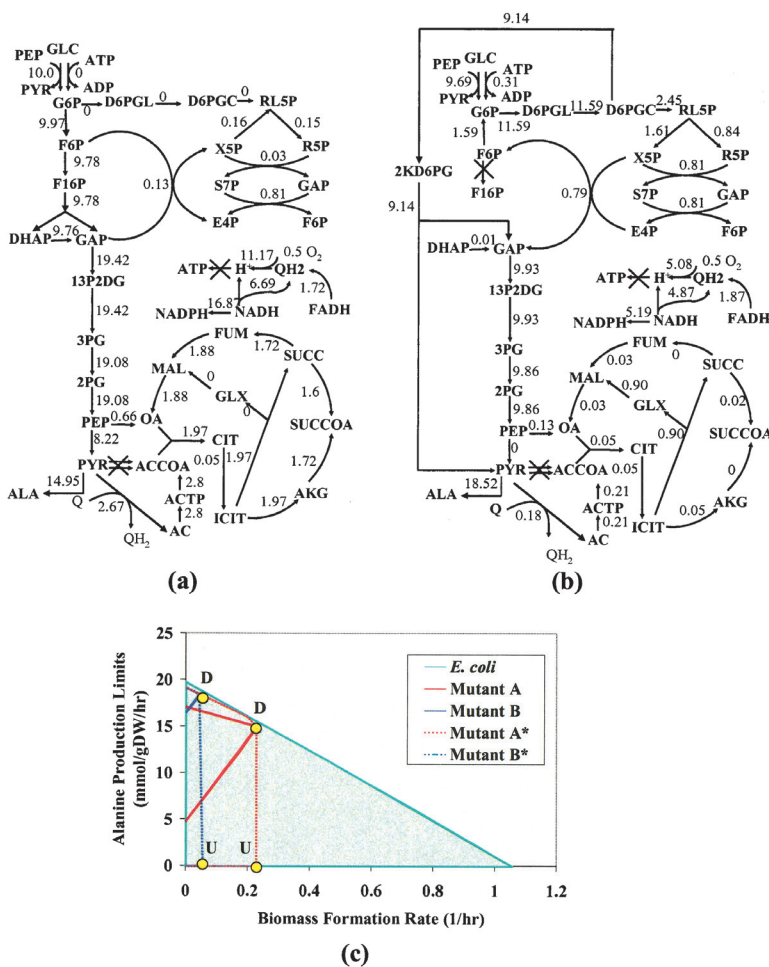
No. of knockouts	ID	Reaction	Enzyme	Growth rate (1/hr)	Uptake		Secretion		
					O <sub>2</sub> <sup>b</sup>	NH <sub>3</sub> <sup>b</sup>	CO <sub>2</sub> <sup>b</sup>	Alanine <sup>b</sup>	
3	A	1	PYR + COA + NAD → NADH + CO <sub>2</sub> + ACCOA	Pyruvate dehydrogenase	0.23	5.58	17.18	6.13	14.95
		2	PYR + COA → ACCOA + FOR	Pyruvate formate lyase					
		3	ATP ↔ ADP + PI + 4 HEXT	ATPase					
4	B	1	F6P + ATP → FDP + ADP	Phosphofructokinase	0.04	2.54	18.97	2.65	18.53
		2	PYR + COA + NAD → NADH + CO <sub>2</sub> + ACCOA	Pyruvate dehydrogenase					
		3	PYR + COA → ACCOA + FOR	Pyruvate formate lyase					
		4	ATP ↔ ADP + PI + 4 HEXT	ATPase					

<sup>a</sup>The maximum theoretical yield of alanine is 19.77 mmol/g DW.hr. In all mutants, the secretion of D-alanine is blocked.

<sup>b</sup>Data expressed as millimoles per gram dry weight per hour.

alanine production, U) is possible (dotted vertical lines corresponding to mutants A\* and B\*) for both mutant networks. This result is similar to the lactate example (Burgard et al., 2003) discussed previously. The availability of a number of

alternative conversion routes gives the network the flexibility of not secreting alanine at maximum biomass yield even after the identified reactions are eliminated. In contrast to the chorismate study, where the fluxes towards ammonia,



**Figure 6.** (a) Flux distribution for mutant A (with three reaction eliminations) and (b) mutant B (with four reaction eliminations) for overproducing alanine. (c) Alanine production limits of the mutant networks obtained by maximizing and minimizing the alanine secretion for different biomass rates. The dotted lines outline the behavior of the network when the carbon dioxide and ammonia transport rates are not fixed. Points D correspond to the desired points of alanine secretion identified by OptKnock. Points U are points at which no alanine is produced by the network. The solid lines (red and dark blue) depict the behavior of the networks for mutants A and B, respectively, when the transport rates have been fixed. The light blue lines denote the alanine production limits of wild-type *E. coli* network.



Figure 7a depicts the flux distribution of the mutant network and Table IV provides the list of deleted reactions. These include the serine degradation reaction, serine deaminase. Commercial strains, not surprisingly, are also deficient in serine degradation (Ikeda, 2003). Also, OptKnock suggests that the reaction catalyzed by phosphoglycerate mutase (*gpm*) or enolase (*eno*) (see Fig. 3a) should be removed. The elimination of the former is quite obvious because 3PG is the key precursor to the formation of serine. The deletion of enolase (which converts 2PG to PEP) has an equivalent effect because it blocks the path by which 3PG is ultimately converted to PEP during glycolysis. The third deleted reaction suggested by OptKnock is the respiratory ATPase reaction, whose removal alters the metabolism, inducing a higher glucose flux through Entner–Doudoroff glycolysis, increasing the formation of 3PG.

As was the case with alanine overproduction, a tight coupling between serine secretion and biomass formation is obtained only after the ammonia and carbon dioxide transport rates are fixed to those determined from the bilevel optimization problem at maximum biomass production. The values of these rates for the serine-secreting mutant are listed in Table IV. The serine production limits of the mutant network, plotted as a function of biomass formation, are shown in Figure 7b. Although the wild-type *E. coli* network may not necessarily secrete serine, the resulting mutant network has to overproduce serine after the elimination of the suggested reactions and the fixation of the transport rates of ammonia and carbon dioxide. The next section elucidates the strategies for aspartate overproduction and demonstrates that not only carbon dioxide and ammonia transport rates, but also oxygen transport rate, plays a vital role in influencing the aspartate-producing potential of the network.

### Aspartate Overproduction (Aspartate Family)

Aspartate, asparagine, lysine (formed via the diaminopimelic acid pathway), methionine, threonine, and isoleucine are members of the aspartate family of amino acids. Aspartate is formed from the citric acid cycle intermediate oxaloacetate, and is subsequently converted to the other five amino acids. However, oxaloacetate is involved in a number of different metabolic reactions and there is no guarantee that increasing the production of this metabolite will eventually lead to the overproduction of any of the six amino acids derived from it. Therefore, in this study we modify the objective function of OptKnock to be the direct maximization of aspartate, which is the key precursor to all other amino acids in the family.

OptKnock identifies a mutant with four deleted reactions that is predicted to reach as high as 75.3% of the maximum theoretical yield of aspartate at maximum growth. Table V lists two different strategies for overproducing aspartate. The flux distribution for mutant A, where three reactions are deleted, is outlined in Figure 8a. The removal of 2-ketoglutarate dehydrogenase (*sucAB*, *lpdA*) redirects more flux toward the formation of aspartate from oxaloacetate, which would otherwise have been utilized in the

TCA cycle. The deletion of acetate kinase (*ack*) or alternatively phosphotransacetylase (*pta*), also allows for an increase in the formation of aspartate. These reactions are reversible and the network tends to convert acetate to ACCOA, so that the flux through the TCA cycle is maximized, leading to maximum biomass formation. However, mutant A shows a reduced flux through the TCA cycle that prevents consumption of oxaloacetate and forces the flux through these reversible reactions to zero. The elimination of ATPase (*atp*) also favors the formation of aspartate. Mutant B involves the removal of pyruvate kinase (*pyk*) in addition to those already eliminated in mutant A, and its flux distribution is shown in Figure 8b. Notably, the exclusion of this additional reaction increases the yield of aspartate in mutant B by approximately 89% over that in mutant A. This glycolytic reaction converts PEP to pyruvate, which can be converted into a number of metabolites. Consequently, by its removal, the network can channel the available PEP to oxaloacetate, the immediate precursor to aspartate, through the phosphoenolpyruvate carboxylase (*ppc*) reaction.

The coupling between aspartate secretion and biomass formation is a strong function of the transport rates of oxygen, carbon dioxide and ammonia. Figure 8c shows the production limits of the networks for mutants A and B after these transport rates have been fixed and compares them with those of the wild-type *E. coli* network. Although the latter network may not secrete aspartate, both mutant networks are required to overproduce aspartate due to the re-engineered network stoichiometry. The aspartate study manifests the significance of oxygen transport rate for ensuring the overproduction of aspartate from the network. Next, the glutamate study shows that not only reaction eliminations and fixing of transport rates, but also blocking the secretion of key metabolites is sometimes needed for achieving the desired overproduction.

### Glutamate Overproduction ( $\alpha$ -Ketoglutarate Family)

The common precursor to the amino acids in this family is  $\alpha$ -ketoglutarate, a compound participating in the TCA cycle. As in the case of oxaloacetate, this metabolite branches into a number of pathways. The reaction scheme for the formation of these amino acids is also similar to the one for the amino acids in the aspartate family as  $\alpha$ -ketoglutarate transaminates to form glutamate, which is subsequently converted to glutamine, proline, and arginine: the other three amino acids in the family. Therefore, we adopt a strategy similar to the one described in the previous example, where the objective function is to directly maximize the secretion flux for glutamate.

The deletion mutants that lead to higher yields of glutamate are listed in Table VI. The yield predicted for the mutant B network is as high as 84% of the maximum theoretical yield of glutamate, which is approximately 78% of the yield based on glucose weight. The current industrial yields of glutamic acid in the industry are in the range of 45% to 55% (Ikeda, 2003). In mutant A,  $\alpha$ -ketoglutarate



**Table V.** The deletion mutants for overproduction of aspartate.

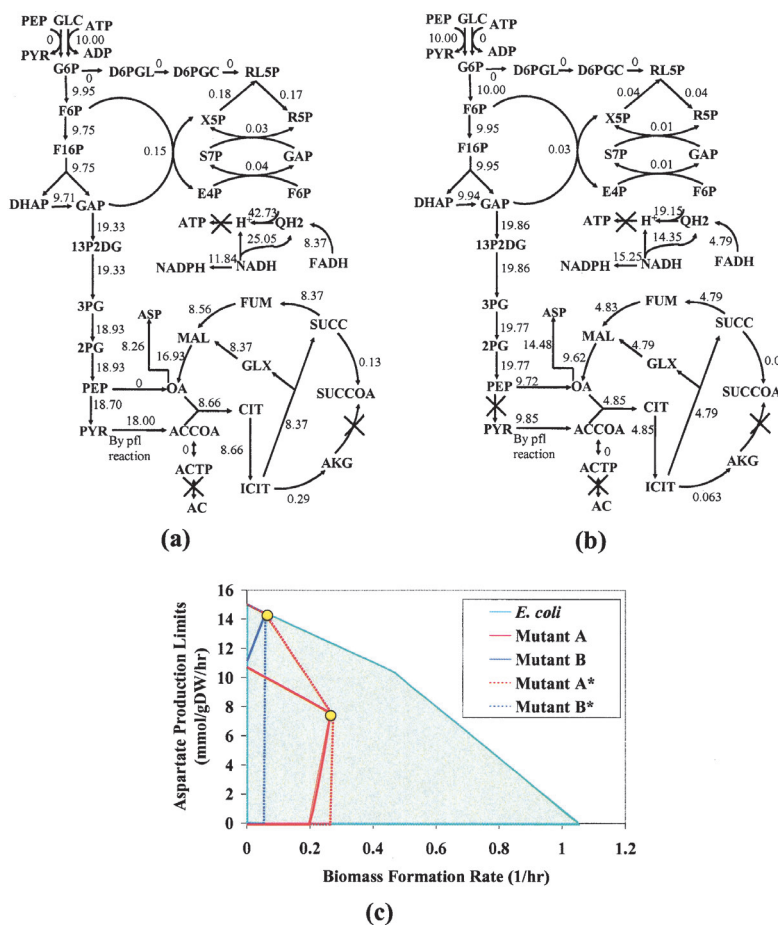
No. of knockouts	ID	Reaction	Enzyme	Growth rate (1/hr)	Uptake		Secretion	
					O <sub>2</sub> <sup>b</sup>	NH <sub>3</sub> <sup>b</sup>	CO <sub>2</sub> <sup>b</sup>	Aspartate <sup>b</sup>
3	A	1 AKG + NAD + COA → CO <sub>2</sub> + NADH + SUCCOA	2-ketoglutarate dehydrogenase	0.26	21.36	10.19	10	7.6
		2 ACCOA + PI ↔ ACTP + COA or ACTP + ADP ↔ ATP + AC	Phosphotransacetylase or Acetate kinase					
		3 ATP ↔ ADP + PI + 4 HEXT	ATPase					
4	B	1 PEP + ADP → PYR + ATP	Pyruvate kinase	0.05	9.57	14.9	9.55 (uptake)	14.34
		2 AKG + NAD + COA → CO <sub>2</sub> + NADH + SUCCOA	2-ketoglutarate dehydrogenase					
		3 ACCOA + PI ↔ ACTP + COA or ACTP + ADP ↔ ATP + AC	Phosphotransacetylase or Acetate kinase					
		4 ATP ↔ ADP + PI + 4 HEXT	ATPase					

<sup>a</sup>The maximum theoretical yield of aspartate is 19.05 mmol/g DW.hr.

<sup>b</sup>Data expressed as millimoles per gram dry weight per hour.

dehydrogenase (*sucAB*, *lpdA*), phosphotransacetylase (*pta*) or acetate kinase (*ack*), and ATPase (*atp*) are deleted. The flux distribution for this mutant network is shown in Figure 9a. The first reaction leads to the conversion of α-ketoglutarate (or 2-oxoglutarate), the reactant for gluta-

mate formation, to succinate. Note that the 2-oxoglutarate dehydrogenase complex, which catalyzes this reaction, is reported to have very low activities in all strains of *Corynebacterium glutamicum* (Kimura, 2003; Kumagai, 2000), which are used for the industrial production of



**Figure 8.** (a) Flux distribution for the aspartate-secreting mutant A (with three reactions eliminated) and (b) for mutant B (with four reactions eliminated). (c) Aspartate-production limits of the mutant networks compared with those of wild-type *E. coli* network. The dotted lines denote the behavior without fixing of the transport rates, and the solid lines show the behavior after fixing the transport rates of carbon dioxide, ammonia, and oxygen.

**Table VI.** The deletion mutants for overproduction of glutamate<sup>a</sup>.

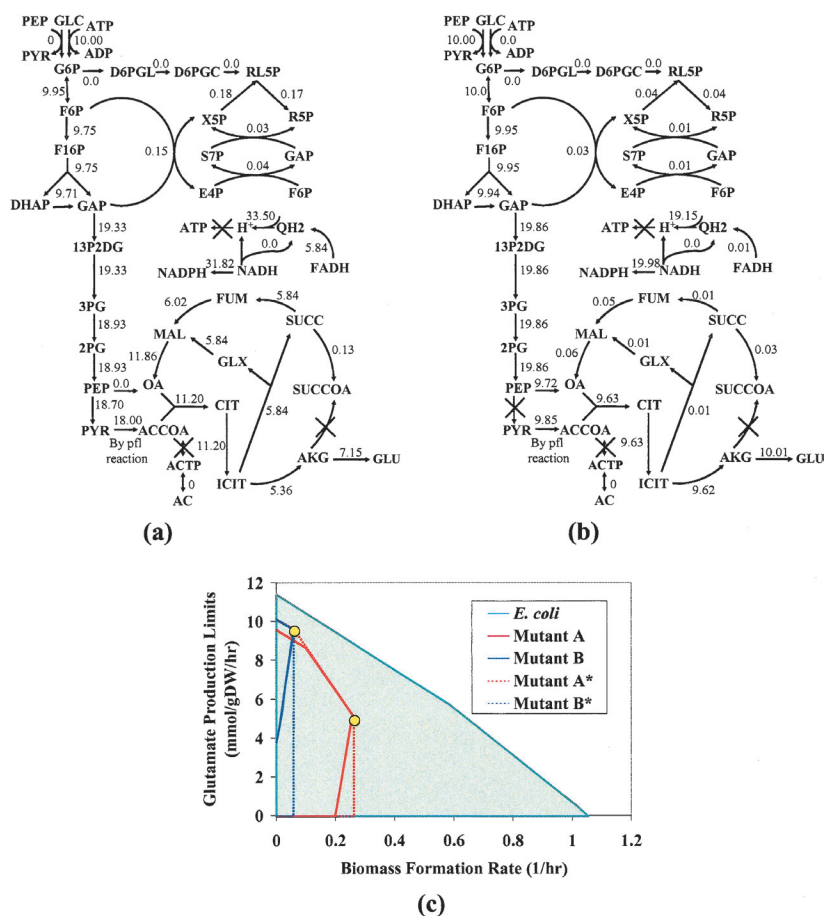
No. of knockouts	ID	Reaction	Enzyme	Growth rate (1/hr)	Uptake (O <sub>2</sub> ) <sup>b</sup>	Secretion (glutamate) <sup>b</sup>
3	A	1 AKG + NAD + COA → CO <sub>2</sub> + NADH + SUCCOA	2-ketoglutarate dehydrogenase	0.26	16.75	5.14
		2 ACCOA + PI ↔ ACTP + COA or ACTP + ADP ↔ ATP + AC	Phosphotransacetylase or Acetate kinase			
		3 ATP ↔ ADP + PI + 4 HEXT	ATPase			
4	B	1 PEP + ADP → PYR + ATP	Pyruvate kinase	0.05	9.58	9.56
		2 AKG + NAD + COA → CO <sub>2</sub> + NADH + SUCCOA	2-ketoglutarate dehydrogenase			
		3 ACCOA + PI ↔ ACTP + COA or ACTP + ADP ↔ ATP + AC	Phosphotransacetylase or Acetate kinase			
		4 ATP ↔ ADP + PI + 4 HEXT	ATPase			

<sup>a</sup>The maximum theoretical yield of glutamate is 11.36 mmol/g DW.hr.

<sup>b</sup>Data expressed as millimoles per gram dry weight per hour.

glutamate. Glutamate-producing mutants of *E. coli* have also been reported with the metabolic pathway blocked between 2-oxoglutarate and succinate (Kimura, 2003). The deletion of phosphotransacetylase or acetate kinase prevents acetyl-CoA (ACCOA) from being converted to acetate. ACCOA is a key metabolite in the TCA cycle during which

α-ketoglutarate is produced. ATPase removal once again favors the production of glutamate. Mutant B, whose distribution of fluxes is shown in Figure 9b, suggests the removal of the glycolytic reaction pyruvate kinase in addition to the eliminations discussed already for mutant A. This prevents the conversion of phosphoenolpyruvate to pyruvate. In



**Figure 9.** (a) Flux distribution for the glutamate-secreting mutant A (with three reactions eliminated) and (b) for mutant B (with four reactions eliminated). (c) Glutamate-production limits of the mutant networks compared with those of the wild-type *E. coli* network. The dotted lines represent the behavior of the mutant networks after elimination of the suggested reactions, and the solid lines represent their behavior after fixing the transport rate of oxygen in addition to blocking the secretion of key metabolites after the suggested reactions have been deleted.

mutant B, the network chooses to directly convert PEP to oxaloacetate, so that maximum flux is directed toward the TCA cycle, without leaking any considerable amount to pyruvate or pyruvate-derived products.

Interestingly, the reactions identified by OptKnock are exactly the same as those for the case of aspartate. However, the specific rate of oxygen transport suggested by OptKnock is different for glutamate overproduction (see Table VI). Also, a strong coupling between biomass formation and the secretion of glutamate can be obtained only after the export routes of D-alanine, acetate, lactate, ethanol, pyruvate, fumarate and malate are blocked. In mutant B, the transport reaction for acetate is left enabled. The extra functionality removed in this case compensates for the elimination of the acetate transport reaction from the network. Figure 9c contrasts the glutamate production limits of the wild-type *E. coli* network from those of the mutant networks. Although the wild-type *E. coli* network does not have to produce any glutamate for any level of biomass formation, mutant A is “forced” to secrete glutamate when the biomass production reaches approximately 75% of its maximum value. Moreover, for mutant B, the modified stoichiometric constraints on the network cause it to secrete glutamic acid for any rate of biomass formation, with the highest yield obtained when the latter rate is 0.057 per hour.

## DISCUSSION

In this study, we modified and extended the bilevel optimization framework OptKnock (Burgard et al., 2003) to predict the reactions whose elimination from the *E. coli* metabolic network (Edwards and Palsson, 2000) may enhance amino acid production at maximum biomass yield. The stoichiometric constraints ensured that the network secreted the desired amino acid or precursor as an obligatory byproduct of growth. We focused on the five families into which the amino acids are categorized. The framework correctly predicted that an aerobic environment is required for overproducing all amino acids. Also, the crucial role that energy balance plays in amino acid formation was computationally verified. Specifically, we found that removal of the ATPase reaction can potentially augment the network’s capacity to produce amino acids such as alanine, serine, aspartate, and glutamate by increasing the flux through glycolysis, which supplies the precursors to these amino acids. This has been observed experimentally in the case of valine and leucine (Ikeda, 2003; Tomita et al., 1996), the other two members of the pyruvate family besides alanine. The decrease in ATPase activity is associated with the low-energy status of the cell. This induces a higher uptake of glucose (Jensen et al., 1993) in experimental cultures and hence, a higher rate of its metabolism by the glycolytic pathway. The rates of transport of oxygen, carbon dioxide, and nitrogen were also included as optimization variables in the formulation due to their importance in ensuring the elimination of any remaining degrees of freedom for the network that lead to the decoupling of biomass formation and

biochemical overproduction. Notably, the transport rates of these compounds are important central variables in commercial bioreactors and are tightly regulated to maintain conditions that preclude the formation of undesirable side products (Ikeda, 2003; Kumagai, 2000).

OptKnock suggested not only straightforward strategies involving elimination of competing pathways, but also identified reactions whose removal leads to indirect channeling of more carbon flux towards amino acid production. A case in point is serine overproduction, where OptKnock suggested the removal of not only the serine degradation reaction, serine deaminase, but also the enolase and the ATPase functionalities. By considering the metabolic network in its entirety, OptKnock has the advantage of assessing the global impact of gene deletions and thus helps to reconcile the puzzle of sometimes rather contradictory strategies suggested for different experimental set-ups. Specifically, we addressed the ambiguity associated with the channeling of flux from PEP to aromatic acid production by the removal of the *ppc* gene and offered a plausible explanation. We found that the removal of *ppc* can lead to the redirection of carbon flux into the formation of chorismate only if it is accompanied by the removal of pyruvate oxidase, pyruvate dehydrogenase and pyruvate lyase reactions.

The OptKnock framework, at present, can only suggest reaction eliminations to improve biochemical yields. This single type of manipulation of the network does not fully describe the range of engineering strategies available for shaping flux distributions in metabolic networks. To this end, we plan to extend OptKnock to account for not only reaction deletions but also different levels (e.g., twofold, threefold, etc.) of repression and activation of various reaction steps. On the modeling end, the inclusion of regulatory interactions, either in the form of Boolean constraints (Covert and Palsson, 2002; Covert et al., 2001) or continuously varying feedback loops (Kompala et al., 1984; Varner and Ramkrishna, 1999), into the model will further improve its predictive power. Given the inherently incomplete nature of the employed models and the fact that the substrate uptake level (i.e., glucose) is fixed, it is important to interpret the OptKnock predictions carefully. For example, blocking of a reaction in many cases is equivalent to the overexpression of a competing pathway. Finally, the model assumes that all metabolites with export mechanisms can be secreted by the network. However, membrane permeability can be an important factor for excretion of some amino acids, as noted for threonine (Debabov, 2003). Despite these limitations, OptKnock, as shown in this work, provides useful suggestions for strain optimization despite its current limitations, and establishes a foundation for further modeling enhancements by the addition of kinetic and regulatory information.

## APPENDIX: COMPUTATIONAL PROTOCOL

The linear programming (LP) model for maximizing the biomass yield of a steady-state metabolic network com-

prising a set of  $N = \{1..N\}$  metabolites and a set of  $M = \{1..M\}$  reactions is given by:

$$\begin{aligned}
 & \text{maximize} && v_{biomass} && (\mathbf{max\ Biomass}) \\
 & \text{subject to} && \sum_{j=1}^M S_{ij}v_j = 0 && \forall i \in N \\
 & && v_{pts} + v_{glk} = v_{glc\_uptake} \text{ mmol/g DW} \cdot \text{hr} \\
 & && v_{ATP} \geq v_{atp\_main} \text{ mmol/g DW} \cdot \text{hr} \\
 & && v_j \geq 0 && \forall j \in M_{irrev} \\
 & && v_j \leq 0 && \forall j \in M_{secr\_only} \\
 & && v_j \in R && \forall j \in M_{rev}
 \end{aligned}$$

where  $S_{ij}$  is the coefficient of metabolite  $i$  in reaction  $j$ , biomass formation is quantified as an aggregate reaction flux,  $v_{biomass}$ , draining biomass components in their appropriate biological ratios (Neidhardt and Curtiss, 1996), and  $v_{atp\_main}$  is the non-growth-associated minimum ATP requirement. The uptake rate of glucose  $v_{glc\_uptake}$  is fixed and encompasses both the phosphotransferase system,  $v_{pts}$ , and glucokinase,  $v_{glk}$ , uptake mechanisms. A basis glucose uptake rate of 10 mmol/g DW·hr was chosen in this study. The set of reactions,  $M$ , is divided into reversible and irreversible reactions. Metabolites that are only secreted by the network are included in the set,  $M_{secr\_only}$ . Note that the forward direction of transport fluxes corresponds to the uptake of the metabolite and the reverse direction corresponds to its secretion. Following from the basis uptake rate of glucose, fluxes are reported in mmol/g DW·hr and the biomass formation rate is given as grams biomass formed/g DW·hr or 1/hr.

To incorporate reaction eliminations into the problem, we use binary variables,  $y_j$  (Burgard et al., 2001; Burgard and Maranas, 2001), which take on a value of one if a particular reaction is active, and a value of zero otherwise. An active reaction has an upper bound,  $v_j^{max}$ , and a lower bound,  $v_j^{min}$ , obtained by maximizing and minimizing each flux subject to the constraints in the **max Biomass** problem. We refer to this problem formulation as (**max/min**).

The following mixed integer bilevel integer programming problem is solved to identify which reactions should be eliminated from the network ( $y_j = 0$ ), such that the secretion of a particular biochemical,  $v_{biochemical}$ , is maximized during maximal growth. This was briefly described in step 1 of the computational procedure:

$$\begin{aligned}
 & \text{maximize } v_{biochemical} && (\mathbf{OptKnock}) \\
 & (\text{over } y_j) \\
 & \text{subject to} \left( \begin{array}{l}
 \text{maximize } v_{biomass} && (\mathbf{max\ Biomass}) \\
 (\text{over } v_j) \\
 \text{subject to} \sum_{j=1}^M S_{ij}v_j = 0, && \forall i \in N \\
 v_{pts} + v_{glk} = v_{glc\_uptake} \text{ mmol/g DW} \cdot \text{hr} \\
 v_{ATP} \geq v_{atp\_main} \text{ mmol/g DW} \cdot \text{hr} \\
 v_{biomass} \geq v_{biomass}^{target} \text{ 1/hr} \\
 v_j^{min}y_j \leq v_j \leq v_j^{max}y_j && \forall j \in M
 \end{array} \right)
 \end{aligned}$$

$$\begin{aligned}
 & \sum_{j \in M} (1 - y_j) \leq K \\
 & y_j \in \{0, 1\} && \forall j \in M
 \end{aligned}$$

where  $K$  is the number of allowable reaction eliminations and the **max Biomass** problem contains additional constraints, ensuring that: (i) a minimal level of biomass formation is attained; and (ii) the fluxes through reactions proposed for elimination ( $y_j = 0$ ) are set to zero.

We employ a solution methodology based on linear programming duality theory to solve this problem efficiently. The details of the solution procedure for this problem have been discussed in detail by Burgard et al. (2003).

To examine the coupling between biochemical production and biomass formation, we solve the following **max/min optimization** problem for various levels of biomass formation,  $v_{biomass}^{target}$ . This forms step 2 of the procedure for all amino acid studies:

$$\begin{aligned}
 & \text{maximize/minimize } v_{biochemical} && (\mathbf{max/min}) \\
 & \text{subject to} \sum_{j=1}^M S_{ij}v_j = 0 && \forall i \in N \\
 & v_{pts} + v_{glk} = v_{glc\_uptake} \text{ mmol/g DW} \cdot \text{hr} \\
 & v_{ATP} \geq v_{atp\_main} \text{ mmol/g DW} \cdot \text{hr} \\
 & v_{biomass} \geq v_{biomass}^{target} \text{ 1/hr} \\
 & v_j \geq 0 && \forall j \in M_{irrev} \\
 & v_j \leq 0 && \forall j \in M_{secr\_only} \\
 & v_j \in R && \forall j \in M_{rev} \\
 & v_j = 0 && \forall j \in M_{knockout}
 \end{aligned}$$

Note that the fluxes through the reactions  $M_{knockout}$  identified by OptKnock for deletion are fixed at zero. If the minimum value of biochemical production is considerably lower (or even zero) than the maximum value for most attainable biomass yields, these are considered to be uncoupled. This was encountered in four of the five cases examined. Therefore, we investigated fixing the ammonia, carbon dioxide, and/or oxygen transport fluxes to their values identified by OptKnock to achieve a stronger coupling between biochemical and biomass production (step 3 of the computational procedure). In the glutamate case, where the fixing of these transport rates was not successful for achieving glutamate overproduction, the export of key metabolites was also blocked from the network. Optimization problems were solved using CPLEX 7.0, accessed via the GAMS modeling environment on an IBM RS6000-270 workstation.

## References

- Backman KC. 1992. U.S. Patent 5169 768.
- 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.

- Burgard AP, Vaidyaraman S, Maranas CD. 2001. Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments. *Biotechnol Progr* 17:791–797.
- Chandran SS, Yi J, Draths KM, Von Daeniken R, Weber W, Frost JW. 2003. Phosphoenolpyruvate availability and the biosynthesis of shikimic acid. *Biotechnol Progr* 19:808–814.
- Chen R, Hatzimanikatis V, Yap WM, Postma PW, Bailey JE. 1997. Metabolic consequences of phosphotransferase (PTS) mutation in a phenylalanine-producing recombinant *Escherichia coli*. *Biotechnol Progr* 13:768–775.
- Covert MW, Palsson BO. 2002. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J Biol Chem* 277:28058–28064.
- Covert MW, Schilling CH, Famili I, Edwards JS, Goryanin II, Selkov E, Palsson BO. 2001. Metabolic modeling of microbial strains in silico. *Trends Biochem Sci* 26:179–186.
- Debabov VG. 2003. The threonine story. *Adv Biochem Eng Biotechnol* 79:113–136.
- Draths KM, Pompliano DL, Conley DL, Frost JW, Berry A, Disbrow GL, Stavarsky RJ, Lievens JC. 1992. Biocatalytic synthesis of aromatics from D-glucose: The role of transketolase. *J Am Chem Soc* 114:3956–3962.
- 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.
- Edwards JS, Palsson BO. 2000. The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proc Natl Acad Sci USA* 97:5528–5533.
- Eggeling L, Sahm H. 1999. Amino acid production: Principles of metabolic engineering. In: Lee SY, Papoutsakis ET, editors. *Metabolic engineering*. New York: Marcel Dekker. p 153–176.
- Flores N, Xiao J, Berry A, Bolivar F, Valle F. 1996. Pathway engineering for the production of aromatic compounds in *Escherichia coli*. *Nat Biotechnol* 14:620–623.
- Gubler M, Jetten M, Lee SH, Sinskey AJ. 1994. Cloning of the pyruvate kinase gene (pyk) of *Corynebacterium glutamicum* and site-specific inactivation of pyk in a lysine-producing *Corynebacterium lactofermentum* strain. *Appl Environ Microbiol* 60:2494–2500.
- Ikeda M. 2003. Amino acid production processes. *Adv Biochem Eng Biotechnol* 79:1–35.
- Jensen PR, Michelsen O, Westerhoff HV. 1993. Control analysis of the dependence of *Escherichia coli* physiology on the H(+)-ATPase. *Proc Natl Acad Sci USA* 90:8068–8072.
- Kimura E. 2003. Metabolic engineering of glutamate production. *Adv Biochem Eng Biotechnol* 79:37–57.
- Kompala DS, Ramakrishna D, Tsao GT. 1984. Cybernetic modeling of microbial growth on multiple substrates. *Biotechnol Bioeng* 26:1272–1281.
- Kumagai H. 2000. Microbial production of amino acids in Japan. *Adv Biochem Eng Biotechnol* 69:71–85.
- Lu JL, Liao JC. 1997. Metabolic engineering and control analysis for production of aromatics: Role of transaldolase. *Biotechnol Bioeng* 53:132–138.
- Miller JE, Backman KC, O'Connor JM, Hatch TR. 1987. Production of phenylalanine and organic acids by phosphoenolpyruvate carboxylase-deficient mutants of *Escherichia coli*. *J Indust Microbiol* 2:143–149.
- Neidhardt FC, Curtiss R. 1996. *Escherichia coli and Salmonella: Cellular and molecular biology*. Washington, DC: ASM Press.
- Patnaik R, Liao JC. 1994. Engineering of *Escherichia coli* central metabolism for aromatic metabolite production with near theoretical yield. *Appl Environ Microbiol* 60:3903–3908.
- Patnaik R, Spitzer RG, et al. 1995. Pathway engineering for production of aromatics in *Escherichia coli*: Confirmation of stoichiometric analysis by independent modulation of AroG, TktA and Pps activities. *Biotechnol Bioeng* 46:361–370.
- Schilling CH, Covert MW, Famili I, Church GM, Edwards JS, Palsson BO. 2002. Genome-scale metabolic model of *Helicobacter pylori* 26695. *J Bacteriol* 184:4582–4593.
- Sprenger G, Siewe R, Sahm H, Karuts M, Sonke T. 1998. Microbial preparation of substances from aromatic metabolism/II, U.S. Patent WO9 818 936.
- Stephanopoulos G, Vallino JJ. 1991. Network rigidity and metabolic engineering in metabolite overproduction. *Science* 252:1675–1681.
- Stephanopoulos GN, Aristidou AA, Nielson J. 1998. *Metabolic engineering principles and methodologies*. New York: Academic Press.
- Tomita F, Yokota A, Hashiguchi K, Ishigooka M, Kurahashi O. 1996. U.S. Patent WO9 606 926.
- Varner J, Ramkrishna D. 1999. Metabolic engineering from a cybernetic perspective. 1. Theoretical preliminaries. *Biotechnol Progr* 15:407–425.