Probing the Performance Limits of the *Escherichia coli* Metabolic Network Subject to Gene Additions or Deletions

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

Received 30 May 2000; accepted 3 January 2001

Abstract: An optimization-based procedure for studying the response of metabolic networks after gene knockouts or additions is introduced and applied to a linear flux balance analysis (FBA) *Escherichia coli* model. Both the gene addition problem of optimally selecting which foreign genes to recombine into *E. coli,* as well as the gene deletion problem of removing a given number of existing ones, are formulated as mixed-integer optimization problems using binary 0–1 variables. The developed modeling and optimization framework is tested by investigating the effect of gene deletions on biomass production and addressing the maximum theoretical production of the 20 amino acids for aerobic growth on glucose and acetate substrates. In the gene deletion study, the smallest gene set necessary to achieve maximum biomass production in *E. coli* is determined for aerobic growth on glucose. The subsequent gene knockout analysis indicates that biomass production decreases monotonically, rendering the metabolic network incapable of growth after only 18 gene deletions. In the gene addition study, the *E. coli* flux balance model is augmented with 3,400 non-*E. coli* reactions from the KEGG database to form a multispecies model. This model is referred to as the Universal model. This study reveals that the maximum theoretical production of six amino acids could be improved by the addition of *only one or two* genes to the native amino acid production pathway of *E. coli*, even though the model could choose from 3,400 foreign reaction candidates. Specifically, manipulation of the arginine production pathway showed the most promise with 8.75% and 9.05% predicted increases with the addition of genes for growth on glucose and acetate, respectively. The mechanism of all suggested enhancements is either by: 1) improving the energy efficiency and/or 2) increasing the carbon conversion efficiency of the production route.
© 2001 John Wiley & Sons, Inc. *Biotechnol Bioeng* **74**: 364–375, 2001.

Keywords: E. coli metabolism; flux balance models; gene deletion; gene recombination

BACKGROUND AND OBJECTIVES

Metabolic pathway engineering has attracted significant interest in recent years, catalyzed by the rapidly increasing number of sequenced microbial genomes. Altogether, as of April 2001, over 45 microbial genomes have been completely sequenced (TIGR microbial database; website http://www.tigr.org) and bioinformatic tools have allowed the functional assignment of 45–80% of their coding regions (Pennisi, 1997). This newly acquired information can be used in conjunction with recent microbial mathematical models to construct algorithmic techniques for maximizing metabolic objectives through systematic gene recombination. In addition, as the prediction capability of metabolic models continues to improve, the effect of multiple gene deletions on network robustness and organism survivability can be studied with increasing confidence.

In general, mathematical models of cellular metabolism fall into two distinct categories, ones that incorporate kinetic and regulatory information and others that include only the stoichiometry of the reaction pathways. The first class of models matches cellular behavior at an original steady state and then employs kinetic and regulatory relations to examine how the cell behaves away from this steady state in the presence of small perturbations brought about by environmental changes or enzyme engineering. Metabolic control analysis (MCA) (Kacser and Burns, 1973; Heinrich and Rapoport, 1974), introduced in the 1970s, enabled the systematic evaluation and description of many metabolic factors governing the control of flux such as enzyme activities and effector concentrations. For example, the application of an MCA based (log)linear kinetic model developed by Hatzimanikatis and Bailey (1997), along with the metabolic optimization framework introduced by Hatzimanikatis et al. (1996a), pointed out experimentally verified ways for enhancing ethanol production in metabolically engineered *E. coli* cells (Hatzimanikatis et al., 1998). Alternatively, synergetic or S-system representation models based on the Biochemical Systems Theory formalism introduced by Savageau in the late 1960s (Savageau, 1969a,b, 1970) cast each metabolite concentration balance as two competing power-law functions describing aggregation and consumption. Upon a logarithmic transformation, this provided the framework for steady state enzyme level optimization through linear programming (Voit, 1992; Regan et al., 1993; Torres et al., 1996). The quantitative effect of uncertainty or im-

**Correspondence to:* Costas D. Maranas
Contract grant sponsor: the NSF Career Award
Contract grant number: CTS-9701771 REU supplement

© 2001 John Wiley & Sons, Inc.

precision in S-system parameters, likely to occur at conditions away from the nominal steady state, was investigated by Petkov and Maranas (1997). Nonlinear kinetic models (e.g., Michaelis-Menten) have been used in place of the linear MCA or S-system models whenever detailed kinetic expressions for each reaction step are known. Both linear and nonlinear kinetic modeling approaches lend themselves well to optimization strategies for exploring possible changes in enzymes activities and/or regulatory structure that optimize a given metabolic objective (Hatzimanikatis et al., 1996a,b).

The second class of models, on the other hand, utilizes only the stoichiometric mass balances of the metabolic network to generate the broadest set of flux distributions potentially available to the cell. By requiring only the stoichiometry of biochemical pathways and cellular composition information, flux balance analysis (FBA) can be used to construct stoichiometric boundaries for the metabolic flux distributions in the absence of detailed kinetic and thermodynamic data. However, this versatility comes at the expense of perhaps unknowingly crossing kinetic or regulatory flux barriers. Therefore, FBA predictions must be treated as upper bounds to the performance of the metabolic network. Flux balance models have been widely used to successfully study many metabolic networks including penicillin production by *Penicillium chrysogenum* (Jorgensen et al., 1995; Henriksen et al., 1996), growth of *Corynebacterium melassecola* on glucose and fructose (Pons et al., 1996), by-product secretion by *E. coli* under various oxygenation rates (Varma et al., 1993a), acetate secretion from *E. coli* under ATP maximization conditions (Majewski and Domach, 1990; Delgado and Liao, 1997), and to improve large-scale cultivation of mammalian cells (Xie and Wang, 1994a,b, 1996a,b,c, 1997). In addition, the analysis of metabolic flux boundaries has provided valuable physiological insight through the determination of maximum theoretical yields (Varma et al., 1993b) and the identification of key principle nodes controlling flux split ratios around branch points (Vallino and Stephanopoulos, 1994a,b).

The utility of flux balance models coupled with the explosion of newly available genetic data motivates the need for a systematic procedure utilizing FBA to determine the most desirable recombinatory choices to undertake. Until now, recombinant DNA technology has been used to add straightforward conversion pathways which produce new and desirable cellular functions. For instance, *E. coli* strains producing biotin (Sabatie et al., 1991) and indigo (Ensley, 1985) were designed by introducing foreign genes obtained directly from the biotin and indigo production pathways of *Bacillus sphaericus* and *Pseudomonas putida*, respectively. Other success stories include enhancing *E. coli*'s ability to overexpress heterologous proteins (Aristidou et al., 1994; Chou et al., 1994; Dedhia et al., 1994) and engineering microorganisms to biodegrade pollutants such as heavy metals (Wang et al., 1997), phosphates (Keasling et al., 1998; Van Dien and Keasling, 1998), and trichloroethylene (Winter et al., 1989). It is the objective of this study to utilize FBA and mixed-integer programming tools to select the mathematically optimal genes for recombination into *E. coli* from a metabolic database encompassing many genes from multiple species. The resulting pathways need not lie directly on main production pathways, as they may enhance production indirectly by either redirecting metabolic fluxes into the production pathways or by increasing the energy efficiency of the present pathways.

The recent upsurge of sequenced genomes has also brought significant attention to the question of which genes are crucial for supporting cellular life. FBA modeling provides a useful tool to help elucidate this question. Although FBA models cannot simulate the regulatory structure alterations associated with gene deletions, these models can capture whether sufficient network connectivity exists to produce metabolites critical to cellular survival. In fact, a recent FBA model proposed by Edwards and Palsson (2000a) was able to qualitatively predict the growth patterns of 86% of the mutant *E. coli* strains examined. This model was also used to identify some of the essential gene products of central metabolism for aerobic and anaerobic *E. coli* growth on glucose (Edwards and Palsson, 2000b). Determining the maximum number of tolerable gene deletions in a given metabolic system, however, requires a discrete optimization strategy in which multiple gene deletions can be simultaneously examined. A related approach utilizing discrete optimization to identify all alternate optima in linear metabolic models has been proposed by Lee et al. (2000).

In this work, we examine how stoichiometric boundaries of cellular performance expand or contract in the presence of multiple gene additions or deletions. A FBA model of the cellular metabolism of *E. coli* is constructed incorporating the reaction pathways provided by Pramanik and Keasling (1997), along with modifications suggested by Karp et al. (1999) based on more recent data. The modifications are either small molecule corrections based on more recent metabolic information or the removal of certain pathways now known to be absent from the *E. coli* genotype. A stoichiometric matrix as suggested by Schilling (1999) containing all metabolic reactions from the Kyoto Encyclopedia of Genes and Genomes (KEGG, 1999, website http://www.genome.ad.jp/kegg/) is compiled and incorporated into the model. We refer to this multispecies stoichiometric matrix as the Universal stoichiometric matrix. A short discussion of FBA will be presented next, followed by the gene addition and deletion formulations and their application to biomass and amino acid production in *E. coli*.

FLUX BALANCE ANALYSIS

FBA requires only the stoichiometry of biochemical pathways and cellular composition information to identify boundaries for the flux distributions available to the cell. Although microorganisms have evolved highly complex

control structures which eventually collapse these available boundaries into single points, FBA models are still valuable in setting upper bounds for performance targets and in identifying "ideal" flux distributions. The underlying principle of FBA is mass balances on the metabolites of interest. For a metabolic network comprised of N metabolites and M metabolic reactions, we have:

$$\sum_{j=1}^{M} S_{ij}v_j = b_i, \forall\, i \tag{1}$$

where S_{ij} is the stoichiometric coefficient of metabolite i in reaction j, v_j represents the flux of reaction j, and b_i quantifies the network's uptake (if negative) or secretion (if positive) of metabolite i. For all internal metabolites, b_i is zero. Reversible reactions are defined simply as two irreversible reactions in opposite directions, constraining all fluxes to non-negative values.

Typically, the resulting flux balance system of equations is underdetermined as the number of reactions exceeds the number of metabolites and additional information is required to solve for the reaction fluxes. Several researchers have measured external fluxes to add as constraints to their underdetermined models, rendering them completely determined or overdetermined (Jorgensen et al., 1995; Vallino and Stephanopoulos, 1993; Papoutsakis and Meyer, 1985a,b; Pons et al., 1996). However, additional assumptions such as removing reaction pathways are often needed before external flux measurements can completely define a system and neglecting potentially active pathways to render a system completely defined may cause large changes in calculated fluxes (Pramanik and Keasling, 1997). A popular technique for investigating metabolic flux distributions is linear optimization (Varma and Palsson, 1994). The key conjecture is that the cell is capable of spanning all flux combinations allowable by the stoichiometric constraints and thus achieving any flux distributions that maximize a given metabolic objective (e.g., biomass production). The linear programming model for maximizing biomass production is:

$$\begin{aligned} \text{Maximize} \quad & Z = v_{biomass} \\ & \sum_{j=1}^{M} S_{ij}v_j = b_i, \forall\, i \\ & b_i \in \Re, \forall\, i \\ & v_j \in \Re^{+}, \forall\, j \end{aligned} \tag{2}$$

where $v_{biomass}$ is a flux drain comprised of all necessary components of biomass in their appropriate biological ratios (Neidhardt, 1997). Other objective functions such as maximizing metabolite production (Varma et al., 1993b), maximizing biomass production for a given metabolite production (Varma et al., 1993b), and maximizing ATP production have also been investigated (Majewski and Domach, 1990; Ramakrishna et al., 2001).

Escherichia coli Stoichiometric Models

Microbial stoichiometric models incorporate collections of reactions known to occur in the studied species for simulating metabolism. The complete sequencing of the *E. coli* genome (Blattner et al., 1997) makes it a model organism for the study presented in this article because extensive knowledge regarding its biochemical pathways is readily available. Varma and Palsson (1993) proposed the first detailed FBA *E. coli* model capable of predicting experimental observations. The stoichiometric matrix included 95 reversible reactions utilizing 107 metabolites for simulating glucose catabolism and macromolecule biosynthesis. This model was used to investigate by-product secretion of *E. coli* at increasingly anaerobic conditions (Varma et al., 1993a) and was able to predict the right sequence of by-product secretion consistent with experimental findings: first, acetate at slightly anaerobic conditions, then formate, and finally ethanol at highly anaerobic conditions. Building on the previous model, Pramanik and Keasling (1997) introduced a model that incorporated 126 reversible reactions (including 12 reversible transport reactions) and 174 irreversible reactions, as well as 289 metabolites. Pramanik and Keasling (1997) correlated the macromolecule composition of *E. coli* as a function of growth rate and verified their model with experimental data. The model successfully predicted several levels of genetic control such as the glycoxylate shunt closing for growth on glucose and the PEP carboxykinase flux tending toward oxaloacetate. Furthermore, the glycoxylate shunt was active during growth on acetate, while the flux through PEP carboxykinase was toward phosphoenolpyruvate.

The stoichiometric *E. coli* model used in this study employs 178 irreversible, 111 reversible, and 12 transport reactions compiled largely from the model published by Pramanik and Keasling (1997). The modifications to the Pramanik and Keasling stoichiometric matrix are given in Table I. They are primarily small molecule corrections (e.g., ATP in place of GTP for succinate thiokinase) or the removal of reactions now known to be absent from *E. coli* based on more recent data (Karp et al., 1999). Note that similar changes were also independently included in the most recently published *E. coli* model of Edwards and Palsson (2000a). The metabolic network is fueled by transport reactions allowing an unconstrained supply of ammonia, hydrogen sulfate, and phosphate, along with a constrained supply of glucose or acetate to enter the system. Oxygen uptake is unconstrained to simulate aerobic conditions. Unconstrained secretion routes for lactate, formate, ethanol, glyceraldehyde, succinate, and carbon dioxide by-products are provided by the transport reaction fluxes. The Universal model is constructed by incorporating 3,400 cellular reactions from the Kyoto Encyclopedia of Genes and Genomes into the modified Keasling stoichiometric model. The Uni-

Table I. Modifications to the Pramanik and Keasling model.*

Enzymes	Reactions
Reactions assumed irreversible	
Phosphofructokinase	Fructose-1,6-bisphosphate → Fructose-6-phosphate + PI
Citrate synthase	Acetyl-CoA + Oxaloacetate → CoA + Citrate
2-Ketoglutarate dehydrogenase	2-Ketoglutarate + NAD + CoA → Succinyl-CoA + CO2 + NADH
PRSCAIM synthetase	RCAIM + ATP + Aspartate → ADP + Pi + PRSCAIM
Glycerol kinase	Glycerol + ATP → Glycerol-3-phosphate + ADP
Reactions removed from model	
Unknown pathway	5′-methylthioadenosine → Adenosine + Methionine
Cystathionase	Homocysteine + Adenosine ↔ s-Adenosyl-homocystine
Sulfotransferase	Adenosine-3,5-diphosphate + sulfite ↔ 3-Phosphoadenylylsulfate
Reactions modified	
Fructose-1,6-bisphosphate aldolase	Fructose-1,6-bisphosphate → Fructose-6-phosphate + Pi
Isocitrate dehydrogenase	Isocitrate + NADP ↔ CO2 + NADPH + 2-Ketoglutarate
Succinate thiokinase	Succinyl-CoA + ADP + PI ↔ ATP + CoA + Succinate
Prephenate dehydrogenase	Prephenate + NAD → CO2 + NADH + para-Hydroxy phenyl pyruvate
Hol dehydrogenase	Histidinol + 3 NAD → 3 NADH + Histidine
RCAIM synthetase	AIR + CO2 + ATP → 5-p-Ribosyl-4-carboxy-5-aminoimidazole + ADP + Pi
GTP cyclohydrolase	GTP → D6RP5P + Formate + Ppi
3,4-Dihydroxy-2-Butanone-4-phosphate synthase	Ribulose-5-phosphate → DB4P + Formate
H2Neopterin triphosphate pyrophosphatase	AHTD → PPi + Pi + DHP
CoA synthase	OIVAL + METTHF + NADPH + ALA + CTP + 4 ATP + CYS → THF + NADP + AMP + 2 PPi + 2 ADP + CO2 + CoA + CDP

*Modifications based on information by Karp et al. (1999).

versal stoichiometric matrix contains all reactions known to occur in *E. coli*, as well as a number of reactions from other organisms.

MATHEMATICAL MODELING OF GENE DELETIONS/ADDITIONS

Practically every metabolic reaction is regulated to some extent by one or more enzymes, produced by the translation of one or more genes. As a result, the removal of certain genes from microbial DNA sequences can be fatal or have little if any effect, depending on the role of the enzymes coded for by these genes. Conversely, the addition of certain genes through recombinant DNA technology can have either no effect or produce novel desirable cellular functionalities. Given a stoichiometric model of *E. coli* metabolism and the Universal stoichiometric matrix encompassing reactions occurring in multiple species, the goal of this section is to formulate a mathematical model that 1) captures cellular robustness in the presence of multiple gene deletions, and 2) identifies additional genes from the Universal dataset having the most profound effect on improving a given metabolic objective.

First, define $K = \{k\} = \{1,\ldots,M,\ldots,T\}$ as the set of all possible genes where M represents the number of *E. coli* genes and T represents the total number of genes in the dataset. This set can be partitioned into two subsets E and NE where subset E represents genes present in *E. coli* and subset NE represents genes present only in non-*E. coli* species:

$$\mathsf{E} = \{k \mid 1 \le k \le M\} \tag{3}$$

$$NE = \{k \mid M + 1 \le k \le T\} \tag{4}$$

Subsequently, let binary variable y_k describe the presence or absence of each gene k:

$$y_k = \begin{cases} 0 & \text{if gene k is not expressed in host organism} \\ 1 & \text{if gene k is present and functional} \end{cases} \tag{5}$$

The selection of the optimal gene choices for deletion or insertion from DNA recombination can be determined by appropriately constraining the number of nonzero elements in $\mathbf{y}$. The case of removing a given number of genes, d, from *E. coli* can be investigated by including the following constraint:

$$\sum_{k \in \mathsf{E}} (1 - y_k) \ge d \tag{6}$$

This ensures that no more than $(M - d)$ genes are available to the metabolic network. Similarly, the effect of introducing any number of additional genes, h, can be investigated by utilizing:

$$y_k = 1, \forall\, k \in \mathsf{E} \tag{7}$$

$$\sum_{k \in NE} y_k \le h \tag{8}$$

Equation (7) allows all *E. coli* genes to be present and functional if necessary, while Eq. (8) sets an upper limit to the number of allowable additions. The optimal genes selected by the model are obtained by determining which elements of NE are equal to one. In addition, since multiple genes often correspond to a single reaction and occasionally multiple reactions are catalyzed by an enzyme coded for by

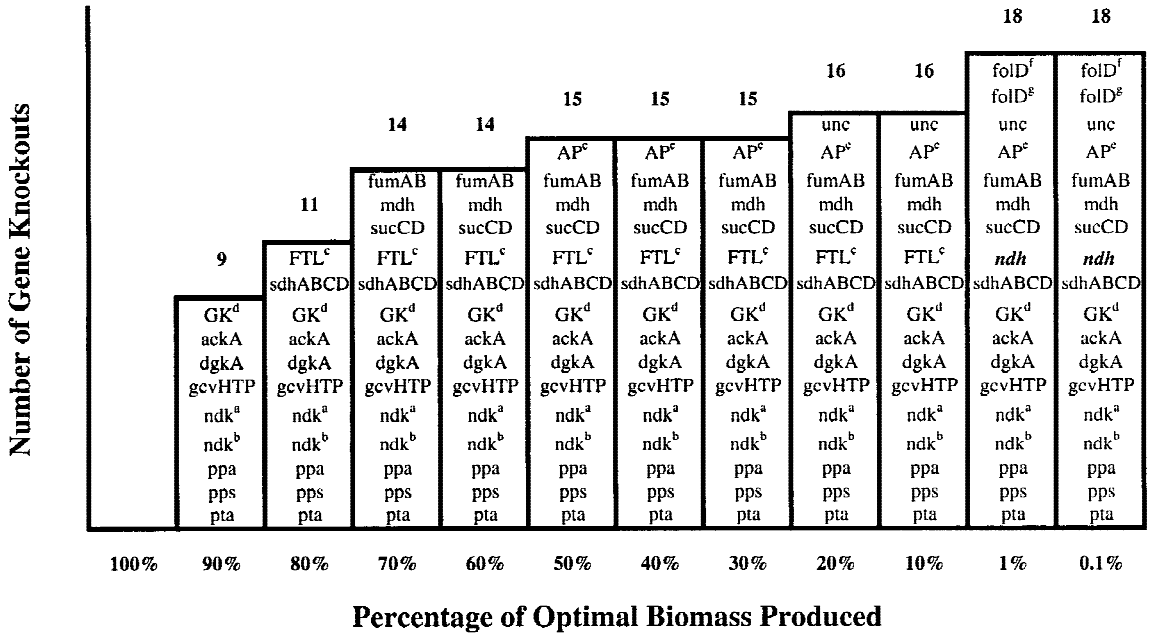

Figure 1. Gene knockouts at various biomass production levels for growth on glucose. [a,b,f,g]Same gene responsible for two intracellular reactions. [c,d,e]No gene has been assigned to these intracellular reactions.

a single gene, the binary parameter a_{jk} is defined to describe which enzymes are coded for by which genes:

$$a_{jk} = \begin{cases} 0 & \text{if gene k has no direct effect on reaction j} \\ 1 & \text{if gene k codes for an enzyme catalyzing reaction j} \end{cases} \quad (9)$$

Parameter a_{jk} establishes links between genetic functional assignments and reactions. In order for a flux v_j to take on a nonzero value, at least one gene must code for an enzyme catalyzing this reaction ($a_{jk} = 1$) and this gene must be present and functional in the host organism ($y_k = 1$). Given that at least one gene must code for every enzyme, we have:

$$\sum_k a_{jk} y_k \begin{cases} = 0 & \text{if no gene coding for the enzyme of reaction } j \text{ is functional} \\ \geq 1 & \text{if at least one gene coding for the enzyme of reaction } j \text{ is functional} \end{cases} \quad (10)$$

Table II. Genes selected for removal by knockout study

Enzymes	Genes	Reactions
3,5-ADP phosphatase	AP[e]	35ADP → AMP + Pi
Acetate kinase	ackA	AC + ATP → ACTP + ADP
CDP kinase	ndk[a]	CDP + ATP → CTP + ADP
CMP kinase	ndk[b]	CMP + ATP → CDP + ADP
F0F1-ATPase	unc	ADP + Pi + H_{ext} → ATP
Formate THF ligase	FTL[c]	THF + FORMATE + ATP → ADP + Pi + FTHF
Fumarase	fumAB	FUM → MAL
Glyceraldehyde kinase	GK[d]	GLAL + ATP → ADP + T3P1
Glycine cleavage system	gcvHTP	GLY + THF + NAD → METTHF + NADH + CO2 + NH3
Malate dehydrogenase	mdh	MAL + NAD → NADH + OA
Methenyl THF cyclohydrolase	foID[f]	METHF → FTHF
Methylene THF dehydrogenase	foID[g]	METTHF + NADP → METHF + NADPH
NADH dehydrogenase I	ndh	NADH + Q → NAD + QH2 + 4 H_{ext}
PEP synthase	pps	PYR + ATP → PEP + AMP + Pi
Phosphatidate phosphatase	dgkA	DGR + Pi → PA
Phosphotransacetylase	pta	ACTP + COA → ACCOA + Pi
Pyrophosphatase	ppa	PPi → 2 Pi
Succinate dehydrogenase	sdhABCD	SUCC + FAD → FADH2 + FUM
Succinate thiokinase	sucCD	SUCCOA + GDP + Pi → GTP + COA + SUCC

[a,b]Same gene responsible for two intracellular reactions.
[f,g]Same gene responsible for two intracellular reactions.
[c,d,e]No gene has been assigned to these intracellular reactions.

This implies that the following constraint:

$$\leq \nu_j \leq V_j^{max} \left(\sum_k a_{jk} y_k \right) \tag{11}$$

ensures that $\nu_j = 0$ if there exists not active gene k capable of supporting reaction j. In this case, $\Sigma_k a_{jk} y_k = 0$, which in turn forces the value of ν_j to zero. Alternatively, if at least one such gene is functional, then $\Sigma_k a_{jk} y_k \geq 1$, allowing ν_j to assume any value between zero and an upper bound V_j^{max}. These upper bounds are set by maximizing the given flux ν_j subject to the stoichiometric constraints. These problems are solved using CPLEX 6.6 accessed via the commercial software package GAMS. Problems with up to 3700 binary variables were solved on an IBM RS6000-270 workstation.

GENE KNOCKOUT STUDY

In this study we explore what is the smallest gene set capable of maximizing biomass production on glucose substrate (uptake basis: 10 mmol) and what is the maximum number of gene deletions from this gene set that still maintains a specified level of biomass production. First, we maximized the biomass production flux, $\nu_{biomass}$. The solution yields the maximum theoretical level of biomass production ($\nu_{biomass}^{\max} = 1.25$ g biomass/gDW · h) achievable by the metabolic network within the stoichiometric constraints. Next, the minimum number of genes that maintains a specified target level of biomass production $\nu_{biomass}^{target}$ (as a percentage of the maximum) is determined. The new objective function minimizes the total number of functional *E. coli* genes available to the cell subject to the constraint of setting biomass production $\nu_{biomass}$ greater than or equal to $\nu_{biomass}^{target}$. This problem is formulated as:

$$\begin{aligned} \text{Minimize } Z &= \sum_{k \in \mathsf{E}} y_k \\ \text{subject to } & \sum_{j=1}^{M} S_{ij} \nu_j = b_i, \ \forall i \\ & 0 \leq \nu_j \leq V_j^{max} \left(\sum_k a_{jk} y_k \right) \\ & \nu_{biomass} \geq \nu_{biomass}^{t\,\arg\,et} \\ & \nu_j \in \Re^+, \forall j \\ & b_i \in \Re, \forall i \\ & y_k \in \{0,1\}, \forall k \in \mathsf{E} \end{aligned} \tag{12}$$

where the nonzero elements of y_k define the minimum gene set capable of attaining the target growth rate. The smallest gene set $\mathcal{M}_{100\%}$, capable of sustaining the maximum theoretical growth rate is obtained by setting $\nu_{biomass}^{target} = 100\% \cdot \nu_{biomass}^{\max}$. The model predicts that 202 nontransport intracellular reactions out of 400 available reactions (111 ×

Table III. Model predictions of maximum theoretical yields of amino acids for growth on glucose and acetate.

	Maximum theoretical yield (mmol/per 10 mmol glucose)				Maximum theoretical yield (mmol/per 10 mmol acetate)			
	Palsson '93	Modified keasling '97	Universal model	% Increase	Palsson '93	Modified keasling '97	Univeral model	% Increase
Alanine	20.00	20.00	20.00	—	3.93	5.29	5.29	—
Arginine	7.74	9.26	10.07	8.75%	1.51	2.43	2.65	9.05%
Asparagine	15.60	18.18	19.23	5.77%	3.24	4.66	4.91	5.45%
Aspartate	18.20	20.00	20.00	—	3.82	5.29	5.29	—
Cysteine	9.75	11.49	11.90	3.57%	1.81	3.29	3.42	3.80%
Glutamate	10.00	13.33	13.33	—	2.68	3.65	3.65	—
Glutamine	10.00	13.33	13.33	—	2.50	3.46	3.46	—
Glycine	20.00	35.33	35.33	—	3.94	9.00	9.00	—
Histidine	7.30	9.77	9.80	0.23%	1.37	2.43	2.54	4.53%
Isoleucine	7.34	8.00	8.07	0.91%	1.44	2.13	2.13	—
Leucine	6.67	8.00	8.00	—	1.59	2.18	2.18	—
Lysine	7.84	8.45	8.45	—	1.55	2.18	2.18	—
Methionine	5.74	7.04	7.19	2.16%	1.11	1.81	1.85	2.46%
Phenylalanine	5.29	5.76	5.76	—	1.00	1.47	1.47	—
Proline	10.00	10.91	10.91	—	2.10	2.90	2.90	—
Serine	20.00	23.04	23.04	—	3.94	5.87	5.87	—
Threonine	12.30	15.00	15.00	—	2.50	3.91	3.91	—
Tryptophan	4.14	4.67	4.73	1.28%	0.76	1.17	1.19	1.32%
Tyrosine	5.48	6.03	6.03	—	1.03	1.54	1.54	—
Valine	10.00	10.00	10.00	—	1.96	2.67	2.67	—

Palsson '93: *E. coli* model proposed by Varma et al. (1993b).

Modified Keasling '97: Modified Palsson and Keasling (1997) *E. coli* model as described in text.

Universal Model: Modified Palsson and Keasling (1997) *E. coli* model augmented with non-*E. coli* reactions compiled by the Kyoto Encyclopedia of Genes and Genomes.

% Increase: Between the modified Palsson and Keasling (1997) model and the Universal model.

Table IV. Model selections of enzymatic reactions that will enhance the theoretical amino acid production capabilities of *Escherichia coli.*

Amino acid	Substrate	EC#	Enzyme	Reaction catalyzed
Arginine	Glucose:	2.7.1.90	6-Phosphofructokinase (pyrophosphate)	Fructose-6-P + PPI → Fructose-1,6-Bisphosphate + Pi
		2.7.2.2	Carbamate kinase	ATP + NH3 + CO2 → ADP + Carbamoyl Phosphate
	Acetate:	2.7.2.2	Carbamate kinase	ATP + NH3 + CO2 → ADP + Carbamoyl Phosphate
		2.7.2.12	Acetate kinase (pyrophosphate)	Acetate + PPi → Pi + Acetyl-Phosphate
	Glucose/		Aspartate—ammonia ligase (ADP-	
Asparagine	Acetate:	6.3.1.4	forming)	ATP + NH3 + L-Aspartate → Pi + ADP + L-Asparagine
	Glucose/			
Cysteine	Acetate:	2.7.7.5	Sulfate adenylyltransferase (ADP)	Sulfate + ADP → Pi + Adenylyl-Sulfate
Histidine	Glucose:	1.4.1.10	Glycine dehydrogenase	NAD + glycine → glyoxylate + NADH + NH3
		2.7.1.90	6-Phosphofructokinase (pyrophosphate)	Fructose-6-P + PPi → Fructose-1,6-Bisphosphate + Pi
	Acetate:	1.4.1.10	Glycine dehydrogenase	NAD + glycine → glyoxylate + NADH + NH3
			Phosphoenolpyruvate carboxykinase	
		4.1.1.38	(pyrophosphate)	PPi + Oxaloacetate → CO2 + Pi + PEP
Isoleucine	Glucose:	many		
Methionine	Glucose:	2.7.7.5	Sulfate adenylyltransferase (ADP)	Sulfate + ADP → Pi + Adenylyl-Sulfate
	Acetate:	1.4.1.10	Glycine dehydrogenase	NAD + glycine → glyoxylate + NADH + NH3
		2.7.7.5	Sulfate adenylyltransferase (ADP)	Sulfate + ADP → Pi + Adenylyl-Sulfate
		2.7.9.1	Pyruvate, phosphate dikinase	Pyruvate + Pi + ATP → AMP + PPi + PEP
			Phosphoenolpyruvate carboxykinase	
		4.1.1.38	(pyrophosphate)	PPi + Oxaloacetate → CO2 + Pi + PEP
Tryptophan	Glucose:	2.7.1.90	6-Phosphofructokinase (pyrophosphate)	Fructose-6-P + Ppi → Fructose-1,6-Bisphosphate + Pi
		2.7.9.1	Pyruvate, phosphate dikinase	Pyruvate + Pi + ATP → AMP + PPi + PEP
	Acetate:	2.7.9.1	Pyruvate, phosphate dikinase	Pyruvate + Pi + ATP → AMP + PPi + PEP
			Phosphoenolpyruvate carboxykinase	
		4.1.1.38	(pyrophosphate)	PPi + Oxaloacetate → CO2 + Pi + PEP

2 reversible reactions + 178 irreversible reactions) are required to sustain $v_{biomass}^{max}$. These reactions include the glycolytic reactions, the pentose phosphate pathway, the TCA cycle, the respiratory reactions, and all other anabolic and catabolic routes necessary for optimal growth.

Given $\mathcal{M}_{100\%}$, the next goal is to determine which of these genes could be knocked out while still allowing the metabolic network to sustain specified suboptimal growth rates. This is accomplished by setting $v_{biomass}^{target}$ equal to various percentages of $v_{biomass}^{max}$ and constraining the intracellular reaction fluxes outside of $\mathcal{M}_{100\%}$ to zero. It must be noted that this assumption prevents the model from activating any genes outside of the $\mathcal{M}_{100\%}$ set and the significance of this assumption will be discussed in the following section. The number of allowable gene knockouts for various biomass production levels are given in Figure 1, while the selected gene removals are presented in Table II. As expected, as the biomass production demands on the network are lessened the model tolerates more gene knockouts. However, the range of allowable knockouts is rather small. Specifically, the model tolerates at most nine gene deletions with a biomass requirement of $90\% \cdot v_{biomass}^{max}$, while 18 gene removals render the network incapable of biomass formation. Thus the subset containing all elements of $\mathcal{M}_{100\%}$ minus the 18 gene knockouts (194 genes) describes the smallest subset of $\mathcal{M}_{100\%}$ capable of sustaining *E. coli* cellular growth for the employed FBA model. Additionally, it must be noted that all subsets include the seven experimentally verified essential gene products of central metabolism identified by the *in silico* gene deletion study of *E. coli* conducted by Edwards and Palsson (2000b).

DISCUSSION OF THE GENE DELETION STUDY

Investigation of the specific gene knockouts provides interesting insight into the effect of various energy generation pathways. The suggested gene deletions imply that the energetic status of the network is improved as the required biomass production demands on the cell are reduced. This is demonstrated by the fact that as the biomass requirements are lessened, the optimization formulation sequentially eliminates pathways responsible for the formation of energy. One such observation involves the gradual degradation of the TCA cycle. When the model is constrained to produce only 80% of the optimal level of biomass, the network no longer utilizes the succinate dehydrogenase enzyme to produce $FADH_2$. Further reducing the biomass production requirement to 70% enables the removal of the fumAB, mdh, and sucCD genes foregoing the formation of one GTP and one NADH per unit reaction flux. The next major energy formation pathway to be eliminated occurs at a biomass production level of 20%. At this point, the energetic state of the cell is such that it no longer requires the formation of ATP from the cellular proton gradient. Finally, at the lowest biomass production levels, the cell no longer requires the oxidation of NADH to force protons across the cellular membrane.

This study provides insight into the dependence of cellular growth on various energy generation pathways and provides an estimate of the minimum number of metabolic genes capable of enabling cellular growth. The prediction of 194 genes is lower than the theoretical estimation of 256 by Mushegian and Koonin (1996), obtained by investigating the complete genomes of *Haemophilus influenzae* and *My-*

coplasma genitalium and assuming genes preserved across large phylogenetic distances are most likely essential. This was expected, considering the inability of this reaction-based framework to account for genes associated with translation, transcription, replication, and repair and the lumping of pathways by the stoichiometric model. A more practical comparison involves considering the number of metabolic genes included in the minimal gene set estimation. In this case, the predicted set of 194 metabolic genes overestimates the 94 metabolic genes included in the minimal gene set proposed by Mushegian and Koonin (1996). This overestimation arises in part because the effect of activating metabolic genes outside of the original optimal gene set was not investigated. Work underway shows that this lowers the minimal gene set estimation by opening additional metabolic routes. Furthermore, this study only allowed glucose to enter the network as organic fuel and limited metabolic capacity can be compensated for by a proportionately greater dependence on the importation of nucleosides, amino acids, and other metabolites (Hutchison et al., 1999). The problem of identifying the mathematically minimum number of metabolic reactions capable of supporting growth on a multiorganic substrate based on the more recent *E. coli* model of Edwards and Palsson (2000a) is addressed in an upcoming article. (Maranas group website: http://fenske/Faculty/CMaranas/index.html).

AMINO ACID PRODUCTION OPTIMIZATION STUDIES

In this section, we identify mathematically optimal reaction pathways to recombine into the *E. coli* metabolic network to optimize amino acid formation for growth on glucose and acetate. We explored the theoretically optimal formation of all 20 amino acids. Each optimization run was performed for two cases: 1) including only the reactions present in *E. coli*, and 2) allowing the model to select all reactions from the Universal stoichiometric matrix. The problem of maximizing the amino acid production is formulated by substituting amino acid accumulation, b_{aa}, in place of $v_{biomass}$ in Eq. [2], while the problem of maximizing the amino acid formation b_{aa}^{UNV} of the Universal network is formulated as:

$$\begin{aligned} \text{Maximize} \quad & Z = b_{aa}^{UNV} \\ \text{subject to} \quad & \sum_{j=1}^{M} S_{ij}\, v_j = b_i, \ \forall\, i \\ & y_k = 1, \forall\, k \in \mathsf{E} \\ & 0 \le v_j \le V_j^{max} \left(\sum_k a_{jk}\, y_k \right) \\ & v_j \in \Re^{+}, \forall\, j \\ & b_i \in \Re, \forall\, i \\ & y_k \in \{0,1\}, \forall\, k \in \mathsf{E} \cup \mathit{NE} \end{aligned} \tag{13}$$

Note that this formulation allows the selection of any number of reactions from the multispecies reaction list. Reactions chosen by the model but absent in *E. coli* (i.e., all nonzero y_k elements of NE) provide routes for manipulating the cellular metabolism through recombinant DNA technology. The theoretical amino acid production capabilities of the *E. coli* metabolic network, with and without the additional reactions from the Universal matrix, are shown in Table III for growth on glucose and acetate. It must be noted that it is the structural pathway changes predicted by the model that are more meaningful than the exact numerical values because these are theoretical maximum yield calculations. Predictions by the Varma and Palsson (1993) model are shown for comparison. As expected, the maximum production capabilities by the Varma and Palsson (1993) model are slightly below the predictions of the more complex employed model due to the additional metabolic routes available for production.

The results show that improvements to seven amino acid production pathways of *E. coli* are theoretically attainable with the addition of genes from various organisms. Manipulation of the arginine pathway shows the most promise, with 8.75% and 9.05% increases with additional genes for growth on glucose and acetate, respectively. The optimal recombinant asparagine pathway shows 5.77% and 5.45% increases over current *E. coli* growth on glucose and acetate, while cysteine production can be raised 3.57% and 3.80%, respectively. The histidine production pathway is revealed as another encouraging target for DNA recombination with 0.23% and 4.53% improvements available as well. The isoleucine, methionine, and tryptophan formation pathways offer the final three genetic objectives for enhancing production.

The enzymes responsible for introducing these various improvements to the *E. coli* amino acid production pathways are shown in Table IV. In most cases, the addition of only one or two genes to the original amino acid production pathway results in an increased maximum theoretical yield, even though the complete list of 3,400 reactions was available for selection. For example, introducing foreign genes coding for carbamate kinase and the pyrophosphate-dependent version of 6-phosphofructokinase further optimizes arginine production for growth on glucose, while adding carbamate kinase and another gene coding for acetate kinase renders the arginine production pathway on acetate stoichiometrically optimal. Expressing the genes coding for aspartate-ammonia ligase and sulfate adenylyltransferase in *E. coli* results in the increase mentioned above in asparagine and cysteine production, respectively. Only the production of isoleucine on glucose and acetate substrates and the production of methionine on acetate require over two additional enzymes to reach optimality according to the model.

DISCUSSION OF THE GENE ADDITION STUDY

Careful examination of these amino acid pathways reveals how these additional enzymes improve the energetic efficiency of the original routes. The original and Universal arginine production pathways for growth on glucose are shown in Figure 2. The two pathways differ in only two reactions—the pyrophosphate-dependent analog of 6-phosphofructokinase in the Universal model replaces the ATP-dependent version present in *E. coli*, and carbamate kinase

in the Universal model replaces carbamoyl phosphate synthetase from the original *E. coli* model. The first improvement to energy utilization occurs because the Universal model 6-phosphofructokinase uses pyrophosphate formed from argininosuccinate synthase reaction instead of ATP to transfer a phosphate group to fructose-6-phosphate in the third step in glycolysis. The *E. coli* model, which sends this pyrophosphate through pyrophosphatase for hydrolytic cleavage, in effect wastes the energy from this energy-rich phosphoanhydride bond. By recapturing this otherwise wasted energy, the pyrophosphate version of 6-phosphofructokinase requires one less ATP phosphoanhydride bond per arginine molecule produced.

The second form of cellular energy savings is realized by the replacement of carbamoyl phosphate synthetase. The native carbamoyl phosphate synthetase creates one mole of carbamoyl phosphate from carbon dioxide at the expense of two ATP phosphoanhydride bonds. This reaction also requires an amino group of one glutamine molecule, which subsequently forms glutamate. Reforming glutamine from glutamate requires yet another ATP; thus, each unit flux through carbamoyl phosphate synthetase requires three ATP. Carbamate kinase, incorporated in the Universal model, forms carbamoyl phosphate from carbon dioxide and ammonia at the expense of only one ATP. Therefore, carbamate kinase requires two less ATP bonds per unit flux of carbamoyl phosphate formed. Overall, the additional genes used by the Universal model save the original pathway three net ATP bonds, increasing arginine production by 8.75%. A similar analysis can be performed on native and Universal arginine production routes from acetate substrate depicted in Figure 3.

The *E. coli* asparagine production pathway is shown in Figure 4 for two modes of glucose entry into the metabolic network—glucokinase and the phosphotransferase system. Interestingly, the *E. coli* model prefers glucokinase to the more common phosphotransferase system for glucose entry during optimal asparagine production. Although glucokinase is known to play a minor role in glucose metabolism under normal conditions, replacement of the phosphotransferase system by this reaction increases asparginine production from 1.560 mol/mol glucose to 1.818 mol/mol glucose. Glucose entry via the phosphotransferase system requires substantial flux through phosphoenolpyruvate (PEP) syn-

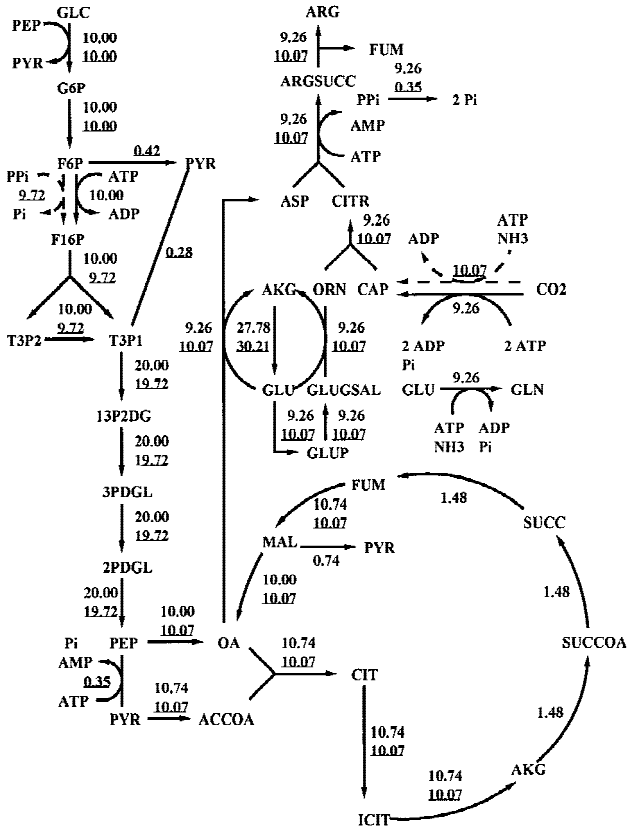

Figure 2. Optimal *E. coli* and Universal* arginine production pathways for growth on glucose. The utilization of carbamate kinase and the pyrophosphate-dependent analog of 6-phosphofructokinase by the Universal arginine production pathway preserves a net of three ATP phosphoanhydride bonds. *Values in figure underlined.

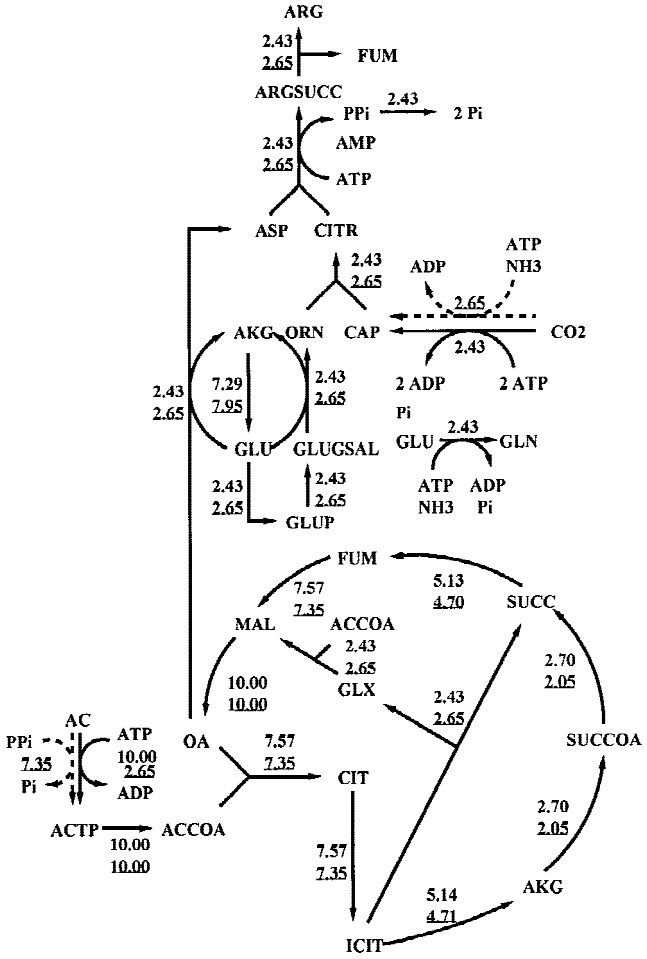

Figure 3. Optimal *E. coli* and Universal* arginine production pathways for growth on acetate. The incorporation of carbamate kinase and the pyrophosphate-dependent analog of acetate kinase by the Universal pathway saves three ATP phosphoanhydride bonds. *Values in figure underlined.

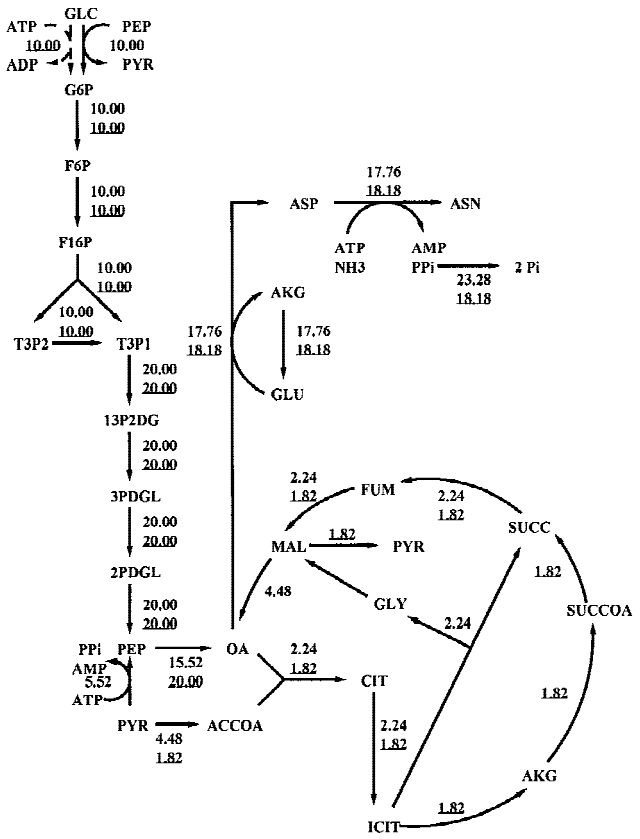

Figure 4. Optimal asparagine production pathway for two modes of glucose utilization: glucokinase* and the phosphotransferase system. *Values in figure underlined.

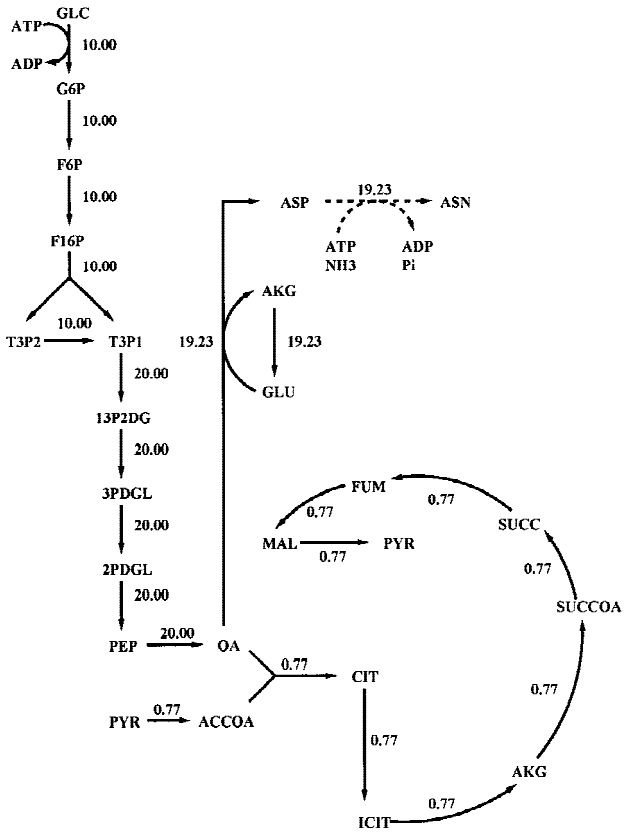

Figure 5. Optimal Universal asparagine production pathway for growth on glucose. The Universal pathway conserves the equivalent of one ATP bond by using an ADP-forming aspartate-ammonia ligase instead of an AMP-forming version as in Figure 4.

thase to regenerate PEP from pyruvate carrying the net expense of one ADP phosphoanhydride bond. Thus, either overexpressing glucokinase in *E. coli* or adding a more active recombinant glucokinase enzyme may improve asparagine production. Figure 5 illustrates the optimal Universal route for asparagine production on glucose. By choosing the ADP-forming aspartate-ammonia ligase enzyme over the AMP-forming version present in *E. coli*, the energy efficiency of this pathway is improved. Presently, no pathways for the conservation of the pyrophosphate bond energy have been identified in *E. coli*; thus, the formation of AMP uses the equivalent of two ATP phosphoanhydride bonds. In contrast, by forming ADP, the Universal pathway requires the breakage of only one phosphoanhydride bond per unit flux. In fact, the energy efficiency of the Universal model is such that the formation of asparagine does not require ATP formation from the transmembrane proton gradient. This gradient is used solely to transport inorganic phosphate into the cell. This mechanism improves asparagine production 5.77% for growth on glucose and 5.45% for growth on acetate.

The optimal histidine production pathways of the *E. coli* and Universal models for growth on acetate are shown in Figure 6. Again, the Universal model selects a reaction to conserve the phosphoanhydride bond energy of pyrophosphate generated in this case by both ATP phosphoribosyltransferase and phosphoribosyl-ATP pyrophosphatase. Thus, the Universal model is at least two ATP more efficient than the *E. coli* model per histidine molecule produced. In addition, the addition of glycine dehydrogenase to the *E. coli* model improves the carbon conversion of the native histidine pathway. Under optimal histidine production conditions in native *E. coli*, intracellular glycine is converted to carbon dioxide and ammonia by the glycine cleavage system. In this process, only one of glycine's carbons is conserved by its transfer to tetrahydrofolate. The Universal model, on the other hand, conserves both carbons by converting glycine to glyoxylate, which subsequently is pumped back into the glyoxylate shunt. Both mechanisms improve the maximum theoretical yield of histidine 4.53%.

CONCLUSIONS

The proposed optimization framework provided the quantitative means to study metabolic network performance limits in response to gene deletions or additions. Metabolic network performance relates to either robustness in the face of gene deletions or flux enhancements through foreign gene recombination from an ever-expanding database of available genes. Although complete gene–enzyme relationships are not currently available, the formulation enables the

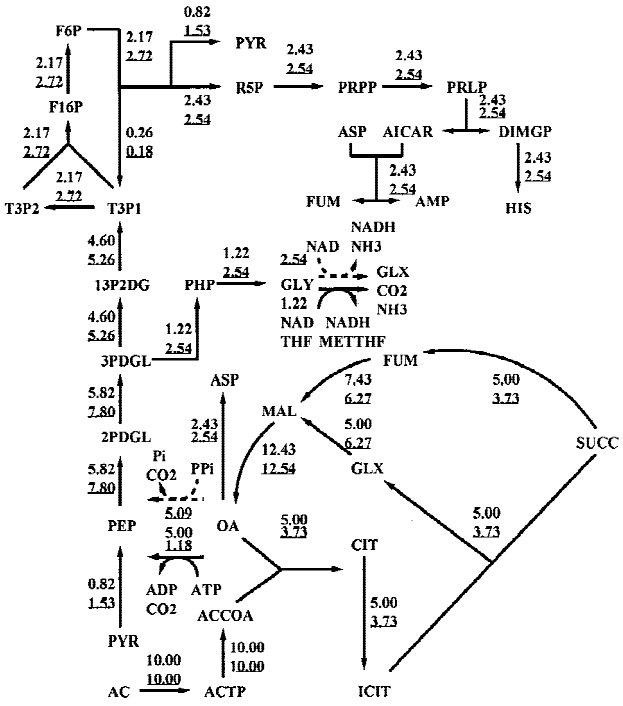

Figure 6. Optimal *E. coli* and Universal* histidine production pathways for growth on acetate. Both the energy efficiency (two ATPs) and carbon conversion efficiency of the Universal pathway are improved by the incorporation of a pyrophosphate-dependent analog of PEP carboxykinase and glycine dehydrogenase, respectively. *Values in figure underlined.

incorporation of this information as it becomes available. The gene knockout analysis revealed that the *E. coli* metabolic network optimized for growth could endure an increasing amount of gene knockouts as its growth demands are lowered. Furthermore, the network could theoretically tolerate at most 18 gene deletions before biomass production is no longer possible. The gene addition studies revealed that adding additional options to the *E. coli* genotype by DNA recombination provided improvements to the maximum theoretical productions of seven amino acids. These improvements occur by one of two mechanisms: 1) by improving energy efficiency, or 2) by increasing the carbon conversion efficiency of the production route.

The reliance of flux balance analysis strictly on stoichiometric characteristics is its greatest strength, but can also be its most prominent weakness. The flux distributions within the cell are ultimately uniquely determined by the regulatory mechanisms within the cell, the kinetic characteristics of cellular enzymes, and the expression of these enzymes. Assuming cells operate in a stoichiometrically optimal fashion yields a wider boundary of metabolic flux distributions than may be available to the cell. Currently, we are incorporating regulatory information into flux balance models with the use of logic constraints. These constraints will ensure that up or down movements in metabolite concentrations are consistent with up or down shifts in reaction flux values. A more tightly constrained model will give additional insight on how overproducing cellular products affects overall metabolic regulation. As the accuracy of metabolic models improves and the amount of information available for FBA grows, the framework introduced in this article can be used to suggest promising gene addition and/or deletion candidate.

We thank Dr. Vassily Hatzimanikatis for useful discussions.

References

Aristidou A, San KY, Bennett GN. 1994. Modification of central metabolic pathway in *Escherichia coli* to reduce acetate accumulation by heterologous expression of the *Bacillus subtilis* acetolactate synthase gene. Biotechnol Bioeng 44:944–951.

Blattner FR, et al. 1997. The complete genome sequence of Escherichia coli K-12. Science 277:1453–1474.

Chou CH, Bennett GN, San KY. 1994. Effect of modified glucose uptake using genetic engineering techniques on high-level recombinant protein production in *Escherichia coli* dense cultures. Biotechnol Bioeng 44:952–960.

Dedhia NN, Hottinger T, Bailey JE. 1994. Overproduction of glycogen in *Escherichia coli* blocked in the acetate pathway improves cell growth. Biotechnol Bioeng 44:132–139.

Delgado J, Liao JC. 1997. Inverse flux analysis for reduction of acetate excretion in *Escherichia coli*. Biotechnol Progr 13:361–367.

Edwards JS, Palsson BO. 2000a. The *Escherichia coli* MG1655 *in silico* metabolic genotype: Its definition, characteristics, and capabilities. P. Natl. Acad. Sci. USA 97:5528–5533.

Edwards JS, Palsson, BO. 2000b. Metabolic flux balance analysis and the *in silico* analysis of *Escherichia coli* K-12 gene deletions. BMC Bioinformatics 1:1.

Edwards JS, Ramakrishna R, Schilling CH, Palsson BO. 1999. Metabolic flux balance analysis. In: Lee SY, Papoutsakis ET, editors. Metabolic engineering. New York: Marcel Dekker. p 13–58.

Ensley BD. 1985. Microbial production of indigo. US Patent 4,520,103.

Hatzimanikatis V, Bailey JE. 1997. Effects of spatiotemporal variations on metabolic control: approximate analysis using (log)linear kinetic models. Biotechnol Bioeng 54:91–104.

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

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

Hatzimanikatis V, Emmerling M, Sauer U, Bailey, JE. 1998. Application of mathematical tools for metabolic design of microbial ethanol production. Biotechnol Bioeng 58:154–161.

Heinrich R, Rapoport TA. 1974. A linear steady-state treatment of enzymatic chains. Eur J Biochem 42:89–95.

Henriksen CM, Christensen LH, Nielsen J, Villadsen J. 1996. Growth energetics and metabolic fluxes in continuous cultures of *Penicillium chrysogenum*. J Biotechnol 45:149–164.

Hutchison CA, et al. 1999. Global transposon mutagenesis and a minimal mycoplasma genome. Science 286:2165–2169.

Jorgensen H, Nielsen J, Villadsen J. 1995. Metabolic flux distributions in *Penicillium chrysogenum* during fed-batch cultivations. Biotechnol Bioeng 46:117–131.

Kacser H, Burns JA. 1973. The control of flux. Symp Soc Exp Biol 27: 65–104.

Karp PD, Riley M, Paley SM, Pelligrinitoole A, Krummenacker M. 1999. EcoCyc: encyclopedia of *Escherichia coli* genes and metabolism. Nucleic Acids Res 27:55.

Keasling JD, Van Dien SJ, Pramanik J. 1998. Engineering polyphosphate metabolism in *Escherichia coli:* implications for bioremediation of inorganic contaminants. Biotechnol Bioeng 58:231–239.

Lee S, Phalakornkule C, Domach MM, Grossman IE. 2000. Recursive

MILP model for finding all the alternate optima in LP models for metabolic networks. Comput Chem Eng 24:711–716.
Majewski RA, Domach MM. 1990. Simple constrained optimization view of acetate overflow in *Escherichia coli*. Biotechnol Bioeng 35: 732–738.
Mushegian AR, Koonin EV. 1996. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. P Natl Acad Sci USA 93:10268–10273.
Papoutsakis E, Meyer C. 1985a. Equations and calculations of product yields and preferred pathways for butanediol and mixed-acid fermentations. Biotechnol Bioeng 27:50–66.
Papoutsakis E, Meyer C. 1985b. Fermentation equations for propionic acid bacteria and production of assorted oxychemicals from various sugars. Biotechnol Bioeng 27:67–80.
Pennisi E. 1997. Laboratory workhorse decoded. Science 277:1432–1434.
Pons A, Dussap C, Pequignot C, et al. 1996. Metabolic flux distribution in *Cornybacterium melassecola* ATCC 17965 for various carbon sources. Biotechnol Bioeng 51:177–189.
Pramanik J, Keasling JD. 1997. Stoichiometric model of *Escherichia coli* metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. Biotechnol Bioeng 56: 398–421.
Ramakrishna R, Edwards JS, McCulloch A, Palsson BO. 2001 Flux-balar analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints. Am J Physiol Regulatory Integrative Comp Physiol 280:R695–R704.
Regan L, Bogle IDL, Dunnill P. 1993. Simulation and optimization of metabolic pathways. Comput Chem Eng 17:627–637.
Sabatie J, et al. 1991. Biotin formation by recombinant strains of *Escherichia coli*: influence of the host physiology. J Biotechnol 20:29–50.
Savageau MA. 1969a. Biochemical systems analysis. I. Some mathematical properties of the arte law for the component enzymatic reactions. J Theor Biol 25:365–369.
Savageau MA. 1969b. Biochemical systems analysis. II. The steady state solutions for an n-pool system using a power-law approximation. J Theor Biol 25:370–379.
Savageau MA. 1970. Biochemical systems analysis. III. Dynamic solutions using a power-law approximation. J Theor Biol 26:215–226.
Schilling CH, Edwards JS, Palsson BO. 1999. Toward metabolic phenomics: analysis of genomic data using flux balances. Biotechnol Progr 15:288–295.
Torres NV, Voit EO, Gonzales-Alcon C. 1996. Optimization of nonlinear biotechnological processes with linear programming: application to citric acid production by *Aspergillus niger*. Biotechnol Bioeng 49: 247–258.
Vallino J, Stephanopoulos G. 1993. Metabolic flux distributions in *Corynebacterium glutamicum* during growth and lysine overproduction. Biotechnol Bioeng 41:633–646.
Vallino, JJ, Stephanopoulos, G. 1994a. Carbon flux distribution at the glucose 6-phosphate branch point in *Corynebacterium glutamicum* during lysine overproduction. Biotechnol Progr 10:327–334.
Vallino, JJ, Stephanopoulos, G. 1994b. Carbon flux distribution at the pyruvate branch point in *Corynebacterium glutamicum* during lysine overproduction. Biotechnol Progr 10:320–326.
Van Dien SJ, Keasling JD. 1998. Optimization of polyphosphate degradation and phosphate secretion using hybrid metabolic pathways and engineering host strains. Biotechnol Bioeng 59:754–761.
Varma A, Palsson BO. 1993. Metabolic capabilities of *Escherichia coli*. II. Optimal growth patterns. J Theor Biol 165:503–522.
Varma A, Palsson BO. 1994. Metabolic flux balancing: basic concepts, scientific and practical use. Bio/Technology 12:994–998.
Varma A, Boesch BW, Palsson BO. 1993a. Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. Appl Environ Microb 59:2465–2473.
Varma A, Boesch BW, Palsson, BO. 1993b. Biochemical production capabilities of *Escherichia coli*. Biotechnol Bioeng 42:59–73.
Voit EO. 1992. Optimization of integrated biochemical systems. Biotechnol Bioeng 40:572–582.
Wang CL, et al. 1997. Cadmium removal by a new strain Pseudomonas aeruginosa in aerobic culture. Appl Environ Microb 63:4075–4078.
Winter RB, Yen KM, Ensley BD. 1989. Efficient degradation of trichloroethylene by a recombinant *Escherichia coli*. Bio/Technology 7: 282–285.
Xie L, Wang D. 1994a. Applications if improved stoichiometric model in medium design and fed-batch cultivation of animal cells in bioreactor. Cytotechnology 15:17–29.
Xie L, Wang D. 1994b. Stoichiometric analysis of animal cell growth and its application in medium design. Biotechnol Bioeng 43:1164–1174.
Xie L, Wang D. 1996a. Material balance studies on animal cell metabolism using stoichiometrically based reaction network. Biotechnol Bioeng 52:579–590.
Xie L, Wang D. 1996b. Energy metabolism and ATP balance in animal cell cultivation using a stoichiometrically based reaction network. Biotechnol Bioeng 52:591–601.
Xie L, Wang D. 1996c. High cell density and high monoclonal antibody production through medium design and rational control in a bioreactor. Biotechnol Bioeng 51:725–729.
Xie L, Wang D. 1997. Integrated approaches to the design of media and feeding strategies for fed-batch cultures of animal cells. Trends Biotechnol 15:109–113.