Research article

Metabolic flux balance analysis and the in silico analysis of Escherichia coli K-12 gene deletions

Jeremy S Edwards[1,2] and Bernhard O Palsson[1]

Address: [1]Department of Bioengineering, University of California, San Diego La Jolla, CA 92093-0412, USA. and [2]Harvard Medical School Department of Genetics Warren Alpert Rm. 514 200 Longwood Ave. Boston, MA02115, USA.
E-mail: [1 2] jedwards@arep.med.harvard.edu [1] palsson@ucsd.edu

Published: 27 July 2000

BMC Bioinformatics 2000, **1**:1

Received: 11 July 2000
Accepted: 27 July 2000

The electronic version of this article is the complete one and can be found online at http://biomedcentral.com/1471-2105/1/1

Abstract

Background:: Genome sequencing and bioinformatics are producing detailed lists of the molecular components contained in many prokaryotic organisms. From this 'parts catalogue' of a microbial cell, *in silico* representations of integrated metabolic functions can be constructed and analyzed using flux balance analysis (FBA). FBA is particularly well-suited to study metabolic networks based on genomic, biochemical, and strain specific information.

Results:: Herein, we have utilized FBA to interpret and analyze the metabolic capabilities of *Escherichia coli*. We have computationally mapped the metabolic capabilities of *E. coli* using FBA and examined the optimal utilization of the *E. coli* metabolic pathways as a function of environmental variables. We have used an *in silico* analysis to identify seven gene products of central metabolism (glycolysis, pentose phosphate pathway, TCA cycle, electron transport system) essential for aerobic growth of *E. coli* on glucose minimal media, and 15 gene products essential for anaerobic growth on glucose minimal media. The *in silico tpi* $^{-}$, *zwf*, and *pta* $^{-}$ mutant strains were examined in more detail by mapping the capabilities of these *in silico* isogenic strains.

Conclusions:: We found that computational models of *E. coli* metabolism based on physicochemical constraints can be used to interpret mutant behavior. These *in silica* results lead to a further understanding of the complex genotype-phenotype relation.

Supplementary information: [http://gcrg.ucsd.edu/supplementary_data/DeletionAnalysis/main.htm]

Introduction

The knowledge of a complete genome sequence holds the potential to reveal the 'blueprints' for cellular life. The genome sequence contains the information to propagate the living system, and this information exists as open reading frames (ORFs) and regulatory information. Computational approaches have been developed (and are continuously being improved) to decipher the information encoded in the DNA [1,2,3,4,5,6,7]. However, it is becoming evident that cellular functions are intricate and the integrated function of biological systems involves many complex interactions among the molecular components within the cell. To understand the complexity inherent in cellular networks, approaches that focus on the systemic properties of the network are also required.

The complexity of integrated cellular systems leads to an important point, namely that the properties of complex biological processes cannot be analyzed or predicted based solely on a description of the individual components, and integrated systems based approaches must be

applied [8]. The focus of such research represents a departure from the classical *reductionist approach* in the biological sciences, and moves toward the *integrated approach* to understanding the interrelatedness of gene function and the role of each gene in the context of multi genetic cellular functions or *genetic circuits* [8,9,10].

The engineering approach to analysis and design of complex systems is to have a mathematical or computer model; e.g. a dynamic simulator of a cellular process that is based on fundamental physicochemical laws and principles. Herein, we will analyze the integrated function of the metabolic pathways, and there has been a long history of mathematical modeling of metabolic networks in cellular systems, which dates back to the 1960s [11, 12]. While the ultimate goal is the development of dynamic models for the complete simulation of cellular metabolism, the success of such approaches has been severely hampered by the lack of kinetic information on the dynamics and regulation of metabolism. However, in the absence of kinetic information it is still possible to assess the theoretical capabilities and operative modes of metabolism using flux balance analysis (FBA) [10, 13,14,15,16,17].

We have developed an *in silico* representation of *Escherichia coli (E. coli in silico)* to describe the bacterium's metabolic capabilities [18]. *E. coli in silico* was derived based on the annotated genetic sequence [19], biochemical literature [20], and the online bioinformatic databases [21,22,23]. The properties of *E. coli in silico* were analyzed and compared to the *in vivo* properties of *E. coli,* and it was shown that *E. coli in silico* can be used to interpret the metabolic phenotype of many *E. coli* mutants [18]. However, the utilization of the metabolic genes is dependent on the carbon source and the substrate availability [24, 25]. Thus, the mutant phenotype is also dependent on specific environmental parameters. Therefore, herein we have utilized *E. coli in silico* to computationally examine the condition dependent optimal metabolic pathway utilization, and we will show that the FBA can be used to analyze and interpret the metabolic behavior of wildtype and mutant *E. coli* strains.

Materials and Methods

Flux balance analysis

All biological processes are subjected to physico chemical constraints (such as mass balance, osmotic pressure, electro neutrality, thermodynamic, and other constraints). As a result of decades of metabolic research and the recent genome sequencing projects, the mass balance constraints on cellular metabolism can be assigned on a genome scale for a number of organisms. Methods have been developed to analyze the metabolic capabilities of a cellular system based on the mass balance constraints and this approach is known as flux balance analysis (FBA) [13, 14, 16] (see the supplementary information for an FBA primer). The mass balance constraints in a metabolic network can be represented mathematically by a matrix equation:

$$S \bullet v = 0 \quad \text{Equation 1}$$

The matrix **S** is the *mxn* stoichiometric matrix, where *m* is the number of metabolites and *n* is the number of reactions in the network (The *E. coli* stoichiometric matrix is available in matrix format in the supplementary information and in a reaction list in Appendices 1-3). The vector v represents all fluxes in the metabolic network, including the internal fluxes, transport fluxes and the growth flux.

For the *E. coli* metabolic network represented by Eqn. 1, the number of fluxes was greater than the number of mass balance constraints; thus, there were multiple feasible flux distributions that satisfied the mass balance constraints, and the solutions (or feasible metabolic flux distributions) were confined to the nullspace of the matrix **S**.

In addition to the mass balance constraints, we imposed constraints on the magnitude of individual metabolic fluxes.

$$\alpha_i \le v_i \le \beta_i \quad \text{Equation 2}$$

The linear inequality constraints were used to enforce the reversibility of each metabolic reaction and the maximal flux in the transport reactions. The reversibility constraints for each reaction are indicated online. The transport flux for inorganic phosphate, ammonia, carbon dioxide, sulfate, potassium, and sodium was unrestrained ($\alpha_i = -\infty$ and $\beta_i = \infty$). The transport flux for the other metabolites, when available in the *in silico* medium, was constrained between zero and the maximal level ($0 \le v_i \le v_i^{max}$). The v_i^{max} values used in the simulations are noted for each simulation (Fig. 1). When a metabolite was not available in the medium, the transport flux was constrained to zero. The transport flux for metabolites capable of leaving the metabolic network (i.e. acetate, ethanol, lactate, succinate, formate, and pyruvate) was always unconstrained in the net outward direction.

Figure 1

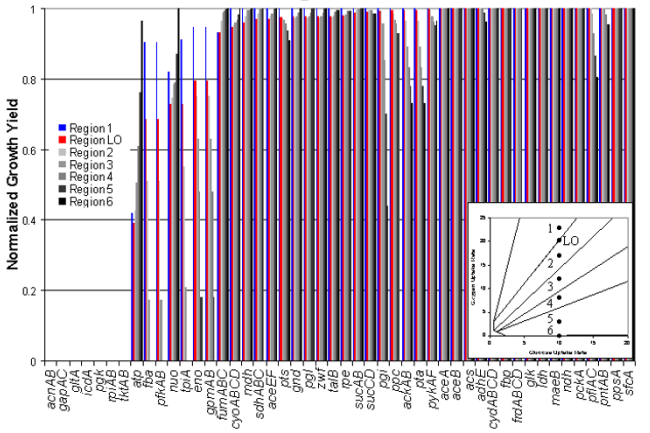

Figure 1
Growth phenotypes of *in silico* deletion strains; maximal biomass yields on glucose for all possible single gene deletions in central intermediary metabolism. The environmental variables (uptake rate/external metabolic fluxes) are set to correspond to a point within each of the phases of the wild-type PhPP (figure inset). The maximal yields were calculated using flux-balance analysis with the objective of maximizing the growth flux. The biomass yields are normalized with respect to the results for the full metabolic genotype. The α and β value for the constraints on the external fluxes for glucose and oxygen uptake are defined as follows (units- mmole g^{-1} hr^{-1}): Phase 1 - v_{glc} = 10, v_{oxy} = 23; LO - v_{gic} = 10, v_{oxy} = 20.3; Phase 2 - v_{glc} = 10, v_{oxy} = 17; Phase 3 - v_{glc} = 10, v_{oxy} = 12; Phase 4 - v_{glc} = 10, v_{oxy} = 8; Phase 5 - v_{glc} = 10, v_{oxy} = 3; Phase 6 - v_{glc} = 10, v_{oxy} = 0.

The intersection of the nullspace and the region defined by the linear inequalities defined a region in flux space that we will refer to as the feasible set, and the feasible set defined the capabilities of the metabolic network subject to the imposed cellular constraints. It should be noted that every vector **v** within the feasible set is not reachable by the cell under a given condition due to other constraints not considered in the analysis (i.e. maximal internal fluxes and gene regulation). The feasible set can be further reduced by imposing additional constraints (i.e. kinetic or gene regulatory constraints), and in the limiting condition where all constraints are known, the feasible set may reduce to a single point.

A particular metabolic flux distribution within the feasible set (vector **v** which satisfies the constraints in Eqns. 1 and 2) was found using linear programming (LP). A commercially available LP package was used (LINDO, Lindo Systems Inc., Chicago, II). LP identified a solution that minimized a metabolic objective function (subject to the imposed constraints- Eqns. 1 and 2) [16, 48, 49], and was formulated as shown below:

Minimize -Z

where $Z = \Sigma\, c_i\, v_i = \langle c \bullet v \rangle$ Equation 3

The vector **c** was used to select a linear combination of metabolic fluxes to include in the objective function [50]. Herein, **c** was defined as the unit vector in the direction of the growth flux, and the growth flux was defined in terms of the biosynthetic requirements:

$$\sum_{all\,m} d_m \cdot X_m \xrightarrow{v_{growth}} Biomass$$

(Equation 4)

where d_m is the biomass composition of metabolite X_m (we used a constant biomass composition defined from the literature [51] (see Appendix 4)), and the growth flux was modeled as a single reaction that converts all the biosynthetic precursors into biomass.

Phenotype Phase Plane Analysis

All feasible *E. coli in silico* metabolic flux distributions are mathematically confined to the feasible set, which is a region in flux space (^{n}), where each solution in this space corresponds to a feasible metabolic flux distribution.

Phenotype Phase Plane (PhPP): A PhPP is a two-dimensional projection of the feasible set, and below we will briefly discuss the formalism for constructing the PhPP. Two parameters that describe the growth conditions (such as substrate and oxygen uptake rates) were defined as the two axes of the two dimensional space. The optimal flux distribution was calculated (using LP) for all points in this plane by solving the LP problem while adjusting the exchange flux constraints (defining the two-dimensional space). A finite number of qualitatively different patterns of metabolic pathway utilization were identified in such a plane, and lines were drawn to demarcate these regions. Each region is denoted by $Pn_{x,y}$, where 'P' indicates that the region was defined by a phenotype phase plane analysis, 'n' denotes the number of the demarcated phase (as shown in a particular PhPP figure), and 'x, y' denotes the two uptake rates on the axis of the PhPP. PhPPs were also generated for mutant genotypes; represented as $P^{gene}n_{x,y}$.

One demarcation line in the PhPP was defined as the line of optimality (LO). The LO represents the optimal relation between exchange fluxes defined on the axes of the PhPP.

Alterations of the genotype

FBA and *E. coli in silico* were used to examine the systemic effects of in *silico* gene deletions. The genes involved in the central metabolic pathways (glycolysis,

pentose phosphate pathway, TCA cycle, electron transport) were subjected to removal from *E. coli in silico*. To simulate a gene deletion, all metabolic reactions catalyzed by a given gene product were simultaneously constrained to zero. Some metabolic reactions were catalyzed by more than one enzyme, and all genes that code for enzymes that catalyze a given reaction were simultaneously removed (i.e. *rpiAB*). Furthermore, all genes that make up an enzyme complex were also simultaneously removed (i.e. *sdhABCD*).

The optimal metabolic flux distribution for the generation of biomass was calculated for each *in silico* deletion strain. The *in silico* gene deletion analysis was performed with the transport flux constraints defined by the wild-type PhPP. The constraints imposed for each simulation are noted in Fig. 1.

For each *in silico* deletion strain, the optimal production of the twelve biosynthetic precursors and the metabolic cofactors was also calculated to identify auxotrophic requirements and impaired functions in the metabolic network (Table 1). The optimal production of the biosynthetic precursors was calculated by setting the objective function to the drain of a single metabolite (i.e., ATP $\rightarrow$ ADP + P_i, or PEP $\rightarrow$). The numerical value of the objective function for each *in silico* deletion strain was reported as a fraction of the wild-type optimal value (Table 1).

Optimal production of the twelve biosynthetic precursors and the metabolic cofactors.

Compound	μmol/g DW	Compound	μmol/g DW
Amino Acids		Phospholipids	
Alanine	488	Phosphatidyl serine	2.58
Arginine	281	Phosphatidyl ethanolamine	96.75
Asparagine	229	Phosphatidyl glycerol	23.22
Aspartate	229	Cardiolypin	6.45
Cysteine	87	Fatty acid composition (% of total fatty acid)	
Glutamate	250	Myristic acid (2.68)	
Glutamine	250	Myristoleic acid (7.70)	
Glycine	582	Palmitic acid (38.23)	
Histidine	90	Palmitoleic acid (10.74)	
Isoleucine	276	Heptadecenoic acid (16.11)	
Leucine	428	cis-Vaccenic acid (0.90)	
Lysine	326	Oleic acid (17.91)	
Methionine	146	Nonadecenoic acid (5.73)	
Phenylalanine	176	Cell Wall Sructures	
Proline	210	Lipopolysaccharide	8.4
Serine	205	Peptidoglycan	27
Threonine	241	Cofactors and other molecules	
Tryptophan	54	5-Methyl-THF	50.0
Tyrosine	131	Putrescine	35.0

Optimal production of the twelve biosynthetic precursors and the metabolic cofactors.

Valine	402	Spermidine	7.0
Protein synthesis/processing (ATP/ Amino Acid)	4.306	NAD	2.15
Ribonucleotides		NADH	0.05
ATP	165	NADP	0.13
GTP	203	NADPH	0.4
CTP	126	UDP-Glucose	3.0
UTP	136	ATP	4.0
RNA synthesis/processing (ATP/Nucleotide)	0.40	ADP	2.0
Deoxyribonucleotides		AMP	1.0
dATP	24.7	CoA	0.03
dTTP	24.7	Acetyl-CoA	0.04
dGTP	25.4	Succinyl-CoA	0.01
dCTP	25.4	Glycogen	154
DNA synthesis/processing (ATP/Nucleotide)	1.372		

The optimal production was calculated for the wild-type strain and the deletion strains. The constraints are set as defined in Figure 1. The color code quantitatively defines effect of the in silico deletion; red corresponds to 0.0, yellow corresponds to 0-50% of the wild-type production, blue corresponds to 50-100% of the wild-type, and no color coding is illustrated when the production is unchanged from the wild-type. Data for other carbon sources is available online. G6P, glucose-6-phosphate; F6P, fructose-6-phosphate; R5P, ribose-5-phosphate; E4P,erythrose 4-phosphate; T3P1, glyceraldehyde 3-phosphate; 3PG, 3-phosphoglycerate; PEP, phosphoenolpyruvate; PYR, pyruvate; ACCOA, acetyl-CoA; AKG, α-ketoglutarate; SUCCOA, succinyl-CoA; OA, oxaloacetate.

Results

We have previously described the construction of phenotype phase planes (PhPPs) (see materials and methods) and the analysis of the glucose-oxygen PhPP. We have previously described the effect of *in silico* 'gene deletions' on the ability of *E. coli in silico* to 'grow' under a single condition [18]. Since the utilization of the metabolic pathways is condition dependent, herein, we have investigated the link between the environmental conditions and the optimal metabolic pathway utilization *in silico* by: 1. studying the effects of gene deletions in all phases of the glucose-oxygen PhPP, and 2. broadening the analysis of *in silico* deletion strains by comparing PhPPs from isogenic *in silico* strains.

Gene Deletions: A point within each phase of the glucose-oxygen PhPP was chosen to define the transport flux constraints (indicated in Fig. 1) for the FBA simulations. At each point, the growth characteristics of all *in silico* gene deletion strains (of central metabolic pathway genes) were examined. Based on the results, the genes were categorized as; *essential* (growth under the defined condition requires the activity of the corresponding gene product), *critical* (growth at a reduced yield (< 95% of wild-type)), or *non-essential* (growth at near wild-type yield (> 95%)). The effects of the *in silico* gene deletions were phase-dependent, allowing us to identify optimal growth phenotypes for each growth condition. Additionally, the optimal production of the 12 biosynthetic precursors, high-energy phosphate bonds, and redox potential was calculated for each *in silico* deletion strain (Table 1) to determine the specific effect of the gene deletion on the metabolic capabilities. For instance, the *in silico acnAB'* strain was unable to synthesize α-ketoglutarate under all simulated growth conditions, and thus, *acnAB* was defined as essential for growth in a glucose minimal media (Table 1).

The optimal utilization of the metabolic pathways was dependent on the specific transport flux constraints, and the qualitative shifts in optimal metabolic behavior as a function of two transport fluxes are shown in the PhPP. The optimal biomass yield and biosynthetic precursor production capabilities were calculated for each *E. coli in silico* deletion strain for a point within each region of the PhPP, and the optimal values were normalized to the wild-type (Fig. 1). The condition dependent metabolic phenotypes were computationally analyzed, and the results are organized by the overall metabolic phenotype; *essential, conditionally essential,* or *non-essential genes*.

Essential genes: The gene products that were essential for growth with conditions defined by the line of optimality (LO) (see materials and methods) were also identified as essential within all other phases (*acnAB, gapAC, gltA, icdA, pgk, rpiAB, tktAB*). Specifically, the *gltA* $^-$, *icdA* $^-$, and *acnAB* $^-$ *in silico* deletion strains were unable to produce one biosynthetic precursor (α-ketoglutarate, Table 1), and retained the capability to synthesize the remaining biosynthetic precursors and cofactors nearly equivalent to the wild-type. This prediction is consistent with the defined media required for the cultivation of *aglt* $^-$ *E. coli* mutant strain (glucose minimal media supplemented with glutamine or proline) [26]. Furthermore, the essential glycolytic gene products (*pgk, gapAC*) were required for the synthesis of oxaloacetate, succinyl-CoA, α-ketoglutarate, pyruvate, phosphoenolpyruvate (PEP), and 3-phosphoglycerate within all conditions, and were unable to synthesize all biosynthetic precursors under anaerobic growth conditions. The remaining two essential gene products were in the pentosephosphate pathway (*tktAB, rpiAB*). The *tktAB* and *rpiAB* gene products were required for the synthesis of erythrose 4-phosphate in all phases (aromatic amino acid supplement required for the cultivation of *tkt* $^-$ *E. coli* mutant strains [27]). Additionally, *rpiAB* $^-$ strains were identified as ribose auxotrophs by the *in silico* analysis, which was consistent with experimental data [28].

Conditionally essential genes: During the growth simulations with external parameters defined by the LO, there were genes defined as critical for growth; however, many of these genes were essential for cellular growth upon oxygen limitations (*fba, pfkAB, tpiA, eno, gpmAB*). These genes were termed conditionally essential. The *fba* $^-$, *pfkAB* $^-$, and *tpiA* $^-$ *in silico* deletion strains had a limited capability to synthesize glyceraldehyde 3-phosphate, 3-phosphoglycerate, phosphoenolpyruvate, pyruvate, acetyl-CoA, α-ketoglutarate, succinyl-CoA, oxaloacetate, and high-energy phosphate bonds in all phases, and were completely unable to synthesize many of the biosynthetic precursors in phases 46 (Table 1) (*tpi* $^-$ *in silico* strain discussed below). The growth potential of the *eno* $^-$ and *gpmAB* $^-$ *in silico* deletion strains was theoretically maintained under aerobic conditions by the synthesis and degradation of serine, and without the serine degradation pathway, the *eno* $^-$ and *gpmAB* $^-$ gene products were defined as essential. However, the *eno* $^-$ and *gpmAB* $^-$ *in silico* deletion strains were limited in their production capability of high-energy phosphate bonds under all conditions, and were unable to produce any of the biosynthetic precursors in phase 6 even with the serine degradation pathway.

Additionally, several LO non-essential gene products were essential (*sdhABCD, ppc, frdABCD*) for growth within other phases. The *in silico* analysis suggested that the *sdhABCD* and *frdABCD* gene products were required for anaerobic pyrimidine biosynthesis. Additionally, the *frdABCD* gene products were essential for the anaerobic synthesis of the NAD cofactor. However, these *in silico* results could be due to inaccurate stoichiometric information with respect to cofactor utilization and should be critically examined. Finally, the *ppc* gene product was required for the anaerobic synthesis of oxaloacetate and α-ketoglutarate, but the *in silico* analysis suggests that this gene product is not essential for growth in aerobic conditions where the glyoxylate by-pass has the potential to replenish the biosynthetic precursors [29].

Non-essential genes: Several genes that are critical for growth in conditions defined by the LO were non-essential for growth in other phases (*nuo*, *cyoABCD, fumABC*). The *in silico nuo* $^-$ and *cyoABCD* $^-$ deletion strains were limited in their production capabilities of high-energy phosphate bonds for aerobic growth; however, under anaerobic conditions high-energy phosphate bonds were produced by substrate level phosphorylation. The production capabilities of the *fumABC* $^-$ *in silico* deletion strain was not limited with respect to the biosynthetic precursors shown in the table (other than a slight limitation of ATP production in $P1_{glucose, oxygen}$). However, the *fumABC* $^-$ *in silico* deletion strain was limited in its production capabilities of several amino acids (arg, gly, his-not shown in table), but under anaerobic conditions, these capabilities were not limited with respect to the wild-type.

Several LO non-essential gene products were critical (*pgi, pta, ackAB*) for growth within other phases. The *in silico pgi* deletion strain had a reduced capacity to produce all the biosynthetic precursors under oxygen limitation, and this resulted in a decreased normalized growth yield of this *in silico* deletion strain. The *pta* and *ackAB* gene products participate in the metabolic pathway leading to the formation of acetate. Acetate was predicted as a metabolic by-product upon oxygen limitations (all phases below the LO). Under conditions defined by $P5\text{-}6_{glucose, oxygen}$, the production capabilities of several of the biosynthetic precursors (glucose 6-phosphate, fructose 6-phosphate, ribose 5-phosphate, erythrose 5-phosphate, glyceraldehyde 3-phosphate) were limited in the *pta* and *ackAB in silico* deletion strains (*pta* $^-$ *In silico* deletion strain discussed below).

This sub-section illustrated the condition-dependent effect of gene deletions on the metabolic genotype-phenotype relation. The results covered the range of substrate uptake rates and defined the optimal metabolic pathway utilization of isogenic strains *in silico* under different combinations of environmental parameters. The optimal utilization of the metabolic pathways was dependent on the metabolic genotype; thus, different metabolic geno-

types are characterized by different PhPPs. The results presented above provide insight into the genotype phenotype relation. Next, we will compare the PhPPs from *in silico* deletion strains to the wild-type to provide a more complete definition of optimal phenotypes.

in silico Deletion Strain Phenotype Phase Plane Analysis: Comparative analysis of the phase planes for several mutant strains (*tpi* ⁻ , *pta* ⁻ , and *zwf*) were performed. These case studies were chosen to further investigate the metabolic genotype-phenotype relation *in silico* and to demonstrate the use of FBA to interpret and analyze cellular metabolism.

tpi: The *tpi* ⁻ PhPP showed 3 distinct optimal metabolic phenotypes- one glucose limited phase ($P^{tpi}2_{glucose,\ oxygen}$), and two futile phases (Fig. 2). Futile phases are characterized by a negative effect of one of the substrates on the objective function. One of the futile phases was due to excess oxygen ($P^{tpi}1_{glucose,\ oxygen}$) and the other was due to excess glucose ($P^{tpi}3_{glucose,\ oxygen}$). Although the *tpi* ⁻ *in silico* metabolic genotype theoretically supported biomass production, the feasible steady states were restricted to a limited phase of the phase plane and the flexibility of the metabolic network was reduced to one dimension.

Figure 2A

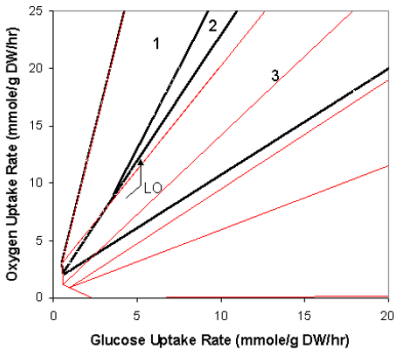

Figure 2
Mutant strain phenotype phase plane analysis. The wild-type strain PhPP is shown in the thin lines for comparative purposes. The glucose-oxygen PhPP was constructed for each of the respective *in silico* deletion strains (A) *tpi* (B) *zwf (C) pta.*

The optimal utilization of *the tpi* ⁻ metabolic network under environmental conditions defined by the LO^{tpi} was characterized by increased PPP fluxes to bypass the TPI block. The PPP operated cyclically; thus, leading to a high production of NADPH. Due to the high NADPH production in the PPP, the TCA cycle flux was optimally reduced and functioned only to produce the biosynthetic precursors.

The *in silico* analysis suggests that the *tpi* ⁻ metabolic network was restricted by the ability to regenerate phosphoenolpyruvate (PE) for the PTS, and the *in silico* analysis identified 3 metabolic 'routes' for the regeneration of PEP. Two of the 'routes' were equivalent (alternate optimal solutions), (1) The PEP was regenerated by the phosphoenolpyruvate synthase (PPS), or (2) the glactose transporter was used for the transport of glucose which was subsequently phosphorylated by the glucokinase reaction. These two routes were equivalent with respect to the objective function (although they were structurally different). The third PEP regeneration route involved the glyoxylate bypass and the phosphoenolpyruvate carboxykinase, and this route was characterized by a 38% reduction in the optimal biomass yield. Furthermore, experimentally it was shown that constitutive expression of the glyoxylate bypass suppressed the PEP deficient phenotype [30, 31]. The PEP regeneration routes (discussed above) theoretically allow the *tpi* ⁻ to grow, and one of these solutions was required for the growth of the *tpi* ⁻ *in silico* strain.

zwf: *zwf* codes for glucose-6-phosphate dehydrogenase (G6PDH), the first enzyme in the oxidative branch of the PPP. *zwf* has been shown to be a non-essential gene for the growth of *E. coli* in glucose minimal media, and *zwf* strains grow at near wild-type growth rates [32]. *zwf* was predicted by FBA to be a non-essential gene for growth in glucose minimal media (Fig. 1). We conducted a phenotype phase plane analysis of the *zwf* strain and examined the systemic metabolic function of *zwf* and its relation to the environmental conditions *in silico* (Fig. 2). The slope of the LO^{zwf} slightly increased (relative to the wild-type), indicating a higher oxygen:glucose ratio for optimal growth. Removing the G6PDH from the metabolic network eliminated all metabolic pathways that utilized the oxidative branch of the PPP. Therefore, the *zwf* PhPP was significantly changed in the phases that utilized the oxidative branch of the PPP ($P^{zwf}2_{glucose,\ oxygen}$ and $P^{zwf}3_{glucose,\ oxygen}$) but was unchanged in phases that did not optimally utilize the *zwf* gene product ($P^{zwf}4_{glucose,\ oxygen}$).

pta: Acetate excretion is a common characteristic of *E. coli* metabolism and several approaches have been applied to reduce acetate production to improve the productivity of engineering strains [33,34,35]. Acetate production can be interpreted using FBA [36, 37], and we have used a phase plane analysis to quantitatively analyze the conditions for which acetate excretion optimally occurs. Acetate was optimally excreted from the cell within all phases of the glucose-oxygen PhPP below the LO. We have generated the *pta* ⁻ PhPP and analyzed the

metabolic characteristics of the in *silico pta* ¯ strain (Fig. 2). The *pta* ¯ PhPP indicated that this mutant strain maintained the potential to support growth (both aerobically and anaerobically). Experimentally, *the pta* ¯ *E. coli* strain has been shown to grow aerobically and anaerobically on glucose minimal media [38]. The *in silico* analysis predicted that *the pta* ¯ strain optimally shifted the carbon flux from acetate to ethanol in $P^{pta}3$. However, in $P^{pta}4$, the optimal metabolic by-products included lactate, ethanol, and pyruvate, and under completely anaerobic conditions, succinate was also optimally produced as a metabolic byproduct. These metabolic byproducts were qualitatively consistent with experimental observations in the *pta* ¯ strain [38].

Discussion

The rapid development of bioinformatic databases is resulting in extensive information about the molecular composition and function of several single cellular organisms. These genetic and biochemical databases [21, 23, 39] have now been developed to the point where the methods of systems science need to be used to analyze, interpret, and predict the integrated behavior of complex multigeneic biological processes. Herein, we have utilized an *in silico* representation of *E. coli* to study the condition dependent phenotype of *E. coli* and central metabolism gene deletion strains. We have shown that a computational analysis of the metabolic behavior can provide valuable insight into cellular metabolism. The results presented herein address a pressing question in the post-genome era; *how can genome sequence information be used to analyze integrated cellular functions?* Given the central importance of this question, we will discuss the general applicability, limitations, and future prospects for FBA and functional genomics.

The FBA metabolic modeling framework is different than other well-known metabolic modeling approaches. FBA can more accurately be defined as a metabolic constraining approach, this is because FBA defines the 'best' the cell can do, rather than predicting the metabolic behavior. To accomplish this, we have constrained metabolic function based on the most reliable information, the metabolic stoichiometry (the stoichiometry is well known for the vast majority of the metabolic processes). However, FBA does have predictive capabilities when a physiologically meaningful objective function can be defined, and the *E. coli* FBA results, with maximal growth rate as the objective function, have been shown to be consistent with experimental data under nutritionally rich conditions [40]. It should be mentioned that FBA does not directly consider regulation, or the regulatory constraints on the metabolic network, but rather FBA assumes that the regulation is such that metabolic behavior is optimal. This assumption produces results that are generally consistent with experimental data, however, this assumption is only valid for a system that has evolved toward optimality. In mutant strains, the regulation of the metabolic network has not evolved to operate in an optimal fashion. Therefore, the optimal utilization of the mutant metabolic network does not necessarily correspond to the *in vivo* utilization of the metabolic network. Computational analysis of metabolic processes, coupled to an experimental program may provide valuable information regarding the regulatory structure of metabolic networks, and will provide a challenge for future computational studies coupled to highly parallel experimental programs, such as large-scale mutation studies [41].

Currently, about one-third of the *E. coli* open reading frames do not have a functional assignment. Thus, the metabolic network studied here is incomplete and does not account for all the metabolic processes carried out by *E. coli*. However, we have used the biochemical literature to refine the *in silico* metabolic genotype and given the long history of *E. coli* metabolic research [20], a large percentage of the *E. coli* metabolic capabilities have likely been identified. However, when additional metabolic capabilities are discovered [42], the *E. coli* stoichiometric matrix can be updated, leading to an iterative model building process. Furthermore, inconsistencies between the model and experimental data may help point to unidentified metabolic functions. Additionally, the *in silico* analysis can help identify missing or incorrect functional assignments; for example, by identifying sets of metabolic reactions that are not connected to the metabolic network by the mass balance constraints.

The study presented herein is an example of the rapidly growing field of *in silico* biology. It is clear that computer modeling and simulations must be used iteratively with an experimental program to continually improve *in silico* models and to develop systemic understanding of cellular functions. Thus, an *in silico* analysis can be used to define an experimental program. For example, the ability to construct well-defined knockout strains of *E. coli* [43] opens the possibility to critically evaluate the relation between the *in silico* representation of mutant behavior and the *in vivo* metabolic network under well-defined genetic and environmental conditions for strategically chosen genes. This possibility is particularly timely, given the increasing number of genome scale measurements that are now possible, through 2D gels [44, 45] and DNA array technology [46, 47].

Conclusions

Herein, we have utilized an *in silico* representation of *E. coli* to study the condition dependent phenotype of *E. coli* and central metabolism gene deletion strains. We have shown that a computational analysis of the metabolic behavior can provide valuable insight into cellular

metabolism. The present *in silico* study builds on the ability to define metabolic genotypes in bacteria and mathematical methods to analyze the possible and optimal phenotypes that they can express. These capabilities open the possibility to perform *in silico* deletion studies to help sort out the complexities of *E. coli* mutant phenotypes.

Additional material

Supplementary material supplied by authors.
Appendix 1: Metabolic map
Appendix 2: List of E. coli metabolic reactions in the model
Appendix 3: Metabolite abbreviations

References

1. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Research* 1997, **25**:3389-402
2. Haussler D: **Computational genefinding.** *Trends Guide to Bioinformatics* 1998, **Supp, 1998**:12-15
3. Thieffry D, Salgado H, Huerta AM, Collado-Vides J: **Prediction of transcriptional regulatory sites in the complete genome sequence of Escherichia coil K-12.** *Bioinformatics* 1998, **14**:391-400
4. Yada T, Totoki Y, Ishikawa M, Asai K, Nakai K: **Automatic extraction of motifs represented in the hidden Markov model from a number of DNA sequences.** *Bioinformatics* 1998, **14**:317-25
5. Roth FP, Hughes JD, Estep PW, Church GM: **Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation.** *Nature Biotechnology* 1998, **16**:939-45
6. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM: **Systematic determination of genetic network architecture.** *Nat Genet* 1999, **22**:281-5
7. Hughes JD, Estep PW, Tavazoie S, Church GM: **Computational identification of cis-regulatory elements associated with groups of functionally related genes in Saccharomyces cerevisiae.** *J Mol Biol* 2000, **296**:1205-14
8. McAdams HH, Shapiro L: **Circuit simulation of genetic networks.** *Science* 1995, **269**:651-656
9. Palsson BO: **What lies beyond bioinformatics?** *Nature Biotechnology* 1997, **15**:3-4
10. Edwards JS, Palsson BO: **How will bioinformatics influence metabolic engineering?** *Biotechnology and Bioengineering* 1998, **58**:162-169
11. Tyson JJ, Othmer HG: **The dynamics of feedback control circuits in biochemical pathways.** *Progress in Theoretical Biology* 1978, **5**:1-62
12. Goodwin BC: **Oscillatory organization in cells, a dynamic theory of cellular control processes.** *Academic Press, New York* 1963,
13. Bonarius HPJ, Schmid G, Tramper J: **Flux analysis of underdetermined metabolic networks: The quest for the missing constraints.** *Trends in Biotechnology* 1997, **15**:308-314
14. Edwards JS, Ramakrishna R, Schilling CH, Palsson BO: **Metabolic Flux Balance Analysis.** *In: Metabolic Engineering Edited by Lee SY, Papoutsakis ET. pp. 13-57: Marcel Deker;* 1999, :13-57
15. Edwards JS, Palsson BO: **Systems Properties of the Haemophilus influenzae Rd Metabolic Genotype.** *Journal of Biological Chemistry* 1999, **274**:17410-17416
16. Varma A, Palsson BO: **Metabolic Flux Balancing: Basic concepts, Scientific and Practical Use.** *Bio/Technology* 1994, **12**:994-998
17. Sauer U, Cameron DC, Bailey JE: **Metabolic capacity of Bacillus subtilis for the production of purine nucleosides, riboflavin, and folic acid.** *Biotechnology and Bioengineering* 1998, **59**:227-238
18. Edwards JS, Palsson BO: **The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities.** *Proc Natl Acad Sci USA* 2000, **97**:5528-33
19. Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y: **The complete genome sequence of Escherichia coli K-12.** *Science* 1997, **277**:1453-74
20. Neidhardt FC: *Escherichia coli and Salmonella: cellular and molecular biology, 2nd edn. Washington, D.C.: ASM Press;* 1996,
21. Karp PD, Riley M, Saier M, Paulsen IT, Paley SM, Pellegrini-Toole A: **The EcoCyc and MetaCyc databases.** *Nucleic Acids Res* 2000, **28**:56-9
22. Kanehisa M: **A database for post-genome analysis.** *Trends in Genetics* 1997, **13**:375-6
23. Selkov E Jr, Grechkin Y, Mikhailova N, Selkov E: **MPW: the Metabolic Pathways Database.** *Nucleic Acids Research* 1998, **26**:43-5
24. Doelle HW, Hollywood NW: **Transitional steady-state investigations during aerobic-anaerobic transition of glucose utilization by Escherichia coli K-12.** *Eur J Biochem* 1978, **83**:479-84
25. Smith MW, Neidhardt FC: **Proteins induced by aerobiosis in Escherichia coli.** *J Bacteriol* 1983, **154**:344-50
26. Lee J, Goel A, Ataai MM, Domach MM: **Flux adaptations of citrate synthase-deficient Escherichia coli.** *Ann N Y Acad Sci* 1994, **745**:35-50
27. Josephson BL, Fraenkel DG: **Transketolase mutants of Escherichia coli.** *J Bacteriol* 1969, **100**:1289-95
28. Fraenkel DG: **Glycolysis.** *In: Escherichia coli and Salmonella Edited by Neidhardt FC, vol. 1. pp. 189 198. Washington, DC.: ASM Press;* 1996, :189-198
29. Vanderwinkel E, De Vlieghere M: **Physiologie et genetique de l'isocitratase et des malate synthetase chez Escherichia coli.** *Eur J Biochem* 1968, **5**:81-90
30. Vinopal RT, Fraenkel DG: **Phenotypic suppression of phosphofructokinase mutations in Escherichia coli by constitutive expression of the glyoxylate shunt.** *Journal of Bacteriology* 1974, **118**:1090-100
31. Kornberg HL, Smith J: **Role of phosphofructokinase in the utilization of glucose by Escherichia coli.** *Nature* 1970, **227**:44-6
32. Fraenkel DG: **Selection of Escherichia coli mutants lacking glucose-6-phosphate dehydrogenase or gluconate-6-phosphate dehydrogenase.** *J Bacteriol* 1968, **95**:1267-71
33. Weikert C, Sauer U, Bailey JE: **Increased phenylalanine production by growing and nongrowing Escherichia coli strain CWML2.** *Biotechnol Prog* 1998, **14**:420-4
34. Aristidou AA, San K-Y, Bennett GN: **Improvement of biomass yield and recombinant gene expression in Escherichia coli by using fructose as the primary carbon source.** *Biotechnology Progress* 1999, **15**:140-145
35. Farmer WR, Liao JC: **Reduction of aerobic acetate production by Escherichia coli.** *Appl Environ Microbiol* 1997, **63**:3205-10
36. Ko Y-F, Bentley W, Weigand W: **A metabolic model of cellular energetics and carbon flux during aerobic Escherichia coli fermentation.** *Biotechnology and Bioengineering* 1994, **43**:847-855
37. Majewski RA, Domach MM: **Simple constrained optimization view of acetate overflow in E. coli.** *Biotechnology and Bioengineering* 1990, **35**:732-738
38. Gupta S, Clark DP: **Escherichia coli derivatives lacking both alcohol dehydrogenase and phosphotransacetylase grow anaerobically by lactate fermentation.** *J Bacteriol* 1989, **171**:3650-5
39. Ogata H, Goto S, Fujibuchi W, Kanehisa M: **Computation with the KEGG pathway database.** *Biosystems* 1998, **47**:119-128
40. Varma A, Palsson BO: **Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110.** *Applied and Environmental Microbiology* 1994, **60**:3724-3731
41. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connelly C, Davis K, Dietrich F, Dow SW, El Bakkoury M, Foury F, Friend SH, Gentalen E, Giaever G, Hegemann JH, Jones T, Laub M, Liao H, Liebundguth N, Lockhart DJ, Lucau-Danila A, Lussier M, M'Rabet N, Menard P, Mittmann M, Pai C, Rebischung C, Revuelta JL, Riles L, Roberts CJ, Ross-MacDonald P, Scherens B, Snyder M, Sookhai-Mahadeo S, Storms RK, S Vr, Voet M, Volckaert G, Ward TR, Wysocki R, Yen GS, Yu K, Zimmermann K, Philippsen P, Johnston M, Davis RW: **Functional Characterization of the S. cerevisiae Genome by Gene Deletion and Parallel Analysis.** *Science* 1999, **285**:901-906

42. Reizer J, Reizer A, Saier MH Jr: **Is the ribulose monophosphate pathway widely distributed in bacteria?** *Microbiology* 1997, **143**:2519-20
43. Link AJ, Phillips D, Church GM: **Methods for generating precise deletions and insertions in the genome of wild-type Escherichia coli: Application to open reading frame characterization.** *Journal of Bacteriology* 1997, **179**:6228-6237
44. Vanbogelen RA, Abshire KZ, Moldover B, Olson ER, Neidhardt FC: **Escherichia coli proteome analysis using the gene-protein database.** *Electrophoresis* 1997, **18**:1243-1251
45. Link AJ, Robison K, Church GM: **Comparing the predicted and observed properties of proteins encoded in the genome of Escherichia coli K-12.** *Electrophoresis* 1997, **18**:1259-1313
46. DeRisi JL, Iyer VR, Brown PO: **Exploring the metabolic and genetic control of gene expression on a genomic scale.** *Science (Washington D C)* 1997, **278**:680-686
47. Wodicka L, Dong H, Mittmann M, Ho M-H, Lockhart D: **Genome-wide expression monitoring in Saccharomyces cerevisiae.** *Nature Biotechnology* 1997, **15**:1359-1367
48. Pramanik J, Keasling JD: **Stoichiometric model of Escherichia coli metabolism: Incorporation of growth-rate dependent biomass composition and mechanistic energy requirements.** *Biotechnology and Bioengineering* 1997, **56**:398-421
49. Bonarius HPJ, Hatzimanikatis V, Meesters KPH, De Gooijer CD, Schmid G, Tramper J: **Metabolic flux analysis of hybridoma cells in different culture media using mass balances.** *Biotechnology and Bioengineering* 1996, **50**:299-318
50. Varma A, Palsson BO: **Metabolic capabilities of Escherichia coli: II. Optimal growth patterns.** *Journal of Theoretical Biology* 1993, **165**:503-522
51. Neidhardt FC, Umbarger HE: **Chemical Composition of Escherichia coli.** *In: Escherichia coli and Salmonella : cellular and molecular biology Edited by Neidhardt FC, vol. 1, 2nd ed. pp. 13/-16. Washington, D.C.: ASM Press* 1996, 13-16