

Metabolic systems cost-benefit analysis for interpreting network structure and regulation

Ross P. Carlson

Department of Chemical and Biological Engineering, Center for Biofilm Engineering, Montana State University, Bozeman, MT 59717, USA

Received on November 2, 2006; revised and accepted on February 27, 2007

Advance Access publication March 7, 2007

Associate Editor: Limsoon Wong

ABSTRACT

Motivation: Interpretation of bioinformatics data in terms of cellular function is a major challenge facing systems biology. This question is complicated by robust metabolic networks filled with structural features like parallel pathways and isozymes. Under conditions of nutrient sufficiency, metabolic networks are well known to be regulated for thermodynamic efficiency however; efficient biochemical pathways are anabolically expensive to construct. While parameters like thermodynamic efficiency have been extensively studied, a systems-based analysis of anabolic proteome synthesis ‘costs’ and the cellular function implications of these costs has not been reported.

Results: A cost-benefit analysis of an *in silico* *Escherichia coli* network revealed the relationship between metabolic pathway proteome synthesis requirements, DNA-coding sequence length, thermodynamic efficiency and substrate affinity. The results highlight basic metabolic network design principles. Pathway proteome synthesis requirements appear to have shaped biochemical network structure and regulation. Under conditions of nutrient scarcity and other general stresses, *E.coli* expresses pathways with relatively inexpensive proteome synthesis requirements instead of more efficient but also anabolically more expensive pathways. This evolutionary strategy provides a cellular function-based explanation for common network motifs like isozymes and parallel pathways and possibly explains ‘overflow’ metabolisms observed during nutrient scarcity.

Contact: alicia@iastate.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

Metabolic network structure is a reflection of ancient environments. A formal understanding of network design principles like the cellular function of parallel pathways and isozymes requires knowledge of evolutionary stresses and the competitive basis of the organism’s adaptive response. The elemental composition of life is vastly different than the elemental composition of the Earth (Stern and Elser, 2002). Scarce elemental resources, competition from other organisms and geochemical cycles are thought to have influenced the elemental make-up of microbes (Baudouin-Cornu *et al.*, 2001; Bragg and Hyder, 2004; Giovanni *et al.*, 2005). Evolutionary success necessitates competitive elemental investment strategies.

For instance, *E.coli* DNA and amino acid sequences are reported to have been influenced by sulfur, nitrogen and energy scarcity (Akashi and Gojobori, 2002; Baudouin-Cornu *et al.*, 2001, 2004; Bragg and Hyder, 2004).

During conditions of nutrient sufficiency, the role of thermodynamic efficiency on metabolic regulation has been well established both theoretically and experimentally (Carlson and Sien, 2004a; Lehninger, 1975; Varma and Palsson, 1993; Varma *et al.*, 1993). Under conditions of nutrient excess and nutrient scarcity, ‘overflow’ metabolisms, which deviate from maximum cellular thermodynamic efficiency, have been described (El-Mansi and Holms, 1989; Majewski and Domach, 1990; Neijssel *et al.*, 1996; Teixeira de Mattos and Neijssel, 1997). It has been suggested that overflow metabolisms during conditions of nutrient excess are the result of loose substrate uptake control, limited citric acid cycle flux capacity or limited respiratory chain capacity (Holms, 1996). The cellular function of an overflow metabolism during times of nutrient starvation or environmental stress has been largely overlooked. While maintenance energy requirements are well known to lower cellular yields at low growth rates, the metabolic shift to enzymes with lowered thermodynamic efficiency has been a paradox (Neijssel *et al.*, 1996; Teixeira de Mattos and Neijssel, 1997). This mystery highlights a major practical problem facing post-genomic biology, which is the lack of a formal understanding of network design principles (Huang, 2000).

Here, a cost-benefit analysis based on process design principles demonstrates that the *E.coli* metabolic network structure and function are a result of evolutionary pressures for strategic proteomic investment of scarce anabolic resources as well as for thermodynamic efficiency. The significance of investment costs, which here are defined as the elemental requirements to construct a functional metabolic pathway, has not been previously identified as an important selection pressure. Earlier network studies of *E.coli* growth focused only parameters like thermodynamic efficiency (for instance: Carlson and Sien 2004a; Varma and Palsson, 1993). This study explicitly elucidates the relationships between metabolic pathway proteome synthesis requirements, pathway thermodynamic efficiency, enzyme-substrate affinity and DNA-coding sequence length for a network comprised of tens of millions of unique, mathematically defined pathways.

2 METHODS

2.1 Economic analysis of metabolic pathways

The trade-off between capital investment costs and operating costs is a classic factory design challenge. As a polarized example, a factory assembly line can be constructed with expensive, automated, energy efficient equipment representing a 'high investment cost-low operating cost' strategy or an assembly line can be constructed with cheap, labor intensive, energy inefficient equipment representing a 'low investment cost-high operating cost' strategy. Optimal resolution of these trade-offs is typically determined through a multidimensional analysis of location-specific parameters like regional tax laws, labor costs and energy prices (Peters and Timmerhaus, 1980). Robust biochemical networks with numerous parallel pathways and isozymes represent an array of investment opportunities. Analogous to factory assembly lines, metabolic pathways can be constructed using different combinations of enzymes resulting in pathways with very different properties. It is hypothesized that the optimal resolution of the metabolic trade-off between pathway investment costs, which are equivalent to proteome synthesis requirements, and pathway operating costs, which reflect thermodynamic efficiency, can be resolved through multidimensional analysis of parameters like growth rate and the availability of anabolic and catabolic resources.

The *E.coli* central metabolism model is presented in Table S1 (Supplementary Material). The reaction stoichiometry, protein assignment for each reaction, enzyme subunit stoichiometry, and the carbon, nitrogen, sulfur, amino acid count and DNA-coding sequence requirements for each protein are listed. Enzyme assignments and subunit stoichiometries are based on data from EcoCyc, EcoGene and current literature (Kessler *et al.*, 2005; Rudd, 2000).

The metabolism model considers glucose as the sole energy source while glucose and CO₂ are potential carbon sources. Seven different *E.coli* biomass compositions are considered. The biomass stoichiometries are listed in Table S2. The macromolecular composition can be found in Carlson and Sreenc (2004a). The presented biomass compositions are used to test the robustness of the analysis to different cellular compositions. The biomass synthesis expression is based on the theory of Neidhardt *et al.* (1990) and considers the metabolite drain off the central metabolism. The individual enzymes involved in these biosynthetic reactions are not considered. The seven biomass expressions are normalized with respect to the ATP requirements so each doubling, time-dependent expression represents a different number of biomass Cmoles as listed in Table S2. These expressions only consider biosynthesis and do not include maintenance energy expenditures.

The *in silico* representation of the *E.coli* central metabolism was decomposed into its simplest, biologically relevant reaction sequences using elementary flux mode analysis (Schuster *et al.*, 1994, 2000; Stelling *et al.*, 2002). The elementary flux modes were identified using FluxAnalyzer (version 5.2) (Klamt *et al.*, 2003). A flux mode is a mathematically defined biochemical pathway. A typical flux mode connects the conversion of substrates into products while balancing system carbon, redox and ATP fluxes. Considering a single growth rate and therefore a single biomass composition, ~3.4 million unique, flux modes were identified.

Flux modes are comprised of a series of metabolite membrane transport steps and chemical reactions. Each enzyme-catalyzed event requires an investment of anabolic resources for enzyme synthesis. The flux mode investment costs were calculated by translating the modes, which synthesize either biomass or ATP as a product, into sets of utilized enzymes. Investment costs were not calculated for other flux modes including futile cycles. These enzymes sets represent theoretical proteomes. Five different investment costs were calculated for each enzyme set. The atomic carbon, nitrogen and sulfur as well as the total number of amino acids required to synthesize each enzyme set were

tabulated using proteomic databases (see Supplementary Material). The total number of amino acids is considered in addition to the three elements because this parameter can be correlated to cellular phosphorus requirements and energy charge (see Supplementary Material). The fifth tabulated cost was the number of DNA nucleotides required to code each enzyme set.

The atomic carbon, nitrogen, sulfur and amino acid investment costs considered enzyme expression levels. Pathway velocity was assumed to follow saturation-type kinetics analogous to enzymatic Michaelis-Menten kinetics and cellular Monod kinetics (Teixeira de Mattos and Neijssel, 1997). For low and high substrate concentrations, the kinetic expression can be simplified with either the first order or the zeroth order approximation, respectively. The presented study considers only environmental conditions that can be described by the first order kinetic approximation ($[S] \ll K_m$) where $0 < \mu \ll \mu_{max}$. See Supplementary Material for a formal treatment of investment cost assumptions. The case for transporter substrate affinity is discussed in a later section. The zeroth order kinetics case will be presented elsewhere.

Flux mode operating costs were defined as the expenditure of substrate, either glucose or oxygen, to synthesize either one Cmole of biomass or one mole of ATP. Glucose was the sole considered energy source, so all flux modes that synthesized either biomass or ATP as a product consumed glucose. Operating costs were not calculated for futile cycles or pathways that did not produce either biomass or ATP. The operating costs were determined for the appropriate flux modes using a MATLAB subroutine which calculated the ratio of the Cmoles of substrate consumed to either the Cmoles of biomass produced or the moles of ATP produced. The biomass operating cost is equivalent to the inverse biomass yield and reflects the thermodynamic efficiency of the pathway (Carlson and Sreenc, 2004a).

3 RESULTS

3.1 Multidimensional costs analysis

The investment and operating costs for each flux mode were plotted as coordinates in a multidimensional space. The coordinate magnitudes reflect fundamental flux mode properties. For instance, Figure 1a illustrates the multidimensional relationship between nitrogen investment cost, glucose operating cost and oxygen operating cost for flux modes synthesizing biomass. Each sphere represents the coordinates of a unique flux mode producing 1 Cmole of biomass. Earlier network studies of *E.coli* growth analyzed oxygen and glucose fluxes that maximized thermodynamic efficiency (for instance: Carlson and Sreenc, 2004a; Varma and Palsson, 1993). Figure 1a graphically illustrates how the current study adds additional dimensions of data to the previous work. Figure 1b–d highlights regions of interest in the 2-dimensional glucose operating cost/nitrogen investment cost plane. The data is broken into three separate plots, each with a different scale, to avoid compression of the high investment cost data points. If factory design principles are relevant to evolutionary design strategies, low costs are desirable. Flux modes with coordinates lying near the axes represent low-cost strategies. Considering Figure 1b, the flux modes labeled '1' and '2' lie nearest the ordinate and have the smallest glucose operating costs and therefore convert glucose into biomass with the highest efficiency. In Figure 1d, the flux mode labeled '12' lies closest to the abscissa and has the smallest nitrogen investment cost. Between the minimal operating cost and the minimal investment cost strategies is a continuous set of strategies

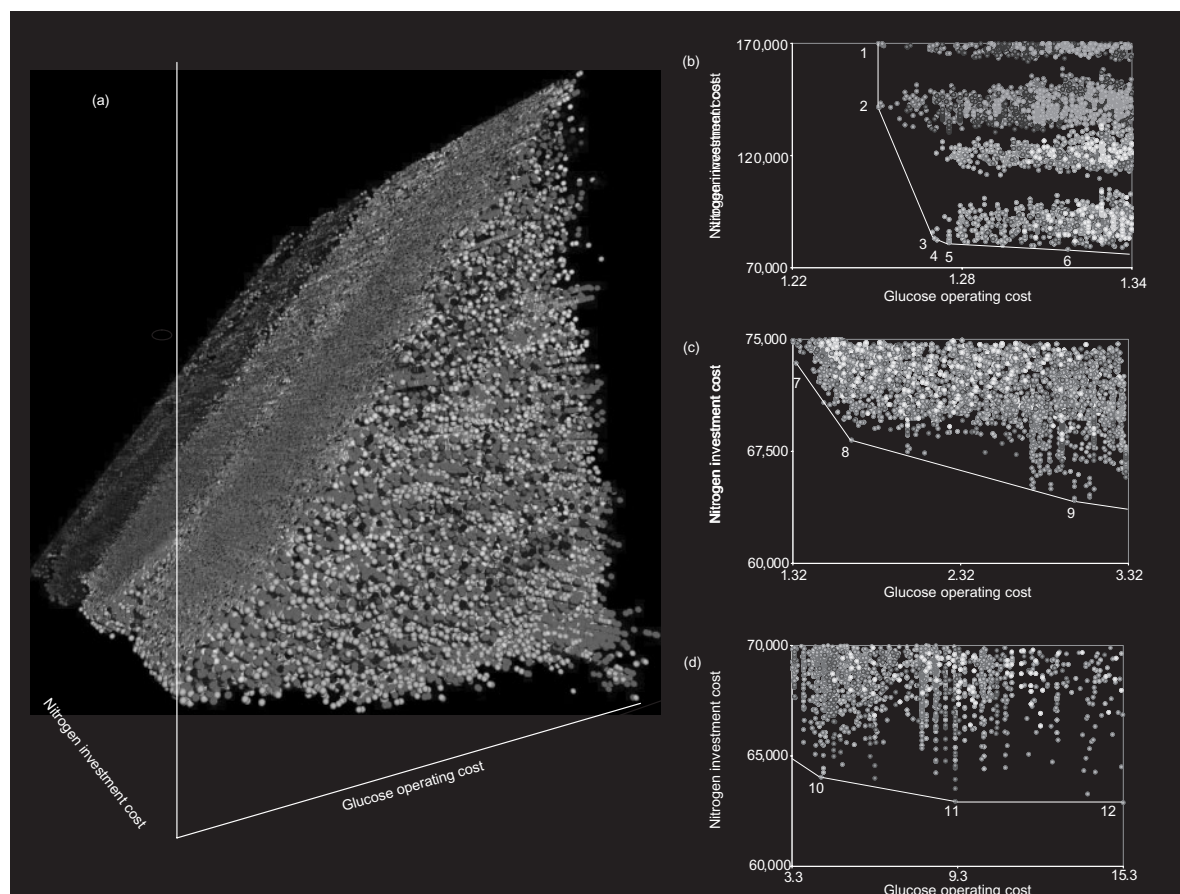


Fig. 1. (a) Multidimensional relationship between *E. coli* metabolic pathway properties. Metabolic pathway nitrogen investment cost, glucose operating cost and oxygen operating cost were plotted as coordinates in a multidimensional space. Each sphere represents the coordinates of a unique mathematically defined metabolic pathway that produces biomass. The spatial relationships illustrate fundamental pathway properties related to the proteome synthesis requirements (investment cost) and thermodynamic efficiency (operating cost). The nitrogen investment cost axis has units of nitrogen atoms required per metabolic pathway with values ranging from 70 000 to 170 000 nitrogen atoms per pathway. The oxygen operating cost axis has units of moles O_2 consumed per Cmole biomass synthesized with values ranging from 0 to 1.75 moles O_2 per Cmole biomass. The glucose operating cost axis has units of Cmoles glucose consumed per Cmole of biomass synthesized with values ranging from 1 to 3.3 Cmole glucose per Cmole biomass. Approximately 800 000 biomass producing modes from the 200-min doubling time simulation are shown in the plot. The data set is truncated to highlight biochemical pathways with coordinates near the origin. (b–d) Minimization envelope for the glucose operating cost and nitrogen investment cost plane. Twelve unique biochemical pathways labeled, 1 through 12, are required to most economically span the metabolic space between the minimal operating cost (pathway No. 1) and the minimal investment cost (pathway No. 12) strategies. The scale of the axes changes between Figure 1b, c, and d to highlight the pathways of interest and to avoid compression of the pathway data points. See text for more detail. The pathways are color-coded for enzyme usage patterns related to the glucose transport system (PtsG, GalP, Mgl) and the conversion of pyruvate to acetyl-CoA. (PDHc, Pfl): red = PtsG + PDHc, gray = PtsG + Pfl, green = GalP + Pfl, black = GalP no PDHc or Pfl, orange = Mgl + PDHc, blue = GalP + PDHc, yellow = Mgl + Pfl.

defined by a line segment envelope that minimizes the combined operating and investment costs. For the presented example, 12 unique flux modes are required to most economically span the possibilities between the minimal operating cost and the minimal investment cost strategies (see Table S5 in the Supplementary Material for an explicit listing of the 12 flux modes along the minimization envelope). The continuous ‘minimization envelope’ is explicitly defined by non-negative linear combinations of the flux modes labeled 1–12. Flux modes with coordinates above the minimization envelope have higher investment or higher operating costs than the strategies on the minimization envelope (Carlson and Sreenc, 2004a).

Additional orthogonal, 2-dimensional projections can be found in the Supplementary Material.

Figure 2 highlights shifts in enzyme usage as a function of biomass composition, four investment costs and glucose operating costs. Isozymes and parallel pathway enzymes are of special interest. The enzyme patterns are from flux modes lying on their respective minimization envelope. The twelve flux modes which define the minimization envelope in Figure 1b–d are labeled in the nitrogen investment cost section for a 200-min doubling time. Each rectangle represents a single flux mode along the minimization envelope. The same minimization envelope flux modes are analyzed for usage of

five different enzyme groupings. Data from seven biomass compositions are presented. The total number of flux modes along each minimization envelope varies between some biomass compositions. This was often due to a linear combination of flux modes which have equal and opposite fluxes in the pentose phosphate pathway reactions. When these flux modes are combined, the reaction flux sums to zero creating a situation where the associated enzyme is not required, creating a unique flux mode (Vijayasankaran *et al.*, 2005). The pentose phosphate pathway fluxes vary depending on the biomass composition.

Flux modes which minimize operating costs typically utilize enzymes with high thermodynamic efficiencies but also require relatively high anabolic resource investments. Flux modes in Figure 2 minimizing operating costs utilize the PEP-dependent PTS glucose transporter, the Embden–Meyerhof–Parnas glycolysis pathway, the pyruvate dehydrogenase complex

(PDHc), cyclic TCA cycle fluxes including the resource intensive α -ketoglutarate dehydrogenase complex, the soluble transhydrogenase enzyme (UdhA) and the high yielding Nuo-cytochrome o electron transport chain. These modes do not secrete any partially oxidized by-products. The majority of the high investment cost-low operating cost strategy predictions have been experimentally established with nutrient sufficient chemostat growth (Alexeeva *et al.*, 2000; Carlson and Srienc, 2004b; Iuchi and Lin, 1991; Sauer *et al.*, 2004; Varma *et al.*, 1993). When substrate uptake is limited to a constant rate, like in a chemostat, maximizing yields, which is equivalent to minimizing operating costs, has been recognized as a competitive strategy (Menlendez-Hevia *et al.*, 1997).

Enzymes associated with minimal investment cost strategies typically have relatively low thermodynamic yields but the enzymes also have relatively small anabolic resource investment requirements. These flux modes utilize the galactose

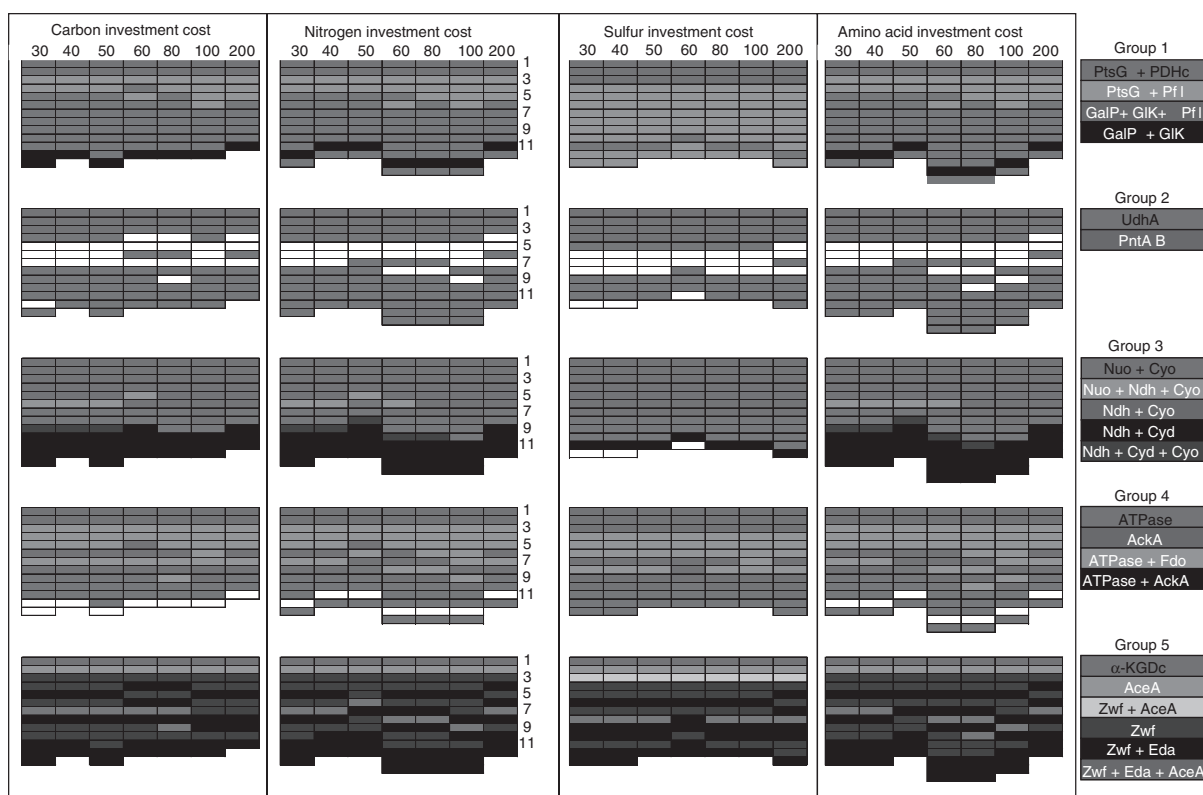


Fig. 2. Central metabolism enzyme usage patterns as a function of biomass composition, carbon, nitrogen, sulfur, and amino acid investment costs, and glucose operating cost. The data is divided into four sections with each section representing the results from a different investment cost analyses. Each colored rectangle represents a single metabolic pathway along the respective investment cost–glucose operating cost minimization envelope. The minimization envelope pathways were analyzed for five different sets of enzymes labeled Group 1 through Group 5. As an example, the twelve biochemical pathways defining the minimization envelope in Figure 1b–d are labeled in the nitrogen investment cost section. These same twelve pathways were analyzed for the five different groupings of enzymes. For each group, moving down the column minimizes metabolic pathway investment cost while moving up each column minimizes metabolic pathway glucose operating cost. Biomass composition changes from left to right. Compositions are consistent with different cell doubling time biomass compositions expressed as minutes. The color key for each group of enzyme data is given on the right. Group 1 examines glucose transporters and kinases (PtsG, Mgl/Glk, GalP/Glk) and the pyruvate dehydrogenase complex (PDHc) and pyruvate formate lyase (Pfl). Group 2 analyzes use of the soluble (UdhA) and membrane-bound (PntAB) transhydrogenase enzymes. Group 3 analyzes two NADH dehydrogenase enzymes (Nuo and Ndh) and the cytochrome o and d systems (Cyo and Cyd). Group 4 analyzes the use of the ATPase enzyme complex, the aerobic formate dehydrogenase (Fdo) and acetate kinase (AckA). Group 5 examines use of the TCA cycle α -ketoglutarate dehydrogenase complex (α -KGDc), the glyoxylate shunt isocitrate lyase enzyme (AceA), the oxidative branch of the pentose phosphate pathway (Zwf) and the Entner–Doudoroff pathway (Eda).

permease/glucose kinase (GalP/GIK) glucose transport and phosphorylation system, the Entner–Doudoroff pathway, the pyruvate formate lyase enzyme (Pfl), branched oxidative and reductive TCA cycle fluxes, the membrane-bound transhydrogenase (PntAB), and the lower proton motive force yielding Ndh-cytochrome d electron transport chain. These flux modes secrete partially oxidized products like formate and acetate. The majority of these enzyme usage patterns have been reported for experimental conditions with low anabolic resource availability or during nutrient starvation (Fischer and Sauer, 2003; Murray and Conway, 2005; Nyström and Gustavsson, 1998; Nyström and Neidhardt, 1993; Nyström *et al.*, 1996). Another common experimental observation consistent with the low investment cost predictions is the reduced substrate yields reported during anabolic resource limitation (Teixeira de Mattos and Neijssel, 1997). Under conditions of resource scarcity, construction of a functional, inexpensive pathway seems to take precedence over thermodynamic efficiency. The inexpensive, less efficient pathways have properties analogous to the observed ‘overflow’ metabolism. In addition to permitting higher fluxes, the ‘overflow’ metabolism requires less resources to synthesize the associated pathway enzymes (Teixeira de Mattos and Neijssel, 1997). Additional references for proteomic studies supporting the predictions can be found in the Supplementary Material.

From the minimal operating cost scenario to the minimal investment cost scenario, the glucose operating costs increase by a factor of about 10 while the atomic investment costs decrease by a factor of ~ 3 .

3.2 General stress responses

The investment cost scenarios in Figure 2 predict general nutrient stress responses. Enzyme usage patterns for the four investment costs are all very similar. *Escherichia coli* possesses a family of general stress regulator proteins known as the universal stress proteins. The universal stress proteins are expressed during starvation for carbon, nitrogen, sulfate, phosphate and amino acids as well as during exposure to heat, oxidants, metals, ethanol, antibiotics and other stresses (Nachin *et al.*, 2005). The enzymatic patterns of the low investment cost-high operating cost strategies are consistent with the reported universal stress protein mediated enzyme usage (Nachin *et al.*, 2005; Nyström and Gustavsson, 1998; Nyström and Neidhardt, 1993; Nyström *et al.*, 1996). Many bacteria are considered homeostatically regulated. The biomass carbon to nitrogen ratio (C:N) is nearly constant regardless of substrate C:N ratio (Sterner and Elser, 2002). Therefore, a limitation of either carbon or nitrogen requires a subsequent down shift in the investment of the other element. This is consistent with the present study. The flux mode enzyme C:N ratio as well as the C:S ratio for the minimization envelope are relatively constant regardless of a minimal operating cost or a minimal investment cost strategy (see Supplementary Material).

3.3 Evolutionary trade-off values

Nutrient-specific stress responses often involve expression of high affinity enzymes (Teixeira de Mattos and Neijssel, 1997). The high affinity *E. coli* glucose transporter (Mgl) and ammonia

assimilation pathway (GOGAT) are two examples. These enzymes have both higher investment costs and higher operating costs than comparable parallel pathways and therefore are not predicted with the presented methodology. However, competitive evolutionary pressures have selected the use of these enzymes suggesting that under some circumstances, the added investment and operating costs are off-set by advantages of high substrate affinity. The relationship between investment costs, operating costs and enzyme affinity was analyzed to quantify the advantage of high substrate affinity in terms of investment costs. The operating cost and investment cost differences between the three glucose uptake systems are shown in Figure 3a. The investment cost difference (ordinate-axis) between the lower affinity GalP system and the higher affinity Mgl system increases as the flux mode carbon investment is minimized. The GalP system has a K_m of $\sim 10 \mu\text{M}$ while the Mgl system K_m is estimated to be less than $1 \mu\text{M}$ (Gosset, 2005; Manche *et al.*, 1999). Increasing glucose affinity by an order of magnitude can translate into an investment cost of an additional ~ 15000 carbon atoms per flux mode. As a comparison, the PtsG system has a K_m of $\sim 5 \mu\text{M}$ (Gosset, 2005). Ammonia assimilation affinity was also examined (Fig. 3b). The glutamate dehydrogenase (Gdh) ammonia assimilation path has a K_m of $\sim 3 \text{mM}$ while the glutamine synthetase-glutamate synthase (GOGAT) ammonia assimilation system has a K_m of $< 0.2 \text{mM}$ (Helling, 1994). Improving the ammonia affinity by approximately one order of magnitude can translate into an additional investment cost of ~ 7500 nitrogen atoms per flux mode. High affinity enzymes, would in theory, permit higher metabolite fluxes than low affinity enzymes under the same low substrate concentration conditions which may justify the additional investment and operating costs.

4 DISCUSSION

Thermodynamically efficient metabolic pathways require more resources to construct than less efficient pathways. The difference in metabolic pathway construction requirements seems to have resulted in regulation patterns that favor less expensive metabolic pathways during nutrient scarcity. A functional, inexpensive pathway seems to take precedence over thermodynamic efficiency during nutrient stress resulting in a lower yielding energy metabolism with properties analogous to an ‘overflow’ metabolism.

The current study supports the widely held view that *E. coli* evolved under nutrient feast and famine conditions (Koch, 1971). *Escherichia coli* can toggle between parallel pathways with high protein investment cost-low operating cost strategies or low protein investment cost-high operating cost strategies with minimal enzyme regulation to adapt to catabolic or anabolic nutrient ‘feasts or famines’. Velicer and Lenski (1999) studied *E. coli* fitness after consecutive rounds of nutrient abundance and nutrient scarcity. Strain fitness under nutrient abundance did not preclude its fitness under nutrient scarcity and vice versa. Evolutionary pressures have created an organism that carries the resource burden of numerous genetic instructions and numerous regulatory systems but has the

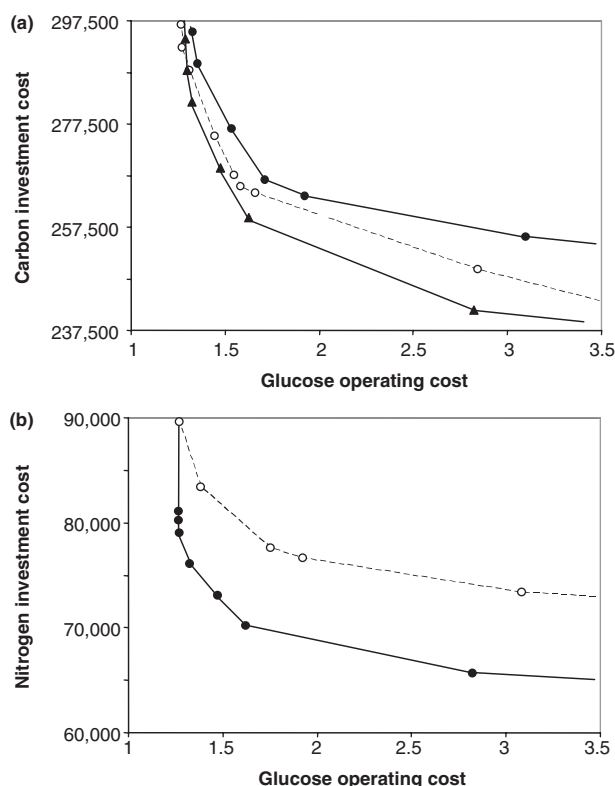


Fig. 3. Investment and operating costs associated with high affinity enzyme systems. The evolutionary trade-off between nutrient investment costs, pathway-operating costs and enzyme affinity was explored using multidimensional analysis. **(a)** The glucose operating cost and carbon investment cost differences between three glucose uptake systems. Each line represents the most efficient use of a single glucose transport system. Note that the actual minimization envelope transitions from the PtsG system to the GalP system at a carbon investment cost of $\sim 290,000$ carbon atoms per pathway. The investment cost difference was analyzed for the lower affinity GalP system ($K_m \sim 10 \mu M$, solid triangles), medium affinity PtsG system ($K_m \sim 5 \mu M$, open circles), and the higher affinity Mgl system (K_m estimated less than $1 \mu M$, filled circles). Increasing glucose affinity by an order of magnitude can require an investment of an additional $\sim 15,000$ carbon atoms per biochemical pathway. **(b)** The glutamate dehydrogenase (Gdh) ammonia assimilation path has a K_m of $\sim 3 mM$ (filled circles) while the glutamine synthetase-glutamate synthase (GOGAT) ammonia assimilation system has a $K_m < 0.2 mM$ (open circles). Improving the ammonia affinity by approximately one order of magnitude can require an additional investment of $\sim 7,500$ nitrogen atoms per metabolic pathway. Data is truncated to highlight regions of interest. The glucose operating cost is expressed as the Cmoles of glucose consumed per Cmole of biomass synthesized. The carbon investment cost is the atomic carbon requirement to synthesize the pathway enzyme set. The nitrogen investment cost is the atomic nitrogen requirement to synthesize the pathway enzyme set.

advantage of being able to compete under a wide variety of nutritional environments.

Previous proteomic studies of central metabolism adaptations to starvation conditions suggest the responses are regulated at the protein synthesis level (Nyström and Gustavsson, 1998). Two major regulators in this starvation metabolic shift are the universal stress protein UspA and ArcA.

ArcA is a well-studied regulator associated with the absence of an external electron acceptor and explains the numerous similarities between starvation and anaerobiosis enzyme usage patterns (Alexeeva *et al.*, 2000; Nyström and Gustavsson, 1998; Nyström and Neidhardt, 1993). The presented, theoretical study also suggests that the nutrient scarcity response would be regulated at the level of protein translation and not necessarily enzyme activity. Synthesizing a protein that is inactive would represent a non-performing investment of resources. While there are cases of non-functioning protein subunits being synthesized, likely as an investment which prepares the cell for rapid changes in environmental conditions, these protein subunits are likely expressed at low levels especially during starvation conditions (Alexeeva *et al.*, 2000).

There is much interest in minimal genome size, the minimal number of genes required for life, and the optimization of industrial organisms (Giovanni *et al.*, 2005; Glass *et al.*, 2006). Minimal genomes are not likely to carry many, if any, isozymes or parallel pathways. A theoretical minimal genome would only carry instructions for a limited number of metabolic strategies and a limited number of regulatory systems. In the present study, pathways that minimize investment costs require $\sim 66\%$ fewer amino acids as compared to the minimal operating cost strategies. Not surprising, minimizing the number of amino acids also reduces the size of the DNA-coding sequences (Fig. S1 in Supplementary Material). However, there is not a simple linear correlation between the number of flux mode amino acids and the coding sequence length due to multimeric enzyme structure. The minimal amino acid investment strategy requires around 33% fewer DNA-coding nucleotides than the minimal operating cost scenario. Minimizing genomes raises an interesting evolutionary consideration. Primitive life likely started with a single metabolic strategy and only evolved additional strategies with time. The central metabolic portal of the *E. coli* minimal investment cost strategies is the Entner–Doudoroff pathway (Fig. 2). This pathway along with features like the branched use of the TCA cycle enzymes, the lack of oxidative phosphorylation, and absence of the PTS glucose transport system have all been cited as examples of a primitive paleometabolism (Menéndez-Hevia *et al.*, 1997; Romano and Saier, 1992). This could suggest that *E. coli* has evolved toward strategies that minimize operating costs, however, at least key portions of primitive metabolic pathways have been retained because they offer competitive advantages during famines or other environmental stresses.

The current study considers 134 *E. coli* genes. For a single biomass composition, these genes can be arranged into $3.4 +$ million unique, mathematically defined biochemical pathways. Genome scale analysis will require improved computer memory architecture like a 64-bit system or more efficient software algorithms. The presented methodology applies cellular function to a metabolic network and then identifies the corresponding patterns of enzyme/gene usage. The approach provides a theoretical basis for studying critical network properties like the possible function of isozymes and parallel pathways. A comparison of low investment cost strategies, where proteome investment is minimized, with oxygen-limited, low operating cost strategies, where thermodynamic yield is maximized for the available oxygen, illustrates

how different environmental stresses can have similar enzymatic usage patterns. For instance, competitive strategies for both environmental conditions utilize branched TCA cycle fluxes, the pyruvate formate-lyase enzyme, and both secrete partially oxidized by-products like acetate and formate (Carlson and Sreenc, 2004b). However, the current study suggests there are metabolic adaptation differences between the two stresses. The Entner–Doudoroff pathway is used with low investment cost strategies instead of the Embden–Meyerhof–Parnas glycolysis pathway which is used with oxygen-limited, low operating cost strategies. This slight difference highlights the challenges facing approaches that first generate experimental patterns of enzyme/gene usage and then try to identify cellular function.

ACKNOWLEDGEMENTS

The author would like to thank R. Hunt for plotting the data in Figure 1a, N. Vijayasankaran, D.A. Fell, P.S. Stewart and an anonymous reviewer for helpful manuscript suggestions, the Montana State University EPSCoR program, the College of Engineering, Vice President for Research, the Department of Chemical and Biological Engineering and the Center for Biofilm Engineering for their generous funding.

Conflict of Interest: none declared.

REFERENCES

- Akashi, H. and Gojobori, T. (2002) Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *PNAS*, **99**, 3695–3700.
- Alexeeva, S. *et al.* (2000) Effects of limited aeration and of the arcAB system on intermediary pyruvate catabolism in *Escherichia coli*. *J. Bacteriol.*, **182**, 4934–4940.
- Baudouin-Cornu, P. *et al.* (2001) Molecular evolution of protein atomic composition. *Science*, **293**, 297–300.
- Baudouin-Cornu, P. *et al.* (2004) Intimate evolution of proteins. *J. Biol. Chem.*, **279**, 5421–5428.
- Bragg, J.G. and Hyder, C.L. (2004) Nitrogen versus carbon use in prokaryotic genomes and proteomes. *Proc. R. Soc. Lond. B*, **271**, S374–S377.
- Carlson, R. and Sreenc, F. (2004a) Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: identification of reactions. *Biotechnol. Bioeng.*, **85**, 1–19.
- Carlson, R. and Sreenc, F. (2004b) Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: Creation of overall flux state. *Biotechnol. Bioeng.*, **86**, 149–162.
- El-Mansi, E.M.T. and Holms, H. (1989) Control of carbon flux to acetate excretion during growth of *E. coli* in batch and continuous culture. *J. Gen. Microbiol.*, **135**, 2875–2883.
- Fischer, E. and Sauer, U. (2003) A novel metabolic cycle catalyzes glucose oxidation and anaplerosis in hungry *Escherichia coli*. *J. Biol. Chem.*, **278**, 46446–46451.
- Giovanni, S.J. *et al.* (2005) Genome streamlining in a cosmopolitan oceanic bacterium. *Science*, **309**, 1242–1245.
- Glass, J.I. *et al.* (2006) Essential genes of a minimal bacterium. *PNAS*, **103**, 425–430.
- Gosset, G. (2005) Improvement of *Escherichia coli* production strains by modifications of the phosphoenolpyruvate:sugar phosphotransferase system. *Microb. Cell Fact.*, **4**, 14–24.
- Helling, R.B. (1994) Why does *Escherichia coli* have two primary pathways for synthesis of glutamate? *J. Bacteriol.*, **176**, 4664–4668.
- Holms, H. (1996) Flux analysis and control of the central metabolic pathways in *Escherichia coli*. *FEMS Microbiol. Rev.*, **19**, 85–116.
- Huang, S. (2000) The practical problems of post-genomic biology. *Nat. Biotechnol.*, **18**, 471–472.
- Iuchi, S. and Lin, E.C.C. (1991) Adaptation of *Escherichia coli* to respiratory conditions: regulation of gene expression. *Cell*, **66**, 5–7.
- Kesseler, I.M. *et al.* (2005) EcoCyc: a comprehensive database resource for *Escherichia coli*. *Nucleic Acids Res.*, **33**, D334–D337.
- Klamt, S. *et al.* (2003) Fluxanalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps. *Bioinformatics*, **19**, 261–269.
- Koch, A.L. (1971) The adaptive responses of *Escherichia coli* to a fast and famine existence. *Adv. Microb. Physiol.*, **6**, 147–217.
- Lehninger, A.L. (1975) *Biochemistry*. 2nd edn. Worth Publishers, Inc., New York, USA.
- Majewski, R.A. and Domach, M.M. (1990) Simple constrained-optimization view of acetate overflow in *E. coli*. *Biotechnol. Bioeng.*, **35**, 732–738.
- Manche, K. *et al.* (1999) Mutational adaptation of *Escherichia coli* to glucose limitation involves distinct evolutionary pathways in aerobic and oxygen-limited environments. *Genetics*, **153**, 5–12.
- Menendez-Hevia, E. *et al.* (1997) Theoretical approaches to the evolutionary optimization of glycolysis: chemical analysis. *Eur. J. Biochem.*, **244**, 527–543.
- Murray, E.L. and Conway, T. (2005) Multiple regulators control expression of the Entner–Doudoroff aldolase (Eda) of *Escherichia coli*. *J. Bacteriol.*, **187**, 991–1000.
- Nachin, L. *et al.* (2005) Differential roles of the universal stress proteins of *Escherichia coli* in oxidative stress resistance, adhesion, and motility. *J. Bacteriol.*, **187**, 6265–6272.
- Neidhardt, F.C. *et al.* (1990) *Physiology of the Bacterial Cell*. Sinauer Associates Inc., Sunderland, MA, USA.
- Neijssel, O.M. *et al.* (1996) In: Neidhardt, F.C. (ed.), *Escherichia coli and Salmonella*. ASM Press, Washington, D.C., pp. 1683–1692.
- Nyström, T. and Gustavsson, N. (1998) Maintenance energy requirement: what is required for stasis survival of *Escherichia coli*. *Biochim. Biophys. Acta*, **1365**, 225–231.
- Nyström, T. and Neidhardt, F.C. (1993) Isolation and properties of a mutant of *Escherichia coli* with an insertional inactivation of the *uspA* gene, which encodes a universal stress protein. *J. Bacteriol.*, **175**, 3949–3956.
- Nyström, T. *et al.* (1996) Bacterial defense against aging: role of the *Escherichia coli* ArcA regulator in gene expression, readjusted energy flux and survival during stasis. *EMBO J.*, **15**, 3219–3228.
- Peters, M.S. and Timmerhaus, K.D. (1980) edn. In: *Plant Design and Economics for Chemical Engineers*, 3rd edn. McGraw-Hill Book Company, New York, USA.
- Romano, A.H. and Saier, M.H. (1992) In: Mortlock, R.P. (ed.), *The Evolution of Metabolic Function*. CRC Press, Boca Raton, pp. 143–170.
- Rudd, K.E. (2000) EcoGene: a genome sequence database for *Escherichia coli* K-12. *Nucleic Acids Res.*, **28**, 60–64.
- Sauer, U. *et al.* (2004) The soluble and membrane-bound transhydrogenase UdhA and PntAB have divergent functions in NADPH metabolism of *Escherichia coli*. *J. Biol. Chem.*, **279**, 6613–6619.
- Schuster, S. *et al.* (1994) Detecting elementary modes of functioning in metabolic networks. *Mod. Trends Bio. Thermo Kinetics*, **3**, 103–105.
- Schuster, S. *et al.* (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.*, **18**, 326–332.
- Stelling, J. *et al.* (2002) Metabolic network structure determines key aspects of functionality and regulation. *Nature*, **420**, 190–193.
- Sterner, R.W. and Elser, J.J. (2002) *Ecological Stoichiometry*. Princeton University Press, Princeton.
- Teixeira de Mattos, M.J. and Neijssel, O.M. (1997) Bioenergetic consequences of microbial adaptation to low-nutrient environments. *J. Biotechnol.*, **59**, 117–126.
- Varma, A. and Palsson, B.O. (1993) Metabolic capacities of *Escherichia coli* I. synthesis of biosynthetic precursors and cofactors. *J. Theor. Biol.*, **165**, 477–502.
- Varma, A. *et al.* (1993) Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. *Appl. Environ. Microbiol.*, **59**, 2465–2473.
- Velicer, G.J. and Lenski, R.E. (1999) Evolutionary trade-offs under conditions of resource abundance and scarcity: experiments with bacteria. *Ecology*, **80**, 1168–1179.
- Vijayasankaran, N. *et al.* (2005) Metabolic pathway structure for recombinant protein synthesis in *Escherichia coli*. *Appl. Microbiol. Biotechnol.*, **68**, 737–746.