Optimization-Based Framework for Inferring and Testing Hypothesized Metabolic Objective Functions

Anthony P. Burgard, Costas D. Maranas

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

Received 18 June 2002; accepted 7 November 2002

DOI: 10.1002/bit.10617

Abstract: An optimization-based framework is introduced for testing whether experimental flux data are consistent with different hypothesized objective functions. Specifically, we examine whether the maximization of a weighted combination of fluxes can explain a set of observed experimental data. Coefficients of importance (CoIs) are identified that quantify the fraction of the additive contribution of a given flux to a fitness (objective) function with an optimization that can explain the experimental flux data. A high CoI value implies that the experimental flux data are consistent with the hypothesis that the corresponding flux is maximized by the network, whereas a low value implies the converse. This framework (i.e., **ObjFind**) is applied to both an aerobic and anaerobic set of *Escherichia coli* flux data derived from isotopomer analysis. Results reveal that the CoIs for both growth conditions are strikingly similar, even though the flux distributions for the two cases are quite different, which is consistent with the presence of a single metabolic objective driving the flux distributions in both cases. Interestingly, the CoI associated with a biomass production flux, complete with energy and reducing power requirements, assumes a value 9 and 15 times higher than the next largest coefficient for the aerobic and anaerobic cases, respectively. © 2003 Wiley Periodicals, Inc. *Biotechnol Bioeng* **82:** 670–677, 2003.

Keywords: flux balance analysis; optimization in metabolic engineering; bilevel programming

INTRODUCTION

Living organisms have evolved to maximize their chances for survival (Darwin, 1899). This is manifested at the level of metabolism with the presence of multiple redundant pathways leading to and from key intermediates so that the removal of a single enzyme will (likely) not prevent an organism's ability to produce key components (Edwards and Palsson, 2000b; Price et al., 2002; Schilling and Palsson, 2000). Furthermore, experimental evidence suggests that organisms have developed control structures to ensure optimal growth in response to environmental constraints (Edwards et al., 2001). Although the existence of a fitness function driving an organism's evolution is widely accepted, it is unclear whether its fingerprint can be detected in the flux distributions of primary metabolism. Specifically, are metabolic networks driven to evolve as optimal biomass producers, maximum ATP generators, or optimal consumers of available substrates? In this investigation, we address the question of whether such a fitness function, or in optimization language, an objective function, can be identified from experimentally determined metabolic fluxes. We also examine how this fitness surrogate varies as environmental conditions change.

There exist two classes of metabolic modeling frameworks that inherently account for the presence of a fitness function that drives the metabolic machinery toward optimal survivability. First, the cybernetic modeling approach assumes that an organism is an optimal strategist utilizing all available resources with maximum efficiency (Ramkrishna et al., 1987). The expression and activity of the enzymes that catalyze network functionality are regulated by cybernetic control variables obtained from the solution of a constrained optimization problem (Dhurati et al., 1985; Kompala et al., 1984). This framework also contends that even genetically altered systems have the same underlying goal of optimal resource allocation, although the set of competing physiological choices open to the network expands or contracts depending upon the nature of the alteration (Varner and Ramkrishna, 1999a, 1999b). Cybernetic models have been successful in modeling the growth dynamics of yeast in batch and continuous cultures (Jones and Kompala, 1999), diauxic growth patterns and simultaneous consumption of substrates of *Escherichia coli* (Ramakrishna et al., 1996) and *Klebsiella oxytoca* (Kompala et al., 1986), and the time evolution of the aspartate family of amino acids in *Corynebacterium lactofermentum* (Varner and Ramkrishna, 1999c). Stoichiometric or flux balance analysis (FBA) models, on the other hand, employ this optimality principal in a slightly different fashion (Edwards et al., 2002). They use only the stoichiometric mass balances of

Correspondence to: C. D. Maranas

Contract grant sponsor: National Science Foundation; Department of Energy

Contract grant number: BES0120277

the metabolic network and cellular composition information to generate the broadest set of reaction flux distributions potentially available to the cell. This assumes that a metabolic network is capable of spanning all flux combinations allowable by the stoichiometric constraints with the maximization of biomass yield typically postulated as the objective function. Stoichiometric models have, in some cases, been successful in predicting the phenotypical characteristics of cells such as growth rates (Edwards et al., 2001; Pons et al., 1996), metabolic byproduct secretion rates (Varma et al., 1993b; Varma and Palsson, 1994), biochemical production rates (Henriksen et al., 1996; Jørgensen et al., 1995), and viability in the presence of gene deletions (Burgard and Maranas, 2001; Burgard et al., 2001; Edwards and Palsson, 2000a).

Although many hypotheses have been put forward as surrogates for cellular fitness functions, substantially less work has been conducted toward systematically validating them with experimentally derived flux distributions of metabolic networks. This is due in part to fact that the complete quantification of fluxes throughout the central metabolic pathways was intractable until recently. Traditionally, the only observables have been the growth rate, the uptake and secretion rates of substrates and products, and those reaction fluxes that can be calculated directly based on uptake or secretion of these external compounds. The development of complete isotopomer models (Christensen and Nielsen 2000; Forbes et al., 2001; Klapa et al., 1999; Park et al., 1999; Zupke et al., 1997), however, enables the calculation of the amount of reaction flux passing through every reaction of central metabolism. Briefly, isotopomer analysis involves the uptake of a ^{13}C-labeled substrate and the subsequent analysis of the labeling state by nuclear magnetic resonance (NMR) and/or gas chromatography/mass spectoscopy (GC/MS) measurements (Christensen and Nielsen, 1999; Schmidt et al., 1999b; Szyperski, 1995; Wiechert and de Graaf, 1996). This allows for a more accurate quantification of intracellular flux distributions providing an additional set of observables to test the various fitness function hypotheses.

In this work, we introduce a mathematically rigorous framework for testing whether experimental flux data are consistent with different hypothesized objective functions. Rather than starting by postulating such an objective function, or even accepting that one exists, we introduce a quantitative framework akin to inverse optimization (Tarantola, 1987) for inferring or disproving the consistency of different hypotheses. Specifically, we examine whether the maximization of a weighted combination of fluxes can explain a set of observed experimental data. For example, the driving force governing cellular metabolism may be a compromise between ATP, redox, and biomass production fluxes depending on the state of the system. Mathematically, deciphering this balance requires identification of the weights or coefficients, c_j, that accept the experimental fluxes, $v_j^{\exp}$, as an optimal solution to the following linear programming (LP) problem:

$$\text{Maximize:} \quad \sum_j^M c_j v_j$$

subject to:

$$\sum_j^M S_{ij} v_j = 0, \qquad \forall\, i \in N$$

$$v_j \geq 0, \qquad \forall\, j \in M$$

where M and N are the total number of reactions and metabolites, respectively, S_{ij} is the stoichiometric coefficient of metabolite i in reaction j, v_j represents the flux of reaction j, and c_j is a weight associated with reaction j. These coefficients, heretofore referred to as coefficients of importance (CoI), are scaled so that their sum is equal to 1. Intuitively, the coefficients of importance, c_j, quantify the fraction of the additive contribution of a given flux, v_j, to the fitness function whose maximization explains the experimental flux data. A high value for c_j implies that the experimental flux data are consistent with the hypothesis that the flux, v_j, is driven toward its maximum allowable value, whereas a low value implies the converse. FBA-based modeling typically assumes that the coefficients of importance define a unit vector in the direction of a growth flux comprised of all necessary components of biomass in their corresponding biological ratios (Ingraham et al., 1983). Other postulated objective functions include: (i) the maximization of metabolite (Varma et al., 1993a) or ATP (Majewski and Domach, 1990; Ramakrishna et al., 2001) production; and (ii) the minimization of the Euclidean norm (i.e., sum of the fluxes) (Bonarius et al., 1996), nutrient uptake, or redox production (Savinell and Palsson, 1992). The approach proposed here, referred to as **ObjFind**, requires the solution of a bilevel optimization problem that minimizes the squared deviations of identified fluxes from experimental data while ensuring that the identified fluxes are the product of an optimization problem. A solution strategy founded upon duality theory concepts is discussed in detail in the "Modeling and Computational Protocol" subsection. Based on this framework, we examine: (i) what is the objective function (if any) of a metabolic network that is the most consistent with experimental flux data; (ii) whether this objective function is unique; and (iii) how the driving forces governing cellular metabolism vary under different environmental conditions.

RESULTS

The **ObjFind** procedure is applied to the central metabolic network of *E. coli*. Experimental flux values, $v_j^{\exp}$, determined from an isotopomer analysis (Schmidt et al., 1999a)

study for both aerobic and anaerobic growth conditions, are used in conjunction with a stoichiometric model of *E. coli* central metabolism (Palsson, 2002) in an effort to pinpoint which underlying driving forces are governing the network's operation. This model, comprised of 62 reactions and 48 metabolites, includes all reactions of glycolysis, the TCA cycle, and the pentose phosphate pathway, as well as a number of respiration reactions. Coefficients of importance are assigned to each reaction flux associated with a metabolite drain, energy dissipation, or redox potential dissipation. In other words, an assignment is made for every flux that consumes, by either draining or dissipating, a resource in the network. The reaction fluxes associated with these coefficients are shown with colored arrows in Figure 1. Note that previously postulated objective functions are encompassed here as linear combinations of these reaction fluxes.

In our first case study, we identify CoIs consistent with the experimental fluxes, v_j^{exp}, being optimal to the LP problem maximizing:

$$\sum_j c_j v_j$$

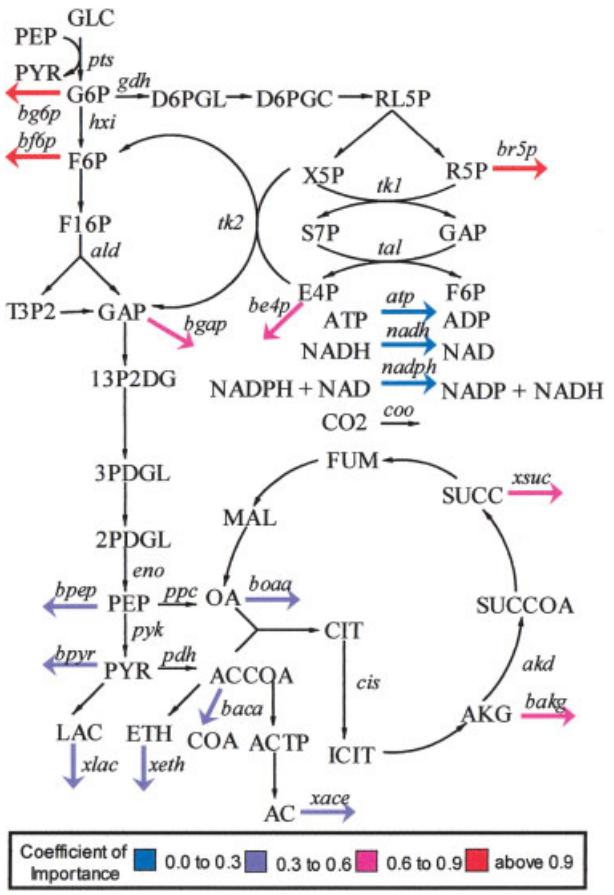

Figure 1. The reaction fluxes allowed to assume nonzero coefficients of importance are shown in color. The reactions with similar CoI magnitudes are denoted by the same colors. Note the magnitudes of the CoIs are similar for both the aerobic and anaerobic growth conditions.

subject to the network stoichiometry. Consequently, if this LP problem is solved using the coefficients c_j, identified by **ObjFind**, an optimal solution, v_j^*, exists, although not necessarily unique, such that:

$$\sum_j (v_j^* - v_j^{exp})^2$$

is minimized. The minimum sums of the squared flux deviations from the experimental data for the aerobic and anaerobic fluxes were found to be 0.016 (mmol/g DW h)^2 and 0.797 (mmol/g DW h)^2, respectively, which were well within the experimental error. The identified CoIs consistent with the aerobic and anaerobic experimental flux distributions are superimposed in Figure 2. Remarkably, the CoIs for both growth conditions are strikingly similar even though the flux distributions (see Table I) for the two cases are quite different. This unexpected convergence is consistent with the presence of a single metabolic objective driving the flux distributions in both cases. This objective is exemplified by the values of the coefficients of importance. It appears that fluxes with similar CoIs cluster within groups that are both topologically and functionally related. Specifically, seven fluxes, shown in purple, are clustered where the glycolysis pathway meets the TCA cycle. In addition, the fluxes with the largest coefficients, shown in red in Figure 1, are associated with drains of metabolites not far from glucose in the metabolic network, whereas the smallest coefficients, shown in blue, are associated with ATP and NADH dissipation. The most notable differences between the two sets of coefficients of importance are associated with the secretion of acetate, ethanol, and succinate, where the anaerobic CoIs are much larger. This is consistent with the fact that these metabolic byproducts are secreted only under anaerobic conditions. It is also noteworthy that the coefficient associated with the ATP dissipation flux is equal to zero for the anaerobic growth case, indicating that the network is more energy deficient under anaerobic growth conditions.

Next we investigate the effect of deviations in the flux distributions from the experimental ones on the robustness of the values identified for the coefficients of importance.

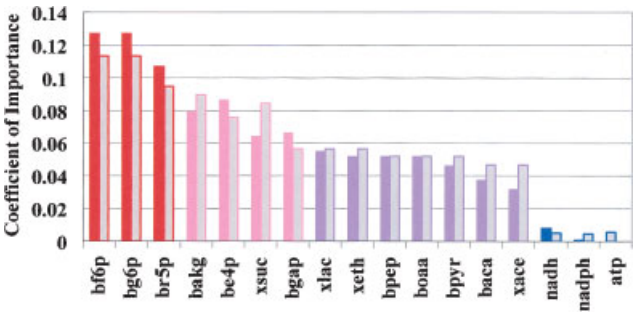

Figure 2. The values of the coefficients of importance for the aerobic (■) and anaerobic (□) experimental flux distributions.

Table I. Experimentally (Schmidt et al., 1999a) determined flux distributions for aerobic and anaerobic conditions.

Reaction[a]	Aerobic	Anaerobic[b]	|% Difference|
pts	115.0	115.0	0.0%
hxi	61.0	30.0	50.8%
gdh	53.1	84.9	59.8%
ald	89.1	84.9	4.8%
tk1	18.9	28.2	49.1%
tal	18.9	28.2	49.1%
tk2	9.6	26.9	179.9%
eno	181.2	194.6	7.4%
pyk	26.6	69.1	159.9%
pdh	126.1	111.4	11.7%
cis	52.0	4.8	90.8%
akd	45.0	4.7	89.6%
ppc	21.4	6.1	71.4%
xace	0.0	34.1	—
xeth	0.0	65.3	—
xlac	0.0	69.8	—
xsuc	0.0	3.9	—
coo[c]	254.9	199.6	21.7%
baca[c]	74.1	5.9	91.3%
bakg[c]	7.0	0.1	98.4%
be4p[c]	9.3	1.3	84.7%
bf6p[c]	0.3	0.2	27.2%
bg6p[c]	0.9	0.3	63.6%
bgap[c]	6.5	1.7	71.4%
boaa[c]	14.4	2	84.8%
bpep[c]	18.1	3.8	77.1%
bpyr[c]	15.3	2.8	80.0%
br5p[c]	5.7	1.5	71.3%

[a]Reactions corresponding to abbreviations are provided in Figure 1.

[b]Flux distributions for the anaerobic case are scaled so that glucose uptake rate is identical to the aerobic case.

[c]Fluxes were obtained from completely defined mass balances around the metabolites of interest.

Specifically, the sum of the squared deviations between the identified and experimental fluxes is allowed to increase from its minimum value by (i) 0, (ii) 1, (iii) 10, and (iv) 100 squared flux units (mmol/g DW h)^2; whereas, at the same time, each CoI is maximized individually to identify its sensitivity to experimental errors. The deviations of the CoIs from their nominal values for both aerobic and anaerobic cases are shown in Figure 3. Clearly, although some changes are present, the maximum extent possible is not high for small experimental errors, implying a substantial robustness in the assignment of values to the coefficients of importance with respect to experimental error. In addition, these deviations seem to be proportional to the nominal values of the CoIs and are approximately 95% larger for the anaerobic versus the aerobic case.

After identifying the CoIs and verifying their robustness to experimental error we turn our attention toward deciphering the biological significance of their values. Specifically, we examine how closely these coefficients of importance track the biomass maximization hypothesis. A biomass reaction flux (Varma and Palsson, 1993), complete with energy and reducing power requirements, is added to the net-

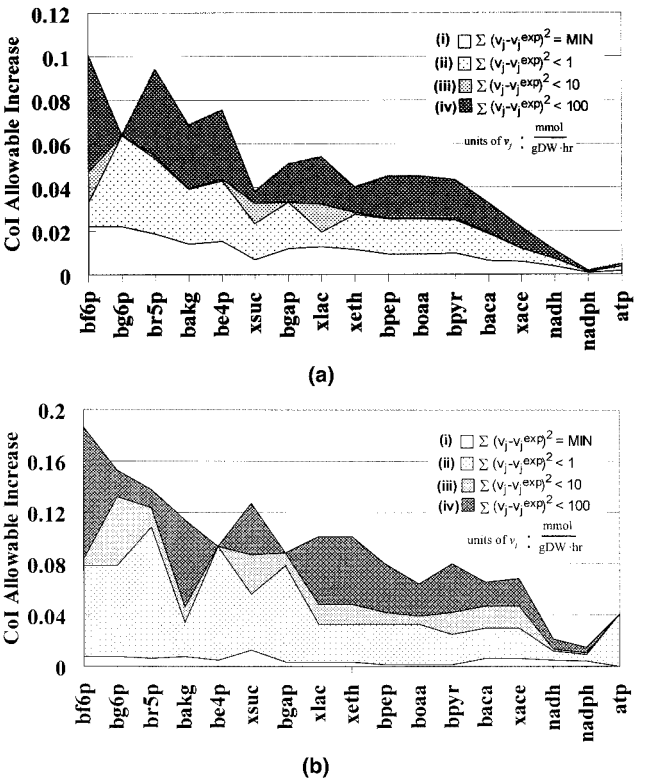

Figure 3. The maximum allowable increase of the coefficients of importance (CoIs) for the (a) aerobic and (b) anaerobic experimental flux distributions when solution optimality is relaxed. MIN = minimum.

work to drain metabolic precursors in their appropriate ratios, as proposed by Ingraham et al. (1983) for biomass formation. A coefficient of importance is assigned to this aggregate biomass flux. Note that an infinite number of solutions exist for the CoIs because the biomass flux is comprised of a linear combination of the other drain fluxes. We thus identify its maximum value as capable of explaining the flux distributions for the aerobic and anaerobic cases, respectively. The value for the coefficient of importance for biomass can then be interpreted as the maximum fraction of cellular resources that are diverted to biomass

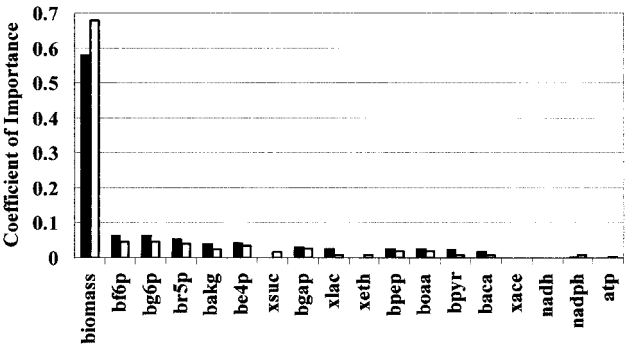

Figure 4. The coefficients of importance for the aerobic (■) and anaerobic (□) experimental flux distributions with the addition of a biomass flux.

formation. We find that the maximum possible values of biomass CoIs for the aerobic and anaerobic cases are 0.58 and 0.68, respectively, as shown in Figure 4. This means that biomass maximization appears to be an important descriptor of the observed flux distributions, but not the unique one, given that 0.58 and 0.68 are not equal to 1. No other flux has a coefficient of importance nearly as high as the one identified for biomass formation. Interestingly, the relative magnitudes of the coefficients for all other reaction fluxes remain similar to their original values except for the metabolic byproduct coefficients, *xsuc* and *xeth*, which are equal to 0 for the aerobic case, and the coefficients for *xace* and *nadh* dissipation, which drop out under both growth conditions.

It is important to note, however, that, while for a given flux distribution the range of allowable values for the coefficients of importance is rather narrow, the converse is not true. In fact, maximization of the sum of CoI-based weighted fluxes, $\sum_j c_j v_j$, subject to stoichiometric balances accepts many different flux distributions as optimal solutions as a consequence of the degeneracy of the LP optimization problem. Therefore, although the experimental fluxes do constitute optimal solutions to the optimization problem, many other alternate optimal solutions exist with or without biological meaning. This implies that the original experimental flux distribution cannot be unambiguously recovered based solely on the values of the coefficients of importance. The same holds true whenever the maximization of biomass, ATP, or any other resource is adopted a priori. Also, note that simple inspection of the ratio $v^{\exp}/v^{\max}$, where $v^{\max}$ is the maximum theoretical value of the specified flux subject to the stoichiometric constraints, does not reveal the trends that are uncovered with the CoIs. For example, the ratio

$$v^{\exp}_{\text{biomass}}/v^{\max}_{\text{biomass}}, \text{ where } v^{\exp}_{\text{biomass}}$$

is the maximum biomass formation with all fluxes equal to their experimental values, is only 0.058 and 0.120 for the aerobic and anaerobic cases, respectively, whereas the CoIs for biomass clearly reveal their importance. This implies that the proposed framework is more robust to deviations in the biomass composition from the Ingraham et al. (1983) approximation.

SUMMARY

In this work, a quantitative framework termed **ObjFind**, based on a bilevel optimization procedure, was developed for testing, disproving, or fine-tuning the consistency of different hypothesized objective functions with experimentally determined flux distributions. This method was applied to identify the coefficients of importance for *E. coli* flux distributions under aerobic and anaerobic growth conditions. These coefficients were remarkably similar, indicating a single cellular driving force governing the distribution of metabolic fluxes. In addition, surprisingly little flexibility was present in the CoIs for both cases among alternate optimal solutions identified when the sum-squared deviation was equal to its minimum value, although some flexibility was observed in these coefficients as the sum-squared deviation was allowed to increase. We also found that the maximization of the aggregate biomass flux to be consistent with the observed experimental flux values. Thus, the maximization of cellular biomass appears to be an important descriptor, although not the unique one, in explaining the observed fluxes. Finally, significant degeneracy was found among optimal solutions to the linear programming problem maximizing the sum (over j) of $c_j v_j$ subject to the network stoichiometry. This implies that the flux distributions through the network cannot be uniquely defined based solely on the identified CoIs, although they do provide insight as to which fluxes, when maximized, are consistent with the experimental flux data. It should be noted that, although one may never prove the existence of a universal objective function, the **ObjFind** procedure provides an unbiased framework for researchers to test the validity of different hypotheses leading to a better characterization of the underlying driving forces of cellular metabolism.

Modeling and Computational Protocol

The linear programming optimization model, referred to as the **Primal**, for a steady-state metabolic network comprised of N metabolites and M metabolic reactions with P potential cellular objectives is:

$$\begin{aligned}
&\text{Maximize:} && Z_P = \sum_{j \in P} c_j v_j && \textbf{(Primal)}\\
&\text{Subject to:} && \sum_{j=1}^{M} S_{ij} v_j = 0, && \forall\, i \in N\\
& && v_{GLC} = \text{uptake}, && \forall\, j \in \text{glucose uptake}\\
& && v_j \geq 0, && \forall\, j \in M
\end{aligned}$$

where S_{ij} is the stoichiometric coefficient of metabolite i in reaction j, v_j represents the flux of reaction j, v_{GLC} is the basis uptake amount of glucose, and c_j is a weight associated with importance of the reaction flux, v_j, referred to as a coefficient of importance. The vector v includes both internal and external fluxes and reversible reactions are defined as two irreversible reactions in opposite directions, constraining all fluxes to positive values. Given a set ($j \in E$) of fluxes, $v_j^{\exp}$, which have been experimentally determined, it is possible to determine which linear combinations of reaction fluxes weighted by c_j are maximized such that the experimental fluxes are optimal with respect to the linear programming problem described by the **Primal**. This requires the solution of the following bilevel optimization problem:

$$\underset{c_j}{\text{Minimize:}} \quad \sum_{j\in E} (v_j - v_j^{\exp})^2 \qquad \textbf{(ObjFind)}$$

Subject to:

$$\left(\begin{array}{ll} \underset{v_j}{\text{Maximize:}} & \sum_{j\in P} c_j v_j \\ \text{Subject to:} & \sum_{j=1}^{M} S_{ij} v_j = 0, \quad \forall\, i \in N \\ & v_{GLC} = \text{uptake}, \quad \forall\, j \in \text{glucose uptake} \\ & v_j \geq 0, \quad \forall\, j \in M \end{array}\right)$$

$$\sum_{j\in P} c_j = 1$$

$$c_j \geq 0, \quad \forall\, j \in P$$

in which the coefficient of importance (c_j) values for the inner problem are adjusted by the outer problem so that the sum-squared difference between the experimental fluxes and the optimal solution, v_j, for the inner problem is minimized. Note that the **ObjFind** problem includes the optimality of the **Primal** problem as a constraint giving rise to two nested optimization problems.

We propose an efficient solution approach borrowing from linear programming (LP) duality theory, which shows that for every LP problem (primal) there exists a unique optimization problem (dual) with an optimal objective value that is equal to that of the primal problem. The dual problem associated with the **Primal** LP problem (Bertsimas and Tsitsiklis, 1997) is:

$$\text{Minimize:} \quad Z_D = (\text{uptake}) \cdot \text{g} \qquad \textbf{(Dual)}$$

$$\text{Subject to:} \quad \sum_{i=1}^{N} u_i S_{ij} \geq c_j, \quad \forall\, j \in P$$

$$\sum_{i=1}^{N} u_i S_{ij} \geq 0, \quad \forall\, j \notin P, \text{ glucose uptake}$$

$$\sum_{i=1}^{N} u_i S_{ij} + g \geq 0, \quad \forall\, j \in \text{glucose uptake}$$

where u_i is the dual variable associated with the first set of constraints in the **Primal**, and g is the dual variable associated with the glucose uptake constraint. The dual variables, u_i and g, indicate the change in the optimal value of Z_P per unit change on the right-hand side of their associated constraint. Likewise, the reaction fluxes, v_j, are the dual variables associated with the constraints of the **Dual** problem.

The concept of strong duality (see Fig. 5) implies that if the primal has an optimal solution, so does the dual, and their respective optimal objective values are equal. Furthermore, the primal and dual problems can be simultaneously feasible only at their respective optimal solutions. Therefore, by constructing an optimization problem formulation that includes both the **Primal** and **Dual** constraints along

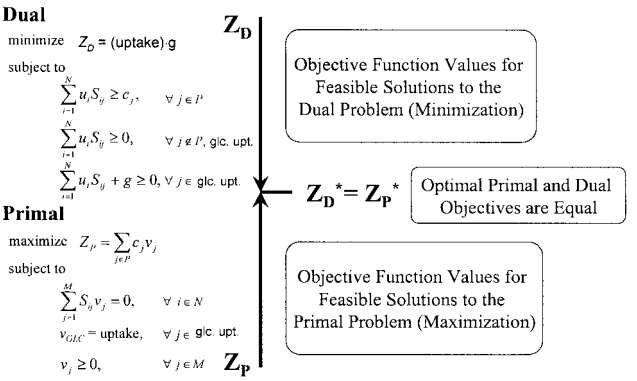

Figure 5. Weak duality states that any feasible dual solution has an objective value that is greater than the optimal primal objective, whereas any feasible primal solution has an objective value that is less than the optimal dual objective. Strong duality states that if the primal has an optimal solution, so does the dual, and their respective optimal objective function values are equal.

with an equality constraint forcing their respective objective function values to be equal to each other, we ensure that any feasible solution (v_j, g, u_i) will be optimal to both the **Primal** and **Dual** problems. Therefore, solution of the following single-level nonlinear optimization problem:

$$\text{Minimize:} \quad \sum_{j\in E} (v_j - v_j^{\exp})^2 \qquad \textbf{(ObjFind)}$$

$$\text{Subject to:} \quad \sum_{j\in P} c_j v_j = (\text{uptake}) \cdot \text{g}$$

$$\sum_{j\in P} c_j = 1$$

$$\sum_{j=1}^{M} S_{ij} v_j = 0, \quad \forall\, i \in N$$

$$v_{GLC} = \text{uptake}, \quad \forall\, j \in \text{glucose uptake}$$

$$\sum_{i=1}^{N} u_i S_{ij} \geq c_j, \quad \forall\, j \in P$$

$$\sum_{i=1}^{N} u_i S_{ij} \geq 0, \quad \forall\, j \notin P, \text{ glucose uptake}$$

$$\sum_{i=1}^{N} u_i S_{ij} + g \geq 0, \quad \forall\, j \in \text{glucose uptake}$$

$$v_j \geq 0, \quad \forall\, j \in M$$

$$c_j \geq 0, \quad \forall\, j \in P$$

systematically characterizes the set of all possible c_j values consistent with the minimization of the sum-squared difference between a subset of observed fluxes, $v_j^{\exp}$, and an optimal solution to the **Primal**. Note that any problems resulting from the presence of multiple optima to the primal are circumvented by including the flux variables directly in the dual minimization. By utilizing LP duality concepts, a method is introduced for transforming the original intrac-

table two-stage optimization problem into a single-stage optimization problem. The related problem of maximizing the value of a particular coefficient, c_j^*, subject to the sum of the squared deviations being below a target value is:

Maximize: c_j^*

Subject to:

$$\sum_{j\in P} c_j v_j = (\text{uptake}) \cdot g$$

$$\sum_{j\in E} (v_j - v_j^{\exp})^2 \le \text{target}$$

$$\sum_{j\in P} c_j = 1$$

$$\sum_{j=1}^{M} S_{ij} v_j = 0, \qquad \forall\, i \in N$$

$$v_{GLC} = \text{uptake}, \qquad \forall\, j \in \text{glucose uptake}$$

$$\sum_{i=1}^{N} u_i S_{ij} \ge c_j, \qquad \forall\, j \in P$$

$$\sum_{i=1}^{N} u_i S_{ij} \ge 0, \qquad \forall\, j \notin P, \text{ glucose uptake}$$

$$\sum_{i=1}^{N} u_i S_{ij} + g \ge 0, \qquad \forall\, j \in \text{glucose uptake}$$

$$v_j \ge 0, \qquad \forall\, j \in M$$

$$c_j \ge 0, \qquad \forall\, j \in P$$

where c_j^* can be the weight associated with any potential cellular objective (i.e., biomass formation, energy production, etc.). It should be noted that the constraint ($Z_P = Z_D$) is nonconvex due to the bilinear $c_j v_j$ terms. Therefore, multiple starting points were used to identify multiple optimal solutions in each case. We observed that, after >100 restarts, only a handful of multiple optima were identified. Problems containing as many as 200 variables were solved in seconds using MINOS 5.0 accessed via the GAMS modeling environment on an IBM RS6000-270 workstation.

References

Bertsimas D, Tsitsiklis JN. 1997. Introduction to linear optimization. Belmont, MA: Athena Scientific.

Bonarius PJB, Hatzimanikatis V, Meesters KPH, de Gooijer CD, Schmid G, Tramper J. 1996. Metabolic flux analysis of hybridoma cells in different culture media using mass balances. Biotechnol Bioeng 50: 299–318.

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

Burgard AP, Vaidyaraman S, Maranas CD. 2001. Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments. Biotechnol Progr 17:791–797.

Christensen B, Nielsen J. 1999. Isotopomer analysis using GC-MS. Metab Eng 1:282–290.

Christensen B, Nielsen J. 2000. Metabolic network analysis. A powerful tool in metabolic engineering. Adv Biochem Eng Biotechnol 66: 209–231.

Darwin C. 1899. The origin of species by means of natural selection. New York: Crowell.

Dhurati PS, et al. 1985. A cybernetic view of microbial growth: modeling of cells as optimal strategists. Biotechnol Bioeng 27:1–9.

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

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

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

Edwards JS, Palsson BO. 2000b. Robustness analysis of the *Escherichia coli* metabolic network. Biotechnol Progr 16:927–939.

Forbes NS, Clark DS, Blauch HW. 2001. Using isotopomer path tracing to quantify metabolic fluxes in pathway models containing reversible reactions. Biotechnol Bioeng 74:196–211.

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.

Ingraham JL, et al. 1983. Growth of the bacterial cell. Sunderland, MA: Sinauer Associates.

Jones KD, Kompala DS. 1999. Cybernetic model of the growth dynamics of *Saccharomyces cerevisiae* in batch and continuous cultures. J Biotechnol 71:105–131.

Jørgensen H, et al. 1995. Metabolic flux distributions in *Penicillium chrysogenum* during fed-batch cultivations. Biotechnol Bioeng 46: 117–131.

Klapa MI, Park SM, Sinskey AJ, Stephanopoulos G. 1999. Metabolite and isotopomer balancing in the analysis of metabolic cycles: I. Theory. Biotechnol Bioeng 62:375–391.

Kompala DS, et al. 1986. Investigation of bacterial growth on mixed substrates: experimental evaluation of cybernetic models. Biotechnol Bioeng 28:1044–1055.

Kompala DS, et al. 1984. Cybernetic modeling of microbial growth on multiple substrates. Biotechnol Bioeng 26:1272–1281.

Majewski RA, Domach MM. 1990. Simple constrained optimization view of acetate overflow in *Escherichia coli*. Biotechnol Bioeng 35: 732–738.

Palsson BO. 2002. *E. coli* central metabolic network. http://gcrg.ucsd.edu/downloads/Pathway_FBA/default.htm.

Park SM, Klapa MI, Sinskey AJ, Stephanopoulos G. 1999. Metabolite and isotopomer balancing in the analysis of metabolic cycles: II. Applications. Biotechnol Bioeng 62:392–401.

Pons A, Dussap CG, Pequignot C, Gros JB. 1996. Metabolic flux distribution in *Corynebacterium melassecola* ATCC 17965 for various carbon sources. Biotechnol Bioeng 51:177–189.

Price ND, Papin JA, Palsson BA. 2002. Determination of redundancy and systems properties of the metabolic network of *Helicobacter pylori* using genome-scale extreme pathway analysis. Genome Res 12:760–769.

Ramakrishna R, Edwards JS, McCulloch A, Palsson BO. 2001. Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints. Am J Physiol Reg Integr Compr Physiol 280:R695–R704.

Ramakrishna R, Ramkrishna D, Konopka AE. 1996. Cybernetic modeling of growth in mixed, substitutable substrate environments: preferential and simultaneous utilization. Biotechnol Bioeng 52:141–151.

Ramkrishna D, et al. 1987. Are microbes optimal strategists? Biotechnol Progr 3:121–126.

Savinell JM, Palsson BO. 1992. Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism. J Theor Biol 154:421–454.

Schilling CH, Palsson BO. 2000. Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. J Theor Biol 203:249–283.

Schmidt K, Nielsen J, Villadesen J. 1999a. Quantitative analysis of meta-

bolic fluxes in *Escherichia coli*, using two-dimensional NMR spectroscopy and complete isotopomer models. J Biotechnol 71:175–189.

Schmidt K, Norregaard LC, Pedersen B, Meissner A, Duus JO, Nielsen JO, Villadsen J. 1999b. Quantification of intracellular metabolic fluxes from fractional enrichment and 13C–13C coupling constraints on the isotopomer distribution in labeled biomass components. Metab Eng 12:166–179.

Szyperski T. 1995. Biosynthetically directed fractional 13C-labeling of proteinogenic amino acids. An efficient analytical tool to investigate intermediary metabolism. Eur J Biochem 232:433–448.

Tarantola A. 1987. Inverse problem theory: methods for data fitting and model parameter estimation. Amsterdam: Elsevier.

Varma A, et al. 1993a. Biochemical production capabilities of *Escherichia coli*. Biotechnol Bioeng 42:59–73.

Varma A, Boesch BW, Palsson BO. 1993b. Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. Appl Environ Microbiol 59:2465–2473.

Varma A, Palsson B. 1993. Metabolic capabilities of *Escherichia coli*. II. Optimal growth patterns. J Theor Biol 165:503–522.

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

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

Varner J, Ramkrishna D. 1999b. Metabolic engineering from a cybernetic perspective. 2. Qualitative investigation of nodal architechtures and their response to genetic perturbation. Biotechnol Progr 153:426–438.

Varner J, Ramkrishna D. 1999c. Metabolic engineering from a cybernetic perspective: aspartate family of amino acids. Metab Eng 11:88–116.

Wiechert W, de Graaf AA. 1996. In vivo stationary flux analysis by 13C labeling experiments. Adv Biochem Eng Biotechnol 54:109–154.

Zupke C, Tompkins R, Yarmush D, Yarmush M. 1997. Numerical isotopomer analysis: Estimation of metabolic activity. Anal Biochem 2472: 287–293.