Combining Pathway Analysis with Flux Balance Analysis for the Comprehensive Study of Metabolic Systems

Christophe H. Schilling,[1] Jeremy S. Edwards,[2] David Letscher,[3] Bernhard Ø. Palsson[1]

[1]Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive, MC 0412, La Jolla, California 92093-0412, USA; telephone: 619-534-5668; fax: 619-822-3120; e-mail: palsson@ucsd.edu
[2]Department of Chemical Engineering, University of Delaware, Newark, Delaware, USA
[3]Department of Mathematics, Oklahoma State University, Stillwater, Oklahoma, USA

Received 4 March 2000; accepted 10 December 2000

Abstract: The elucidation of organism-scale metabolic networks necessitates the development of integrative methods to analyze and interpret the systemic properties of cellular metabolism. A shift in emphasis from single metabolic reactions to systemically defined pathways is one consequence of such an integrative analysis of metabolic systems. The constraints of systemic stoichiometry, and limited thermodynamics have led to the definition of the flux space within the context of convex analysis. The flux space of the metabolic system, containing all allowable flux distributions, is constrained to a convex polyhedral cone in a high-dimensional space. From metabolic pathway analysis, the edges of the high-dimensional flux cone are vectors that correspond to systemically defined "extreme pathways" spanning the capabilities of the system. The addition of maximum flux capacities of individual metabolic reactions serves to further constrain the flux space and has led to the development of flux balance analysis using linear optimization to calculate optimal flux distributions. Here we provide the precise theoretical connections between pathway analysis and flux balance analysis allowing for their combined application to study integrated metabolic function. Shifts in metabolic behavior are calculated using linear optimization and are then interpreted using the extreme pathways to demonstrate the concept of pathway utilization. Changes to the reaction network, such as the removal of a reaction, can lead to the generation of suboptimal phenotypes that can be directly attributed to the loss of pathway function and capabilities. Optimal growth phenotypes are calculated as a function of environmental variables, such as the availability of substrate and oxygen, leading to the definition of phenotypic phase planes. It is illustrated how optimality properties of the computed flux distributions can be interpreted in terms of the extreme pathways. Together these developments are applied to an example network and to core metabolism of *Escherichia coli* demonstrating the connections between the extreme pathways, optimal flux distributions, and phenotypic phase planes. The consequences of changing environmental and internal conditions of the network are examined for growth on glucose and succinate in the face of a variety of gene deletions. The convergence of the calculation of optimal phenotypes through linear programming and the definition of extreme pathways establishes a different perspective for the understanding of how a defined metabolic network is best used under different environmental and internal conditions or, in other words, a pathway basis for the interpretation of the metabolic reaction norm. *Biotechnol Bioeng* **71**: 286–306, 2000/2001.
Keywords: metabolic modeling; convex analysis; linear programming; flux balance analysis; metabolic pathways

INTRODUCTION

Automated sequencing and high-throughput experimental techniques to analyze cellular functions at the genomic level are beginning to reveal the full complexity of living systems. The implication of dealing with complex systems is that their properties are not predictable from the complete description of their components, leading to the concept of emergent properties of living systems (Bhalla and Iyengar, 1999; Palsson, 1997; Weng et al., 1999). To understand the complexity inherent in cellular networks, approaches need to be implemented that focus on the systemic properties of the network. Subsequently, these approaches must be applied to complete cellular systems. Our focus then shifts from a reductionist to a holistic approach for understanding the interdependence of gene function and the role of the gene in the context of multigenetic cellular functions or genetic circuits.

The multigeneic nature of cellular functions presents a significant challenge in extracting phenotypic information from the genotype. Both experimental and theoretical methods that overcome this challenge will contribute significantly toward extracting valuable information from genomic data. Nowhere is this contribution more apparent

Correspondence to: B. Ø. Palsson.
Contract grant sponsors: National Institutes of Health; National Science Foundation
Contract grant numbers: GM57089; 9873384; 9814092

than in cellular metabolism where it is now possible to reconstruct detailed metabolic networks from annotated genome sequences, and experimentally characterize metabolic phenotypes using functional genomic technologies such as whole-genome expression profiling and protein characterization. Furthermore, there exists a history of studying the systemic properties of metabolic networks. This history includes approaches such as metabolic control analysis (Fell, 1996; Heinrich and Rapoport, 1974; Kacser and Burns, 1973), flux balance analysis (Bonarius et al., 1997; Edwards et al., 1999; Varma and Palsson, 1994b), pathway analysis (Clark, 1988; Liao et al., 1996; Mavrovouniotis et al., 1990; Schilling et al., 2000a; Schuster et al., 1999; Seressiotis and Bailey, 1988), cybernetic modeling (Ramakrishna et al., 1996; Varner and Ramkrishna, 1998, 1999), biochemical systems theory (Savageau, 1969a, 1969b, 1970), and temporal decomposition (Palsson et al., 1987).

Although the ultimate goal may be the development of dynamic models for the complete simulation of metabolic systems, the success of such approaches is severely hampered by the current lack of kinetic information on the dynamics and regulation of metabolic reactions. In the absence of kinetic information it is still possible to accurately assess the theoretical capabilities and operative modes of metabolic systems using steady-state analysis. Steady-state analysis is based on the stoichiometry of the metabolic reactions, which is a structural invariant of the metabolic network.

In recent years, an approach known as flux balance analysis has been developed to describe metabolic physiology in a quantitative manner (Bonarius et al., 1997; Edwards and Palsson, 1998; Edwards et al., 1999; Sauer et al., 1998; Schilling et al., 1999a; Varma and Palsson, 1994b). Flux balance analysis is based on the fundamental law of mass conservation and the application of optimization principles to predict the optimal distribution of metabolic resources within a network. It thus provides insightful information about the systemic constraints placed on metabolic function. The analysis is performed under steady-state conditions and it only requires information about the stoichiometry of metabolic pathways and on metabolic demands. This approach is particularly applicable for postgenomic analysis, because the stoichiometric parameters can be defined from the annotated genome sequence of a particular organism (Edwards and Palsson, 1999, 2000; Schilling and Palsson, 2000b). Flux balance analysis is distinguished from metabolic flux analysis (MFA) where fluxes are experimentally measured and used to reduce the underdetermined nature of a metabolic system to allow for the calculation of all the unknown fluxes under steady-state conditions without invoking optimization principles (Stephanopoulos et al., 1998).

Flux balance analysis has recently been applied to analyze and predict the metabolic genotype–phenotype relation from a reaction-based perspective, in which flux activities for individual reactions are examined (Edwards and Palsson, 1998, 1999, 2000; Schilling et al., 1999a). The ability to interpret predicted metabolic phenotypes from a pathway perspective in addition to an individual reaction-based perspective might shed new light on the complexity of the genotype–phenotype relationship. A pathway perspective will provide flux activities for systemic metabolic pathways that are comprised of balanced sets of reactions. These pathways are characterized based on their functional and structural relevance within the system as opposed to their pattern of historical discovery. Recent approaches have been developed to perform detailed pathway decompositions of metabolic networks that hold promise in providing the pathway perspective that is currently sought to complement flux balance analysis and related optimization techniques.

The pathway analysis of metabolic networks has progressed itself over the years from simply analyzing wall charts of metabolic reactions to the recent development of detailed theories for the in-depth study of the structure of metabolic networks (for a recent review, see Schilling et al. [1999b]). The definition of systemic pathways in metabolism is essential for providing a pathway-oriented perspective to overall metabolic functions and phenotypes. One of the more promising approaches for performing this identification and analysis has capitalized on the principles of convex analysis. These principles include the closely related concepts of elementary modes (Schuster et al., 1996, 1999) and extreme pathways (Schilling et al., 2000a), which can both be used to define the limitations and production capabilities of metabolic systems. These approaches to pathway analysis share a common underlying mathematical framework with flux balance analysis. Although this connection has been noted by a number of investigators, a clear definition and illustration of the formal relationships between these two approaches is lacking.

In this article we illustrate the precise relationships that bind metabolic pathway analysis to flux balance analysis, and illustrate their combined use to interpret shifts in metabolic routing that may occur in response to environmental and internal/genetic challenges. The combination of these two approaches allows for a clear interpretation of metabolic phenotypes and flux distributions from the traditional reaction-based perspective and now from a pathway-oriented perspective. Many of the relations are illustrated initially through the use of a hypothetical reaction network to assist in developing a core understanding of both approaches and their connections. We then apply the combined approach for the detailed study of the central metabolic network of *Escherichia coli*. The comprehensive collection of calculations, and illustrations generated for the work presented is available on-line along with additional supportive material to assist the reader (http://gcrg.ucsd.edu/supplementary_data/Pathway_FBA/default.htm). The complementarities of flux balance and pathway analysis for studying metabolic systems along with the combination of emerging genome-scale experimental technologies should serve to form a useful platform from which to study the complexity inherent in cellular metabo-

lism for both basic scientific and applied purposes of metabolic engineering (Stephanopoulos, 1999; Yarmush and Berthiaume, 1997) and biocommodity engineering (Lynd et al., 1999).

DESCRIBING METABOLIC SYSTEMS

A metabolic network is a collection of enzyme-catalyzed reactions and transport processes that serve to dissipate substrate metabolites and generate final metabolites. To describe metabolic networks in a quantitative manner, dynamic mass balances are written for each metabolite in the network, which generates a system of ordinary differential equations that describe the transient behavior of metabolite concentrations:

$$\frac{dX_i}{dt} = \sum_j S_{ij} v_j \tag{1}$$

where v_j corresponds to the jth metabolic flux, $[X_i]$ represents the concentration of the metabolite, and the stoichiometric coefficient, S_{ij}, stands for the number of moles of metabolite i formed (or consumed) in reaction j. As we are often interested in the structural characteristics or invariant properties of the reaction network (Reder, 1988), it is reasonable to place the metabolic system into a steady state, thus reducing the set of differential equations to a set of linear homogeneous equations, written in matrix notation as:

$$S \cdot \mathbf{v} = 0 \tag{2}$$

Although the system is effectively closed to the passage of certain metabolites, others are allowed to enter or exit the system via exchange fluxes (or pseudoreactions [Clarke, 1980]). These fluxes generally do not represent biochemical conversions or transport processes like those of internal fluxes, but can be thought of as representing the inputs and outputs to the system. For example, the demand for a metabolite for incorporation into cellular biomass creates an exchange flux on the intracellular metabolite from the metabolic network. The availability of a substrate in the extracellular environment creates an exchange flux on the extracellular metabolite that represents a source. Thus, all metabolites are internal to the system and a distinction is made between intracellular and extracellular metabolites within the system, closing the material balance to all metabolites, as indicated in Eq. (2).

To complete the description of the metabolic network we must include the constraints that are placed on fluxes due to reaction thermodynamics and the systemic characteristics (input/output) of the network. To simplify matters, we can decompose all reversible internal reactions into a separate forward and reverse flux. This decomposition will create a general constraint that all internal fluxes must be greater than zero, such as in the directed graph of Figure 1a. For exchange fluxes we can set upper and lower bounds on the constraints that represent the corresponding metabolites' ability to enter or exit the system. (For a more detailed discussion see Schilling et al. [2000a].)

At this point, we introduce the reaction network shown in Figure 1a to illustrate these mathematical network descriptions and to provide a simplified example for many of the theoretical concepts that will be illustrated in what follows. The system is comprised of five metabolites that are operated on by seven internal fluxes. Note that fluxes v_3 and v_4 can be viewed as the decomposition of a reversible reaction into a forward and a backward reaction. Also, the stoichiometry on flux v_7 converts 2 mol of D into 1 mol of E. Therefore, metabolite C can be directly converted into E

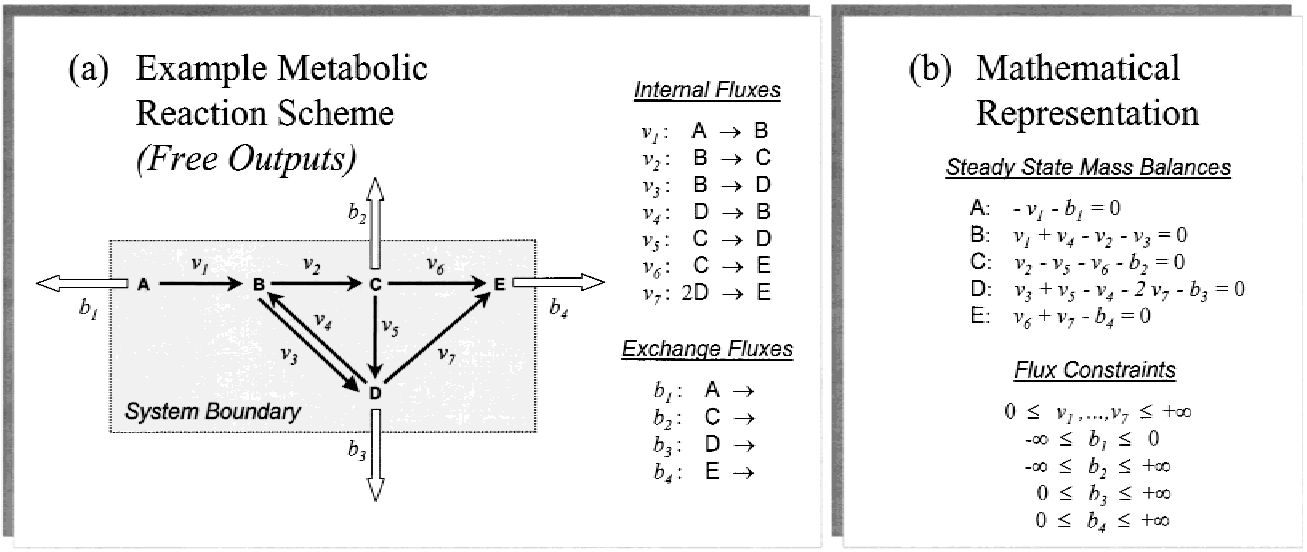

Figure 1. (a) Example metabolic reaction network consisting of 5 metabolites and 11 fluxes (7 internal fluxes, 4 exchange fluxes) listed along with the individual reactions. (b) Mathematical description of the reaction network in the form of steady-state mass balances (linear equalities) and flux constraints imposed on the network (linear inequalities). Together this system of linear equalities/inequalities defines the region of admissible flux vectors.

through v_6; however, if v_5 and v_7 are used, the efficiency is cut in half, analogous to a physiological situation in which carbon may be lost as CO_2 due to an alternative route to produce a metabolite. We allow metabolites C, D, and E to exit the system as products of the network (either as biomass precursors or final metabolic products). To represent substrates available to the system we allow metabolites A and C to enter the system. These potential sources and sinks on the system create four exchange fluxes operating on the system. Note that the exchange fluxes are defined as outward, thus the value of the flux is negative when the metabolite is entering the system and positive when it is formed as a product of the system. The steady-state mass balances are provided in Figure 1b along with the general constraints that are placed on both the internal and exchange fluxes. Each of the steady-state mass balances corresponds to a row of the stoichiometric matrix. This set of linear homogeneous equations and linear inequalities describes the metabolic system under steady state.

CONVEX ANALYSIS

As a result of the inequality constraints placed on various fluxes traditional linear algebra can no longer be used to handle such a mathematical system of equalities/inequalities, forcing the use of convex analysis to study the properties of the solution space for this problem. From convex analysis the solution space for any system of linear homogeneous equations and inequalities is a convex polyhedral cone emanating from the origin. We refer to this region of admissible steady-state flux vectors as the flux cone **(*C*)**. All possible solutions, and hence flux distributions, that the system can operate in a steady state are confined to the interior of the flux cone. To uniquely describe a convex cone we must identify the set of extreme rays/generating vectors $(\mathbf{p}_i)$ that span the cone (similar to the edges of a pyramid). In the context of metabolic systems we refer to these edges as extreme pathways, as each vector corresponds to a particular pathway or active set of fluxes that satisfy Eq. (2) and the inequality constraints placed on the system. Every point **(v)** within the cone can be written as a nonnegative linear combination of these extreme pathways as:

$$C = \left\{ \mathbf{v} : \mathbf{v} = \sum_{i=1}^{k} w_i \mathbf{p}_i,\ w_i \geq 0\ \forall i \right\} \tag{3}$$

Recently, we presented the detailed theoretical framework and algorithm used for the rapid identification of the set of extreme pathways for any metabolic system (Schilling et al., 2000a). It addition to being unique for a particular system the set of extreme pathways is the absolute minimal set of pathways that can be used to span the flux cone in a convex manner. This set of pathways generating the flux cone may also be referred to as a convex basis, and is more properly said to form the conical hull of the flux cone. As an aside, for those familiar with elementary modes, the extreme pathways are a subset of the elementary modes of a reaction network (Schuster et al., 1999). For a more thorough explanation of the similarities and differences between these approaches one is referred to Pfeiffer et al. (1999) and Schilling et al. (2000a).

The set of extreme pathways for the reaction network described in Figure 1 is provided in vector form in Table I with each pathway illustrated in Figure 2. There are eight extreme pathways ($\mathbf{p}_1, \ldots, \mathbf{p}_8$). Two of the pathways ($\mathbf{p}_7$, $\mathbf{p}_8$) display no active exchange fluxes, and thus correspond to internal cycles within the network of which $\mathbf{p}_8$ is simply a result of decomposing a reversible reaction into two opposite reactions. Using the classification scheme developed previously (Schilling et al., 2000a), these pathways correspond to type III pathways, whereas the remaining six pathways all describe functional pathways through the network (type I) and together represent the true systemic functions or capabilities of the network. Their combined activity can be used to describe any flux distribution of the network as in Eq. (3).

The set of extreme pathways is conically or systemically independent, in analogy to the concept of linear independence from linear algebra. Mathematically this means that the pathways cannot be formed by a positive combination of any other vectors or pathways in the flux cone. From a functional point of view this means that every flux distribution can be reached by either switching pathways off or turning them on to a certain level (this is analogous to having a dial controlling the activity of each extreme pathway). We will illustrate this notion in more detail in the following sections.

EFFECTS OF BALANCED DEMANDS ON PATHWAY STRUCTURE

Under changing substrate/supply conditions metabolic networks are continuously faced with a balanced set of biosynthetic demands (i.e., production of amino acids, nucleotides, phospholipids, as well as energy and redox potential). This means that the network must generate a balanced ac-

Table I. The eight extreme pathway vectors for the example network with free outputs.[a]

Pathway number	Internal fluxes							Exchange fluxes			
	v_1	v_2	v_3	v_4	v_5	v_6	v_7	b_1	b_2	b_3	b_4
$\mathbf{p}_1$	1	1	0	0	0	0	0	–1	1	0	0
$\mathbf{p}_2$	1	0	1	0	0	0	0	–1	0	1	0
$\mathbf{p}_3$	2	0	2	0	0	0	1	–2	0	0	1
$\mathbf{p}_4$	0	0	0	0	1	0	0	0	–1	1	0
$\mathbf{p}_5$	0	0	0	0	2	0	1	0	–2	0	1
$\mathbf{p}_6$	0	0	0	0	0	1	0	0	–1	0	1
$\mathbf{p}_7$	0	0	1	1	0	0	0	0	0	0	0
$\mathbf{p}_8$	0	1	0	1	1	0	0	0	0	0	0

[a]The first six pathways are type I pathways (see Fig. 2). The last two pathways are type III pathways that represent internal cycles in the network (i.e., zero weights on all exchange fluxes). For definition of the pathway classification scheme see Schilling et al., (2000a).

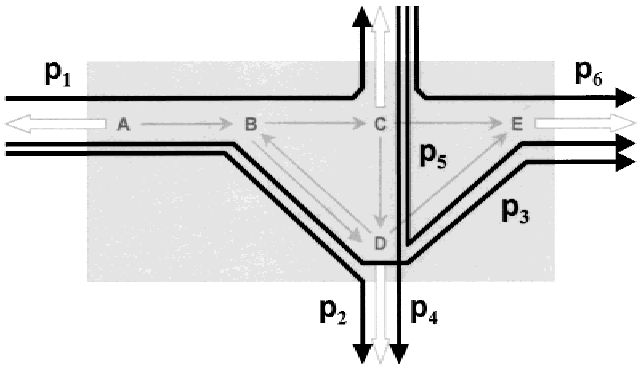

Figure 2. Routing patterns for the six type I extreme pathways calculated for the example reaction network. Detailed pathway vectors indicating the individual flux levels in each pathway are provided in Table I.

tivity through the exchange fluxes for the particular metabolites required to meet these demands. What are the implications of these aggregate demands on the pathway structure of a network?

To assess the systemic performance of a network in meeting biosynthetic demands an exchange flux is introduced into a network representing the aggregate relative demand of the biomass precursors (analogous to an objective function [Varma and Palsson, 1993a]). Additional constraints must also be added to the network to effectively close the material balances on the metabolites participating in the biosynthetic demand (or growth) flux. The introduction of a new exchange flux and the associated restriction of existing exchange fluxes will alter some of the mass balances and linear inequalities of the network. A new pathway structure can be determined based on these changes to the systemic constraints that is functionally related to the original set of pathways, as demonstrated in what follows. To distinguish between the two different forms of output for a system [(1) all material balances closed with an aggregate demand/growth flux included versus (2) no aggregate demand/growth flux and material balances not closed on biosynthetic precursor outputs] we consider the system without a biosynthetic demand flux (as discussed in the previous section) to have *free* outputs, and the system with balanced network demands is referred to as having *linked* outputs.

For the example system in Figure 1, we introduce an exchange flux that represents an aggregate demand on the network requiring 1 mol of metabolite C and D along with 2 mol of metabolite E. This flux serves as a balanced drain on the network and is written:

$$b_z: C + D + 2E \rightarrow \qquad (4)$$

Note that the growth flux only involves metabolites that were previously allowed to exit in the system with free outputs and subsequently to have a preexisting exchange flux. Otherwise, the input and output characteristics of the two representations are no longer compatible and the pathway results will not be complementary. Furthermore, the exchange fluxes for C, D, and E are no longer allowed to serve as direct outputs for the system. This condition will force b_2 to take on a negative value while constraining b_3 and b_4 to zero (see Fig. 3).

The set of extreme pathways for this system is provided in Table II (the schematic illustrations of each pathway are available on-line). We can see that there are ten pathways in total (eight functional pathways of type I, and two type III cyclic pathways identical to those from the free output system in Table I). All functional pathways utilize either A or C, or A and C, to produce the balanced set of demands represented by the growth flux.

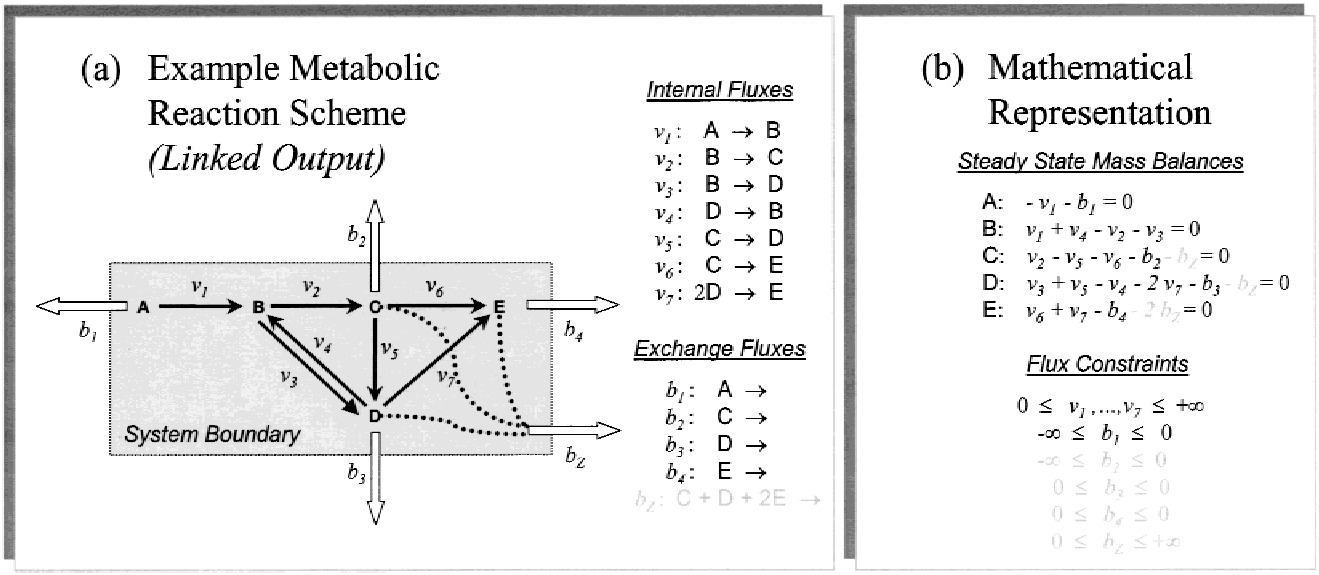

Figure 3. (a) The example network presented in Figure 1 with the addition of an aggregate demand flux that requires 1 mol of metabolite C and D for every 2 mol of metabolite E. (b) Mathematical description of the network with the changes from Figure 1b indicated in gray type. All changes are a result of the inclusion of the demand exchange flux (a linked output) and the associated alterations to the exchange flux constraints of the participating metabolites.

Table II. The ten extreme pathway vectors for the example network in Figure 1 with a linked output.[a]

Pathway number	Internal fluxes							Exchange fluxes					Pathway equivalences linked ~ free
	v_1	v_2	v_3	v_4	v_5	v_6	v_7	b_1	b_2	b_3	b_4	b_z	
$\mathbf{p}'_1$	5	0	5	0	0	0	2	–5	–1	0	0	1	$\mathbf{p}'_1 \sim \mathbf{p}_2 + 2\mathbf{p}_3$
$\mathbf{p}'_2$	1	0	1	0	0	2	0	–1	–3	0	0	1	$\mathbf{p}'_2 \sim \mathbf{p}_2 + 2\mathbf{p}_6$
$\mathbf{p}'_3$	0	0	0	0	5	0	2	0	–6	0	0	1	$\mathbf{p}'_3 \sim \mathbf{p}_4 + 2\mathbf{p}_5$
$\mathbf{p}'_4$	0	0	0	0	1	2	0	0	–4	0	0	1	$\mathbf{p}'_4 \sim \mathbf{p}_4 + 2\mathbf{p}_6$
$\mathbf{p}'_5$	6	1	5	0	0	0	2	–6	0	0	0	1	$\mathbf{p}'_5 \sim \mathbf{p}_1 + \mathbf{p}_2 + 2\mathbf{p}_3$
$\mathbf{p}'_6$	4	3	1	0	0	2	0	–4	0	0	0	1	$\mathbf{p}'_6 \sim 3\mathbf{p}_1 + \mathbf{p}_2 + 2\mathbf{p}_6$
$\mathbf{p}'_7$	6	6	0	0	5	0	2	–6	0	0	0	1	$\mathbf{p}'_7 \sim 6\mathbf{p}_1 + \mathbf{p}_4 + 2\mathbf{p}_5$
$\mathbf{p}'_8$	4	4	0	0	1	2	0	–4	0	0	0	1	$\mathbf{p}'_8 \sim 4\mathbf{p}_1 + \mathbf{p}_4 + 2\mathbf{p}_6$
$\mathbf{p}'_9$	0	0	1	1	0	0	0	0	0	0	0	0	$\mathbf{p}'_9 \sim \mathbf{p}_7$
$\mathbf{p}'_{10}$	0	1	0	1	1	0	0	0	0	0	0	0	$\mathbf{p}'_{10} \sim \mathbf{p}_8$

[a]The first eight pathways correspond to type I pathways, whereas the last two pathways are type III. Pathway equivalencies between the free and linked output representations are provided for each pathway.

If only the growth flux is added to the free output system, and no additional constraints are placed on the exchange fluxes, the set of extreme pathways would include all of the extreme pathways listed in Tables I and II (16 total). By adding one additional flux the dimension in which the flux cone resides has increased by one (from $\mathbb{R}^{11}$ to $\mathbb{R}^{12}$), and the cone has gained six additional edges. If we project the flux cone down onto the original dimensions in $\mathbb{R}^{11}$ it is seen that the six additional pathways all reside within the interior of the flux cone described by the pathway vectors in Table I. Importantly, these new pathways can be interpreted as positive combinations of the extreme pathways from the free output representation that generates the cone as described in Eq. (3).

To determine the appropriate combination of extreme pathways we have to redistribute the values for the growth flux back to the original exchange fluxes of the various metabolites. This redistribution can be achieved by multiplying the growth flux value by the stoichiometric coefficient of each of the metabolites in the reaction and then adding this value to the appropriate exchange fluxes. The values for the internal fluxes remain unchanged. This operation is purely mathematical and does not affect the functionality of the system. Thus, for pathway $\mathbf{p}'_5$ the value for the exchange fluxes would shift from [–6,0,0,0,1] for b_1, b_2, b_3, b_4, and b_z to [–6,1,1,2,0]. From this we can determine that pathway ($\mathbf{p}'_5$) for the linked output system is equivalent to the following combination of pathways for the free output system: $\mathbf{p}_1 + \mathbf{p}_2 + 2 * \mathbf{p}_3$. The equivalencies between all of the free and linked output pathways are provided in Table II.

Outlining the pathway equivalencies offers the ability to interpret flux distributions in terms of the pathways for the linked outputs that represent global metabolic routing patterns used to meet balanced sets of demands, and then further decomposes these patterns into the individual pathways that comprise such a distribution pattern. These complementary perspectives for interpreting metabolic function will be illustrated in the following two sections where we focus on the interpretation of optimal flux distribution calculated using flux balance analysis.

CALCULATING OPTIMAL FLUX DISTRIBUTIONS

The performance capabilities of any metabolic network reside in the flux cone that we have now seen described by the set of extreme pathways. In fact, the answer to any question related to the general structure and fitness of the network lies within this cone. The pathways offer a convenient way of interpreting metabolic function, but how do we best explore the capabilities and functioning of a metabolic network?

One approach that has recently been used to explore the relationship between the metabolic genotype and phenotype for a number of organisms is flux balance analysis (Edwards and Palsson, 1998, 1999, 2000; Schilling et al., 1999a, 2000). This approach uses linear optimization techniques to determine the optimal flux distributions within a network so as to maximize/minimize a particular objective function. The standard form of a linear programming problem is as follows, where a linear objective function is maximized or minimized subject to a series of linear equalities and inequality constraints:

$$\begin{aligned} &\text{Maximize/minimize} \quad Z = \sum_j c_j v_j \\ &\text{subject to} \quad \sum_{i,j} S_{ij} v_j = b_i, \quad \alpha_j \le v_j \le \beta_j \end{aligned} \tag{5}$$

Note that Eq. (5) is analogous to the system of linear equalities/inequalities that forms the general description of the metabolic network in Figure 1, where all the components of the **b** vector are zero. In fact, this system of equations is merely a special case of the general system defined by the mass balances and physicochemical constraints (as shown in Fig. 1b). Linear programming will determine the optimal value for an objective that lies within the flux cone spanned by the extreme pathways. Furthermore, the foundations of linear programming are built on the principles of convex analysis, which can be used to link the pathway analysis of metabolic networks outlined earlier with flux balance analysis.

For metabolic applications, the linear objective function

(Z) to be maximized or minimized can correspond to a number of diverse objectives ranging from particular metabolic engineering design objectives (energy and metabolite production) to the maximization of cellular growth (in a manner similar to that seen in the previous section). Regardless of the objective function chosen, the optimal solutions will lie within the flux cone that is defined by the mass balance and physicochemical constraints (i.e., linear inequalities) placed on the system. Otherwise, the solution to the optimization problem will be infeasible. This solution space is virtually the exact same flux cone that we have just seen described in the previous section in terms of the set of extreme pathways for the linked output system. The only difference is that the flux cone must become bounded (a bounded polytope) to calculate an optimal solution.

We consider the growth exchange flux introduced in Eq. (4) as the objective function. To determine the optimal utilization of the metabolic network (so as to maximize the objective function), additional physicochemical constraints must be considered. Specifically, in this example system, the upper bound on the exchange fluxes for metabolites A and C must be set to limit the amount of metabolites A and C available for intake by the network. Physiologically, these constraints may correspond to limited substrate availability or maximal uptake rates or even maximum rates of diffusion-mediated transport. These maximum capacity constraints serve to place certain bounds on the flux cone such that it is no longer an unbounded polyhedral cone. Regions of the flux cone become bounded by the hyperplanes defined from these constraints. Without these constraints the maximum value of any flux in the network could be infinite, and the linear programming problem is unbounded.

Mathematically, the solution space is the sum of a convex polytope (closed surface) and a new polyhedral cone (unbounded) formed by a subset of the original extremal rays (see Appendix for details). The conical hull of the polyhedral cone now corresponds to the pathways that are unaffected by the added constraints, and the convex hull of the polytope is essentially generated from the remaining pathways that have now been constrained. Importantly, all of the vertices of the polytope correspond to bounded feasible solutions of the linear programming problem, whereas the edges of the flux cone correspond to the unbounded feasible solutions. Both the polytope and the polyhedral cone are contained within the original flux cone, which allows for the interpretation of any point in the solution space using the original set of extreme pathways from the linked output system as well as the free output system through their equivalencies, as we will now illustrate.

LINKING FLUX BALANCE ANALYSIS WITH PATHWAY ANALYSIS

Consider three cases that represent different environmental and internal challenges that the example network may face in meeting its demands: case 1, only metabolite A is available to the network to meet the demands of the growth flux in Eq. (4); case 2, only metabolite A is available, but the internal flux, v_6, is not functional and is thus constrained to equal zero; and case 3, only metabolite C is available to the network to meet the demands of the growth flux. In each case, we will arbitrarily constrain the input of the available substrate metabolite to a value of 1 (to obtain substrate normalized distributions). We can consider the shift from case 1 to case 2 as a sudden challenge to the internal structure of the network due to a loss of function, and the shift from case 1 to case 3 represents an environmental challenge, namely a change in substrate availability.

Using a commercially available linear programming package (Lindo Systems Inc., Chicago, IL) the optimal flux distributions for these three cases were calculated and are presented in Figure 4. We see that the loss in function of v_6 in case 2 decreases the maximum value of the objective function by 33% from 0.25 to 0.17, essentially forcing the network to perform suboptimally. There is a subsequent shift in the utilization of the metabolic pathways that occurs between case 1 and case 2, and to a more severe extent in the shift from case 1 to case 3 due to the change in substrate. It is these shifts in metabolic routing that we seek to interpret from a pathway perspective using the set of extreme pathways and equivalencies provided in Table II.

In each case, the flux distribution corresponds to an exact scalar multiple of one of the extreme pathways for the linked output system in Table II, implying that the solutions are on the extreme edges of the flux cone. As the environmental/genetic conditions change, the corresponding optimal pathway utilizations shift accordingly across faces of the flux cone. If a reaction is eliminated or constrained to zero, all of the pathways that utilize this reaction will be eliminated. From basic principles of vector addition it becomes apparent that the utilization or weight (w_i) on an extreme pathway in comprising a flux distribution must equal zero if the pathway utilizes the internal flux/reaction that has been eliminated. Therefore, every extreme pathway in the system that has a positive flux value for the eliminated reaction will have $w_i = 0$ in Eq. (3). This is the situation in case 2 where all of the pathways that can produce metabolite E optimally have been eliminated by the loss of v_6, leading to the utilization of other pathways to meet the demands of the system.

The changes in pathway utilization allow for the interpretation of metabolic shifts in terms of the underlying pathways of the network. These example cases offer an illustration of how pathways may be switched on and off to achieve different flux distributions in a network. Using the equivalencies in Table II we can generate pathway signatures or profiles based on the utilization of pathways for the free output representation as shown in the lower portion of Figure 4. These signatures offer the visual interpretation of pathways being "switched" on and off in order to establish the optimal flux distributions. Taken together, this illustrates how flux distributions and associated metabolic phenotypes can be interpreted from a pathway-based perspec-

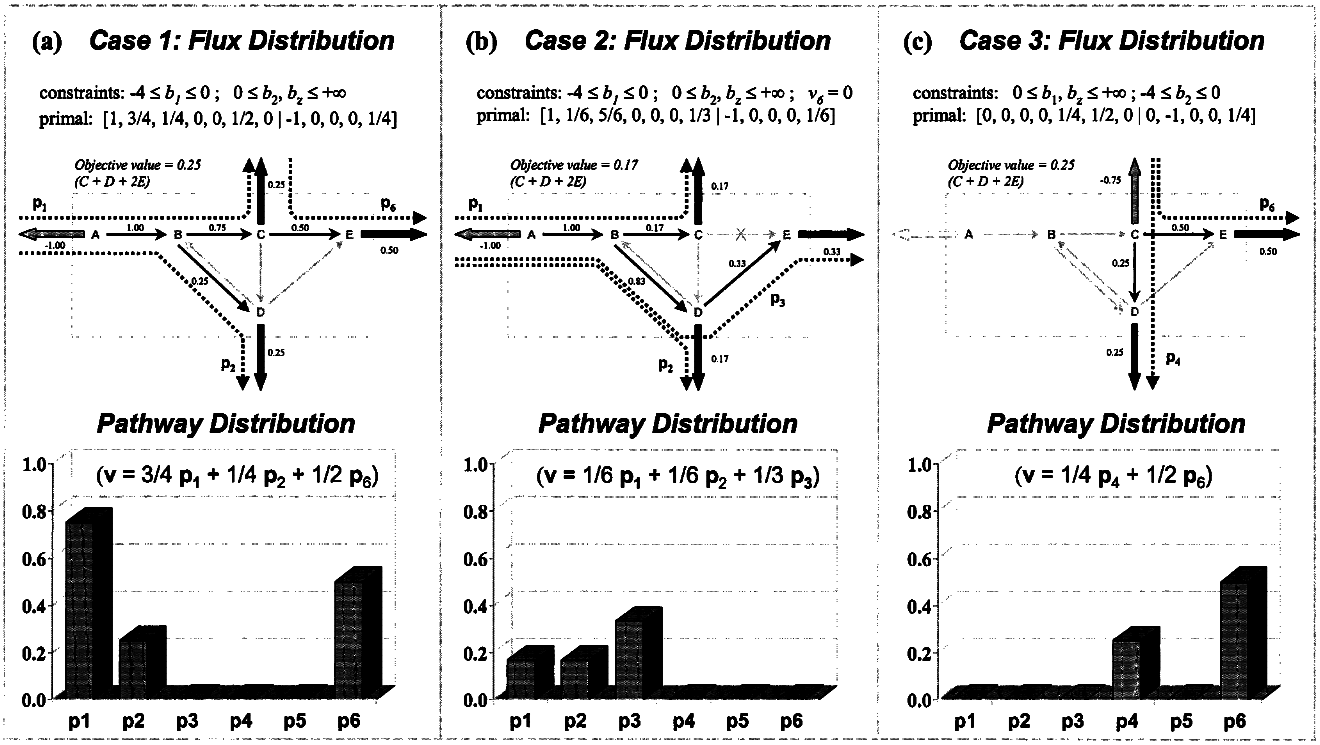

Figure 4. Flux distributions and pathway signatures for the three cases considered. (a) For case 1, only metabolite A is allowed to enter the system leading to an optimal yield of 0.25 with the complete primal solution to the optimization provided. Flux levels through b_2 are redistributed back to the original exchange fluxes and the flux distribution is illustrated. The extreme pathways for the free output system utilized to achieve the flux distribution are overlaid onto the reaction network with their relative magnitudes provided in the pathway signature charts. (b) In case 2, the internal flux, v_6, is restricted to zero with metabolite A remaining as the only substrate. (c) For case 3, the activity of v_6 is restored while the substrate availability is switched from metabolite A to C.

tive. To return to a reaction-based description, it is only necessary to perform a coordinate transformation by multiplying the weights of the pathways by the pathway vectors.

From pathway analysis we arrive at a fundamentally different perspective from which to view metabolic behavior such as that predicted through the use of flux balance analysis. In addition to viewing the activity of individual reactions we can effectively interpret metabolic function in terms of the activity of individual pathways. The aforementioned example is useful in illustrating many of the theoretical concepts for a pathway interpretation of metabolic function and, in what follows, we extend these concepts to analyze complex metabolic networks. Although linear optimization is used to calculate flux distributions in all of the subsequent examples, the concepts of pathway representations of flux distributions are general to any flux distribution regardless of whether it is derived experimentally or theoretically.

APPLICATION TO *E. COLI* CENTRAL METABOLISM

The reconstruction of metabolic networks from annotated genome sequence information is currently an active area of research. From these reconstructions it is possible to study the system characteristics and capabilities of metabolic networks and explore the metabolic genotype–phenotype relationship in fully sequenced organisms using both flux balance analysis and genome-scale pathway analysis. Using flux balance analysis, in silico studies on the systemic properties of the *Haemophilus influenzae* (Edwards and Palsson, 1999) and *E. coli* (Edwards and Palsson, 2000) metabolic networks have recently been completed. In addition, pathway analysis has now been successfully applied to assess the production capabilities and general fitness of a reconstructed metabolic network for *H. influenzae* (Schilling and Palsson, 2000b). In what follows, we will consider a subsystem of the *E. coli* metabolic genotype comprised of the central metabolic pathways. Flux balance analysis will be used to explore the metabolic capabilities and predicted functions of this network under various substrate conditions. Then flux balance analysis will be combined with pathway analysis to provide a complete view of the capabilities and steady-state behavior of the metabolic network from a pathway-oriented perspective.

The schematic illustration of the reactions comprising the central metabolic network of *E. coli* examined in what follows is shown in Figure 5 (the complete table of reactions is provided on-line). Briefly, the network is comprised of the complete set of glycolytic reactions, the pentose phosphate shunt, and the tricarboxylic acid (TCA) cycle without the glyoxylate shunt, along with the necessary transport reac-

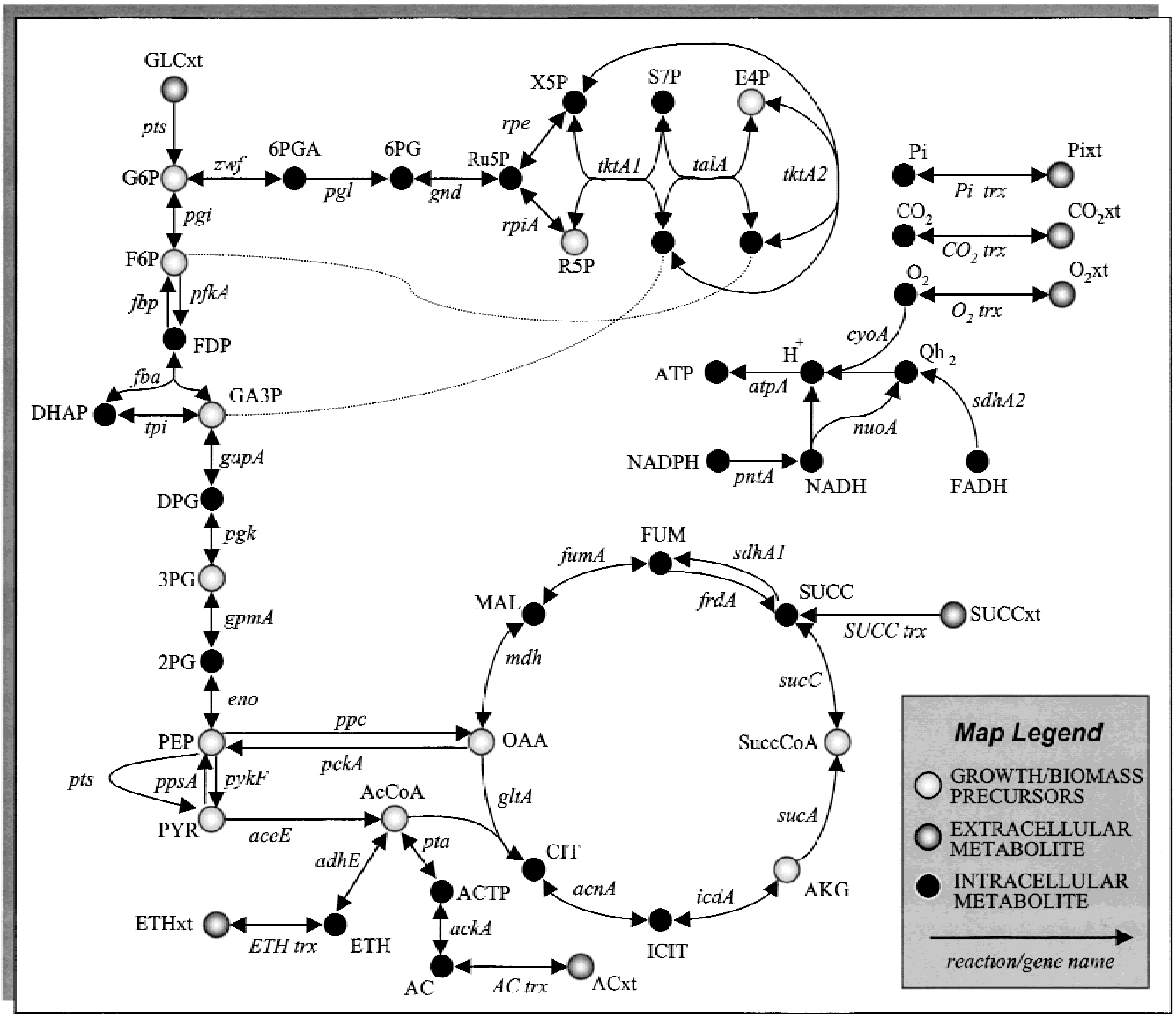

Figure 5. Central metabolic network of *E. coli* considered in the pathway and flux balance analysis. The system is comprised of 53 metabolites, 78 internal fluxes, and 8 exchange fluxes. A complete list of the net reaction capable of occurring in the network is provided on-line with complete reaction names.

tions required to import and export metabolites from the extracellular space. The necessary reactions of the electron transport chain are included with the P/O ratio set to 4/3. The system boundary is drawn around all of the reactions in the network including the transport reactions. Exchange fluxes have been included to allow for the supply of glucose and succinate as carbon sources available to the system, while ethanol and acetate are allowed byproducts of the system. Oxygen, carbon dioxide, and inorganic phosphate are free to enter and exit the system as needed. (Note that this network does not completely describe central metabolism in *E. coli*. We have chosen this representation as a compromise between successful representation of the basic aspects of central metabolism and providing a useful example of the salient features of this combined approach for studying metabolic systems.)

Pathway Analysis for the Free Output System

To study the pathway structure of the system with free outputs it is necessary to include exchange fluxes for all of the necessary precursors and cofactors that are utilized to generate biomass. In this example we define the metabolic demands as those provided in Table III, which were studied in an earlier model of *E. coli* central metabolism (Varma and Palsson, 1993a). The metabolic demands considered for each metabolite are used to generate the growth flux used when considering the linked output system. We focus first on the free output system without the aggregate demand flux.

For the case of free outputs we introduce an exchange flux for each of the biosynthetic precursors that is constrained only to allow the metabolite to exit the system.

Table III. Metabolic demands of precursors and energy/redox cofactors required for the generation of 1-g biomass yield as used in Varma and Palsson (1993a) and estimated from cellular composition data (Ingraham et al., 1983).[a]

Metabolite	Demand (mmol)
ATP	41.3
NAD^+	3.5
NADPH	18.2
G6P	0.2
F6P	0.1
R5P	0.9
E4P	0.4
GA3P	0.1
3PG	1.5
PEP	0.5
PYR	2.8
ACCOA	3.7
OXA	1.8
AKG	1.1
SUCCOA	(Trace)

[a]These demands do not include energy requirements for non–growth-associated maintenance. The growth flux is a summation of the drains of each metabolite multiplied by the demand weights given in the table.

Unconstrained exchange fluxes are added for currency metabolites such as ATP, ADP, AMP, NADH, and NADPH. Therefore, we expect to generate extreme pathways that lead directly from a substrate to a particular precursor or metabolic byproduct (type I pathways) along with a number of pathways and cycles that interconvert only cofactors and currency metabolites (type II pathways). In the case where glucose is considered as the only available substrate the total number of extreme pathways is 1093 (1053 type I, 10 type II, and 30 type III). For the case with succinate as the only available substrate, the number of pathways reduces to 195 extreme pathways (155 type I, 10 type II, 30 type III). These sets of pathways contain all of the optimal as well as suboptimal conversion ratios of each substrate to all of the precursors and can be used to interpret the specific capabilities of the network and other systemic properties, as previously demonstrated for *H. influenzae* (Schilling and Palsson, 2000b). However, our interest is in studying the phenotype or the utilization of the network to meet the balanced set of precursor demands required to generate biomass, rather than focusing on the production of particular precursors of interest. Therefore, we now examine the linked output system using pathway analysis initially, and then combine this with flux balance analysis.

Pathway Analysis for the Linked Output System

For the case of linked outputs a growth flux is introduced that represents the metabolic demands presented in Table III for generating 1 g of dry weight (DW) biomass. We also remove the exchange fluxes for the precursors and currency metabolites that were introduced in the free output system, thus closing the mass balances on all of these metabolites. The system now has eight exchange fluxes (extracellular glucose, succinate, ethanol, acetate, carbon dioxide, oxygen, inorganic phosphate, and the intracellular biomass). In all further calculations, glucose and succinate are constrained to only enter the system, whereas ethanol, acetate, and biomass are constrained to only exit the system. The remaining exchange fluxes are unconstrained (oxygen, carbon dioxide, and inorganic phosphate). With the decomposition of reversible internal fluxes into a forward and reverse reaction the system is comprised of 78 internal fluxes and 8 exchange fluxes. The extreme pathways for the linked output system will involve flux distribution patterns that the system can use to generate biomass, completely oxidize the carbon source, or convert it into one of the byproducts.

First, we consider glucose to be the only substrate available to the network. In this case, there are a total of 115 extreme pathways (85 type I, 30 type III). The type III pathways are all identical to those of the free output system and no type II pathways remain as all of the cofactors are now forced to balance. There are four futile cycles in the system (sets of reactions that can utilize and convert ATP to ADP and inorganic phosphate), those sets being (*pfkA/fbp*), (*pckA, ppc*), (*pykF, ppsA, adk*), and the ATP drain flux, which is used to represent non–growth-associated maintenance.

After careful inspection of the 85 type I pathways it is apparent that many of them are in sets of four pathways that are identical, except for the utilization of a different futile cycle for the dissipation of ATP so as to balance the net production and consumption of ATP. All of these pathways in a set exhibit the same values for the exchange fluxes, illustrating their functional equivalence. To generate a reduced set of pathways that represents the full capabilities of the network we retain only pathways from these sets that utilized the ATP drain flux instead of one of the three futile cycles listed earlier. The type III pathways, which are mainly a consequence of the decomposition of reversible reactions into a forward and a reverse reaction, are also removed from consideration, as they show no activity in the exchange fluxes. Following these reductions we are left with 30 pathways. Table IV provides the details for the exchange fluxes of these pathways normalized to the input value for the substrate glucose.

The pathway analysis was also performed with succinate as the sole carbon source for the system, generating the complete set of 66 extreme pathways (36 type I, 30 type III). Following the same simplification procedure just described, a reduced set of 12 pathways is generated from the complete set (Table V). All of the detailed pathway vectors for the complete and reduced sets in both cases are provided online.

For both glucose and succinate as the sole carbon source there are pathways that generate biomass and either produce CO_2, CO_2 and acetate, or CO_2 and ethanol as byproducts. Similar pathways also exist for generating these byproducts without generating biomass. It is important to note that the network does not allow for the generation of formate or succinate as metabolic byproducts, due to the increased

Table IV. Functional characteristics of the reduced set of 30 extreme pathways calculated for glucose as the sole carbon for the *E. coli* metabolic network.[a]

Pathway number	Exchange fluxes[b]							Net pathway reaction balance
	GLCxt/GLCxt	ETHxt/GLCxt	ACxt/GLCxt	GRO/GLCxt	PIxt/GLCxt	CO2xt/GLCxt	O2xt/GLCxt	
2	–1.000	0	0	0.105	–0.390	1.498	–1.376	GLCxt + 0.39 PIxt + 1.376 O2xt ⇒ 0.105 GRO + 1.498 CO2xt
24	–1.000	0	0	0.105	–0.388	1.522	–1.402	GLCxt + 0.388 PIxt + 1.402 O2xt ⇒ 0.105 GRO + 1.522 CO2xt
21	–1.000	0	0	0.104	–0.383	1.575	–1.455	GLCxt + 0.383 PIxt + 1.455 O2xt ⇒ 0.104 GRO + 1.575 CO2xt
23	–1.000	0	0.125	0.102	–0.376	1.410	–1.293	GLCxt + 0.376 PIxt + 1.293 O2xt ⇒ 0.102 GRO + 1.41 CO2xt + 0.125 ACxt
38	–1.000	0	0.178	0.101	–0.373	1.342	–1.226	GLCxt + 0.373 PIxt + 1.226 O2xt ⇒ 0.101 GRO + 1.342 CO2xt + 0.178 ACxt
4	–1.000	0	0	0.099	–0.367	1.760	–1.646	GLCxt + 0.367 PIxt + 1.646 O2xt ⇒ 0.099 GRO + 1.76 CO2xt
42	–1.000	0.199	0	0.096	–0.357	1.482	–1.172	GLCxt + 0.357 PIxt + 1.172 O2xt ⇒ 0.096 GRO + 1.482 CO2xt + 0.199 ETHxt
39	–1.000	0	0.275	0.095	–0.353	1.377	–1.267	GLCxt + 0.353 PIxt + 1.267 O2xt ⇒ 0.095 GRO + 1.377 CO2xt + 0.275 ACxt
44	–1.000	0.151	0.124	0.095	–0.353	1.377	–1.116	GLCxt + 0.35 PIxt + 1.12 O2xt ⇒ 0.095 GRO + 1.38 CO2xt + 0.15 ETHxt + 0.12 AC
22	–1.000	0.255	0	0.093	–0.344	1.516	–1.154	GLCxt + 0.344 PIxt + 1.154 O2xt ⇒ 0.093 GRO + 1.516 CO2xt + 0.255 ETHxt
43	–1.000	0.310	0	0.091	–0.336	1.503	–1.089	GLCxt + 0.336 PIxt + 1.089 O2xt ⇒ 0.091 GRO + 1.503 CO2xt + 0.31 ETHxt
29	–1.000	0.330	0	0.088	–0.326	1.577	–1.145	GLCxt + 0.326 PIxt + 1.145 O2xt ⇒ 0.088 GRO + 1.577 CO2xt + 0.33 ETHxt
32	–1.000	0	0.330	0.088	–0.326	1.577	–1.476	GLCxt + 0.326 PIxt + 1.476 O2xt ⇒ 0.088 GRO + 1.577 CO2xt + 0.33 ACxt
35	–1.000	0	0	0.088	–0.326	2.238	–2.137	GLCxt + 0.326 PIxt + 2.137 O2xt ⇒ 0.088 GRO + 2.238 CO2xt
45	–1.000	0.595	0	0.078	–0.287	1.493	–0.808	GLCxt + 0.287 PIxt + 0.808 O2xt ⇒ 0.078 GRO + 1.493 CO2xt + 0.595 ETHxt
26	–1.000	0	0	0.070	–0.261	2.993	–2.912	GLCxt + 0.261 PIxt + 2.912 O2xt ⇒ 0.07 GRO + 2.993 CO2xt
3	–1.000	0	0	0.068	–0.251	3.105	–3.027	GLCxt + 0.251 PIxt + 3.027 O2xt ⇒ 0.068 GRO + 3.105 CO2xt
18	–1.000	0	0	0.066	–0.243	3.200	–3.125	GLCxt + 0.243 PIxt + 3.125 O2xt ⇒ 0.066 GRO + 3.2 CO2xt
10	–1.000	0	0	0.063	–0.235	3.289	–3.216	GLCxt + 0.235 PIxt + 3.216 O2xt ⇒ 0.063 GRO + 3.289 CO2xt
14	–1.000	0	0	0.062	–0.229	3.359	–3.288	GLCxt + 0.229 PIxt + 3.288 O2xt ⇒ 0.062 GRO + 3.359 CO2xt
46	–1.000	1.000	0	0	0	4.000	–3.000	GLCxt + 3.0 O2xt ⇒ 4.0 CO2xt + 1.0 ETHxt
49	–1.000	0	1.000	0	0	4.000	–4.000	GLCxt + 4.0 O2xt ⇒ 4.0 CO2xt + 1.0 ACxt
52	–1.000	0	0	0	0	6.000	–6.000	GLCxt + 6.0 O2xt ⇒ 6.0 CO2xt
57	–1.000	2.000	0	0	0	2.000	0	GLCxt ⇒ 2.0 CO2xt + 2.0 ETHxt
61	–1.000	0	2.000	0	0	2.000	–2.000	GLCxt + 2.0 O2xt ⇒ 2.0 CO2xt + 2.0 ACxt
65	–1.000	0	0	0	0	6.000	–6.000	GLCxt + 6.0 O2xt ⇒ 6.0 CO2xt
69	–1.000	0	0	0	0	6.000	–6.000	GLCxt + 6.0 O2xt ⇒ 6.0 CO2xt
72	–1.000	0	1.667	0	0	2.667	–2.667	GLCxt + 2.667 O2xt ⇒ 2.667 CO2xt + 1.667 ACxt
75	–1.000	1.667	0	0	0	2.667	–1.000	GLCxt + 1.0 O2xt ⇒ 2.667 CO2xt + 1.667 ETHxt
84	–1.000	0	0	0	0	6.000	–6.000	GLCxt + 6.0 O2xt ⇒ 6.0 CO2xt

[a]Pathways are ordered based on activity of the growth flux normalized to the glucose uptake. Pathway numbers coincide with the original numbers of the pathway vectors retained from the complete set (available on-line). All exchange flux values are normalized to the glucose intake level (negative values are relative uptake ratios, positive values are production ratios).

[b]GLC, glucose; ETH, ethanol; AC, acetate; PI, inorganic phosphate; CO2, carbon dioxide; O2, oxygen; GRO, biomass/growth flux (see Table III).

number of pathways that would be generated based on these conditions. This increase in pathway number is sufficient enough to distract from the main theoretical issues being illustrated herein, and was therefore not considered. For glucose we see that the maximal yield of biomass in any pathway is 0.105 g DW/mmol glucose (0.583 g DW/g glucose), whereas the maximal yield of succinate is 0.051 g DW/mmol (0.432 g DW/g succinate). These yields cannot be exceeded, as the combination of any set of extreme pathways according to Eq. (3) will not surpass the ratio of biomass/substrate of the optimal pathway. In each case, the minimum biomass yield is zero, as pathways do exist that do not produce biomass. For glucose, the range of oxygen consumption to substrate uptake varies from 0.0 to 6.0 mol O_2/mol glucose, whereas, for succinate, the range narrows from 0.5 to 3.5 mol O_2/mol succinate. Any oxygen/substrate consumption ratios outside of these ranges are infeasible. Other similar systemic constraints can be defined from these pathways that will set boundaries on the functional capabilities of the network. Recall that these sets of extreme pathways define the edges of the flux cone that contain all of the possible metabolic phenotypes possibly displayed by the network. Therefore, it is possible to bracket the behavior of the network from the pathway analysis. Next, we illustrate how the pathways can be used to gain a deeper understanding of shifting metabolic behaviors as determined using flux balance analysis.

FLUX BALANCE ANALYSIS AND PHENOTYPIC PHASE PLANES

Flux balance analysis can be used to examine quantitatively the linked output system detailed earlier. Geometrically, the constraints that are imposed on the input values of the exchange fluxes for the substrates will bound the flux cone defined by the extreme pathways; thus resulting in a bounded polyhedron. Optimal solutions within this space will then lie on a vertex of the polyhedron. These are the bounded feasible solutions of the linear programming problem. First, we calculate the optimal flux distributions for growth on glucose and succinate normalized to 1 mmol of substrate. Illustrations of the optimal flux distributions as calculated for both conditions are provided on-line. The optimal biomass yields determined with flux balance analysis are 0.105 g DW/mmol glucose and 0.051 g DW/mmol succinate, which are identical to the optimal yield calculated from the pathways (pathway #2 for glucose, and pathway #1 for succinate). This result reveals that the optimal solution lies directly on the vertex of the polyhedron that is defined by the extreme pathway containing the optimal ratio of biomass/substrate. In each case, the inclusion of any other pathway into the solution will decrease the yield from the maximum ratios indicated earlier. (For comparison to unpublished data, the experimental optimal yield that we have achieved for growth of *E. coli* K-12 in batch culture on glucose is 0.096 g DW/mmol [n = 6, SD = 0.010]. For growth on succinate the optimal yield is 0.050 g DW/mmol [n = 13, SD = 0.006].)

All metabolic phenotypes (metabolic flux distributions) attainable from a defined metabolic system are mathematically confined to the flux cone. Flux balance analysis is used to search through the flux cone for a solution that maximizes/minimizes a given objective function. It has been shown that, under nutritionally rich growth conditions (i.e., cell is not starved for phosphate or nitrogen), *E. coli* grows in a stoichiometrically optimal manner (Varma and Palsson, 1994a); thus, setting the growth flux as the objective function produces physiologically meaningful results. The optimal flux distribution is only meaningful when interpreted in terms of the specific environmental conditions. Therefore, phenotype phase planes (PhPPs) have been calculated to define the range of optimal phenotypic behavior in the context of a number of environmental parameters (Edwards and Palsson, 1999; Varma et al., 1993b). These phase planes illustrate how optimal phenotypic behavior is dependent on the environmental conditions, and can be used to delineate fundamentally different regions of metabolic behavior in a multiparameter space. In what follows, we show that the demarcations on the PhPP that are generated by flux balance analysis (FBA) and linear programming are identical to the projections of the extreme pathways onto the relevant subspace.

The methodology for defining PhPPs has been described by Edwards et al. (currently in review by B & B). Herein, we briefly describe the construction of PhPPs in the context of the metabolic pathway analysis. The exchange flux or nutrient uptake rates for two metabolites (such as the carbon substrate and oxygen) can form two axes on an (x,y)-plane (these exchange fluxes were two unit vectors in $\mathbb{R}^n$). The optimal metabolic flux distribution is calculated for all points in this plane using linear programming. Each of the two exchange fluxes for the two metabolites considered in the phase plane is set to equal a precise value based on the coordinates in the plane, and optimal solutions are calculated. In other words, the maximum value of the objective function is found as the position of the hyperplanes that bound the flux cone in the respective directions is moved.

Through an examination of the shadow prices (linear programming dual variable), it is determined that there is a finite number of fundamentally different optimal metabolic flux distributions present in such a plane (Varma et al., 1993b). The shadow prices are measurements associated with every metabolite in the network, and indicate the intrinsic value of the metabolite. They are calculated from the dual problem of a linear optimization and are defined as follows:

$$\gamma_i = -\frac{dZ}{db_i} \tag{6}$$

The shadow prices (γ_i) essentially define the sensitivity of the objective function (Z) to changes in the availability of each metabolite/substrate, whereas b_i defines the violation of the mass balance constraint and is equivalent to an uptake reaction. The shadow price can be negative, zero, or posi-

tive, depending on the value of the metabolite to the objective function. Through observations of the shadow prices it can be determined when a new basic feasible solution has been introduced into the optimal solution. Because the extreme pathways of the linked output system are equivalent to the basic feasible solutions, this situation indicates a shift in the metabolic behavior corresponding to a new extreme pathway being utilized. Thus, the weight (w) on a particular extreme pathway increases from zero to a positive value in describing the flux distribution, as in Eq. (3). This procedure leads to the demarcation of distinct regions, or "phases," in the plane where the same pathways are always being utilized to varying extents. In each phase, the optimal use of the metabolic network is fundamentally different, corresponding to different optimal phenotypes.

In terms of the pathway analysis, the PhPP is basically generated by the projection of selected edges of the flux cone into a two-dimensional or, in some cases, a three-dimensional space spanned by the fluxes used to generate the particular phase plane. The boundaries of the phases within the PhPP correspond to particular extreme pathways. The entire region corresponding to a particular phase is subsequently the positive combination of the two extreme pathways that bound the phase. Thus, as we traverse through the phase plane we are essentially moving along the faces of the flux cone. In addition, the internal fluxes that will define the distinct phenotypes in each phase can be identified directly from the extreme pathways. The difference between the FBA phenotype phase plane and the extreme pathway projection is that the phase plane analysis only identifies the pathways that are used under optimal conditions, whereas the extreme pathway analysis maps all the pathways onto the same two-dimensional space. The remaining pathways that do not generate the boundaries of the PhPP represent suboptimal solutions or alternate optimal solutions. We now illustrate these concepts for growth on succinate using the succinate–oxygen PhPP.

SUCCINATE–OXYGEN PHENOTYPIC PHASE PLANE

The succinate–oxygen PhPP contains four distinct metabolic phenotypes (Fig. 6). The dark lines define the boundaries of each region in the phase plane as calculated using linear programming, and the lighter lines are the additional extreme pathways projected onto the two- and three-dimensional spaces. There are several distinguishing characteristics of the succinate–oxygen PhPP. First, the PhPP illustrates that the metabolic network is unable to utilize succinate as the sole carbon source for growth in an anaerobic environment, as evidenced by the PhPP only covering a region of the positive quadrant and the fact that all extreme pathways that generate biomass contain input exchange fluxes for succinate and oxygen. Second, the PhPP has three futile regions. (These are defined as regions where increased uptake of one of the substrates has a negative effect on the ability of the metabolic network to support growth [Edwards et al., currently in review by B & B]: one where oxygen is in excess [region 1]; and two where the carbon source is in excess [regions 3 and 4].) The futile regions can be identified as regions in which the slope of the isoclines is positive, where the isocline is defined as the locus of points for which the combination of the two exchange fluxes will provide for the same optimal value of the objective function. The succinate–oxygen PhPP also contains one dual-substrate limitation region (region 2).

Region 1 (Fig. 6a) is an oxygen-excess-futile region. Additional carbon is consumed to eliminate the oxygen that is provided in excess. Therefore, this region is wasteful, and less carbon is available for biomass production because it is oxidized to eliminate the excess oxygen. The ability of the metabolic network to produce high-energy phosphate bonds is not limiting growth in the futile region. The pathways in Table V are projected onto the phase planes based on the ratio of oxygen uptake to substrate uptake (Fig. 6b). It is seen that pathways #23 and #35, which project onto the same line, bound region 1 at a maximum oxygen/succinate uptake ratio of 6. These pathways do not generate biomass and, subsequently, the biomass yield decreases as the oxygen/succinate ratio increases moving the position in region 1 closer to the upper boundary than the lower boundary.

Separating region 1 from region 2 is a line defined as the line of optimality. This line represents the optimal relation between the two exchange fluxes. The biomass yield on the line of optimality is 0.051 g DW/mmol succinate. The line of optimality corresponds to the extreme pathway #1 in Table V. The optimal utilization of the metabolic pathways involves the cyclic operation of the TCA cycle, no flux in the oxidative branch of the pentose phosphate pathway (PPP), the utilization of the malic enzyme to produce pyruvate and NADPH, and CO_2 production as the only metabolic byproduct. The metabolic flux distribution for a point on this line is available on-line.

In region 2, the availability of succinate and oxygen is limiting the biomass production. Thus, this region is considered a dual-substrate-limited region, and the dual limitation is immediately apparent by the negative slope of the isoclines in region 2. The optimal metabolic flux distribution in this region leads to acetate production, and thus an extreme pathway that includes the acetate exchange flux is turned on. The activated pathway is the extreme pathway separating region 2 from region 3 and corresponds to pathway #20 in Table V, which produces acetate and biomass with a yield of 0.047 g DW/mmol succinate. Thus, every point in region 2 is a positive combination of pathway #1 and pathway #20, which results in a biomass yield ranging from 0.047 to 0.051 g DW/mmol succinate, and an acetate production level ranging from 0 to 0.158 mmol acetate/mmol succinate.

In regions 3 and 4, the uptake of additional succinate has a negative effect on the objective function. The succinate utilized by the metabolic network beyond the demarcation is not utilized to increase the objective function. Rather, cellular resources are required to eliminate the excess succi-

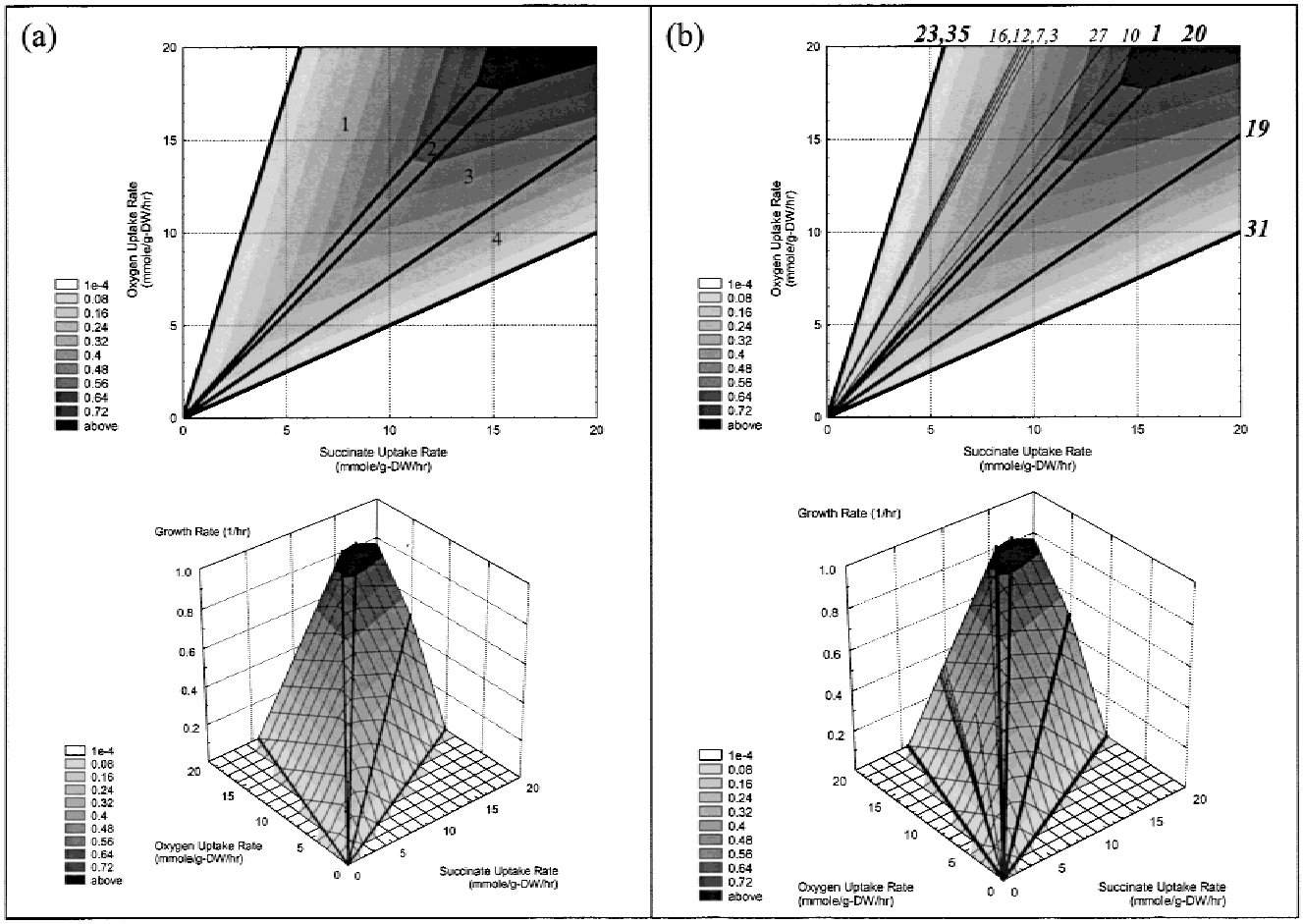

Figure 6. Succinate–oxygen phenotypic phase plane. (a) Four regions of metabolic behavior are shown on the two- and three-dimensional phase portraits. The isocline shading indicates regions in which the objective function (biomass yield) remains constant. (b) Projection of the 12 pathways in Table V onto the phase planes. Pathways are plotted based on the slope of the line as determined from the ratio of the oxygen exchange flux versus the succinate exchange flux. Bold lines and the corresponding bold pathway numbers indicate the pathways corresponding to the boundaries of the phases. Thin lines and corresponding pathway numbers reveal the additional pathways that project onto either the interior of the cone or a face of the cone as evidenced in the three-dimensional phase plane.

nate. Region 3 is bounded by pathway #20 and pathway #19, which produces ethanol as a byproduct. Therefore, any combination of these two pathways will lead to the combined production of acetate (from pathway #20) and ethanol (from pathway #19). Starting in region 2 and traversing into region 3 the metabolic network will turn off pathway #1 and turn on pathway #19. From the complete pathway vectors that provide the flux values for the internal fluxes (provided on-line), it is seen that the *sucA* and *sucC* reactions are predicted to shut off and the *adhE* reaction is turned on, while a number of other sets of reactions will carry increasing or decreasing flux levels. Traversing into region 4, the metabolic network will turn off pathway #20 and turn on pathway #31, which generates ethanol and CO_2 with no biomass production, thus serving to decrease the objective function. When the oxygen/succinate uptake ratios decreases to 2.00 the network will utilize pathway #31 only, and biomass production will reach zero. An oxygen/succinate uptake ratio below 2.00 is considered infeasible for network processing (these points will lie outside of the flux cone).

By determining the extreme pathways that demarcate the regions of the PhPP we are able to determine the precise theoretical range of biomass yield, oxygen/substrate uptake ratios, and the exact combination of pathways that characterize the predicted optimal metabolic phenotypes in each region. This pathway perspective may assist in the identification of pathways and associated reaction sets that are predicted to be switched on/off during a metabolic transition into an alternative functional modality.

GLUCOSE–OXYGEN PHENOTYPIC PHASE PLANE

The glucose–oxygen PhPP contains five regions (Fig. 7a) also bounded by sets of extreme pathways (Fig. 7b). Region 1 is an oxygen-excess-futile region, similar to region 1 for succinate. It is unlikely that the network would ever operate in this region, as it is futile. Based on the positive slope of the isoclines in this region we would expect the network to decrease the oxygen uptake rate, which will increase the biomass yield and bring the behavior closer to the line of

Table V. Functional characteristics of the reduced set of 12 extreme pathways calculated for succinate as the sole carbon source for the *E. coli* metabolic network.[a]

Pathway number	Exchange fluxes[b]							Net pathway reaction balance
	SUCCxt/ SUCCxt	ETHxt/ SUCCxt	ACxt/ SUCCxt	GRO/ SUCCxt	PIxt/ SUCCxt	CO2xt/ SUCCxt	O2xt/ SUCCxt	
1	−1.000	0	0	0.051	−0.188	1.825	−1.267	SUCCxt + 0.188 PIxt + 1.267 O2xt ⇒ 0.051 GRO + 1.825 CO2xt
10	−1.000	0	0	0.049	−0.182	1.895	−1.338	SUCCxt + 0.182 PIxt + 1.338 O2xt ⇒ 0.049 GRO + 1.895 CO2xt
20	−1.000	0	0.158	0.047	−0.172	1.696	−1.142	SUCCxt + 0.172 PIxt + 1.142 O2xt ⇒ 0.047 GRO + 1.696 CO2xt + 0.158 ACxt
3	−1.000	0	0	0.034	−0.125	2.553	−2.014	SUCCxt + 0.125 PIxt + 2.014 O2xt ⇒ 0.034 GRO + 2.553 CO2xt
7	−1.000	0	0	0.033	−0.121	2.600	−2.062	SUCCxt + 0.121 PIxt + 2.062 O2xt ⇒ 0.033 GRO + 2.6 CO2xt
12	−1.000	0	0	0.032	−0.117	2.644	−2.108	SUCCxt + 0.117 PIxt + 2.108 O2xt ⇒ 0.032 GRO + 2.644 CO2xt
16	−1.000	0	0	0.031	−0.114	2.679	−2.144	SUCCxt + 0.114 PIxt + 2.144 O2xt ⇒ 0.031 GRO + 2.679 CO2xt
19	−1.000	0.549	0	0.025	−0.092	1.837	−0.759	SUCCxt + 0.092 PIxt + 0.759 O2xt ⇒ 0.025 GRO + 1.837 CO2xt + 0.549 ETHxt
23	−1.000	0	0	0	0	4.000	−3.500	SUCCxt + 3.5 O2xt ⇒ 4.0 CO2xt
27	−1.000	0	1.000	0	0	2.000	−1.500	SUCCxt + 1.5 O2xt ⇒ 2.0 CO2xt + 1.0 ACxt
31	−1.000	1.000	0	0	0	2.000	−0.500	SUCCxt + 0.5 O2xt ⇒ 2.0 CO2xt + 1.0 ETHxt
35	−1.000	0	0	0	0	4.000	−3.500	SUCCxt + 3.5 O2xt ⇒ 4.0 CO2xt

[a]Pathways ordered based on the activity of the growth flux normalized by the succinate uptake. Pathway numbers coincide with the original numbers of the pathway vectors retained from the complete set (availabe on-line). All exchange flux values are normalized to the succinate intake level (negative values are relative uptake ratios, positive values are production ratios).
[b]SUCC, succinate; ETH, ethanol; AC, acetate; PI, inorganic phosphate; CO2, carbon dioxide; O2, oxygen; GRO, biomass/growth flux (see Table III).

optimality. The line of optimality corresponds to pathway #2 from Table IV yielding 0.105 g DW/mmol glucose. The optimal flux distribution is provided on-line.

An oxygen consumption decreases to below 1.376 mol O_2/mol glucose (the slope of pathway #2 in Fig. 7b), the network moves into region 2, where acetate begins to be secreted. Acetate secretion occurs due to the activation of pathway #38 that bounds the right side of region 2. Traversing through region 2 yields only a modest decrease in the biomass yield from 0.105 to 0.101 g DW/mmol glucose. As the oxygen consumption continues to decrease to below 1.226 mol O_2/mol glucose, the phenotype moves from region 2 into region 3 and the network begins to secrete ethanol in addition to acetate, and the biomass yield decreases further. In terms of the pathways, region 3 is bound to the left by pathway #38 and to the right by pathway #44 (Fig. 7b). Therefore, moving from region 2 into region 3 shifts the pathway utilization by inactivating pathway 2 completely while switching pathway #44 on to combine with pathway #38. From the pathways it is seen that the necessary reactions for ethanol production and transport are all activated, whereas many of the reactions in the TCA cycle (*sucA, sucC, sdhA1, fumA, mdh, sdhA2*) are inactivated as the cycle is no longer used to generate energy and reductive potential because the decreased oxygen availability has limited the use of the electron transport chain to generate ATP.

Region 4 is bracketed by pathway #44 and #45. As every point in this region is a positive combination of these two pathways the biomass yield will range from 0.078 to 0.095 g DW/mol glucose for an oxygen consumption range of 0.808 to 1.116 mol O_2/mole glucose. The *pykF* reaction will be activated while the *ppsA* reaction will be inactivated. Moving below an oxygen consumption rate of 0.808 mol O_2/mol glucose places the network in region 5 where it is completely oxygen limited, and the biomass yield drops sharply as the oxygen availability decreases. Region 5 is bounded by pathway #45 and #57 in Table IV. Note that when the oxygen consumption rate is zero the network is unable to produce biomass, and therefore the network predicts that completely anaerobic growth is unachievable. This result is due to the fact that no oxygen is available to consume the redox equivalents generated to produce the biomass (as part of the objective function). If we include the pyruvate formate lyase reaction in the network and allow formate to be an allowed byproduct the network can support anaerobic growth. In this case, the complete set of extreme pathways would increase from 85 to 192 simply from the addition of this one reaction. To avoid excessive numbers of pathways in this introductory and conceptual investigation we chose not to include this reaction. Therefore, the predicted inability to grow anaerobically is attributed to the model construction that was simplified so as to demonstrate many of the concepts just described without drowning the reader in excessive information and excessive numbers of pathways.

As in the previous section it is again seen that the extreme pathways that span the flux cone demark regions of optimal

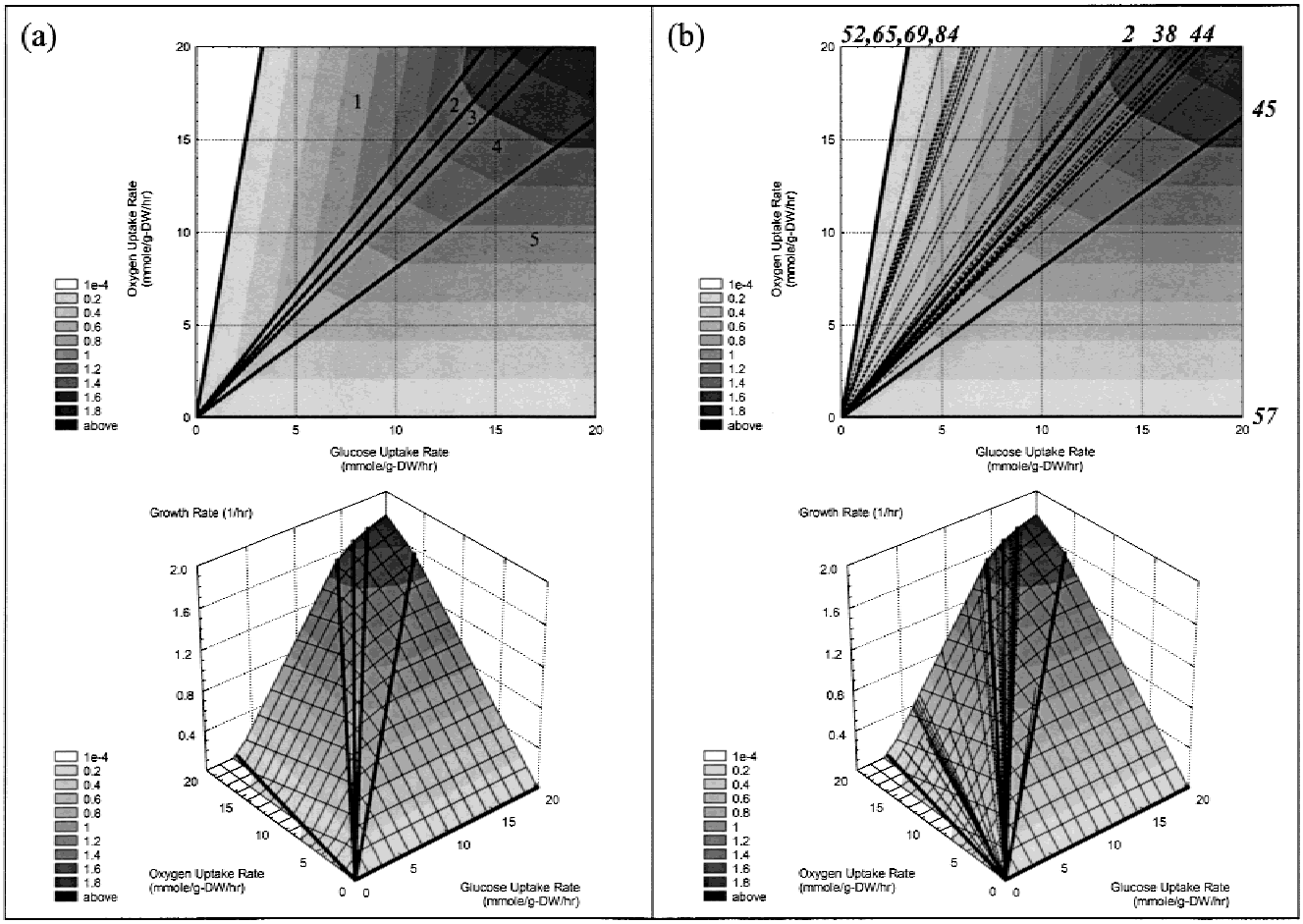

Figure 7. Glucose–oxygen phenotypic phase plane. (a) Five regions of metabolic behavior are shown on the two- and three-dimensional phase portraits. The isocline shading indicates regions in which the objective function (biomass yield) remains constant. (b) Projection of the 30 pathways in Table IV onto the phase planes. Pathways are plotted based on the slope of the line as determined from the ratio of the oxygen exchange flux versus the glucose exchange flux. Bold lines and the corresponding bold pathway numbers indicate the pathways corresponding to the boundaries of the phases. Thin-dashed lines correspond to additional pathways that project onto either the interior of the cone or a face of the cone as evidenced in the three-dimensional phase plane.

metabolic utilization as calculated using linear programming. Using phenotypic phase planes, it is realized how pathways can be interpreted as being switched on and off as environmental parameters change in time. This interpretation made use of the extreme pathways for the linked system, which correspond to the basic feasible solutions of the linear programming problem created through the introduction of maximum capacity constraints. As demonstrated earlier, the extreme pathways for the linked output system that are determined to comprise a flux distribution may be decomposed into pathways calculated for free output description, which provides a more detailed look at pathway utilization.

ALTERATIONS TO INTERNAL STRUCTURE OF THE NETWORK

In the previous two sections, environmental challenges to the metabolic network were considered through examination of the optimal network performance across a complete parameter space for exchange fluxes. In this section the focus shifts to assessing the effects of internal challenges to the network on overall performance and pathway utilization. All pathways in Tables IV and V that do not form the boundary of a region in the PhPP for either glucose or succinate (as shown in Figs. 6 and 7) are suboptimal pathways and will not be used under optimal conditions. These suboptimal extreme pathways will be utilized when alterations are made to the reaction network/metabolic genotype that lead to the removal of one or more of the optimal extreme pathways. Removal of internal reactions through either gene deletions or a loss in enzymatic function has the effect of completely eliminating any extreme pathway that utilizes the reaction being removed.

To illustrate the correlation between removal of a reaction and the corresponding impact on pathway structure and optimal performance of the network, each reaction in Figure 5 was individually eliminated from the network and optimal flux distributions were calculated using flux balance analysis. This was performed for both substrate growth conditions, and optimal biomass yields are provided in Table VI. To determine the effects on the pathway structure, all extreme pathways generating biomass that utilize the disabled

reaction are indicated in Table VI. These pathways can no longer be used by the system. If there are no extreme pathways remaining that can be utilized to generate biomass then the network is completely disabled. However, in many cases, there are viable pathways that remain intact despite the loss of a reaction. From the remaining pathways, the one with the highest biomass yield can be determined from the information in Tables IV and V, and often it will be suboptimal with respect to a pathway that has been eliminated. In comparing the optimal yields as calculated using linear programming versus the extreme pathways available to the system in the face of the internal challenges, it is seen that the predicted optimal growth yield and flux distribution corresponds precisely with one of the remaining extreme pathways that has the highest ratio of biomass to substrate uptake. This gain provides a clear representation of the connection between pathway analysis and flux balance analysis. For a number of reactions the effects of removal are identical due to the fact that a number of reactions belong to the same reaction/enzyme subset (Pfeiffer et al., 1999; Schilling and Palsson, 2000b). These subsets correspond to reactions that always occur together in the same ratio in all pathways in which they are active (i.e., they are dependent on each other). Table VII lists the enzyme subsets for the *E. coli* central metabolic network considered in both cases.

Internal challenges will also have a profound effect on how the network responds in the face of environmental challenges. To examine these combined challenges, the

Table VI. Pathway and flux balance analysis results associated with removal of individual reactions or genes.[a]

Gene Name	(Succinate) Pathways available								Optimal Biomass Yield	(Glucose) Pathways available																						Optimal Biomass Yield
	1	3	7	10	12	16	19	20		2	3	4	10	14	18	21	22	23	24	26	29	32	35	38	39	42	43	44	45	
aceE									0																						0
ackA	1	3	7	10	12	16	19		0.051	2	3	4	10	14	18	21	22		24	26	29		35			42	43		45	0.105
acnA									0																						0
adhE	1	3	7	10	12	16		20	0.051	2	3	4	10	14	18	21		23	24	26		32	35	38	39						0.105
adk	1	3	7	10	12	16	19	20	0.051		3	4	10	14	18						29	32	35		39	42	43	44	45	0.099
atpA									0																						0
cyoA									0																						0
eno									0																						0
fba									0											26											0.070
fbp									0	2	3	4	10	14	18	21	22	23	24	26	29	32	35	38	39	42	43	44	45	0.105
frdA	1	3	7	10	12	16	19	20	0.051	2	3	4	10	14	18	21	22	23	24	26	29	32	35	38	39	42	43	44	45	0.105
fumA									0							21	22	23		26	29	32		38	39		43	44	45	0.104
gapA									0																						0
gltA									0																						0
gnd						16			0.031					14																	0.062
gpmA									0																						0
icdA									0																						0
mdh									0							21	22	23		26	29	32		38	39		43	44	45	0.104
nuoA									0																						0
pckA									0	2	3	4	10	14	18	21	22	23	24	26	29	32	35	38	39	42	43	44	45	0.105
pfkA	1	3	7	10	12	16	19	20	0.051											26											0.070
pgi									0								22	23	24		29	32	35								0.105
pgk									0																						0
pgl						16			0.031					14																	0.062
pntA1	1	3	7		12	16	19	20	0.051	2	3	4	10	14	18									38	39	42		44	45	0.105
ppc	1	3	7	10	12	16	19	20	0.051																						0
ppsA	1	3	7	10	12	16	19	20	0.051		3	4	10	14	18						29	32	35		39	42	43	44	45	0.099
pta	1	3	7	10	12	16	19		0.051	2	3	4	10	14	18	21	22		24	26	29		35			42	43		45	0.105
pykF									0	2		4				21	22	23	24	26	29	32	35	38	39	42	43	44		0.105
rpe			7						0.033						18																0.066
rpiA									0																						0
sdhA1									0							21	22	23		26	29	32		38	39		43	44	45	0.104
sdhA2									0							21	22	23		26	29	32		38	39		43	44	45	0.104
sucA				10			19	20	0.049							21	22	23		26	29	32		38	39		43	44	45	0.104
sucC				10			19	20	0.049							21	22	23		26	29	32		38	39		43	44	45	0.104
talA					12				0.032				10																		0.063
tktA1					12				0.032				10																		0.063
tktA2		3							0.034		3																				0.068
tpiA									0											26											0.070
zwf						16			0.031					14																	0.062

[a]Pathways that utilized the removed reaction are indicated in dark gray boxes, with the X indicating that they are no longer available due to the removal of the reaction. Light gray boxes are pathways still available that generate a biomass yield. The pathway with the highest yield in each case is indicated by the double-bordered boxes. Optimal biomass yields as calculated using linear programming are indicated for each deletion on both substrates in units of grams dry weight per millimole of substrate.

Table VII. Enzyme subsets calculated for the *E. coli* central metabolic network considering both glucose succinate as possible substrates.

Reaction/enzyme subsets
(*AC trx, ackA, pta*)
(*ETH trx, adhE*)
(O_2 *trx, cyoA*)
(*fba, tpiA*)
(*gapA, pgk*)
(*gpmA, eno*)
(*ppsA, adk*)
(*zwf, pgl, gnd*)
(*tktA1, talA*)
(*gltA, acnA, icdA*)
(*sucA, sucC*)
(*sdhA1, fumA, mdh, sdhA2*)

PhPP studies examined in the previous section can be repeated for each of the network deficiencies examined in Table VI. In Figure 8, the 12 extreme pathways for growth on succinate are projected into the oxygen–succinate plane and the regions of the phase plane are calculated using flux balance analysis to determine the qualitatively different regions of metabolic function. In comparing the wild-type situation (Fig. 8a) to the deletion of any one of the genes comprising the oxidative branch of the pentose phosphate shunt (Fig. 8b), there is only one pathway that remains capable of generating biomass, and this pathway forms the boundary between the two regions of the phase plane. The other two pathways that generate the boundaries do not produce biomass. Figure 8c represents the loss of either the *sucA* or *sucC* reaction. In this case, many of the original pathways remain intact with the exception of the optimal pathway utilized in the standard wild-type network. Only one of the boundaries to the phase plane shifts, indicating the utilization of a different pathway that is now optimal due to the loss of the original optimal pathway. Taken together, this illustrates how an internal challenge to a network forces the utilization of a different set of pathways under changing environmental conditions.

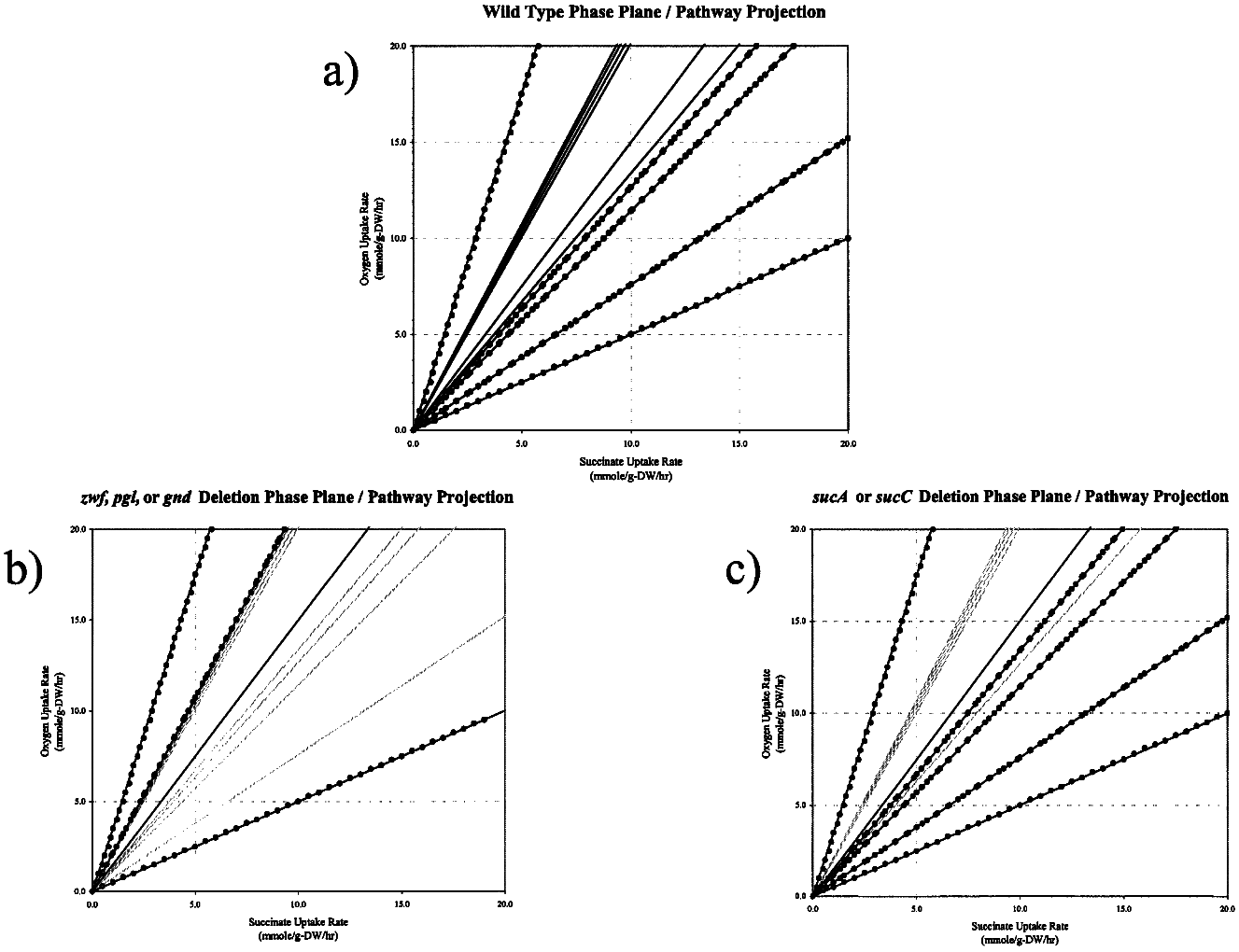

Figure 8. Pathway projections and phase plane outlines for mutant strains with succinate as the substrate. For (a) wild-type, (b) *zwf, pgl,* or *gnd* deletion, (c) *sucA* or *sucC* deletion, each of the pathways available to the network are indicated by the solid black lines, whereas those no longer available due to the deletion are indicated by the solid gray lines. Points outlining the different regions of the phase plane for each case are indicated by the filled circles as calculated using linear programming optimizing for the biomass yield.

DISCUSSION

Using a simple example network and a detailed network of *E. coli* central metabolism, we have illustrated how metabolic phenotypes (flux distributions) and shifts in metabolic behavior can be interpreted from a pathway-based perspective. From the physical principles of mass conservation and the constraints of reaction thermodynamics, the mathematical concepts of convex analysis are used to define sets of extreme pathways that span the entire flux space/cone for complex metabolic networks. These pathways represent balanced sets of metabolic reactions whose functional characteristics together represent the capabilities of the metabolic network. From a pathway-oriented perspective, systems can be studied in terms of their ability to generate particular metabolic precursors (free outputs) or examined for their ability to generate a balanced set of these biosynthetic demands (linked outputs). For both approaches, network flux distributions, such as those generated using optimization techniques (e.g., flux balance analysis), can be interpreted as nonnegative combinations of the extreme pathways as indicated mathematically in Eq. (3). These different approaches allow for two different pathway perspectives (free versus linked outputs), which are related to each other by precise pathway equivalencies (not to be confused with equivalencies among reactions).

Using flux balance analysis, the complete range of optimal metabolic phenotypes can be examined under defined environmental and genetic conditions through the use of phenotypic phase planes (PhPPs). For the *E. coli* network, the PhPPs were generated for growth on succinate and glucose spanning particular regions and dimensions of the flux cone. Projecting the set of extreme pathways onto the PhPP reveals the precise pathway utilizations that are occurring in each region of metabolic behavior. The pathways serve to bracket the behavior in each region, which creates ranges of biomass yield along with oxygen and substrate uptake limitations. In the face of internal challenges to the structure of the network the set of viable extreme pathways is reduced, in turn limiting the manner in which a network operates under changing environmental conditions, leading to changes in the phase plane. The flux cone becomes smaller due to the loss of extreme pathways, which imposes further constraints on its functional capabilities.

Using the pathways it is possible to translate back to a reaction-based interpretation of the predicted metabolic phenotypes. This reveals sets of reactions that are completely inactivated or activated as well as subtle changes in either the increased or decreased utilization of various reactions. Combining flux balance analysis and pathway analysis provides the ability to examine metabolic capabilities and interpret optimal metabolic phenotypes from both reaction- and pathway-oriented perspectives. Together these approaches can be used to examine quantitatively the constraints placed on metabolic activity due to the structure and connectivity of the reaction network and the physicochemical principles that govern biological systems.

Ideally, we would like to study the pathway structure and utilization for an entire organism with this combined approach. Flux balance analysis is easily capable of handling large networks as linear programming has been designed over the years to handle problems with millions of variables. The same cannot be said for convex analysis algorithms associated with enumerating pathways in networks in any field. For small networks, such as the one in Figure 1, computing the set of extreme pathways is quite trivial. However, as the size of the network increases in a linear fashion, the time to calculate the extreme pathways as well as the number of pathways increases in an exponential fashion. In the traditional language of theoretical computer science the problem of enumerating the vertices of a polyhedral cone is referred to as an NP-hard problem, such as the "traveling salesman" problem (Papadimitriou and Steiglitz, 1998). Although it is feasible to construct comprehensive flux balance models for completely reconstructed metabolic genotypes, it is not feasible to enumerate all pathways present in such a network for interpreting metabolic function. A very conservative estimate would place the number of pathways in *E. coli* well over 10^6. Although it is important to calculate these complete sets for the identification of more detailed systems properties related to the capabilities of the network, the shear number of pathways may complicate the combined pathway/flux balance analysis approach to an extent that makes the interpretation overwhelming. This problem of excessive pathways can be partially dealt with by either focusing on particular situations (both environmental and genetic) that would reduce the system as we have done in this article, or to properly subdivide the metabolic system to solve for pathways in each subsystem that may be pieced together to interpret metabolic capabilities (Schilling and Palsson, 2000b).

Neither flux balance analysis nor pathway analysis incorporates information on reaction kinetics and regulation, limiting their insight into dynamic responses. From these approaches it is possible to realize some of the fundamental constraints that metabolic systems are faced with and define the flux cone that contains all admissible steady-state flux vectors. As detailed information on in vivo reaction dynamics is accumulated, the ability to analyze, interpret, and perhaps predict dynamic responses of metabolic networks to environmental and genetic perturbations using dynamic modeling approaches may become more feasible. For the optimal phenotypes calculated in a wild-type strain to be realized in vivo a regulatory network that achieves this state must be present. This would then assume that a regulatory scheme has evolved to optimize the same parameters optimized in the linear programming (i.e., growth). Although this may or may not be the case, in general, regulatory schemes and reaction dynamics will serve to further constrain metabolic behavior to operate in confined subspaces of the flux cone. Identifying these regions from both the theoretical and experimental side represents a challenge for the future.

A number of experimental technologies have now made the holistic study of biological systems feasible beyond the theoretical realm. The ability to assimilate DNA chip-based

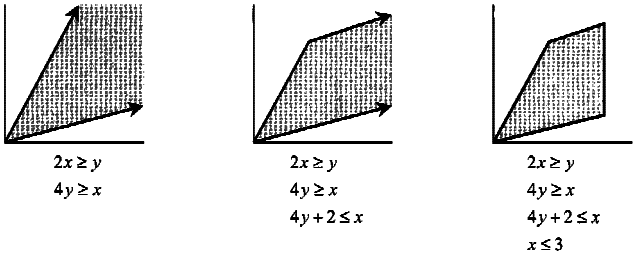

Figure A1. A convex cone, a generic polyhedron, and a polytope.

and protein expression technologies providing genome-scale information with computational methods for metabolic network analysis will become important in advancing the study of metabolic physiology and the practice of metabolic engineering. Currently, the interpretation of high-throughput experimental information on systemic behavior is limited by a lack of analysis capabilities. Can systems-based quantitative in silico approaches, such as flux balance and pathway analysis, be used to assist in understanding this flood of data? This is a question that will need to be answered as interest builds in the genomics community for quantitative systems analysis.

The authors thank the Whitaker Foundation for their support through graduate fellowships in bioengineering.

APPENDIX

A convex polyhedron is a region in $\mathbb{R}^n$ determined by linear equalities and inequalities (Fig. A1). One of the properties of a polyhedron is convexity: Any line segment connecting two points in the polyhedron is completely contained in the polyhedron. If a polyhedron is bounded then it is called a polytope. It is called a polyhedral cone if every ray through the origin and any point in the polyhedron are completely contained in the polyhedron.

A point $\vec{v}$ in a polyhedron is called a vertex if it is not the convex combination of two other points in the polyhedron. [A convex combination of the points $\vec{p}$ and $\vec{q}$ has the form $t\vec{p} + (1 - t)\vec{q}$, where $0 < t < 1$.] A ray $\vec{r}$ is called an extremal ray if no point on the ray is a positive linear combination of points in the polytope but not on the ray. Note that a cone only has extremal rays and a polytope only has vertices (Fig. A2).

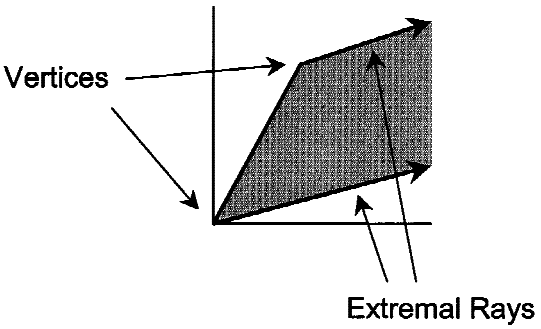

Figure A2. Vertices and extremal rays of a polyhedron.

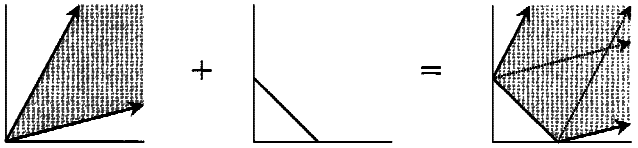

FIgure A3. The sum of a polytope and a cone.

The set of vertices and extremal rays is unique for any polyhedron. In fact, any polyhedron (P) can be written uniquely in the form:

$$P = \mathrm{Conv}(\vec{v}_1, \ldots, \vec{v}_k) + \mathrm{Cone}(\vec{r}_1, \ldots, \vec{r}_l) \quad \text{(A-1)}$$

where the convex hull and the conical hull are defined as:

$$Conv(\vec{v}_1,\ldots,\vec{v}_k) = \{t_1\vec{v}_1 + \ldots + t_k\vec{v}_k | t_i \geq 0, t_1 + \ldots + t_k = 1\}$$
$$Cone(\vec{r}_1,\ldots,\vec{r}_l) = \{t_1\vec{r}_1 + \ldots + t_l\vec{r}_l | t_i \geq 0\} \quad \text{(A–2)}$$

Note that $\{\vec{v}_1, \ldots, \vec{v}_k\}$ are the vertices, and $\{\vec{r}_1, \ldots, \vec{r}_l\}$ are the extremal rays of P. An alternate interpretation is that very polyhedron is the sum of a polytope and a cone (Fig. A3). Both vertices and rays can be found efficiently and have significance in a variety of situations extending well beyond biochemical reaction networks. In linear programming, the vertices and rays are known as basic feasible solutions. They have the property that any linear function on a polyhedron achieves its maximum value at a vertex or it is unbounded and lies on an extremal ray.

References

Bhalla US, Iyengar R. 1999. Emergent properties of networks of biological signaling pathways. Science 283:381–387.

Bonarius HPJ, Schmid G, Tramper J. 1997. Flux analysis of underdetermined metabolic networks: the quest for the missing constraints. Trends Biotechnol 15:308–314.

Clarke BL. 1980. Stability of complex reaction networks. Adv Chem Phys 43:1–215.

Clarke BL. 1988. Stoichiometric network analysis. Cell Biophys 12: 237–253.

Edwards J, Palsson B. 1999. Properties of the *Haemophilus influenzae* Rd metabolic genotype. J Biol Chem 274:17410–17416.

Edwards JS, Palsson BØ. 1998. How will bioinformatics influence metabolic engineering? Biotechnol Bioeng 58:162–169.

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

Edwards JS, Ramakrishna R, Palsson BØ. (Accepted). Characterizing phenotypic plasticity: a phase plane analysis.

Edwards JS, Ramakrishna R, Schilling CH, Palsson BØ. 1999. Metabolic flux balance analysis. In: Lee SY, Papoutsakis ET, editors. Metabolic engineering. New York: Marcel Dekker. p 13–57.

Fell D. 1996. Understanding the control of metabolism. London: Portland Press.

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

Ingraham JL, Maaloe O, Neidhardt FC. 1983. Growth of the bacterial cell. Sunderland, MA: Sinauer Associates.

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

Liao JC, Hou S-Y, Chao Y-P. 1996. Pathway analysis, engineering, and physiological considerations for redirecting central metabolism. Biotechnol Bioeng 52:129–140.

Lynd LR, Wyman CE, Gerngross TU. 1999. Biocommodity engineering. Biotechnol Progr 15:777–793.

Mavrovouniotis ML, Stephanopoulos G, Stephanopoulos G. 1990. Computer-aided synthesis of biochemical pathways. Biotechnol Bioeng 36:1119–1132.

Palsson BØ. 1997. What lies beyond bioinformatics? Nature Biotechnol 15:3–4.

Palsson BØ, Joshi A, Ozturk SS. 1987. Reducing complexity in metabolic networks: making metabolic meshes manageable. Fed Proc 46: 2485–2489.

Papadimitriou CH, Steiglitz K. 1998. Combinatorial optimization: algorithms and complexity. New York: Dover.

Pfeiffer T, Sanchez-Valdenebro I, et al. 1999. METATOOL: for studying metabolic networks. Bioinformatics 15:251–257.

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.

Reder C. 1988. Metabolic control theory: a structural approach. J Theor Biol 135:175–201.

Sauer U, Cameron DC, Bailey JE. 1998. Metabolic capacity of *Bacillus subtilis* for the production of purine nucleosides, riboflavin, and folic acid. Biotechnol Bioeng 59:227–238.

Savageau MA. 1969a. Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions. J Theor Biol 25:365–369.

Savageau MA. 1969b. Biochemical systems analysis. II. The steady state solutions for an n-pool system using a power-law approximation. J Theor Biol 25:370–379.

Savageau MA. 1970. Biochemical systems analysis. 3. Dynamic solutions using a power-law approximation. J Theor Biol 26:215–226.

Schilling CH, Edwards JS, Palsson BØ. 1999a. Towards metabolic phenomics: analysis of genomic data using flux balances. Biotechnol Progr 15:288–295.

Schilling CH, Letscher D, Palsson BØ. 2000a. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. J Theor Biol 203:229–248.

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

Schilling CH, Schuster S, Palsson BØ, Heinrich R. 1999b. Metabolic pathway analysis: basic concepts and scientific applications in the postgenomic era. Biotechnol Progr 15:296–303.

Schuster S, Dandekar T, Fell DA. 1999. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. Trends Biotechnol 17:53–60.

Schuster S, Hilgetag C, Woods J, Fell D. 1996. Elementary modes of functioning in biochemical networks. In: Cuthbertson R, Holcombe M, Paton R, editors. Commutation in cellular and molecular biological systems. London: World Scientific. p 151–165.

Seressiotis A, Bailey JE. 1988. MPS: an artificially intelligent software system for the analysis and synthesis of metabolic pathways. Biotechnol Bioeng 31:587–602.

Stephanopoulos G. 1999. Metabolic fluxes and metabolic engineering. Metab Eng 1:1–11.

Stephanopoulos G, Aristidou AA, Nielsen J, editors. 1998. Metabolic engineering: principles and methodologies. San Diego, CA: Academic Press.

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

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

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

Varma A, Palsson BØ. 1994b. Metabolic flux balancing: basic concepts, scientific and practical use. Bio/Technology 12:994–998.

Varner J, Ramkrishna D. 1998. Application of cybernetic models to metabolic engineering: investigation of storage pathways. Biotechnol Bioeng 58:282–291.

Varner J, Ramkrishna D. 1999. Mathematical models of metabolic pathways. Curr Opin Biotechnol 10:146–150.

Weng G, Bhalla US, Iyengar R. 1999. Complexity in biological signaling systems. Science 284:92–96.

Yarmush ML, Berthiaume F. 1997. Metabolic engineering and human disease. Nature Biotechnol 15:525–528.