

# The Cell as a Decision-Making Unit

LORENZO CASTELLI<sup>1</sup>, RAFFAELE PESENTI<sup>2</sup>, AND DANIEL SEGRÈ<sup>3</sup>

<sup>1</sup>Dipartimento di Ingegneria e Architettura, Università degli Studi di Trieste, 31427 Trieste, Italy

<sup>2</sup>Dipartimento di Management, Università Ca' Foscari Venezia, 30121 Venice, Italy

<sup>3</sup>Bioinformatics Program, Departments of Biology, Biomedical Engineering, and Physics, Boston University, Boston, MA 02215 USA

CORRESPONDING AUTHOR: D. SEGRÈ (dsegre@bu.edu).

**ABSTRACT** Each living cell needs to solve a resource allocation problem, in which multiple inputs (uptake fluxes) and outputs (secretion fluxes) are the outcome of the stoichiometry of biochemical pathways and the regulation of metabolic enzymes. Quantifying the efficiency with which a cell solves this resource allocation problem constitutes a basic question in “cellular economics.” In this letter, we propose the use of data envelopment analysis (DEA) to define multidimensional yields that can capture the multidimensional nature of cell input–output processes. The DEA, by treating cells as decision-making units, enables one to introduce the concept of efficiency frontier that is both intimately connected to the shadow prices of flux balance analysis and useful to estimate the phenotypic phase space from experimental measurements of fluxes.

**INDEX TERMS** Data envelopment analysis, flux balance analysis, cell efficiency, linear programming.

## I. INTRODUCTION

CELLULAR metabolism is comprised of hundreds of biochemical reactions, forming a highly interconnected network [6]. This network is responsible for guaranteeing a supply of energy and building blocks to the cell. The usage of each reaction in this network depends on environmental and internal parameters, such as availability of nutrients or interactions with other organisms. Understanding the detailed time-dependent orchestration of enzyme levels for directing metabolic rates (or fluxes) under different conditions constitutes a dauntingly complex problem. However, simplifying assumptions have been successfully employed to understand how, at steady state, a cell may be able to balance incoming resources and metabolic tasks to efficiently maintain itself into a balanced homeostatic state. In particular, by using a steady-state approximation to constrain the space of possible metabolic fluxes, and by employing optimality principles to search this space for biologically meaningful states [6], it has been possible to model the behavior of biochemical networks at the genome scale, with useful applications in metabolic engineering, cancer research, and environmental studies.

Despite these successes, constraint-based models of metabolism remain a coarse approximation of a very complex biological reality. In particular, due to the nature of flux balance analysis (FBA) [3], and in line with standard definitions of yields in metabolic engineering, FBA calculations often end up estimating the optimal value of one specific cellular output (usually the biomass production, or growth rate) relative to the most limiting environmental resource (e.g., the carbon source). This focus on a single yield does not

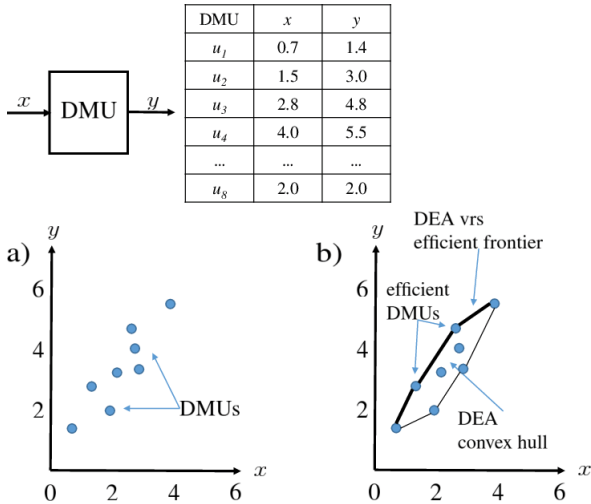
accurately reflect the fact that living systems generally face decisions about the simultaneous management of multiple input sources (e.g., uptake of carbon, nitrogen, and sulfur), and multiple output tasks (e.g., growth and storage of excess carbon).

Here, we apply to metabolic network modeling a linear programming (LP) method called data envelopment analysis (DEA), which has been extensively used to study complex decision-making strategies in economical systems, even in the presence of multiple inputs and outputs [5]. We will first introduce the main DEA concepts, and map the DEA formalism onto a metabolic network optimization problem. Next, we will demonstrate the power of the DEA in this context, by showing how it establishes new connections between experimental measurements of uptake/secretion fluxes and the FBA feasible space.

## II. BACKGROUND

A *decision-making unit (DMU)* is any entity that uses  $M$  inputs  $x \in R_+^M$  to produce  $N$  outputs  $y \in R_+^N$ . Two DMUs are *homogeneous*, if they use the same set of inputs and produce the same set of outputs. Given a set  $\Gamma$  of homogeneous DMUs, on one side, the DEA provides an estimate of the performance (or efficiency) of each DMU in  $\Gamma$  relative to the best observed practice in this set of DMUs. On the other side, DEA infers the expected performances of DMUs that, although not in  $\Gamma$ , have input/output values similar to the ones of the DMUs in  $\Gamma$ .

DEA describes the DMUs as *black boxes*, as it assesses a DMU efficiency only on the basis of the DMU input and



**FIGURE 1.** Top: representation of a single-input single-output DMU, and table exemplifying DMUs' input/output values. Bottom- (a) Visualization of DEA DMUs on the input/output plane. (b) DMUs convex hull and efficient frontier.

output values without requiring the knowledge of the processes that the DMU uses to transform its inputs into outputs [Fig. 1(a)]. Specifically, DEA defines the efficiency of the  $i$ th DMU  $u_i$  in  $\Gamma$  as the weighted ratio of the outputs over the inputs, that is

$$e_i = \mu^T y_i / \nu^T x_i \quad (1)$$

where  $(x_i, y_i)$  are the input and output vectors of  $u_i$ , and  $(\nu, \mu)$  is the associated weights' vectors. Under the efficiency definition (1), each DMU  $u_0$  in  $\Gamma$  exhibits constant return to scale characteristics: for every positive scalar  $t$  that multiplies the input vector  $x_0$ , DMU  $u_0$  proportionally increases its outputs, i.e., it produces  $t \cdot y_0$ . If, instead, the outputs of DMUs in  $\Gamma$  rise more than or less than proportionately with inputs, we say that they exhibit variable, respectively, increasing or decreasing, returns to scale (VRS). The DEA deals with this latter situation by introducing a *scale* variable  $q$  in the definition of efficiency, that now becomes

$$e_i = \mu^T y_i + q / \nu^T x_i \quad (2)$$

where  $q$  assumes a positive, negative, or null value if DMU  $u_i$  exhibits increasing, decreasing, or constant returns to scale.

When DMU  $u_0$  is under evaluation, the DEA determines nonnegative weights  $(\nu, \mu)$ , and possibly  $q$ , that maximize  $e_0$ , subject to the constraint that the efficiency of every DMU in  $\Gamma$  is bounded above by 1, that is,

$$\max_{\nu, \mu \geq 0, q} e_0 \quad \text{s.t. } e_i \leq 1, \quad i \in \Gamma. \quad (3)$$

Problem (3) can be formulated as a dual pair of LP problems, as first shown in [2] for the CRS case, and in [1] for the VRS case.

We say that DMU  $u_0$  is *more efficient* than  $u_1$  if it uses no more inputs to produce no fewer outputs and is doing strictly better in at least one dimension. In this situation,  $e_0$  is greater than  $e_1$ . If no other DMU in  $\Gamma$  is more efficient

than  $u_0$ , we simply say that  $u_0$  is *efficient*. In this situation,  $e_0$  is equal to 1. The *convex hull* of  $\Gamma$  is the smallest convex polyhedron  $H$  that includes all the DMUs in  $\Gamma$ . Each point of  $H$  represents a possible DMU, whose inputs and outputs are convex combinations of the corresponding inputs and outputs of DMUs in  $\Gamma$ . DMUs in  $H$  are similar to the ones in  $\Gamma$  in the sense that they assume intermediate values between the values assumed by the corresponding inputs and outputs of the DMUs in  $\Gamma$ . The *efficient frontier* is the set of efficient DMUs in  $H$  [1]. The black line in Fig. 1(b) delimits the set  $H$  of the DMUs in Fig. 1(a). The thick part of the line is the efficient frontier  $F$ . Note that, by construction, the vertexes of  $F$  are efficient DMUs in  $\Gamma$ .

### III. CELLS as DMUs

In evaluating cellular metabolism, we may think of a cell as a DMU. The weighted DMU input and output variables may be viewed as corresponding to cellular uptake/secretion fluxes, i.e., the fluxes that cross the cellular boundaries. For the case of a single input and a single output, the DEA efficiency converges to a classical yield. More complex DEA efficiency functions may correspond to other physical quantities (e.g., biomass production yield relative to total carbon consumed). However, in general, the whole point of the DEA is to enable an estimate of the importance of different inputs and outputs for the efficiency of a system, even if the inputs and outputs do not have the same units, and their linear combinations through the weights do not necessarily represent any physical quantity.

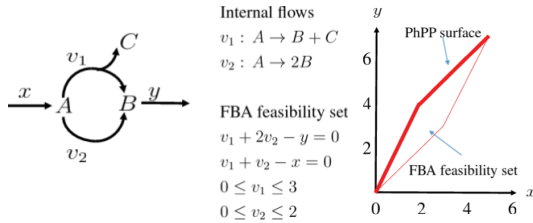
Since the cellular metabolic response to changing conditions may be nonlinear, an increase of available resources may correspond to a nonproportional increase of growth rate and other metabolite secretion rates. In applying the DEA to cellular metabolism, we will therefore focus on variable return-to-scale models.

Observe also that cellular metabolism properties, as computed by DEA, should be thought of as averages over time and over a whole population, that do not aim to capture faster dynamic, e.g. from cell cycle oscillations and fluctuations. This approximation, implicitly assumed also in most FBA calculations, implies that, in what follows, we will focus on comparing populations of cells rather than single cells. Thus, each DMU can be thought of as representative of one of these populations.

### IV. DEA and FLUX BALANCE ANALYSIS

In this section, we investigate the relationship between DEA and FBA. In particular, we will demonstrate that the DEA VRS efficient frontier is closely related to the phenotypic phase plane (PhPP) surface, a geometrical entity useful for representing different discrete states expected in the FBA calculations of metabolism [3]. Notably, standard DEA derivations provide a natural framework for inferring an approximate PhPP surface from experimental data, even in the absence of any knowledge about the stoichiometry of the underlying metabolic network. We will first illustrate this

concept with a simple yet nontrivial toy model. From the DEA perspective, this model corresponds to a set of single-input single-output DMUs/cells in  $\Gamma$  (Fig. 1). From the FBA viewpoint, these cells are feasible steady states of a metabolic network defined by a set of reactions and constraints (Fig. 2). The connection between DEA and FBA is made by assuming that the FBA model describes the metabolic states of the cells in  $\Gamma$ : the input and output of each the DMU in the DEA model correspond, respectively, to the uptake of metabolite  $A$  (flux  $x$ ) and the secretion of metabolite  $B$  (flux  $y$ ).



**FIGURE 2.** Left: toy metabolic network with its steady state and capacity constraints. Right: corresponding FBA feasibility set and PhPP surface.

### A. FLUX BALANCE ANALYSIS

The structure of a cellular metabolic network involving  $L$  metabolites and  $R$  reactions can be described by a stoichiometric matrix  $S \in R^{L \times R}$ , whose element  $S_{lr}$  corresponds to the amount of metabolite  $l$  that participates in reaction  $r$ .  $S_{lr}$  is positive if metabolite  $l$  is a product in reaction  $r$ , and negative if it is a reactant. Let  $v = \{v_r : r = 1, \dots, R\}$  be the vector of metabolic fluxes. FBA identifies the steady-state flux vector  $v$  maximizing a given linear objective function and is generally formulated as the following LP problem:

$$\max_v \sum_{r=1}^R c_r v_r \quad (4a)$$

$$\sum_{r=1}^R S_{lr} v_r = 0 \quad l = 1, \dots, L \quad (4b)$$

$$a_r \leq v_r \leq b_r \quad r = 1, \dots, R \quad (4c)$$

where  $c = \{c_r : r = 1, \dots, R\}$  is the objective function vector, and  $a_r$  and  $b_r$  define capacity constraints. One of the most commonly used objective functions is the maximization of the growth rate, based on the hypothesis that microbial organisms have evolved toward an optimally efficient growth capacity [4]. Overall, the steady state (4b) and capacity (4c) constraints define the feasible set of the FBA problem, which we denote by  $\Phi$ .

The PhPP surface  $F^\Phi$  is defined as the set of points in  $\Phi$  that optimize the objective function (4a) for a fixed set of values of the nutrient uptake rates (i.e., the input fluxes). The PhPP has been shown to provide a useful geometrical illustration of how cells may adopt different metabolic strategies based on environmental conditions, including evolutionary trajectories [4], [6]. Due to the LP structure of model (4), the PhPP is composed of some facets of  $\Phi$ , which can be computed using the shadow prices of the FBA problem. In the presence of multiple inputs, we can identify isoclines on the facets (or phases) of the PhPP,

i.e., sets of points (lines) on a facet of the PhPP whose inputs lead to the same value of the objective function. The slope of the isoclines can be calculated as the ratio of shadow prices of (4) and is different on each facet of the PhPP.

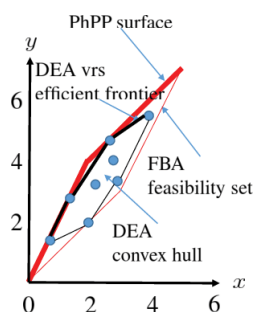
For our toy model, the steady state of the metabolic network translates into a set of metabolite conservation constraints (equalities in Fig. 2). These, together with additional capacity constraints (inequalities in Fig. 2), define the feasibility set  $\Phi$ . The projection of  $\Phi$  on the subspace of the two fluxes  $x$  and  $y$ , which we call  $\Phi_\perp$ , corresponds, in Fig. 2, to the polyhedron enclosed by red lines, with vertexes in the points (0,0), (2,4), (5,7), and (3,3).

To better grasp the structure of  $\Phi_\perp$  in the toy model, consider, as an example, an input flux  $x = 1$ . If all this flux is consumed by reaction  $v_1$ , we observe an output flux  $y = 1$ . Conversely, if all this flux is consumed by reaction  $v_2$ , we observe  $y = 2$ . For any intermediate distribution of the input flux, we observe  $1 \leq y \leq 2$  (due to the different stoichiometries of B in  $v_1$  and  $v_2$ ). If we assume that the objective of hypothetical cells with this toy metabolism is to maximize the output flow  $y$  for each fixed input  $x$ , then we expect that cellular input/output flux values measured in hypothetical experiments would be either on or close to the polyhedron facets delimited by vertexes (0,0) and (2,4) for  $0 \leq x \leq 2$ , delimited by vertexes (2,4) and (5,7) for  $2 \leq x \leq 5$ . These two facets are the projection  $F_\perp^\Phi$  of the PhPP surface  $F^\Phi$  on the input/output space. In any case, we expect that input/output values of any toy cell will fall within the  $\Phi_\perp$ ; otherwise, their metabolic fluxes would violate either the stoichiometric equations or the capacity constraints. It is also worth noting that the output-to-input ratio  $y/x$  is not constant for the cells in  $\Phi_\perp$ , not even for the cells that lay on the two facets (0,0)-(2,4) and (2,4)-(5,7). Indeed,  $y/x = 2$  for the cells that lays on the facet (0,0)-(2,4) as they produce 2 units of outputs for each unit of input  $x$ , whereas the ratio  $y/x$  declines with an increase in  $x$  for cells on the facet (2,4)-(5,7) as they increase their output production only by 1 for each additional unit of input  $x$ . In other words, the output-to-input ratio presents a variable (decreasing) return-to-scale.

This example translates the description of a simple metabolic network in the language routinely used in the DEA theory. Intuitively, by comparing Figs. 1(b) and 2, one can see the similarity between the DEA efficiency frontier and the FBA PhPP surface. In Section IV-B, we formalize the above-mentioned intuition in a theorem and show that the DEA frontier can serve as an approximation of the PhPP, whose accuracy increases with the number of DMUs available. We also see how the efficiency definition as in (2) can describe DMUs' variable return to scale.

### B. DEA VRS CAN HELP ESTIMATE the PhPP

*Theorem 1:* If an FBA model describes the metabolic network of the cells/DMUs in a DEA set  $\Gamma$ , then the DEA convex hull  $H$  is included in or equal to the FBA set  $\Phi_\perp$  (e.g., see Fig. 3).



**FIGURE 3.** Relation between the FBA feasibility set and the DEA convex hull.

*Proof:* If the FBA model correctly describes the cells in  $\Gamma$ , the inputs/outputs of these cells should fall in the FBA set  $\Phi_{\perp}$ , i.e.,  $\Gamma \subset \Phi_{\perp}$ , and either lay on or close to the projection of the PhPP surface. The theorem is then proved by observing that both  $\Phi_{\perp}$  and  $H$  are convex polyhedra, and the vertexes of the latter are in  $\Gamma$ .  $\square$

*Corollary 1:* a) If  $\Gamma$  includes the vertexes of  $F_{\perp}^{\Phi}$ , then the DEA efficient frontier  $F$  is equal to  $F_{\perp}^{\Phi}$ . b) If DMUs in  $\Gamma$  are uniformly randomly sampled from  $\Phi$ , then the DEA efficient frontier  $F$  is equal, with probability equal to 1, to  $F_{\perp}^{\Phi}$  when the number of DMUs goes to infinity.

Theorem 1 and Corollary 1 point out that DEA, using only the knowledge of the input/output values of a sample  $\Gamma$ , can provide an approximate description of  $\Phi_{\perp}$ . Thus, a description of the PhPP can be obtained either from the complete knowledge of the stoichiometry and subsequent FBA calculations, or, in an approximated way, through the DEA frontier, which can be derived solely based on a set of input/output measurements. The larger this set of measurements, the better the DEA frontier approximation.

### C. DEA ESTIMATE OF PhPP ISOCLINES

We finally show an additional correspondence between the DEA VRS efficient frontier and the FBA PhPP. In particular, the slopes of isoclines used to characterize the PhPP (which are derived from the FBA shadow prices) can be estimated using the weights computed in the DEA model. Considering again our toy model, we assume that  $F = F_{\perp}^{\Phi}$ , which means that all DMUs are efficient (see Corollary 1). By solving, as in [1], the LP formulation of model (3), the optimal weights are  $(\mu^*, v^*, q^*) = (1/2, 1, 0)$  which imply  $e_x = \frac{\mu^*(2x)+q^*}{v^*x} = 1$  for all DMUs  $(x, 2x)$  on the  $\Phi_{\perp}$  facet  $(0,0)-(2,4)$ , and  $(\mu^*, v^*, q^*) = (1, 1, -2)$  which imply  $e_x = \frac{\mu^*(x+2)+q^*}{v^*x} = 1$  for all DMUs  $(x, x+2)$  on the  $\Phi_{\perp}$  facet  $(2,4)-(5,7)$ .

The key result is that, for each input value  $x$ , the ratio  $v^*/\mu^*$  is equal to the slope of the corresponding facet of  $F_{\perp}^{\Phi}$ , i.e., it defines a marginal productivity, which can also be viewed as the extra output gained by adding one unit of input. A similar concept can be easily extended for multiple-inputs multiple-outputs DMUs. In this latter situation, the ratio between the weights of two different inputs defines a

marginal rate of substitution. It represents the combination of the two metabolite uptake rates that leads to the same value of the outputs, that is the slope of the isocline of  $F_{\perp}^{\Phi}$ . Furthermore, it can be proved that when  $F$  is different from  $F_{\perp}^{\Phi}$  and  $\Gamma$  includes a DMU  $u_i$  lying on a facet of  $F_{\perp}^{\Phi}$ , then the slope of the isoclines of this phase is bounded by the values of the slopes of the isoclines of the hyperplanes enveloping  $H$ , which have  $u_i$  as a vertex.

### V. DISCUSSION

Borrowing concepts from economics and operations research, we have examined the possible benefits of treating cells and their metabolism as DMUs, which consume resources to produce outputs. The DEA definitions and results can then be used to measure and compare the metabolic efficiency of different cell populations. Notably, DEA can handle the presence of multiple inputs (e.g., uptake of different carbon sources) and multiple outputs (e.g., biomass and secretion of waste products), appropriately comparing the performance of different organisms.

DEA makes it possible to partition a set of observed cell populations into efficient and inefficient ones. The set of efficient cells defines a concave surface called efficient frontier that is related to the surface of the FBA PhPP. We show that the DEA production frontier approximates the PhPP. This is noteworthy because it enables an estimation of the PhPP from experimental measurements of input/output fluxes, instead of necessarily requiring complete knowledge of reaction stoichiometry and detailed FBA calculations.

While the aim of this letter was to introduce for the first time the correspondence between DEA and FBA, and illustrate the potential applications of DEA approaches in studying cellular metabolism, many other concepts introduced in the vast DEA literature could be similarly applied to cellular systems. More advanced approaches, for example, could help deal with imperfect information about the value of the input/output fluxes or with data coming from different environments in which different external factors may influence the behavior of DMUs (see, e.g., [5]).

### REFERENCES

- [1] R. D. Banker, A. Charnes, and W. W. Cooper, "Some models for estimating technical and scale inefficiencies in data envelopment analysis," *Manage. Sci.*, vol. 30, no. 9, pp. 1078–1092, 1984.
- [2] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," *Eur. J. Oper. Res.*, vol. 2, no. 6, pp. 429–444, 1978.
- [3] J. S. Edwards, R. U. Ibarra, and B. O. Palsson, "In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data," *Nature Biotechnol.*, vol. 19, no. 2, pp. 125–130, 2001.
- [4] J. S. Edwards, R. Ramakrishna, and B. O. Palsson, "Characterizing the metabolic phenotype: A phenotype phase plane analysis," *Biotechnol. Bioeng.*, vol. 77, no. 1, pp. 27–36, 2002.
- [5] H. O. Fried, C. A. K. Lovell, and S. S. Schmidt, Eds., *The Measurement of Productive Efficiency and Productive Growth*. New York, NY, USA: Oxford Univ. Press, 2008.
- [6] C. H. Schilling, S. Schuster, B. O. Palsson, and R. Heinrich, "Metabolic pathway analysis: Basic concepts and scientific applications in the post-genomic era," *Biotechnol. Prog.*, vol. 15, no. 3, pp. 296–303, 1999.