# Multi-Objective Optimization of Biological Networks for Prediction of Intracellular Fluxes

José-Oscar H. Sendín, Antonio A. Alonso, and Julio R. Banga

Process Engineering Group
Instituto de Investigaciones Marinas (CSIC)
C/ Eduardo Cabello 6, 36208, Vigo, Spain
osendin@iim.csic.es, antonio@iim.csic.es, julio@iim.csic.es

**Abstract.** In this contribution, we face the problem of predicting intracellular fluxes using a multi-criteria optimization approach, i.e. the simultaneous optimization of two or more cellular functions. Based on Flux Balance Analysis, we calculate the Pareto set of optimal flux distributions in *E. coli* for three objectives: maximization of biomass and ATP, and minimization of intracellular fluxes. These solutions are able to predict flux distributions for different environmental conditions without requiring specific constraints, and improve previous published results. We thus illustrate the usefulness of multi-objective optimization for a better understanding of complex biological networks.

**Keywords:** Multi-objective optimization, Pareto front, Flux Balance Analysis.

## 1 Introduction

Intracellular fluxes in biochemical networks can be calculated *in silico* under the assumption that cellular systems operate in an optimal way with respect to a certain biological objective. Network capabilities and flux distributions have thus been predicted by using, for example, Metabolic Flux Balance Analysis (FBA), the fundamentals of which can be found in e.g. (Varma and Palsson 1994). FBA only requires the stoichiometric model of the network, but since the linear system of mass balance equations at steady-state is generally under-determined, appropriate cellular functions (objectives) must be defined, as well as other possible additional constraints, to find a unique solution. Successful applications of FBA include the prediction of *E. coli* metabolic capabilities (Edwards et al. 2001) and the genome-scale reconstruction of the metabolic network in *S. cerevisiae* (Forster et al. 2003).

In this context, a particularly interesting question which have been addressed recently in detail (Schuetz et al. 2007; Nielsen 2007) concerns the principles behind the optimal biochemical network operation, i.e.: "which are the criteria being optimized in these systems?" By far, the most common objective considered is the maximization of growth (or biomass yield), although other criteria, such as maximization of ATP yield (van Gulik and Heijnen 1995) or minimization of the overall intracellular flux (Bonarios et al. 1996), have been proposed for different systems and conditions.

Since neither we nor nature have a single goal, a more desirable and realistic approach is to consider the simultaneous optimization of two or more criteria, often conflicting. As a consequence, the solution will not be unique but instead this strategy

J.M. Corchado et al. (Eds.): IWPACBB 2008, ASC 49, pp. 197–205, 2009.
springerlink.com 

will result in a set of solutions representing the optimal trade-offs between the different objectives. Multi-objective (or multi-criteria) optimization is better able to cope with the complexity of models from systems biology (Handl et al. 2007), but few applications are found in literature in comparison with other scientific and engineering fields (Sendín et al. 2006).

In this work we face the solution of multi-objective optimization (MO) problems derived from FBA. By simultaneously optimizing several common cellular functions, the aim of this study is to test the capabilities of this approach for predicting intracellular fluxes independently from the environmental conditions and without imposing additional, case-dependent and potentially artificial constraints conditioning the final solution. After presenting the basic concepts and methods in MO, we will consider the central carbon metabolism in *Escherichia coli* as a case study to assess whether optimality principles can be generally applied.

## 2 Multi-Objective Flux Balance Analysis (MOFBA)

### 2.1 Problem Formulation and Basic Concepts

Assuming a biological network operating at steady-state, and if a stoichiometric model is available, the Multi-Objective Flux Balance Analysis problem can be stated as finding the flux distribution which optimizes simultaneously two or more objective functions subject to the mass balance equations:

$$\underset{\mathbf{v}}{\text{Max/Min}}\ \mathbf{Z} = \left[ Z_1(\mathbf{v}) \quad Z_2(\mathbf{v}) \quad \ldots \quad Z_n(\mathbf{v}) \right]^{\mathrm{T}} \tag{1}$$

Subject to:

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

$$\mathbf{v}^L \leq \mathbf{v} \leq \mathbf{v}^U \tag{3}$$

**Z** is the vector of *n* objective functions (linear as well as non-linear); **S** is the (*m* x *r*) stoichiometric matrix, where *m* is the number of intracellular metabolites and *r* the number of reactions; **v** is the vector of *r* fluxes, with lower and upper bounds $\mathbf{v}^L$ and $\mathbf{v}^U$, respectively. Additional constraints can be imposed depending on the problem and the available experimental data and the knowledge about the system.

Simultaneous optimization of multiple objectives differs from traditional single-objective optimization in that if the objectives are in conflict with each other, there will not be a unique solution which optimizes simultaneously all of them. The key concept here is that of Pareto-optimal solution.

A point **v*** in the solution space is said to be Pareto-optimal if there does not exist another feasible point **v** such that $Z_i(\mathbf{v}) \leq Z_i(\mathbf{v}^*)$ for all $i=1,\ldots,n$ and $Z_j(\mathbf{v}) < Z_j(\mathbf{v}^*)$ for some *j*. In other words, **v*** is optimal in the sense that improvement in one objective can only be achieved by worsening one or more of the others. Thus, the solution of a MO problem is a family of potentially infinite points, none of which can be said to be better than another. This family is known as Pareto-optimal set or Pareto front.

### 2.2 Methods for Multi-Objective Optimization

Traditionally, multiple objectives are optimized simultaneously by defining a composite function combining different criteria. The most widely used approach consists in optimizing a weighted sum of the objectives, where each weight represents the relative importance of the associated objective. Within FBA, this type of utility functions has also been proposed, as e.g. maximization of ATP yield per flux unit (Dauner and Sauer 2001; Schuetz et al. 2007). However, this approach will yield only one optimal solution, overlooking the trade-off between the objectives.

In this work we have combined two well-known techniques for generating the complete Pareto-front (or at least a good representation of it):

- **ε-Constraint (EC):** This is also a common and intuitive method for solving a MO problem. In this approach, the original MO problem is transformed into a single-objective linear programming (LP) problem (if the objective functions and the constraints are linear) or a non-linear programming (NLP) problem by optimizing one of the objectives while the others are incorporated as inequality constraints. By changing the value of the parameter ε (i.e. the bounds on the objectives converted to constraints), different Pareto-optimal solutions can be obtained. Its main drawback is the difficulty to choose appropriate values for the parameters of the method to obtain a good picture of the Pareto front, so that no regions are over- or under- represented.
- **Normal Boundary Intersection (NBI):** This technique (Das and Dennis 1998) was developed to overcome the drawbacks of methods like the weighted sum approach in which it is difficult to obtain a complete representation of the Pareto-optimal set. Starting from the individual optima for each objective, NBI also converts the original MO problem into a set of LPs/NLPs in such a way that a systematic change in the method parameters generates an even spread of points on the Pareto front. Thus, the complete trade-off between the objectives can be captured by solving a lesser number of optimization problems. However, some regions of the Pareto surface can be missed in problems with more than two objectives.

It should be noted that global optimization (GO) solvers will be needed for both approaches if the associated single-objective NLPs are non-convex.

## 3 Case Study

Here we consider the central carbon metabolism in *Escherichia coli*, which has been studied in (Schuetz et al. 2007) to examine the predictive capacity of 11 linear and non-linear network objectives. The stoichiometric model consists of 98 reactions and 60 metabolites, and 10 split ratios $R_i$ ($i$=1,…,10) at pivotal branch points were defined (Figure 1).

Taking as reference the above mentioned work, we address the problem in which three relevant cellular functions are optimized simultaneously:

$$\text{Find } \mathbf{v} \text{ to } \left\{ \begin{array}{l} \max Z_1(\mathbf{v}) = v_{Biomass} \\ \max Z_2(\mathbf{v}) = v_{ATP} \\ \min Z_3(\mathbf{v}) = \sum_{i=1}^{r} v_i^2 \end{array} \right\} \tag{4}$$

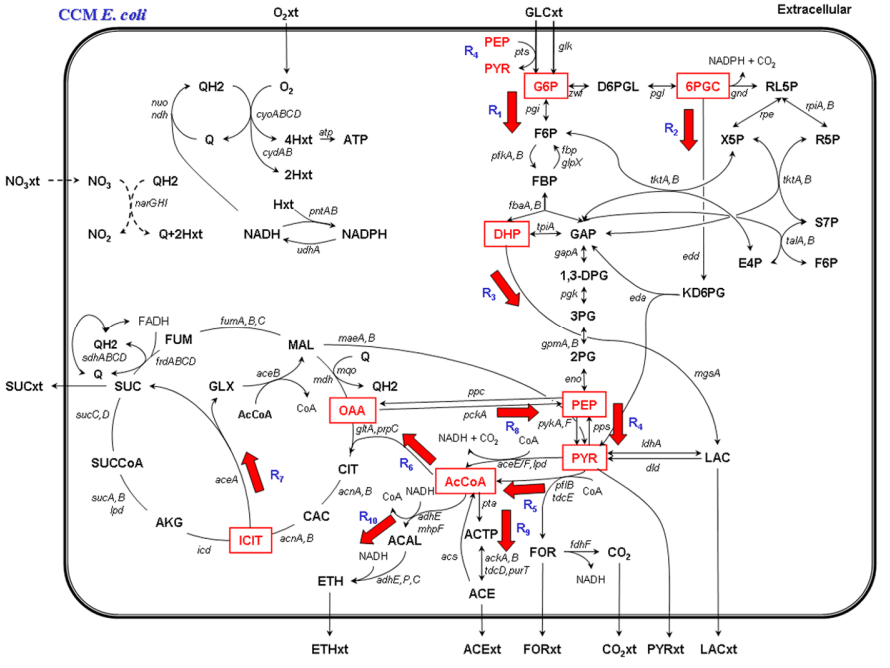


**Fig. 1.** Central carbon metabolism pathways in *Escherichia coli.* Red arrows represent the split ratios which describe the systemic degree of freedom of the network (for further explanation and abbreviations see Schuetz et al. 2007).

**Table 1.** Experimental flux split ratios (R) for the conditions considered

| | Aerobic | | | | Anaerobic |
|---|---|---|---|---|---|
| | Batch | Continuous C-limited | Continuous C-limited | Continuous N-limited | Batch $NO_3$ resp. |
| $v_{Biomass}$ | 8.3 mM/g·h | 5.0 mM/g·h | 7.0 mM/g·h | 4.0 mM/g·h | 1.77 mM/g·h |
| | ExpC1 | ExpC2 | ExpC3 | ExpC4 | ExpC5 |
| R1 | 0.70 ± 0.02 | 0.69 ± 0.12 | 0.64 ± 0.05 | 0.96 ± 0.14 | 0.82 ± 0.02 |
| R2 | 0.13 ± 0.06 | 0.23 ± 0.20 | 0.19 ± 0.11 | 0.00 ± 0.05 | 0.00 ± 0.05 |
| R3 | 0.00 ± 0.05 | 0.00 ± 0.05 | 0.00 ± 0.05 | 0.00 ± 0.05 | 0.00 ± 0.05 |
| R4 | 0.78 ± 0.02 | 0.84 ± 0.14 | 0.70 ± 0.06 | 0.72 ± 0.10 | 0.96 ± 0.02 |
| R5 | 0.81 ± 0.03 | 0.91 ± 0.21 | 0.84 ± 0.14 | 0.90 ± 0.15 | 0.96 ± 0.02 |
| R6 | 0.24 ± 0.02 | 0.64 ± 0.13 | 0.85 + 0.09 | 0.50 ± 0.06 | 0.02 ± 0.01 |
| R7 | 0.00 ± 0.05 | 0.46 ± 0.13 | 0.00 ± 0.05 | 0.00 ± 0.05 | 0.00 ± 0.05 |
| R8 | 0.00 ± 0.05 | 0.35 ± 0.08 | 0.12 ± 0.03 | 0.01 ± 0.01 | 0.00 ± 0.05 |
| R9 | 0.58 ± 0.03 | 0.00 ± 0.05 | 0.00 ± 0.05 | 0.04 ± 0.01 | 0.65 ± 0.01 |
| R10 | 0.00 ± 0.05 | 0.00 ± 0.05 | 0.00 ± 0.05 | 0.00 ± 0.05 | 0.30 ± 0.02 |

subject to the mass balance equations and the upper and lower bounds on fluxes (Eqs. 2-3). No additional constraints are imposed. It should be noted that objective functions $Z_1$ and $Z_2$ are linear, and the overall intracellular flux ($Z_3$) is non-linear, but convex.

Pareto-optimal solutions obtained with a combination of the methods described above will be compared with experimental flux data from *E. coli* (Table 1) under five environmental conditions (oxygen or nitrate respiring batch cultures and aerobic chemostats). The overall agreement is quantified using a standardized Euclidean distance between the computed split ratios and the experimental ones.

## 4 Results and Discussion

### 4.1 Optimization Settings

The three-objective optimization problem defined above is solved using a combination of the ε-constraint technique and NBI. The solution strategy consists of the following steps:

1. Maximize $v_{Biomass}$ using LP
2. Choose different values $bm_i$ for $v_{Biomass}$ in the range $[0, v_{Biomass}^{max}]$
3. For each value $bm_i$, the following bi-objective optimization problem is solved using NBI: maximization of ATP and minimization of the overall intracellular flux subject to the ε-constraint: $v_{Biomass} \geq bm_i$

The resulting NLPs from application of NBI are solved by means of a multi-start clustering algorithm, *GLOBALm* (Sendín et al. 2008). This is a global optimization method which can detect the potential existence of multiple optima (i.e. solutions with the same value of the objective function and different flux profiles). For the sake of comparison with the results reported in (Schuetz et al. 2007), we have made use of the solvers included in the MATLAB® Optimization Toolbox (The MathWorks, Inc.): *linprog* for the LPs and *fmincon* as local solver within *GLOBALm* for the NLPs.

### 4.2 Pareto-Optimal Sets

The resulting Pareto surfaces (interpolated) for both aerobic and anaerobic conditions are showed in Figures 2 and 3, respectively. The trade-off between ATP yield and the overall intracellular flux is also depicted for each one of the biomass fluxes corresponding to the experimental conditions. Both Pareto-optimal sets obtained using the hybrid approach εConstraint-NBI are represented in Figure 4.

From inspection of these figures is clearly evident the existing conflict between ATP production and the overall intracellular flux for a given biomass flux. Maximum ATP yields (higher in the aerobic case) are achieved at the expense of an increase in the enzyme usage and with low growth rates. On the other side, biomass can be maximized while maintaining the overall intracellular flux at low levels. The cost to pay in this case is a decrease in the ATP yield.

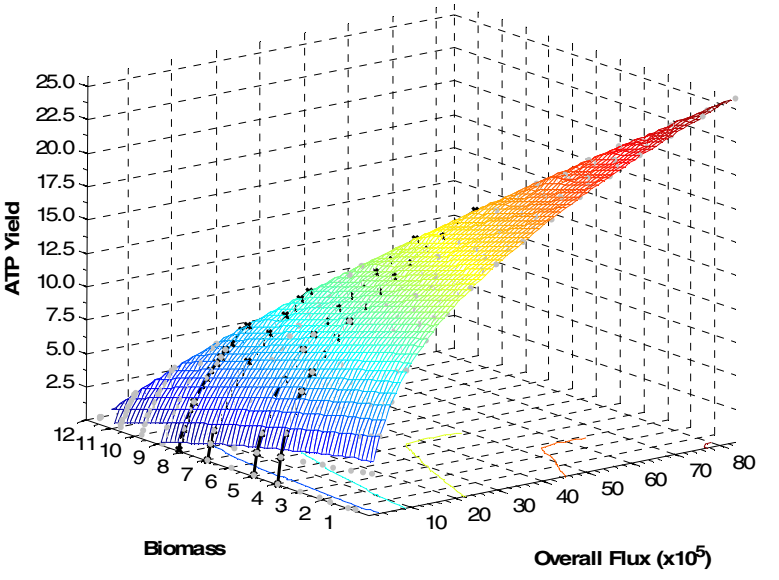


**Fig. 2.** Pareto front in aerobic conditions

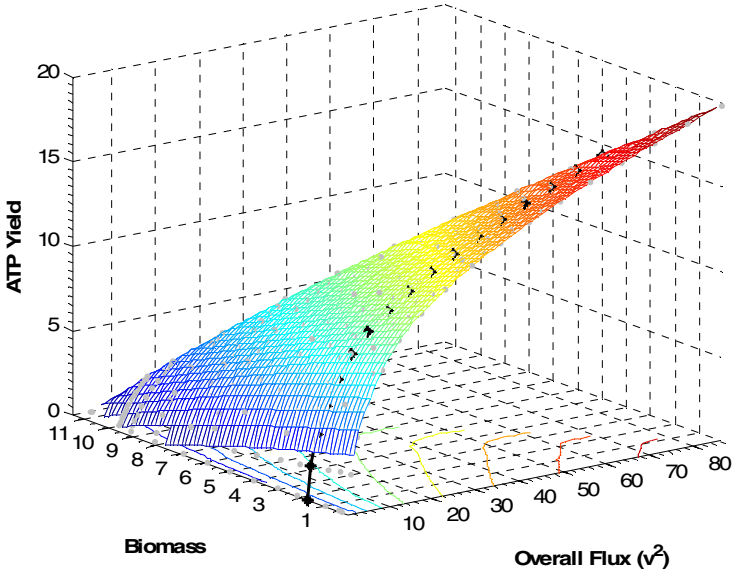


**Fig. 3.** Pareto front in anaerobic conditions

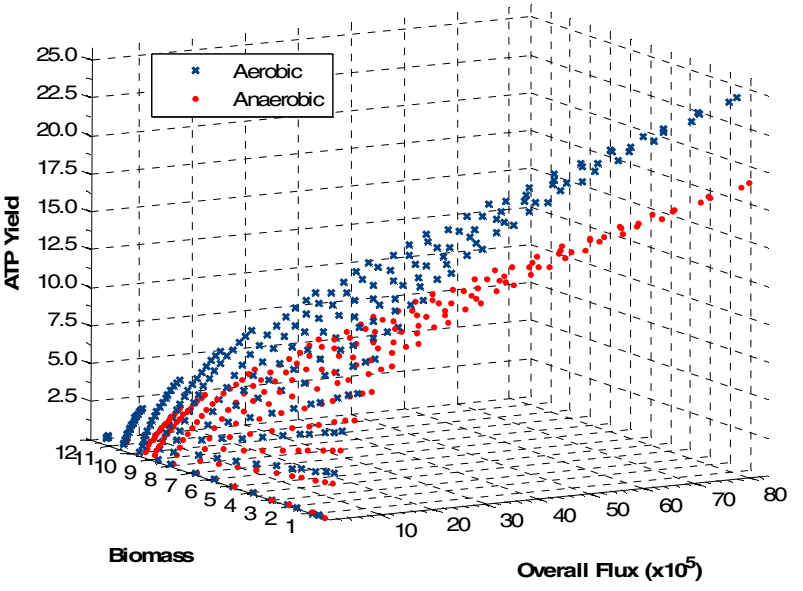


**Fig. 4.** Comparison of Pareto-optimal sets

### 4.3 Analysis of Solutions

Split ratios for each Pareto solution are compared with the experimental data, selecting those which yield the closest flux predictions (Table 2).

**Table 2.** Selected Pareto-optimal points

| | **ExpC1** A | **ExpC2** B | **ExpC3** C | **ExpC4** D | **ExpC5** E |
|---|---|---|---|---|---|
| $v_{Biomass}$ | 8.3 | 5.0 | 11.0 | 7.0 | 1.75 |
| R1 | 0.74 | 0.98 | 0.64 | 0.98 | 0.55 |
| R2 | 0.42 | 0.00 | 0.09 | 0.00 | 1.00 |
| R3 | 0.00 | 0.00 | 0.0 | 0.00 | 0.00 |
| R4 | 0.80 | 0.98 | 0.72 | 0.85 | 0.97 |
| R5 | 0.81 | 0.92 | 0.70 | 0.86 | 0.52 |
| R6 | 0.31 | 0.79 | 0.59 | 0.77 | 0.01 |
| R7 | 0.00 | 0.12 | 0.0 | 0.00 | 0.00 |
| R8 | 0.00 | 0.00 | 0.0 | 0.00 | 0.00 |
| R9 | 0.50 | 0.04 | 0.05 | 0.09 | 0.74 |
| R10 | 0.00 | 0.00 | 0.0 | 0.00 | 0.22 |

For the continuous cultures, the best predictions were found in Schuetz et al. (2007) when maximizing biomass or ATP yield coupled with several constraints. Somewhat similar flux distributions were obtained here, but we want to stress the fact that no additional, case-specific, constraints were imposed. For example, solution B (for C-limited continuous cultures) is similar to that resulting from maximization of ATP subject to an overproduction of 35% of NADPH relative to the NADPH requirement for biomass production, and the flux profile C maximizes biomass while satisfying a constraint on intracellular fluxes (limited to a 200% of the glucose uptake rate), and an upper bound on the oxygen uptake of 150% of the glucose uptake. For N-limited continuous cultures, point D also improves the prediction obtained when only one single objective is considered (with or without additional constraints).

## 5 Conclusions

In this work we have addressed the question of whether intracellular fluxes can be predicted considering optimality principles. The assumption here is that fluxes are distributed to optimize not only one single cellular function but several objectives simultaneously (multi-objective optimization).

In general terms, Pareto-optimal flux distributions improve the best predictions obtained with traditional FBA using different combinations of objective functions and constraints. The advantage of the multi-objective approach is that no additional, case-specific, constraints are needed, and it can be a powerful tool for a better understanding of the factors that influence the metabolic flux.

## Acknowledgments

This work has been supported by EU project BaSysBio LSHG-CT-2006-037469.

## References

Bonarios, H.P.J., Hatzimanikatis, V., Meesters, K.P.H., de Gooijer, C.D., Schmid, G., Tramper, J.: Metabolic flux analysis of hybridoma cells in different culture media using mass balances. Biotechnology Bioengineering 50, 299–318 (1996)

Das, I., Dennis, J.E.: Normal Boundary Intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems. SIAM J. Optimization 8, 631–657 (1998)

Dauner, M., Sauer, U.: Stoichiometric growth model for riboflavin-producing Bacillus subtilis. Biotechnology Bioengineering 76, 132–143 (2001)

Edwards, J.S., Ibarra, R.U., Palsson, B.O.: In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. Nature Biotechnology 19(2), 125–130 (2001)

Forster, J., Famili, I., Fu, P., Palsson, B.O., Nielssen, J.: Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. Genome Research 13(2), 244–253 (2003)

Handl, J., Kell, D.B., Knowles, J.: Multiobjective optimization in bioinformatics and computational biology. IEEE-ACM Transactions on Computational Biology and Bioinformatics 4(2), 279–292 (2007)

Nielsen, J.: Principles of optimal metabolic network operation. Molecular Systems Biology 3, 2 (2007)

Schuetz, R., Kuepfer, L., Sauer, U.: Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. Molecular Systems Biology 3, 119 (2007)

Sendín, J.O.H., Vera, J., Torres, N.V., Banga, J.R.: Model-based optimization of biochemical systems using multiple objectives: a comparison of several solution strategies. Mathematical and Computer Modelling of Dynamical Systems 12(5), 469–487 (2006)

Sendín, J.O.H., Banga, J.R., Csendes, T.: Extensions of a multistart clustering algorithm for constrained global optimization problems. Industrial & Engineering Chemistry Research (accepted for publication, 2008)

van Gulik, W.M., Heijnen, J.J.: A metabolic network stoichiometry analysis of microbial growth and product formation. Biotechnology Bioengineering 48, 681–698 (1995)

Varma, A., Palsson, B.O.: Metabolic flux balancing: Basic concepts, scientific and practical us. Bio/Technology 12(10), 994–998 (1994)