# Design of Robust Metabolic Pathways

R. Umeton
Dept. of Biological Engineering
MIT
Cambridge, MA 02139
umeton@mit.edu

G. Stracquadanio
Dept. of Biomedical Engineering
Johns Hopkins University
Baltimore, MD 21218
stracquadanio@jhu.edu

A. Sorathiya
Computer Laboratory
U. of Cambridge
Cambridge, UK CB3 0FD
as833@cam.ac.uk

A. Papini
Dept. of Evolutionary Biology
U. of Florence
Firenze, Italy I-50121
alpapini@unifi.it

P. Liò
Computer Laboratory
U. of Cambridge
Cambridge, UK CB3 0FD
pl219@cam.ac.uk

G. Nicosia
Dept. of Mathematics & Computer Science
U. of Catania
Catania, Italy I-95125
nicosia@dmi.unict.it

## ABSTRACT

In this paper we investigate plant photosynthesis and microbial fuel cells. We report the following: 1) we introduce and validate a novel multi-objective optimization algorithm, PMO2; 2) in photosynthesis we increase the yield of 135%, while in Geobacter sulfurreducens we determine the trade-off for growth versus redox properties; 3) finally, we discuss Pareto-Front as an estimator of robust metabolic pathways.

## Categories and Subject Descriptors

G.1.6 [**Optimization**]: Global Optimization; J.3 [**Life and Medical Sciences**]: Biology and genetics

## General Terms

Algorithms, Design, Reliability

## Keywords

Metabolic engineering, Mutational Robustness, Geobacter sulfurreducens, Carbon metabolism, Multi-objective Optimization.

## 1. INTRODUCTION

Pareto Optimality is one of the most fruitful and powerful approach where optimization of conflicting objectives is concerned[14, 11]. In this work we report a new methodology, whose focus is on the Pareto-optimality, to explore the performance space of a given biological pathway. Pareto-optimality conditions are those in which it is impossible to make a function better off without necessarily making some other function worse off. Multi-objective optimization problems (MOPs) tackle sets of competing and conflicting objective functions having (strong or weak) nonlinear interdependence; the conflicting behavior emerges when the improvement of one objective function results in the degradation of at least one other objective function. MOPs generally are characterized by a set of solutions that are known as Pareto-optimal; the set of Pareto optimal solutions defines a multi-dimensional space, which is called Pareto-Front. Depending on the problem, the Pareto-Front could be non convex and discontinuous and, hence, finding solutions lying on this space is particularly challenging. We have modeled the $C3$ photosynthetic carbon metabolism in terms of simultaneous optimization of two conflicting biological strengths: maximization of $CO_2$ Uptake and contextual minimization of the total protein-nitrogen employed to gain that property (representative of the biochemical effort the leaf has to devote to gain that $CO_2$ Uptake rate). We have inspected the problem at three $CO_2$ concentrations ($C_i$) in the atmosphere or stroma (25M years ago environment, nowadays one, and the one predicted for the end of the century) and two triose-P (PGA, GAP, and DHAP) export rates: low and high. In this context, our analysis has detected Pareto-optimal configurations in these six $C_i$/triose-P conditions studied. Among the others, two promising candidate solutions for leaf re-engineering are further inspected and compared with the natural leaf enzyme configuration. For the first time, it has been individuated a reasonably small set of key enzymes whose targeted tuning gives rise to a robust maximization of the photosynthetic rate, contextually with an efficient protein-nitrogen employment. Geobacter sulfurreducens, a bacterium capable of using biomasses to produce electrons to be transferred directly to an electrode is a useful model for real optimization since its genome is completely sequenced and a model of its metabolic network is available[10]. The bacterial biomass growth needs to be related to the electron transfer rate. In fact, Geobacteraceae are a family of microorganisms known for their remarkable electron transfer capabilities, which allow them to be very effective in bioremediation of contaminated environments and in harvesting electricity from waste organic matter. Bioengineering a mutant strain in order to reach faster rates in

*DAC 2011,* June 5-10, 2011, San Diego, California, USA.

electron transport yield is highly desirable and could represent a breakthrough for massive application in biotech industry. In this research paper we have given more space to photosynthesis optimization because the smaller dimensionality of the parameter space of enzyme concentrations would turn into a more challenging and insightful analysis.

This paper is organized as follows. Section 2 describes the design methodology and the PMO2 multi-objective optimization algorithm. Section 3 presents experimental results for the Carbon metabolism of C3 plants and Geobacter sulfurreducens. Section 4 concludes the work.

# 2. THE DESIGN METHODOLOGY

This section outlines the PMO2 algorithm responsible for the approximation of the Pareto-Frontiers and the robustness framework for the analysis of the candidate enzyme partitions.

## 2.1 PMO2 Algorithm

Parallel Multi-Objective Optimization (PMO2) algorithm is a multi-objective optimization framework that exploits coarse-grained parallelism to let a pool of non- dominated solutions exchange promising candidate solutions in an archipelago fashion. Using multi-objective optimization algorithms, such as NSGA-II, the framework completes with migration its approach based on islands. The *Non-dominated Sorting Genetic Algorithm II* (NSGA-II,[4]) is a *multi objective evolutionary algorithm* that is designed to assure an efficient approximation of the Pareto optimal set. It is important to note that this algorithm is derivative-free and, in particular, it does not make any assumption on the convexity or discontinuity of the Pareto-Front. Each island is a virtual place where a pool of candidate solutions (e.g., unfeasible, feasible and non-dominated solutions) is let evolve with a specific multi-objective optimization algorithm; communications among islands, in terms of solutions evolved by potentially different algorithms (or different setting of the same optimization algorithm), are arranged through a chosen archipelago topology. The island model outlines a multi-objective optimization environment in which different niches containing different populations (each population is a set of candidate solutions) are evolved by different algorithms and periodically some candidate solutions migrate into another niche to spread their building blocks, increasing the diversity of target island candidate solutions. In this archipelago approach different topology choices can raise to completely different overall solutions, introducing then another parameter that has to be chosen for each algorithm on each island. The PMO2 framework actually encloses two optimization algorithms and many archipelago topologies but its simplest configuration has been used to have a comprehensible comparison with other adopted strategies and to better understand the optimization capabilities of this approach. The adopted configuration has two islands with 2 distinct instance of NSGA-II that exchange candidate solutions every 200 generations with an all-to-all (broadcast) migration scheme at a 0.5 probability rate. Even in its simplest configuration, this approach has shown enhanced optimization capabilities and an improved convergence speed. Moreover, after selecting the Pareto-Front, it seems necessary to automatically select the "best" non-dominated solution looking at their robustness as well[9]. Additionally, when the set of Pareto optimal solutions is rich in terms of number of solutions, a screening strategy is mandatory to ensure a fair comparison of the results provided by each algorithm. Hence, we have added mining trade-off selection strategies based on the geometric considerations of the Pareto-Frontiers: non-dominated solution closest to Ideal point and the minimum (maximum) achieved for each objective function *Pareto Relative Minimum* (PRM), as defined in Sections 2.2 and 2.3.

## 2.2 Pareto-Front Mining and Analysis

Multi-objective optimization algorithms give as result a set of non-dominated solutions, instead of a single optimum (or an individual sub-optimal solution) as in single-objective optimization. In real world applications, it is useful to provide a strategy to select automatically the best trade-off solution; when the set of Pareto optimal solutions is huge, a screening strategy is mandatory. In literature, there are many trade-off selection strategies[3] typically based on the geometric notion of Pareto optimality, or heuristics based on the experimental evidence. A natural strategy is the one that selects the Pareto optimal solution that is closest to the ideal minimum of each objective. Let $P$ a set of non-dominated solutions. The *closest-to-ideal* point is defined as:

$$x \in P : \nexists y \in P : d(y, I_p) < d(x, I_p)$$

where $d : \mathbb{R}^p \to \mathbb{R}$ is a distance metrics and the ideal point is $I_p = \{\min f_1(x), \cdots, \min f_p(x)\}$. It is important to note that it is not required to know the real minimum for each objective; it is possible to use as $I_p$ the minimum achieved for each objective by the algorithm, that we call the *Pareto Relative Minimum* (PRM). Finally, the last selection criterion is the *shadow minimum* selection; according to this strategy, we select those $p$ points that achieve the lowest values on the $k$ objectives considered. It is always useful to select these points because we can gain more information on the best possible values achievable for each objective. The analysis of multi-objective optimization algorithms requires the definition of ad-hoc metrics; first, we adopt the hypervolume indicator[21]. Let $X = (x_1, \cdots, x_k) \subset \mathbb{R}^k$ a k-dimensional decision vectors; the hypervolume function $V_p : \mathbb{R}^k \to \mathbb{R}$ provides the volume enclosed by the union of polytopes $p_1, \cdots, p_i, \cdots, p_k$, where $p_i$ is formed by the intersections of the following hyperplanes arising from $x_i$ along with the axes. In order to assess the quality of Pareto optimal sets obtained by different algorithms, it is important to compare non-dominated solutions obtained: this would estimate which algorithm is able to cover effectively the front and which solutions are globally Pareto optimal. According to these considerations, we introduce two metrics; the *global* and *relative Pareto coverage*. Let $P_A = \cup_{i=1}^{m} P_i$ where $P_i$ is a Pareto-Front; $P_A$ is the Pareto-Front defined by the union of $m$ Pareto-Frontiers. We define the *global Pareto coverage* of the $i-th$ front as follows:

$$G_p(P_i, P_A) = \frac{|x \in P_i \bigwedge x \in P_A|}{|P_A|} \quad (1)$$

$G_p$ provides the percentage of Pareto optimal points of $P_i$ belonging to $P_A$; it is important to note that this metrics provides only a quantitative measure of the performance of the algorithm, since it strongly rewards large Pareto-Frontiers. This metrics gives a qualitative information if and only if the Pareto-Frontiers have a similar dimension. Although it is important to understand the composition of

$P_A$, it is also important to estimate how come that many solutions of a Pareto-Front are dominated by solutions belonging to other frontiers considered; a solution $v \in P_i$ is called *globally Pareto optimal* if it belongs to $P_A$. Let $P_A$ a global Pareto-Front, we define the *relative Pareto coverage* as follows:

$$R_p(P_i, P_A) = \frac{|x \in P_i \bigwedge x \in P_A|}{|P_i|} \quad (2)$$

$R_p$ measure the relative importance of the $P_i$ front in $P_A$. If $R_p \to 1$, two aspects are considered; the algorithm is able to find $R_p \times |P_i|$ globally Pareto optimal solutions, or it has found $R_p \times |P_i|$ solutions in a region of the front not covered by the other methods. However, it is worth noting that algorithms that are able to generate large Pareto-Frontiers are important, especially in real world application, where human experts make the final decision choosing among trade-off points. For this reason, we strongly believe that considering jointly the two metrics could effectively compare the quality of a Pareto-Front.

## 2.3 Robustness Analysis

Finding enzyme concentrations that maximize the $CO_2$ Uptake rate of a metabolic pathway is a complex optimization problem, but it can be tackled effectively by means of derivative-free optimization algorithms. Although it is possible to find optimal concentrations, the synthesis process and the changing ground conditions can alter these enzymes; it is clear that it is crucial to estimate how well the achieved $CO_2$ *Uptake* is preserved under perturbation at enzyme level. Robustness is the persistence of a system property with respect to perturbations[8, 15]. In order to evaluate the robustness of enzyme partitions, the *robustness condition* $\rho$ and the *uptake yield* $\Gamma$ have been defined. Let $\bar{x} \in \mathbb{R}^{23}$ an enzyme partitioning and $f : \mathbb{R}^{23} \to \mathbb{R}$ a function computing the expected $CO_2$ *Uptake* of $\bar{x}$. Given an enzyme partition $\bar{x}_*$ obtained by perturbing $\bar{x}$, the robustness condition $\rho$ is defined as follows;

$$\rho(\bar{x}, \bar{x}_*, f, \epsilon) = \begin{cases} 1 & if \quad \mid f(\bar{x}) - f(\bar{x}_*) \mid \leq \epsilon \\ 0 & otherwise \end{cases} \quad (3)$$

where the robustness threshold $\epsilon$ denotes the maximum percentage of variation from the reference $CO_2$ *Uptake* value (nominal value).

Let $\bar{x} \in \mathbb{R}^n$ an enzyme partitioning and $f : \mathbb{R}^{23} \to \mathbb{R}$ a function computing the expected $CO_2$ *Uptake* of $\bar{x}$. Given an ensemble $\mathcal{T}$ of perturbed enzymatic concentrations obtained by perturbing $\bar{x}$, the uptake yield $\Gamma$ is defined as follows;

$$\Gamma(\bar{x}, f, \epsilon) = \frac{\sum_{\tau \in \mathcal{T}} \rho(\bar{x}, \tau, f, \epsilon)}{|\mathcal{T}|} \quad (4)$$

where the robustness threshold $\epsilon$ denotes the maximum percentage of variation from the nominal $CO_2$ *Uptake* value. The ensemble $T$ has been generated using a Monte-Carlo algorithm; mutations occurring on all the enzymes (global analysis) and one enzyme at time (local analysis) have been considered[15]. It has been fixed a maximum perturbation of 10% on each enzyme concentration, and then it is generated an ensemble of $5 \times 10^3$ trials for the global analysis and 200 trials, for each enzyme, for the local analysis. All of the analysis assume $\epsilon = 5\%$ of the nominal uptake rate.

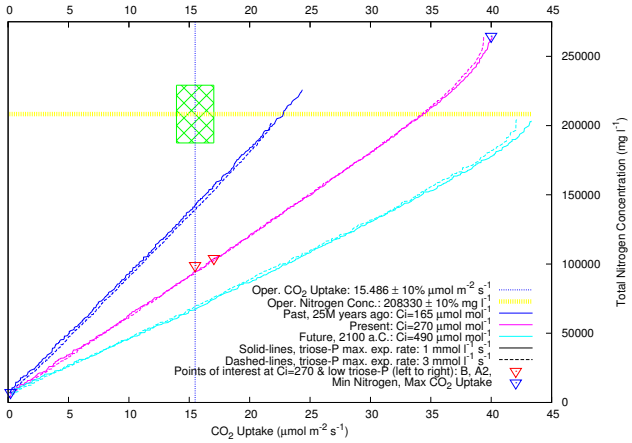


**Figure 1: Multi-objective optimization of two conflicting biological pressures: leaf $CO_2$ Uptake rate versus protein-nitrogen consumption.**

# 3. EXPERIMENTAL RESULTS

## 3.1 Maximizing photosynthetic activity and minimizing nitrogen

The aim of this section is to suggest biotechnological solutions that do not aim to change a protein sequence in order to modify system activity, but rather to act on the relative amounts of proteins in a biochemical pattern in order to increase a particular task. In this study the goal aim is to increase the $CO_2$ Uptake rate maintaining (and in other cases minimizing) the actual amount of total nitrogen contained in the enzymes.

The computational simulation of the carbon metabolism requires the definition of a set of linked ODEs; in our research work, it is considered the model proposed by[20]. The model takes into account rate equations for each discrete step in photosynthetic metabolism, equations for conserved quantities (i.e., nitrogen concentration) and a set of ODEs to describe the rate of concentration change in time for each metabolite. Reactions introduced in the model were categorized into equilibrium and non-equilibrium reactions; equilibrium reactions were inter-conversion between Glyceraldehyde 3-P (GAP) and Dihydroxyacetone-P (DHAP) in stroma and cytosol, xylulose-5-P (XuP5), Rib-5-P (Ri5P), ribulose-5-P (Ru5P) and Fru-6-P (F6P), Glc-6-P (G6P), and Glc-1-P (G1P). All non-equilibrium reactions were assumed to obey Michaelis-Menten kinetics, modified as necessary for the presence of inhibitors or activators.

The capability of reducing the amount of nitrogen necessary to fix $CO_2$ in biomass is an important goal for biotechnology. A large increase in the efficiency of nitrogen usage, will be necessary to maintain or increase current food production in a sustainable manner[17]. Intensive high-yield agriculture is dependent on the addition of fertilizers, especially industrially produced $NH_4$ and $NO_3$[17]. Figure 1 shows that the optimization may largely improve nitrogen usage in photosynthesis without affecting $CO_2$ Uptake rate. Moving beyond the natural operative area (area checked in green), we found leaf configurations that expose a Pareto-optimality in the six conditions considered (three $C_i$ atmosphere values and two triose-P export rates). The candi-

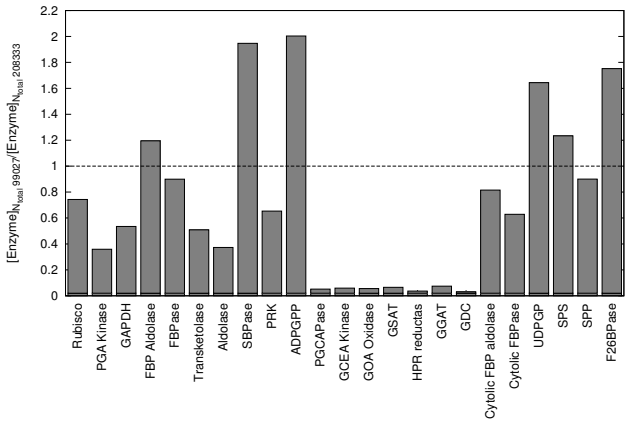


**Figure 2: Comparison among the Pareto-optimal re-engineering candidate B (that uses a total concentration of Nitrogen equal to 99 g $l^{-1}$) and the natural leaf (whose total concentration of Nitrogen is 208 g $l^{-1}$). These Nitrogen concentrations are computed, for a leaf partitioning x, as $\sum_i x_i \cdot (\text{molecular weight})_i \cdot (\text{catalytic number})_i^{-1}$.**

date highlighted as B represents a leaf with a natural $CO_2$ Uptake ability, but employs 47% of the naturally needed protein-nitrogen. The A2 candidate is interesting as well: it needs exactly 50% of the naturally employed protein-nitrogen to gain up to 10% $CO_2$ Uptake capacity, when compared to the natural leaf. The enzymes involved in concentration variation are almost always the same: Rubisco provides nitrogen to increase the concentration of other enzymes. A slight reduction in Rubisco corresponds potentially to a large amount of protein nitrogen available for increasing concentration of the other enzymes. As a matter of fact the high concentration of Rubisco in the leaves was considered to have a possible function also as nitrogen reservoir[7]. Figure 2 shows the concentration of the enzymes in the $B$ leaf with respect to natural concentrations. From a re-engineering point of view, the two leaves are similar; in fact, each enzyme involved shows a growth/reduction in concentration that is within the range 0.05x-2x ca. Despite this relatively small metric distance and the equal Uptake rate, the biochemical effort paid by the two leaf designs is substantially different. SBPase and ADPGPP confirm their leading role in the leaf engineering. These results show that re-engineering the nitrogen partitioning among well determined enzymes (individuated by our framework) can lead to theoretical leaves capable of reducing significantly the general amount of nitrogen without affecting the potential biomass production. It is interesting to observe that the enzymes of the photorespiration, a process acting against the general photosynthetic yield, are not kept at zero as in other models. Photorespiration has a major impact on carbon uptake, particularly under high light, high temperatures, and $CO_2$ or water deficits[5]. Nevertheless, although the functions of photorespiration remain controversial, it is widely accepted that this pathway influences a wide range of processes from bioenergetics, photosystem II function, and carbon metabolism to nitrogen assimilation and respiration. For instance photorespiration is a major source of $H_2O_2$ in photosynthetic cells. Through $H_2O_2$ production and pyridine nucleotide interactions, photorespiration makes a key contribution to cellular redox homeostasis. Doing so, it influences multiple signaling pathways, particularly those that govern plant hormonal responses controlling growth, environmental and defense responses, and programmed cell death[5]. Photosynthesis and particularly the biochemical pathway of carbon fixation (the Calvin Cycle) has been object of many studies (for a review see for instance[18, 6, 13]) and some journals are directly entitled to this fundamental biological process. In this research work we have identified key enzymes to target in order to maximize $CO_2$ Uptake rate and minimize the protein-nitrogen in $C3$ plants. The final aim is to re-arrange resource allocation enzyme-wise in order to obtain a robust trade-off between $CO_2$ Uptake and the total amount of protein-nitrogen. The designed methodology, including multi-objective optimization, unraveled that Rubisco, Sedoheptulosebisphosphatase (SBPase), ADP-Glc pyrophosphorylase (ADPGPP) and Fru-1,6-bisphosphate (FBP) aldolase are the most influential enzymes in carbon metabolism model where $CO_2$ Uptake maximization is concerned (our methodology robustly raises this rate up to 36.382 $\mu$ $mol$ $m^{-2}s^{-1}$; natural leaf value is 15.486 $\mu$ $mol$ $m^{-2}s^{-1}$ ca.). Interesting insights include the fact that the Rubisco enzyme participates with a very high concentration; additionally, some of the photorespiratory enzymes that should be almost switched off to reach the best configurations known[20] cannot be effectively switched off because they are involved in other processes carried by $C3$ plants. The pathway enzymes that lead to sucrose and starch synthesis were shown not to affect $CO_2$ Uptake rate if maintained at their natural concentration levels. The importance of SBPase has already been pointed out by antisense transgenic plants studies[13]. The optimization performed using the PMO2 algorithm provides a large set of trade-off solutions; in particular, 755 Pareto optimal concentrations have been found, that are the 1.83% of the total enzymes partitions explored by the algorithm. In order to assess the quality of the Pareto-Frontiers ($C_i$ of 270 $\mu mol$ $mol^{-1}$ and maximal rate of triose-P (PGA, GAP, and DHAP) export of 3 $mmol$ $L^{-1}$ $s^{-1}$), we compute the above-defined metrics for PMO2 and MOEA-D, another state-of-the-art evolutionary multi-objective optimization algorithm[19]. The results reported in Table 1 confirm the quality of the candidate solutions obtained by PMO2. Successively, from the Pareto-Front, we select the *shadow minima* for each objective and the *closest-to-ideal* solutions; successively, we compute the global robustness of these concentrations. Moreover, in addition to these solutions, we picked 50 Pareto optimal points equally spaced on the Pareto-Front and we estimated their robustness. In Table 2, it is possible to note that the three

| Algorithm | Points | $R_p$ | $G_p$ | $V_p$ |
|---|---|---|---|---|
| PMO2 | **775** | **1.0** | **1.0** | **0.976** |
| MOEA-D | 137 | 0 | 0 | 0.376 |

**Table 1: Pareto-Front analysis. For each algorithm, we report the number of Pareto Optimal points (non-dominated points), the relative Pareto coverage indicator ($R_p$), the global Pareto coverage indicator ($G_p$), and the hypervolume indicator ($V_p$).**

| Selection | $CO_2$ Uptake | Nitrogen | Yield |
|---|---|---|---|
| *Closest-to-ideal* | 21.213 | $1.270 \times 10^5$ | 67 |
| *Max $CO_2$ Uptake* | 39.968 | $2.641 \times 10^5$ | 65 |
| *Min Nitrogen* | 5.7 | $3.845 \times 10^4$ | 50 |
| *Max Yield* | 37.116 | $2.291 \times 10^5$ | 82 |

**Table 2: Pareto-Front analysis. For each Pareto optimal solution, we report the selection criterion, the $CO_2$ Uptake rate ($\mu$ mol $m^{-2}$ $s^{-1}$), the nitrogen amount (mg $l^{-1}$) and the yield value (in percentage).**

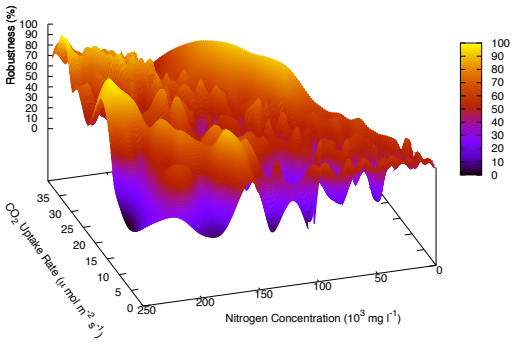


**Figure 3: Photosynthetic Pareto-Surface. Robustness vs. $CO_2$ Uptake and Nitrogen consumption.**

concentrations selected by the automatic criterion are quite robust (Yield column), even if they greatly differ in terms of $CO_2$ Uptake rate and nitrogen concentration; this experimental evidence seems to confirm that trade-off concentrations represent robust pathway configurations despite the changes in their uptake capability and nitrogen required. However, by inspecting the Pareto-Front it is possible to find a new enzyme partition that achieves a slightly worse uptake rate but a remarkable increase in terms of robustness; from this analysis, it is clear that the yield is another conflicting objective and, hence, an inherent trade-off emerges.

More in detail, to inspect the relation between $CO_2$ Uptake, Nitrogen consumption and the inherent solution robustness, we have used our framework to assess the fitness landscape with respect to these three objectives. Figure 3 presents the results of this analysis by means of a 3D Pareto-Surface. Despite the rugged aspect of the surface, that highlights how far from an ideal world and how real is the problem we are tackling, it is clear that Pareto relative minima are highly unstable points, while if we accept a slightly lower optimization in the functional objectives, we can obtain a significantly more reliable solution.

Finally, looking at the concentrations of the *closest-to-ideal* solutions, some more interesting results are observable; except for the GOA Oxidase, each algorithm maintains a concentration close to the natural concentrations. Remarkable increases are observable for GAP DH, GGAT, Cytolic FBP Aldolase, SPP and F26BPase enzymes. At this point, it is possible to infer that these enzymes are the

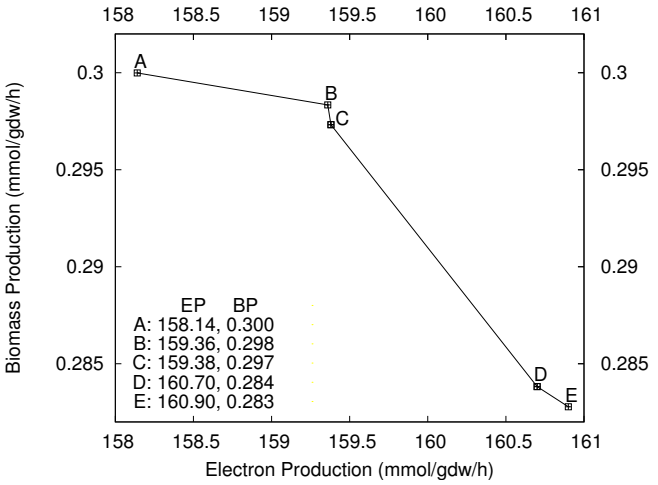


**Figure 4: Pareto Front of Geobacter sulfurreducens: maximization of biomass production versus maximization of electron production. The unit for flux values is mmol/gDW/h (DW is dry weight).**

best candidates for a trade-off performance leaf. Clearly, it is important to remark that modest increments of other enzymes are plausible since they have a higher molecular weight. It should be observed that even if some of the considered enzymes fall to zero in main photosynthesis models in the optimized leaf, such a low concentration could influence other important biochemical pathways. For instance photorespiration-related enzymes as Glu Glyoxylate Aminotransferase and GOA oxydase fall considerably in concentration at the optimized state. Photorespiration is by far the fastest $H_2O_2$ -producing system in photosynthetic cells under many conditions[12]. $H_2O_2$ is an important intracellular signal[5]. Moreover the photorespiratory pathway metabolizes glycolate-2-P to Glycerate-3-P and is considered important to avoid photoinhibition of photosystem II, particularly in $C3$ plants[16]. Photorespiratory mutants of Arabidopsis with inactivation of some of the enzymes of the photorespiratoy pathway did not show negative effects at high level of external $CO_2$, but $CO_2$ fixation rates declined drastically at current atmospheric $CO_2$ concentration [16]. That means that models based only on the photosynthetic pathways leading to strong decrease in concentration of the photorespiratory pathway enzymes should take into consideration that this pathway is necessary to the plant for aspects that have not been considered in current models.

## 3.2 Geobacter sulfurreducens: Maximizing Biomass and Electron Productions

Constraint-based modeling of metabolism has laid the foundation for the development of computational algorithms, which allow more efficient manipulations of metabolic networks. One established approach, OptKnock, has already yield good results in suggesting gene deletion strategies leading to the overproduction of biochemicals of interest in E. Coli[2]. These increments are accomplished by dropping some redundancy in the metabolic pathways in order to eliminate reactions competing directly, or in a more complex way, with those we are interested in. Geobacter sulfurreducens, modeled as an in-silico organism[10], has been

optimized by perturbing its 608 reaction fluxes with the constraint that steady state solutions are preferred (i.e., $S \cdot \bar{x} = 0$, where $S$ is the stoichiometric matrix, $\bar{x}$ the perturbed flux vector and 0 is the null vector). The optimization has been designed to move towards those solutions where two crucial fluxes are maximized: Electron Production and Biomass Production. Five non-dominated solutions ($A$-$E$) are reported in Figure 4 as best trade-offs. In particular, in our multi-objective constrained optimization, the solution $A$ presents a significant slope in the constraint violation reduction: $3.4 \cdot 10^4$ is roughly 1/26.47 when compared to the initial guess solution (that showed a violation in the order of $10^6$) and it keeps decreasing towards steady state solutions. To our knowledge this is the first time that a multi-objective optimization that faces both electron and biomass production is implemented for Geobacter sulfurreducens. Our approach brought a set of Pareto-optimal solutions such that: (i) an enhanced electron and biomass productions are achieved, (ii) the constraint violation is minimized by the algorithm that rewards less violating solutions, and (iii) all of the biological constraints highlighted by the Flux Balance Analysis pointed out by Cobra toolbox[1] on this pathway are intrinsically enforced because they define the search space boundaries in our algorithm. An important bound that is worth mentioning is the ATP: its flux is kept fixed at 0.45 as highlighted in[10] as best value assessed.

Our explorations in Pareto-Front analysis suggest that its shape may reflect the amount of epistasis (where the effects of one gene are modified by one or several other genes) and pleiotropy (when a single mutation or gene affects multiple distinct phenotypic traits) in the metabolic pathway, so that simpler independent traits may generate simpler Pareto-Frontiers. It is known that complexity and in particular fitness traits such as energy balance, growth and survival, depend on both the epistatic and pleiotropic structure of a metabolic pathway and therefore it strongly influences evolutionary predictions.

# 4. CONCLUSION

The multi-objective formulation of the redesign process poses a serious algorithmic challenge, since the defined Pareto Front is not easily analyzable; for this reason, a derivative-free multi-objective optimization algorithm, PMO2, has been designed with the aim of producing a good approximation of Pareto optimal concentrations. Moreover, the computational redesign process has been completed with the introduction of the robustness analysis, which assures an estimation of the persistence of the designed properties under perturbation. The analysis of the results has shown that it is possible to obtain a gain of the uptake rate while minimizing the amount of nitrogen required; the yield analysis has shown a clear propensity of remaining in a robust state of the great majority of the solutions, even if, for a real implementation, the robustness results should be taken into account as an additional selection criterion. Finally, we have applied the PMO2 algorithm to the Geobacter sulfurreducens as well and with respect to that we have obtained a computational model that maximizes the electron and biomass productions while preserving those bounds that ensure a biological significance. To our knowledge this is the first time that Geobacter sulfurreducens has been modeled as a multi-objective optimization problem where the search moves automatically towards steady state solutions, contextually with biological boundaries observance and functional optimization (i.e., biomass and electron productions).

# 5. REFERENCES


[1] S. A. Becker, A. M. Feist, M. L. Mo, G. Hannum, B. O. Palsson, and M. J. Herrgard. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA toolbox. *Nat. Protocols*, 2(3):727–738, Mar. 2007.

[2] A. P. Burgard, P. Pharkya, and C. D. Maranas. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and Bioengineering*, 84(6):647–657, Dec. 2003.

[3] V. Cutello, G. Narzisi, and G. Nicosia. A multi-objective evolutionary approach to the protein structure prediction problem. *J. of the Royal Society Interface*, 3(6):139–151, 2006.

[4] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. on Evol. Comp.*, 6(2):182–197, 2002.

[5] C. H. Foyer, A. J. Bloom, G. Queval, and G. Noctor. Photorespiratory metabolism: genes, mutants, energetics, and redox signaling. *Ann Rev of Plant Biol.*, 60:455–484, 2009.

[6] D. R. Geiger and J. C. S. JC. Diurnal regulation of photosynthetic carbon metabolism in c3 plants. *Annu Rev Plant Phys Plant Mol Biol*, 45:235–256, 1994.

[7] B. Hirel, J. L. Gouis, B. Ney, and A. Gallais. The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches. *J. Exp. Bot.*, 58(9):2369–2387, 2007.

[8] M. Jose, Y. Hu, R. Majumdar, and L. He. Rewiring for robustness. In *Proceedings of the 47th Design Automation Conference*, DAC '10, pages 469–474, New York, NY, USA, 2010. ACM.

[9] Y. Liu, M. Yoshioka, K. Homma, T. Shibuya, and Y. Kanazawa. Generation of yield-embedded pareto-front for simultaneous optimization of yield and performances. In *Proceedings of the 47th Design Automation Conference*, DAC '10, pages 909–912, New York, NY, USA, 2010. ACM.

[10] R. Mahadevan, D. R. Bond, J. E. Butler, A. Esteve-Nuñez, M. V. Coppi, B. O. Palsson, C. H. Schilling, and D. R. Lovley. Characterization of metabolism in the Fe(III)-Reducing organism geobacter sulfurreducens by Constraint-Based modeling. *Applied and Environmental Microbiology*, 72(2):1558–1568, Feb. 2006.

[11] T. McConaghy, P. Palmers, G. Gielen, and M. Steyaert. Simultaneous multi-topology multi-objective sizing across thousands of analog circuit topologies. In *Proceedings of the 44th annual Design Automation Conference*, DAC '07, pages 944–947, New York, NY, USA, 2007. ACM.

[12] G. Noctor, S. D. Veljovic-Jovanovic, S. Driscoll, L. Novitskaya, and C. H. Foyer. Drought and oxidative load in the leaves of c3 plants: a predominant role for photorespiration? *Ann. Bot.*, 89:841–50, 2002.

[13] C. A. Raines. The Calvin cycle revisited. *Photosynthesis research*, 75(1):1–10, 2003.

[14] W. Sheng, L. Xiao, and Z. Mao. Soft error optimization of standard cell circuits based on gate sizing and multi-objective genetic algorithm. In *Proceedings of the 46th Annual Design Automation Conference*, DAC '09, pages 502–507, New York, NY, USA, 2009. ACM.

[15] G. Stracquadanio and G. Nicosia. Computational energy-based redesign of robust proteins. *Computers and Chemical Engineering*, 35(3):464–473, 2011.

[16] S. Takahashi, H. Bauwe, and M. Badger. Impairment of the photorespiratory pathway accelerates photoinhibition of photosystem II by suppression of repair but not acceleration of damage processes in Arabidopsis. *Plant Physiology*, 144(1):487–494, 2007.

[17] D. Tilman, K. G. Cassman, P. A. Matson, R. Naylor, and S. Polasky. Agricultural sustainability and intensive production practices. *Nature*, 418:671–677, 2002.

[18] I. E. Woodrow and J. A. Berry. Enzymic regulation of photosynthetic co2 fixation in c3 plants. *Annu Rev Plant Physiol Plant Mol Biol*, 39:533–594, 1988.

[19] Q. Zhang and H. Li. Moea/d: A multiobjective evolutionary algorithm based on decomposition. *IEEE Trans. on Evol. Comp.*, 11(6):712–731, 2007.

[20] X.-G. Zhu, E. de Sturler, and S. P. Long. Optimizing the distribution of resources between enzymes of carbon metabolism can dramatically increase photosynthetic rate: A numerical simulation using an evolutionary algorithm. *Plant Physiology*, 145:513–526, 2007.

[21] E. Zitzler, D. Brockhoff, and L. Thiele. The hypervolume indicator revisited: On the design of pareto-compliant indicators via weighted integration. *Lecture Notes in Computer Science*, 4403:862, 2007.