Pareto Optimality in Organelle Energy Metabolism Analysis

Claudio Angione, Giovanni Carapezza, Jole Costanza, Pietro Lió, and Giuseppe Nicosia

Abstract—In low and high eukaryotes, energy is collected or transformed in compartments, the organelles. The rich variety of size, characteristics, and density of the organelles makes it difficult to build a general picture. In this paper, we make use of the Pareto-front analysis to investigate the optimization of energy metabolism in mitochondria and chloroplasts. Using the Pareto optimality principle, we compare models of organelle metabolism on the basis of single- and multiobjective optimization, approximation techniques (the Bayesian Automatic Relevance Determination), robustness, and pathway sensitivity analysis. Finally, we report the first analysis of the metabolic model for the hydrogenosome of Trichomonas vaginalis, which is found in several protozoan parasites. Our analysis has shown the importance of the Pareto optimality for such comparison and for insights into the evolution of the metabolism from cytoplasmic to organelle bound, involving a model order reduction. We report that Pareto fronts represent an asymptotic analysis useful to describe the metabolism of an organism aimed at maximizing concurrently two or more metabolite concentrations.

Index Terms—Mitochondrion, chloroplast, hydrogenosome, sensitivity analysis, multiobjective optimization, robustness analysis

1 Introduction

In recent years, the fields of mitochondria, chloroplasts, and other mitochondrion-like organelles (MLO), for example, hydrogenosome and mitosome, have witnessed an extraordinary expansion. This is mainly due to the identification of the pivotal role that mitochondria play in human disease and ageing [1], to the synergy shown by chloroplasts and mitochondria in energy output [2], to the development of new organelle proteomics tools [3], and to the discovery of novel factors involved in organelle division, movement, signaling, and adaptation to varying environmental conditions [4]. Remarkably, no examples of examined eukaryotes lacking a mitochondrion-related organelle exist [5].

It is now widely accepted that all extant eukaryotes are descended from an ancestor that had a mitochondrion. In anaerobic eukaryotes, this organelle has been modified into either a hydrogenosome, which continues to generate energy for the host cell, or a mitosome, which does not. Therefore, the evolutionary history of chloroplasts and mitochondria are intertwined, and a better knowledge of the pathways in both organelles and vestigial organelles (e.g., apicoplast [6] and hydrogenosome [7]) may lead to drug discovery. Interestingly, when a mitochondrion and other organelles coexist in the same cell, their volume and activity can be strongly related with one another. In these cells, mitochondria exhibit behaviors that are absent in other eukaryotic cells [8].

- *C. Angione and P. Lió are with the Computer Laboratory, University of Cambridge, United Kingdom. E-mail: {claudio.angione, pietro.lio}@cl.cam.ac.uk.*
- *G. Carapezza, J. Costanza, and G. Nicosia are with the Department of Mathematics & Computer Science, University of Catania, Italy. E-mail: {carapezza, costanza, nicosia}@dmi.unict.it.*

Manuscript received 20 Aug. 2012; revised 30 June 2013; accepted 9 July 2013; published online 1 Aug. 2013.
For information on obtaining reprints of this article, please send e-mail to: tcbb@computer.org, and reference IEEECS Log Number TCBB-2012-08-0212. Digital Object Identifier no. 10.1109/TCBB.2013.95.

Chloroplasts, mitochondria, hydrogenosomes, and other MLO share many common features. For instance, the hypothesis of Colocation for Redox Regulation (CoRR) predicts that the regulatory coupling operated continuously before, during, and after the transition from prokaryote to eukaryotic organelle [9]. According to the CoRR hypothesis, these organelles would lose their genomes in case of loss of the redox and proton-motive machinery of oxidative phosphorylation. Moreover, they are "smart" energy-transducing devices and make decisions on the basis of environmental changes affecting redox poise. CoRR operates today and will continue to operate in living cells, being a necessary condition for the compatibility of energy conversion with genome function. Finally, these organelles exploit a common pathway to harness energy for biological purposes. This process is called chemiosmotic coupling, since it involves both the chemical bond-forming reactions that generate ATP (chemi) and the membrane-transport processes (osmotic).

Although mitochondria and chloroplasts are two different organelles with different functions, it is useful to compare the electron-transport processes in mitochondria, which convert energy from chemical fuels, with those in chloroplasts, which convert energy from sunlight. Chloroplasts produce glucose from sunlight energy. This glucose then transfers to the mitochondrion for aerobic respiration. The function of chloroplasts is basically to make food, by trapping light energy to convert water and carbon dioxide to form oxygen and glucose (this process is called photosynthesis). It is noteworthy that the plant produces most of its ATP from aerobic respiration, not from photosynthesis. Conversely, the function of mitochondria is to provide energy for the cell's use. Mitochondria extract the energy from the glucose molecules, and store it in ATP molecules; the ATP molecules will then diffuse out of the mitochondrion for the cell's use.

We aim at investigating and comparing the complexity of these organelles through a common framework that

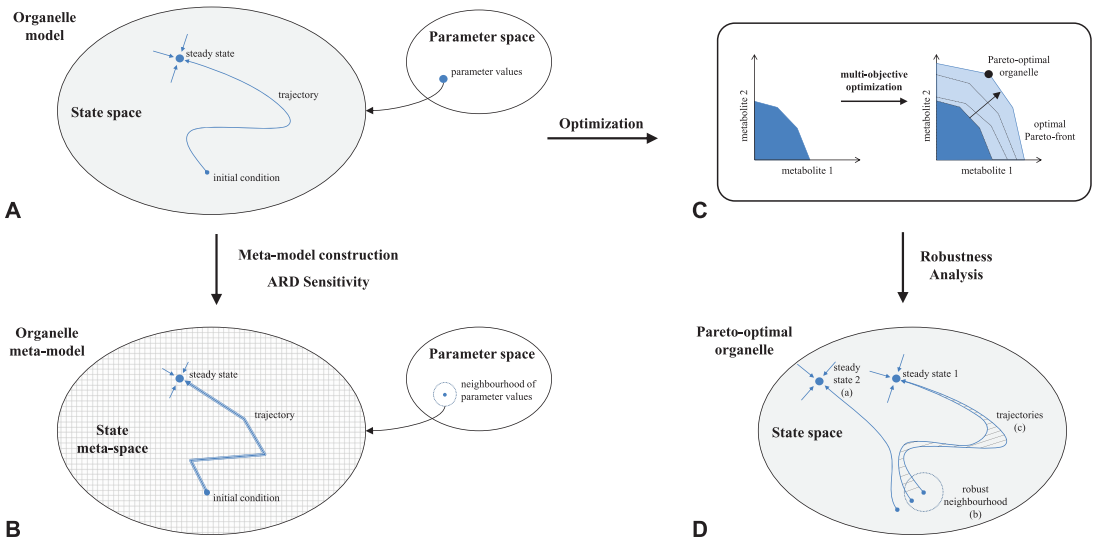


Fig. 1. The common framework to analyze organelles. We start from the complete model of the organelle, whose state space reflects its metabolism. We reduce the organelle model (A) to a metamodel (B) in which we can analyze its high-dimensional parameter space and evaluate the sensitivity of all its parameters by perturbing them in a neighborhood of the original values. Then, we perform a multiobjective optimization (C) on the organelle metabolism, to find the Pareto-optimal front involving two or more metabolites of interest (e.g., ATP and NADH); the genetic algorithm underlying the optimization allows us to reach the optimal Pareto front, i.e., to move the front toward the optimal point (e.g., maximum ATP and NADH), which is unfeasible if the two objective are negatively correlated with one another. Finally, we evaluate the robustness (D) of the Pareto-optimal solutions. Kitano [17] has remarked on the need for a general theory of biological robustness. According to him, a system is robust if it maintains its functionality, even if it transits through a new steady state or if it is unstable (a). According to Gunawardena [18], the robustness to change of initial conditions is called dynamical stability; for instance, one can evaluate the differences in the dynamics of the system (c). According to Stracquadanio and Nicosia [14], the robustness of a system is the number of robust trials over the total number of trials; a perturbation trial is said to be robust when the perturbation is in the robust neighborhood (b) such that the output remains in a given interval.

includes single- and multiobjective optimization, robustness analysis, and sensitivity analysis. In addition to the possible goals and methods reviewed by Handl et al. [10], the optimization of biological networks has been recently performed using mixed integer approaches [11]. The multiobjective optimization in organelles such as the mitochondrion may be related to the different tasks of maximizing the ATP or the heat, or intermediate compounds of the Krebs cycle to provide input for biosynthetic pathways (e.g., the amino acids synthesis). Therefore, we propose the investigation of the evolution from the ancestral bacterium in light of their multiobjective optimization. Pareto fronts are related to the sensitivity and robustness of an organism, and represent a key tool for the metabolic analysis.

The ODE models we take into account are the chloroplast model by Zhu et al. [12] and the mitochondrion model by Bazil et al. [13]. First, we perform a sensitivity analysis to identify the most important components of both organelles. Second, we optimize the ability of these organelles to produce CO_2 and ATP, respectively. The maximization of these metabolites allows for the optimization of both metabolic networks. Finally, we assess the fragileness of the multioptimized metabolic networks using the robustness analysis [14]. Our framework focuses on the fundamental properties of these organelles, and their integration into broader physiological processes. The paper will emphasize the commonalities and the differences in the energy production in mitochondria, chloroplasts, and hydrogenosomes, in light of their evolution from bacterial endosymbionts. A comprehensive analysis of these organelles through a multiobjective optimization approach highlights engineered endosymbiotic relationships between different species and, therefore, represents a valuable tool for synthetic biology [15].

Finally, we propose a model for the hydrogenosomes in *Trichomonas vaginalis*, with the aim of understanding similarities and differences relative to other organelles (e.g., which pathways are retained from mitochondria, and which ones are lost). The hydrogenosome was first noted in *Tritrichomonas foetus* as a component that produced hydrogen and ATP. The origin of this organelle remains a matter of debate, and an extensive analysis on its energy production could give new hints in understanding its ancestral relationship to mitochondria. Thus, using our model, we carry out a flux balance analysis (FBA) [16] taking into account the ATP production as one of the objective functions to optimize.

2 A Common Optimization Framework for Organelles

In Fig. 1, we propose the steps of a common pipeline to analyze organelles. Although not reported in the figure, it is noteworthy that organelles are not closed systems. Indeed, they receive input signals from the nuclear genome, thus several proteins of the cell are transported to them. In every parameterized dynamical system with a given initial condition, each value assumed by the vector of parameters in the parameter space corresponds to a trajectory of the system in the state space. The system converges to the steady state through a dynamics, thus the steady state acts as an attractor [18]. In the ODE models described in this paper, the parameter and state spaces are Euclidian spaces. As regards the size of the state space, we

expect the following inequalities to hold: 1) ancestral bacterium > mitochondrion, and 2) chloroplast > hydrogenosome > apicoplast.

To apply the sensitivity analysis in the whole parameter space, we build an organelle metamodel using SUMO toolbox [19]. In this paper, we apply it to the mitochondrial model proposed by Bazil et al. [13] to obtain a less accurate model that can be deeply analyzed even if the number of inputs is very large. In our case, we are able to investigate the combined effect of 42 input parameters on the ATP yield and to perform a sensitivity analysis by means of the Bayesian Automatic Relevance Determination (ARD). Since the parameter space is 42-dimensional, through a metamodel we avoid the computational cost of running the whole mitochondrial model in the parameter space. The parameters of the metamodel belong to a neighborhood of the parameter values of the original model. Since the metamodel is an approximation of the real model, given the same initial condition, the trajectories in the state metaspace differ from those in the original space. Thanks to the adaptive sampling of the original model, the attractor (i.e., the steady-state point) is the same as that of the original model. Therefore, we can easily investigate the sensitivity of the parameters and obtain a rank that indicates to what extent a change in each parameter affects the behavior of the system.

The multiobjective optimization exploits the concept of Pareto optimality to maximize two or more desired metabolites in the organelle, thus obtaining new artificial strains. The choice of the target metabolites depends on the organelle under investigation. For instance, in mitochondria we will analyze ATP and NADH, the energy molecules in the cell, while in chloroplasts we will focus on the CO_2 uptake, which drives the photosynthesis. They are the most important products in these organelles and, when maximized, allow them to work at the optimal configuration. In particular, the multiobjective optimization turns out to be very useful when we need to maximize simultaneously two or more metabolites responsible for the energy production.

Finally, the robustness analysis assesses the fragility of a strain obtained through the optimization when it undergoes small perturbations, both external (changes in the nutrients) and internal (changes in the metabolism). One can define the robustness region, in which a change in the initial condition does not prevent the algorithm to reach the same steady state.

We apply this methodology to mitochondria and chloroplasts not only to provide meaningful insights into the production or uptake of their key metabolites, but also to obtain a comparison based on Pareto fronts between these two organelles.

3 Comparison between Mitochondria and Chloroplasts Optimization Models

In this section and in the following ones, we compare the behavior, the sensitivity, the optimization, and robustness of the mitochondrion and chloroplast models.

In mitochondria, electrons from a carbohydrate food molecule degrading to CO_2 are transferred through the membrane by a chain of electron carriers, and finally reduce oxygen gas (O_2) to form water. Since they flow down this path from a high-energy state to a low-energy state, electrons release free energy, which allows to drive a series of three H+ pumps in the inner mitochondrial membrane. The third H+ pump catalyzes the transfer of the electrons to O_2. The mechanism of electron transport between two sites is carried out by diffusible molecules that can pick up electrons at one location and deliver them to another location [20]. In mitochondria, the first electron carrier, NAD+, is able to take up two electrons (plus an H+) to become NADH, a water-soluble small molecule that ferries electrons from the sites where food molecules are degraded to the inner mitochondrial membrane. All the proteins in the membrane and all the small molecules involved in the electron transfers form an electron-transport chain.

In chloroplasts, the transfer of electrons is driven by photosystems. In these membrane components, a pigment called chlorophyll captures the energy from the light and, through the C3 cycle, produces the energy-storage molecules ATP and NADH. This energy allows to transfer electrons. Remarkably, while the mitochondrion consumes O_2, the chloroplast generates it. This is due to the fact that the direction of transfer in chloroplasts is opposite to that in mitochondria. Indeed, electrons are taken from water to produce O_2, and they are donated to CO_2 to synthesize carbohydrate (the donation occurs through NADPH, which is closely related to NADH).

The mitochondrial model [13] consists of 73 DAEs, each of which represents either a constraint or the rate of variation of a metabolite involved in bioenergetic reactions of mitochondria. The state variables were initialized to achieve the fully oxidized state [13]. In this work, we calculated the metabolites content that leads to maximize the matrix content of ATP and NADH, maintaining constant oxidized cytochrome c, reduced cytochrome c, ubiquinone, ubiquinol, NAD_{mtx}, $NADH_{mtx}$, GTP_{mtx}, GDP_{mtx} (mtx = matrix), the mitochondrial membrane potential (1 mV), the matrix O_2 (0.0652 nmol/mg), the total CO_2 (21.4762 nmol/mg) and Ca^{2+}. We initialized Ca^{2+} with five different values to evaluate the behavior of mitochondria. First, we used $Ca^{2+} = 10^{-5}$, i.e., the standard condition, and then 10^{-4}, 10^{-6}, $10^{-5} \times 1.5$, and $10^{-5}/1.5$ nmol/mg.

As regards chloroplasts, in the model by Zhu et al. [12] the photosynthetic metabolism is modeled through equations for conserved quantities (e.g., total leaf nitrogen) and a set of 31 linked ODEs, each of which describes the rate of change in a metabolite. The C3 metabolism takes place in four compartments: thylakoid membranes, thylakoid lumen, chloroplast, stroma, and cytosol. Photoreactions and electron carriers are embedded in the membrane, while protons accumulate in the lumen. Carbon reduction and starch synthesis pathways are located in the stroma with sucrose synthesis in the cytosol. The phosphate translocator transports metabolites through the chloroplast membrane. The molar concentrations of the metabolites required in the kinetic equations are the state variables. If necessary, the total amount of a specific compound per unit leaf area is expressed by multiplying its concentration by the volume of the compartment per unit leaf area.

4 Sensitivity Analysis

It is extremely important to focus the in silico design on a set of restricted significant parameters, to decrease the complexity of a future biological implementation. In this research work, we perform sensitivity analysis both on mitochondria and on chloroplasts. We infer that the most sensitive enzymes govern the key reactions in the metabolism, and they are directly linked with the energy production.

4.1 Pathway Sensitivity Analysis in a Chloroplast Model

The representation of the chloroplast as a set of linked ODEs gives a mathematical description of the chemical process and, successively, the Morris analysis gives useful insights on linear and nonlinear contribution of enzymes to the Carbon metabolism. In the set of ODEs of the model by Zhu et al. [12], we have evaluated the most sensitive components through a one-factor-at-a-time (OAT) method proposed by Morris [21]. According to the OAT method, only one input is perturbed while the others are kept at their nominal value. We consider our pathway as a black box with certain inputs and certain outputs. Each step-variation computed for each input is an elementary effect calculated as $u_i = (P(x_1, \ldots, x_i + \Delta_i, \ldots, x_k) - P(x_1, \ldots, x_i, \ldots, x_k))/\Delta_i$, where P represents the pathway, $(x_1, \ldots, x_i, \ldots, x_k)$ is the nominal vector, and Δ_i is the perturbation affecting the ith input. For each factor, we collect an ensemble of elementary effects and compute their mean μ_i and the standard deviation σ_i. We expect a high global influence from those inputs with a high μ_i, and a highly nonlinear or counterintuitive behavior from those inputs with a high σ_i.

For each enzyme (i.e., input), we use the concentration reported in Table SI1, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TCBB.2013.95, as nominal value [22], computing 20 different factor levels, each of which is altered 10 times. We perturb each nominal value by a factor of 10 (both up and down). The sensitivity analysis shows that 11 (out of 23) enzymes are extremely sensitive [23], namely RuBisCO, PGA kinase, GAP dehydrogenase, FBP aldolase, FBPase, SBP aldolase, SBPase, Phosphoribulose kinase, ADPGPP, Phosphoglycolate phosphatase, and GDC. These enzymes showed also high values of σ_i (i.e., $1 < \sigma_i < 15$) when compared to all of the others (i.e., $10^{-4} < \sigma_i < 1$).

To validate these results, we take into account the interaction map defined by the photosynthesis pathway. In particular, we expect that the sensitive enzymes represent the hubs of the pathway. Indeed, the Roswall's community detection method [24] confirmed our assumptions, detecting RuBisCO and GAP dehydrogenase as the most strongly regulated enzymes of the C3 cycle (both enzymes are light regulated). Another key enzyme is Transketolase because it uses as substrates Fructose-6-P (which otherwise would exit from the cycle toward the starch biosynthetic pathway) and 3-P-Glyceraldehyde (produced by GAP dehydrogenase). These enzymes correspond to the main nodes of the C3 cycle leading to the other biosynthetic pathways [22].

4.2 Sensitivity Analysis in a Mitochondrial Model

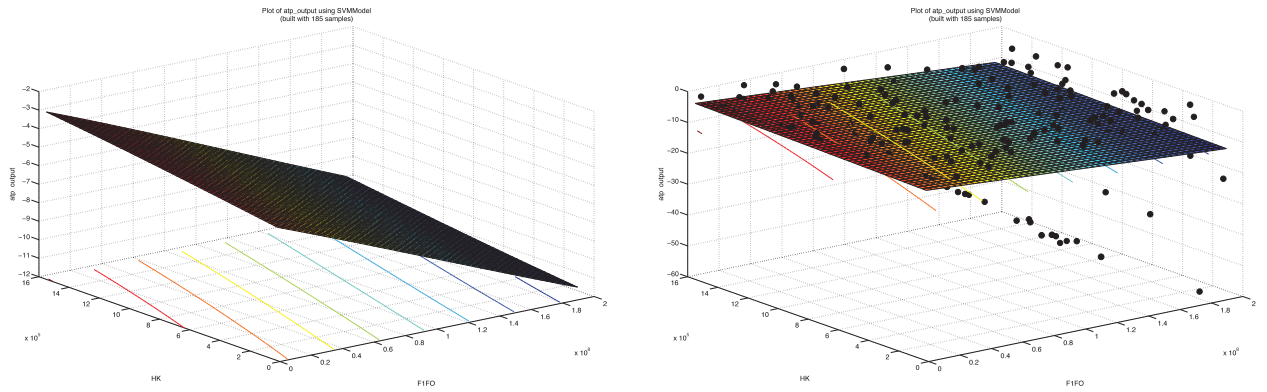
As in the sensitivity analysis carried out on the chloroplast model, the representation of the mitochondrion as 73 Differential Algebraic Equations (DAEs) and, successively, the Morris analysis gives useful insights into linear and nonlinear contribution of the metabolite concentrations involved in the bioenergetic reactions of mitochondria. For each metabolite (i.e., input), we use the fully oxidized state as nominal values, computing 20 different factor levels, each of which is altered 10 times. We perturb each nominal value by a factor of 10 (both up and down). The sensitivity analysis results are shown in supplementary information, available online, (Table SI2, third and fourth column). The metabolites are sorted from the most sensitive to the least sensitive. Only the total adenosine monophosphate (AMP) content and the total fumarate (FUM) content can be considered insensitive variables as they have values of μ and σ near zero. The other variables are sensitive or extremely sensitive. Notably, the most sensitive variable is the free potassium (K^+) content.

4.3 Bayesian ARD Sensitivity on Mitochondrial Parameters

After the sensitivity analysis performed on the initial concentrations, our idea is to investigate the model by Bazil et al. [13] by evaluating how the choice of the adjustable parameters influences the output. The right value of the parameters is hard to determine experimentally. After the publication of the mitochondrial model, the authors proposed an update of the vast majority of the model parameters after the modifications outlined by Wu et al. [25] due to the change to the function used to compute the ionic strength. These corrections witness the fact that both the estimation and the sensitivity of parameters deserve a closer look.

Currently, the most commonly used estimation procedures rely only on nonlinear least squares or on Monte Carlo techniques based on optimization algorithms (both have been used by Bazil et al. in the mitochondrial model). These approaches are computationally expensive and often poorly suited for statistical inference. Alternative methods that use the P-spline theory to estimate the state functions and the ODE parameters were proposed by Ramsay et al. [26]. When the number of parameters is high, the sensitivity analysis can detect the most important ones. Therefore, one can put more effort in estimating the key parameters, rather than using the same technique for all the parameters. Nevertheless, because of the high-dimensional parameter space of the mitochondrial model, typical approaches (e.g., the OAT method adopted for the chloroplast) may not accurately assess the sensitivity of the system; furthermore, the simulation code needed to evaluate the output in the whole parameter space is computationally expensive. Using a surrogate model in the way described in this paper, we are able to globally investigate a high-dimensional parameter space, to identify all the sensitive and uncertain points.

Using the SUMO toolbox [19], we build a surrogate mathematical model that mimics the behavior of the mitochondrion over the complete parameter space. The metamodel adaptively selects as few data points as



(a) (b)

Fig. 2. Both (a) and (b) refer to the ATP production [nmol/mg] (times 10^6) as function of the Hexokinase max rate [nmol/mg/min] and the F_1F_0 ATP synthase activity [nmol/M/min/mg]. The minus sign indicates that the mitochondrion is consuming ATP. In (b), we highlight that the 185 points sampled by SUMO include some outliers not considered in the construction of the metamodel.

possible in the parameter space, and reproduces the output as accurately as possible. This way, we are able to explore the design space and investigate how the combined effect of all the parameters influences the output of the model. We use a least squares support vector machine (LS-SVM) to build a continuous model from the discrete ATP output in the 42-dimensional parameter space. LS-SVM is a learning method that recognizes patterns from given data points [27]. Support vector machines, introduced within the statistical learning theory, is a powerful methodology for solving problems such as function estimation, which has also led to kernel-based learning methods.

Given the parameters p_i, $i = 1, \ldots, 42$, proposed by Bazil et al. we consider the parameter space $\prod_i [0, 2p_i]$. LS-SVM adaptively samples the parameter space to build a continuous model of the output, by minimizing the least squares error between the model and the output evaluated on the points already sampled in the parameter space. The regression is carried out by weighting the inputs so as to obtain the most probable model for the given data points. The first 20 points are selected randomly in the parameter space so as to start the training of the model. Then, the parameter space is sampled dynamically; thus, the surrogate model is built and refined as more points are available. The procedure is halted when the least squares error becomes smaller than a fixed tolerance.

Starting from the surrogate model, we perform a sensitivity analysis by means of the Bayesian Automatic Relevance Determination (ARD) [28]. This allows to investigate the sensitivity of the parameters and understand how a change in a parameter affects the behavior of the whole system. ARD is used to determine the subset of the most relevant inputs for the proposed model. We assign a different weighting parameter to each dimension in the kernel inference. Each input is assigned a value σ^2. In each step, LS-SVM performs a backward selection by removing the input with the largest optimal σ^2. For every step, the Bayesian cost criterion is computed, based on the singular value decomposition (SVD) of the kernel matrix.

To rank the parameters, the effect of each parameter must be compared to the effect of all the others. To this end, in each step of the ARD algorithm a parameter is removed according to a minimal cost criterion. Although at each step a parameter is removed with the smallest cost at that step, it may happen that at a step after the smallest cost is larger. This way, we obtain a list of inputs in decreasing order of relevance (see Table SI3, available online).

According to the Bayesian ARD, the Hexokinase max rate and the F_1F_0 ATP synthase activity are the parameters having the largest influence on the production of ATP in mitochondria. To analyze more thoroughly the role of these two key players, in Fig. 2, we plot the concentration of ATP over the total concentration of adenine nucleotides, according to our metamodel. The minus sign indicates a production of ATP. Since the mitochondrial species for adenine nucleotides are conserved in the model, the following algebraic equations of conservation holds: $[ATP] + [ADP] = A_{tot}$. On the z-axis we plot $10^6([ATP]/A_{tot}) = 10^6(1 - [ADP]/A_{tot})$.

In Fig. 2b, we show the metamodel curve along with the 185 points sampled by SUMO to create the metamodel. The plot confirms that F_1F_0 ATP synthase activity is a highly sensitive parameter. In particular, there are some regions of the parameter space of F_1F_0 ATP synthase activity, and specifically the neighborhood of 1×10^9, in which SUMO finds points considered as outliers that cause a fast decrease of the ATP production.

To have a closer look at the model behavior as function of the parameter values, we build a second-order and a third-order polynomial surrogate models using 250 and 1,028 samples (respectively) in the parameter space. Let us denote by p the vector of the parameter values. For the second-order approximation, the ATP is given by

$$a_0 + c^\top p + p^\top A p, \tag{1}$$

where $a_0 = -5.98$, while c, p, and A are defined in Figs. SI1 and SI3, available in the online supplemental material, and the $\top$ symbol denotes the transpose operation. The heat map of A is shown in Fig. SI2 in supplementary information, available online. For the sake of completeness, in Fig. SI4, available online, we report the curve for the third-order polynomial metamodel.

The sensitivity applied to the mitochondrial model allows to draw the same conclusion as that applied to the

chloroplast model. In both cases, the sensitive enzymes are the key players in the pathway of these organelles, i.e., the enzymes directly associated with the production of energy.

5 Multiobjective Optimization

The methodology we propose in this section allows to optimize simultaneously several biotechnological targets, i.e., the output of the computation carried out by organelles. This approach highlights the complex behavior that may arise in metabolic networks, and hints at future investigation of the concept of optimality in organelles (see online supplemental material). A multiobjective optimization is needed when the system performs multiple tasks, and a given phenotype cannot be optimal at all of them (e.g., when two tasks are in contrast with each other).

Let us assume to have r objective functions $f_1, \ldots, f_r$ to optimize. The problem of multiobjective optimization can be formalized as follows:

$$\max_{x} \; (f_1(x), f_2(x), \ldots, f_r(x))^{\top}, \tag{2}$$

where x is the variable in the search space. Without loss of generality, we have assumed that all the functions have to be maximized; indeed, minimizing a function f_i can be thought of as maximizing $-f_i$.

The solution of a multiobjective problem is a potentially infinite set of points, called Pareto optimal solutions or *Pareto front*. A point y^* in the solution space is said to be Pareto optimal if there does not exist a point y such that $f(y)$ dominates $f(y^*)$, i.e., $f_i(y) > f_i(y^*), \forall i = 1, \ldots, r$, where f is the vector of r objective functions to optimize in the objective space. We use the Pareto-front concept to find the set of designs that represent the best tradeoff between two or more requirements. The Pareto front can also show interesting tradeoff knee points, considered by Higuera et al. [29] as preferred solutions in metabolic networks. The Pareto front is the set of all the phenotypes that remain after eliminating all the feasible phenotypes dominated on all tasks [30].

5.1 Optimization of the Chloroplast Model

Once the 11 sensitive enzymes have been identified through the sensitivity analysis, we employ the multiobjective optimization algorithm on the "sensitive domain" made up of the 11 most sensitive enzymes ($x \in \mathbb{R}^{11}$). We use parallel multiobjective optimization (PMO2) [31] to find all those sensitive enzyme concentration vectors $\hat{x} = (c_1, c_2, \ldots, c_{11})$ such that when $\hat{x}$ has the other 12 enzyme values are kept at their nominal value, the resulting CO_2 uptake function is maximized and the nitrogen consumption is minimized:

$$\max_{\hat{x} \in \mathbb{R}^{11}} \; (f_1(\hat{x}), -f_2(\hat{x}))^{T}, \tag{3}$$

where f_1 represents the CO_2 uptake, while $f_2 = (\sum_{i=1}^{11} \frac{\hat{x}[i] * M_i}{K_i})$ represents the nitrogen consumption, K_i is the catalytic number (or turnover number), and M_i the molecular weight of the ith enzyme. Gaining higher CO_2 uptake rates while employing less nitrogen means absorbing more CO_2 while consuming less "leaf-fuel"; this renders the metabolism cycle more efficient. Hence, our search for $\hat{x}$ must take into account a tradeoff between maximal CO_2 uptake rate and minimal nitrogen employment. Fig. 3 shows the Pareto front made up of the solutions that simultaneously maximize f_1 and minimize f_2. In the simulation we adopt the actual CO_2 atmospheric concentration, i.e., 270 μmol mol^{-1}.

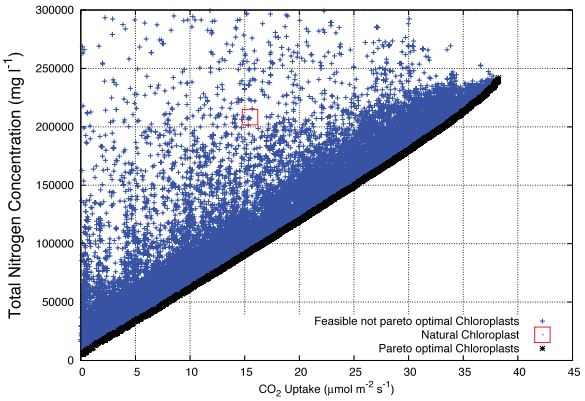


Fig. 3. Multiobjective optimization of two conflicting pressures in the chloroplast: maximization of the CO_2 uptake rate versus minimization of the protein-nitrogen consumption.

5.2 Optimization of the Mitochondrial Model

To optimize multiple objectives, we used the Nondominated Sorting Genetic Algorithm II, also known as NSGA-II [32], that exploits evolution among a population of individuals in the search space to obtain the best offspring.

Here, we optimize the concentration of Adenosine triphosphate (ATP) and Nicotinamide adenine dinucleotide (NADH) in the matrix compartment in the model of the mitochondrial bioenergetics [13]. Since NADH and ATP represent the most important energy molecules, by means of our computational framework we seek and investigate their tradeoff and the conditions (basal concentrations of metabolites in mitochondria) such that ATP and NADH are maximized.

Mitochondria convert free energy in substrates (carbohydrate and fatty acid derived) into the free energy of the ATP hydrolysis reaction. The Gibbs free energy of ATP hydrolysis reaction (ATP $\Longleftrightarrow$ ADP + Pi) is equal to $\Delta G_0 = -RTln(\frac{[ATP]}{[ADP]\times[Pi]})$, and is negative, therefore the ATP hydrolysis is an exoergonic reaction, i.e., the energy content of the products is less than that of reactants. During the reaction there is a release of energy. In this work, we investigate the ATP concentration in the matrix compartment of the mitochondria because this molecule has an high potential to transfer the phosphoric group. Therefore, optimizing ATP concentration in the matrix compartment would mean optimize the free energy transduction.

The model we adopt consists of 73 DAEs to represent the mitochondrial bioenergetics [13]. In particular, the model accounts for 35 biochemical reactions, including the oxidative phosphorylation, the electron transport system, the tricarboxylic acid cycle and related reactions, the Na^+/Ca^{2+} cycle and the K^+-cycle.

The variable space is defined as the space of the feasible initial concentrations of metabolites. We first initialize the population and compute the fitness score, based on

the output functions we aim at optimizing. The individuals of the initial population can be initialized in different ways, for example, randomly or assigning specific values to all the concentrations. In our analysis, we assign specific values—the fully oxidized state reported in the original work [13]—from which we start searching optimal points. We maintain fixed all the other variables and constants of the model (such as temperature to $25°C$ and the water volume in buffering, intermembrane and matrix space and thermodynamics constraints) varying only the initial concentration of metabolites when solving the DAEs system.

An individual is a feasible vector of concentrations of metabolites. Once a vector of concentrations of metabolites is fixed, the objective functions are calculated by solving the DAEs system. For each ATP and NADH curve, we evaluate its integral, considering a single measure. Each individual is assigned a rank, and between two solutions with different nondomination ranks we prefer the one with the lowest rank. After sorting the individuals according to the level of nondomination, the fitness score of each individual is computed by evaluating the objective functions associated with it.

Successively, we carry out three steps in a loop: 1) In a binary tournament selection process, two individuals are selected at random, their fitness is compared and the individual with the best fitness is selected as a parent for the next population; 2) the algorithm selects a number of parents (i.e., the best individuals) equal to the half of the population, and then mutates them using a mutation operator and a crossover operator; 3) a novel population of the same size of the initial population is formed selecting the best individuals from the parents and the offspring. Each individual of the final population will be a point of the Pareto front in the objective space.

Before the optimization, at the fully oxidized state we obtain $\mathrm{NADH} = 1.5987 \cdot 10^{-10}$ nmol/mg (formation) and $\mathrm{ATP} = -0.0014$ nmol/mg (consumption). Then, each metabolite involved in the metabolic network varies in concentration depending on the reactions. The metabolite undergoes several processes, i.e., formation and degradation, transport, and cellular utilization. For every metabolite, a mass balance equation depends on the fluxes of input and output in the compartment. The variation of the metabolite concentration depends on the amount of mass formed and consumed. Therefore, when we maximize the concentration of ATP in the matrix, we can obtain positive values (indicating a formation in the compartment of that molecule) or negative values (indicating that the molecule is needed from outside the system). After the optimization, we obtain the Pareto-optimal points shown in Fig. 4. The Pareto fronts and the reported values of NADH and ATP represent an asymptotic analysis. They describe how the mitochondrial metabolism evolves when it needs to optimize simultaneously the concentrations of ATP and NADH. The Pareto curve can be useful to analyze the trend of the mitochondrial processes that have to satisfy specific constraints.

Furthermore, we analyze more thoroughly two particular Pareto-optimal solutions, i.e., the point with maximum ATP synthesis (and lowest NADH formation) and

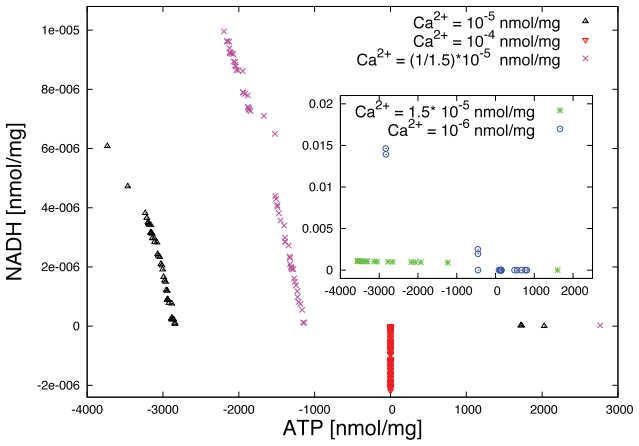


Fig. 4. Pareto fronts of the maximization of NADH versus ATP with different Ca^{2+} concentration in the mitochondrion. The inset plot reports results for $\mathrm{Ca}^{2+} = 10^{-6}$ and $\mathrm{Ca}^{2+} = 1.5 \times 10^{-5}$ nmol/mg and has a different y scale.

the point with maximum NADH formation (and lowest ATP synthesis). After setting $\mathrm{Ca}^{2+} = 10 \cdot 10^{-6}$ nmol/mg, the first solution provides $\mathrm{NADH} = 6.17 \cdot 10^{-15}$ nmol/mg (that can be considered null) and $\mathrm{ATP} = 2027.34$ nmol/mg, with overproduction of SUC_{mtx}, SCoA_{mtx}, CoASH_{mtx}, H^{+}_{mtx}, and ATP_{ims} (ims = intermembrane space, mtx = matrix) and underproduction of ISOC_{mtx}, aKG_{mtx}, MAL_{mtx}, CIT_{ims}, ISOC_{ims}, aKG_{ims}, SUC_{ims}, MAL_{ims}, and GLU_{cyt}, ASP_{cyt} (cyt = cytosolic space).

The second solution provides $\mathrm{NADH} = 6.07 \cdot 10^{-6}$ nmol/mg and $\mathrm{ATP} = -3734.6$ nmol/mg (consumption), overproducing the following metabolites: H^{+}_{mtx}, ISOC_{mtx}, SUC_{mtx}, and ATP_{ims}, whereas CIT_{mtx}, MAL_{ims} and AMP_{ims}, PYR_{ims}, $\mathrm{GLU}_{ims,cyt}$ and aKG_{ims} are totally consumed. These results indicate that the mitochondrion is requesting an amount of ATP that is not available in the matrix. With respect to the state where no ATP is available in the matrix, a negative value of concentration in the model can be thought of as the change of ATP concentration needed to produce the corresponding NADH. Both the extreme solutions are shown in Fig. 5(top).

Increasing the matrix calcium content from 10^{-5} to 10^{-4} nmol/mg causes ATP synthesis and NADH formation to stop, and both molecules are consumed by the metabolism (see Fig. 4, red signs). As in the previous case, we analyze the two extreme Pareto-optimal points in Fig. 5(bottom). This achievement can demonstrate that a perturbation in mitochondrial Ca^{2+} homeostasis has major implications for cell function at the level of ATP synthesis and NADH generation.

If Ca^{2+} increases by a little step, i.e., it is fixed at 1.5×10^{-5} nmol/mg, we obtain an increase in NADH formation, while ATP remains constant (see Fig. 4, green signs). If Ca^{2+} drastically decreases to 10^{-6} nmol/mg, there is a lower ATP synthesis (see Fig. 4, blue circles). Conversely, with $\mathrm{Ca}^{2+} = \frac{1}{1.5} \times 10^{-5}$ nmol/mg, both objectives are maximized.

6 Robustness Analysis

The concept of robustness is frequent in nature, and it seems to be one of the driving forces of evolution [14]. The

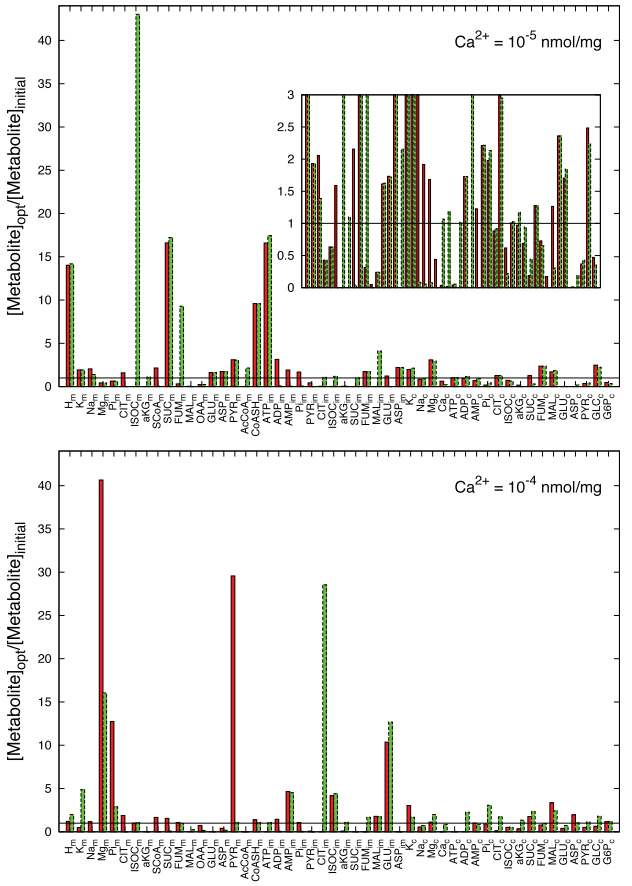


Fig. 5. Ratio of metabolite concentrations optimized by the multiobjective algorithm compared to the initial concentrations for maximum ATP synthesis, in red, and maximum NADH formation, in green. The calcium content is $Ca^{2+} = 10^{-5}$ nmol/mg (top) and $Ca^{2+} = 10^{-4}$ nmol/mg (bottom), respectively. In the inset, we focus on the $[0, 3]$ y-axis range.

ability of a system to preserve its behavior despite internal or external perturbations is a crucial design principle for any biological and synthetic system [17], [18]. The basic principle of the robustness analysis consists of defining the perturbation as a function $\gamma(\Psi, \sigma)$, where γ applies a stochastic noise σ to the system Ψ and generates a trial sample τ. We assume that the noise is defined by a random distribution and we generate a set T of trial samples $\tau = \gamma(\Psi, \sigma)$. Let us define:

$$\rho(\Psi, \tau, \phi, \epsilon) = \begin{cases} 1 & \text{if } |\phi(\Psi) - \phi(\tau)| \leq \epsilon \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

where Ψ is the *reference system*, ϕ is a *metric* (or property), τ is a trial sample of the set T, and ϵ is a *robustness threshold*. The definition of this condition makes no assumptions about the function ϕ, which is not necessarily related to properties or characteristics of the system; however, it is implicitly assumed that it is quantifiable. Each element $\tau \in T$ is said to be robust to the perturbation, due to stochastic noise σ, for a given property (or metric) ϕ, if the condition $\rho(\Psi, \tau, \phi, \epsilon) = 1$ holds.

The robustness of a system Ψ is defined as the number of robust trials $\tau \in T$, with respect to the property ϕ, over the total number of trials ($|T|$). Formally, the robustness of a system is a dimensionless quantity defined as

$$\Gamma(\Psi, T, \phi, \epsilon) = \frac{\sum_{\tau \in T} \rho(\Psi, \tau, \phi, \epsilon)}{|T|}. \quad (5)$$

We note that Γ is a function of ϵ, so the choice of this parameter is crucial and a reasonable value is often chosen after conducting several computational experiment.

6.1 Robustness in the Chloroplast Model

During the "in vitro" implementation, our chloroplast must be robust with respect to possible experimental errors or unpredicted unknowns. Although it is possible to find maximal values for the CO_2 uptake, efficiency of transcription promoters cannot be easily foreseen. Moreover, transcription, translation, and enzyme efficiency can vary depending on many factors, even of environmental origin. In other words, the enzyme concentration is subjected to noise, making it difficult to ensure a certain concentration of each enzyme of the C3 cycle at a given time. For this reason, it is fundamental to know the extent to which the achieved CO_2 uptake will be preserved in case of enzyme level perturbation. Applying the concept of robustness to the C3 cycle allows to calculate the limits of enzyme perturbations at which the system property of interest (a given level of CO_2 uptake) is maintained. Among read-out-equivalent designs, one should prefer the one with the most CO_2 uptake read-out preserved when the enzyme input noise increases.

Let $\bar{x} \in \mathbb{R}^{23}$ be an enzyme partitioning and $f : \mathbb{R}^{23} \to \mathbb{R}$ a function computing the expected CO_2 uptake rate value of $\bar{x}$. Given an enzyme partition $\bar{x}_*$ obtained by perturbing $\bar{x}$, the condition ρ for the robustness of enzymes partitions is defined by adapting the general definition (4) as

$$\rho(\bar{x}, \bar{x}_*, f, \epsilon) = \begin{cases} 1 & \text{if } \mid f(\bar{x}) - f(\bar{x}_*) \mid \leq \epsilon \\ 0 & \text{otherwise,} \end{cases} \quad (6)$$

where the robustness threshold ϵ denotes the maximum percentage of variation allowed from the nominal CO_2 uptake value.

The ensemble T has been generated using a Monte-Carlo algorithm. We consider both mutations occurring on all the enzymes (global robustness analysis) and mutations on one enzyme at time (local robustness analysis) [14]. We adopt a maximum perturbation of 10 percent for each enzyme concentration. We generate an ensemble of $5 \cdot 10^3$ trials for the global robustness analysis, and 200 trials, for each enzyme, for the local robustness analysis. In all the experiments, we fix $\epsilon = 5\%$ of the nominal uptake rate value. In Fig. 6, we show the robustness landscape of each enzyme. Remarkably, most enzymes do not change significantly the uptake rate value, despite the presence of noise. This analysis seems to confirm that the C3 cycle lies in a robust configuration.

The results of our methodology, in terms of new enzyme concentrations for maximal CO_2 uptake, are presented in Table SI1, available online. Our optimization results take into account enzyme kinetic parameters, largely obtained as an average of different plant values, as done in [12]. Such results consider, hence, an ideal plant. In a future research effort, we plan to consider the preparation of specific optimization models for a given C3 crop, as rice, soybean, wheat or barley by taking into account the specific enzyme

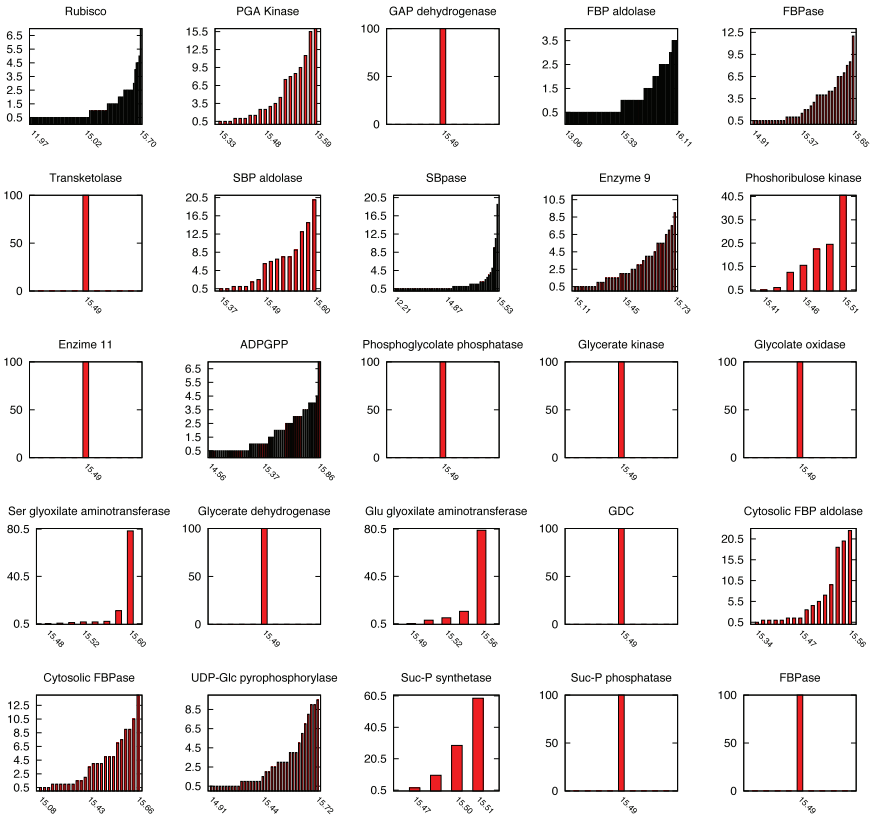


Fig. 6. Robustness analysis. Robustness of the individual enzymes for the natural chloroplast. The enzymes are shown in the same order as in Table SI1, available online. For each enzyme, as mentioned in Section 6.1, we create 200 trials by perturbing the corresponding enzyme concentration, and we show the distribution of the CO_2 uptake rate values obtained. The y values are in percentage (percent), that is, the percentage (out of 200 trials) that obtains the corresponding value (± 0.1) on the x-axis. The chloroplast is robust to perturbations of the enzyme concentration if the values of CO_2 uptake rate are close to the nominal value ($15.48 \frac{\mu mol}{m^2 s}$) in the majority of the perturbation trials. For example, it is more robust with respect to perturbation of GAP dehydrogenase concentration and it is less robust with respect to perturbation of RuBisCO concentration.

kinetic parameters in these plants, and the relative condition of temperature and water availability typical of their areal distribution. Moreover, the study and the use in optimization models of enzymes kinetic parameters of plants adapted to specific climatic conditions would be a starting point for further biotechnological targets, for example, to improve photosynthetic efficiency of crops in specific environmental conditions. As a matter of fact, the C4 photosynthesis type represents itself an adaptation to lower CO_2 atmospheric concentration. The possibility of redesigning the energy circuits of the cells, such as those in the chloroplast and in the mitochondrion, will make it possible to obtain biotechnological solutions for energy demand and organelle-related diseases.

6.2 Robustness in the Mitochondrial Model

Also in the case of "in vitro" implementation of mitochondrial models, the parameters must be robust with respect to possible experimental errors or unpredicted unknowns. Consequently, we apply the concept of robustness also to the optimized mitochondria. In this case, among read-out-equivalent designs, one should prefer the one that ensures the largest NADH formation and ATP consumption read-out preserved when the noise in the input concentrations of metabolites increases.

As in the previous section, the problem can be formalized as follows: Let $\bar{x} \in \mathbb{R}^{50}$ be the array of the concentrations of metabolites and $f : \mathbb{R}^{50} \rightarrow \mathbb{R}^2$ a function computing the expected NADH formation and ATP consumption values corresponding to $\bar{x}$. Given a metabolite partition $\bar{x}_*$ obtained by perturbing $\bar{x}$, the condition ρ for the robustness of metabolite partitions is defined by adapting the general definition (4) as made in the (6), where the robustness threshold ϵ denotes the maximum percentage of variation allowed from the nominal NADH formation and ATP consumption values. The remarks on the ensemble T, the mutations, the perturbation, and the threshold in chloroplasts apply also to mitochondria. The results of our methodology, in terms of new metabolite concentrations for maximal NADH formation and ATP consumption, are presented in Table SI2, available online. Our results highlight that the natural mitochondrion (C1) is the most robust, since it features a global value of 26.94 percent and a local value of 9.00 percent (referred to the concentration of free potassium). The C2 and the C3 mitochondria are less robust than the natural mitochondrion, both globally and locally.

Remarkably, we need to find a tradeoff between the performance of the optimization and the robustness of the solution. Since the multiobjective optimization returned thousands of solutions (the number can decrease if the number of objectives increases) and the robustness analysis is computationally expensive, the choice of the most robust and optimal solution can require a high computational time. A good tradeoff workaround could be of thinking the robustness index as another objective of the optimization.

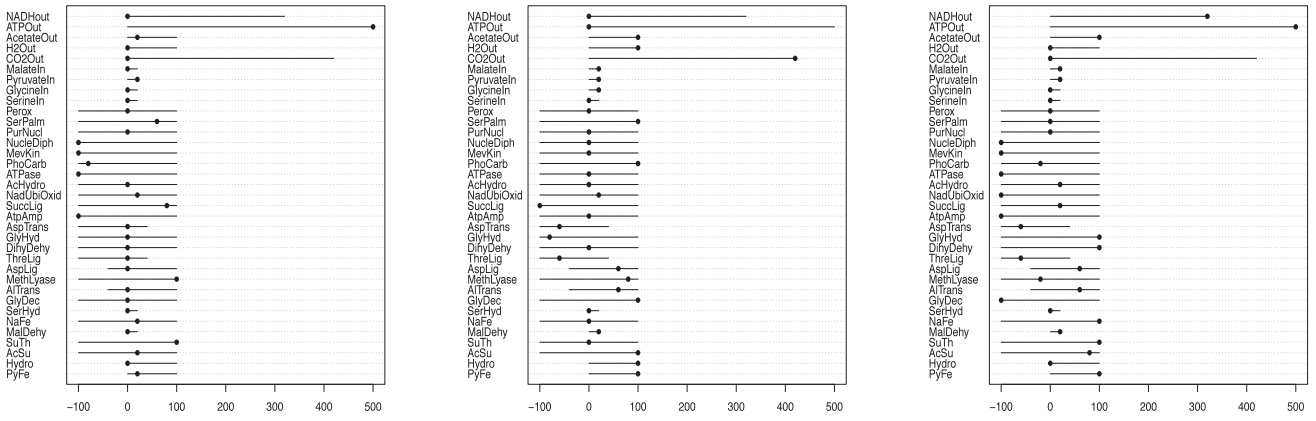


Fig. 7. Hydrogenosome. We report the reactions of its metabolism and their fluxes ($\mathrm{mmol\,h^{-1}\ gdW^{-1}}$) on the x-axis. A line next to a reaction represents all the feasible fluxes for that reaction, while a point represents the value of the specific solution proposed in this configuration of the hydrogenosome. The maximization of the ATP yield (a) implies the consumption of all the H_2, CO_2, and NADH in the organelle. The maximization of CO_2 (b) implies also the maximization of H_2, but requires the maximum import of malate, pyruvate, and glycine. Moreover, it impairs the NaFe reaction, thus keeping the NADH at the initial value. The two-objective optimization of ATP and NADH (c) causes the impairment of the Hydro reaction, thus the hydrogenosome cannot produce H_2. It is, thus, evident the need for a Pareto-optimal tradeoff between energy and H_2 production.

In addition to this, we conducted an additional robustness analysis on the natural mitochondrion perturbing the 42 parameters listed in Table SI3, available online. The formulation of the problem is identical, where the vector $\bar{x}$ in (6) represents the parameters. The results are presented in the last column of Table SI3, available online.

7 A Model of the Hydrogenosome

In several unicellular eukaryotes, including ciliates, fungi, and trichomonads, instead of the traditionally studied mitochondria, there is an alternate organelle called *hydrogenosome*. Hydrogenosomes are anaerobically functioning ATP-producing organelles of mitochondrial origin that represent a particular adaptation of mitochondrial metabolism, and possess the ability of producing molecular hydrogen by using protons as electron acceptors [7]. One of the best studied hydrogenosomes, and thus the one that we consider in our analysis, is that of the sexually transmitted human parasite *Trichomonas vaginalis*.

To investigate the hydrogenosome with the same techniques introduced before, we propose a model of the hydrogenosome metabolism in *T. vaginalis* and we evaluate the ATP flux, comparing it with that of mitochondria. The model we propose is a FBA model [16] that contains all the main reactions occurring in the organelle [33], [34] (see Table SI4, available online), as well as reactions dealing with the import of serine, glycine, pyruvate, and malate into the hydrogenosome. Although our model already shows interesting behaviors, we plan to extend it by adding new reactions. To our knowledge, no hydrogenosome models are present in literature; thus, we believe our model will have great value as the best description that we have to date of the hydrogenosome, and will provide a foundation upon which more accurate models can be arrived at. A complete model of the hydrogenosomal metabolism would be of great biological relevance in the design of new antiparasitic drugs [35].

We specify the reaction network of the hydrogenosome through the LIM package of R, which allows to generate the mass balance for each component. Moreover, we estimate the optimal reaction rates in the flux balance analysis approach [16]. (See the online supplemental material for details on the modeling technique for the hydrogenosome. The source code of the model and all the required libraries are available from the authors at request.) In Fig. 7, we show the multiobjective optimization carried out in the flux balance analysis framework.

To obtain the optimal solution in all the feasible two-objective optimizations, we randomly sample the solution space. In Fig. 8, we display the pairs plot and mark the optimal solution with a red dot. Most reactions are coupled together; it happens frequently that a reaction can occur

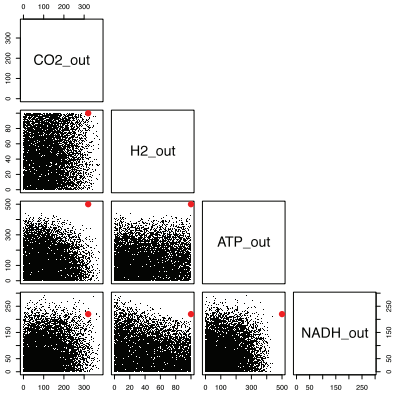


Fig. 8. Hydrogenosome. Two-objective optimization carried out on metabolites and reaction fluxes. The points represent a MCMC sampling of the reaction network carried out with 20,000 iterations, and allow to find tradeoffs among the maximizations carried out in Fig. 7. When two reactions are mostly uncorrelated with one another, all the solutions are feasible, thus the sampling tends to a filled square (e.g., the output of H_2 and CO_2). Conversely, when two reactions show correlation, the sample of the square reveals their relationship. For instance, the NADH output and the H_2 output are in contrast with each other. The red dot is the optimal point.

only when the substrates, which are products of another reaction, are available. For instance, the ATP production depends on the Acetate production, since ATP requires SuccinylCoA, which is a product of the same reaction that produces Acetate. It is noteworthy that a few reactions are in conflict with all the other reactions. For instance, the NaFe reaction, which produces NADH, is in conflict with Hydro, which produces H_2. Indeed, the last row of the pairs plot shows that the NADH production is in conflict with the H_2 production. The presence of a feasible point in this plot does not imply that the hydrogenosome metabolism, even if optimized, is able to reach that point. Notably, the CO_2-NADH plot shows that the hydrogenosome is versatile and can produce both, but it cannot specialize in producing only one metabolite, since there are few points near the axes and the Pareto front exhibits a higher curvature than the other fronts.

8 Conclusions

Both mitochondria and chloroplasts are energy-converting organelles in the cytoplasm of eukaryotic cells; while chloroplasts capture and convert energy of sunlight in plants, mitochondria synthesize energy (ATP). The genetic and the energy-converting networks of mitochondria and chloroplasts are descended, with little modification, from those of their ancestor bacteria. In this regard, we can explore how the Pareto-front analysis related to the energy metabolism can provide interesting insights into the evolutionary dynamics leading to the formation of organelle structures in the single- and multicelled life. Pareto fronts combined with sensitivity and robustness are useful tools to understand the steps of the cellular evolution and the engulfments and specialization of organelles. In Fig. 9, we show the putative evolution through the Pareto-front analysis. The whole figure should be seen in an evolutionary perspective, where the passage from the engulfment to the full functionality of the organelle has required a number of adaptive evolutionary steps and a similar less optimized system working outside the organelle. During the optimization, i.e., the maximization of a metabolite, the Pareto front of an organism undergoes both expansion and contraction phases. These steps can be numerically assessed using the hypervolume [36], to quantify the evolution of the Pareto front. We intend to make use of machine learning algorithms to predict whether a protein enters a membrane according to the transient peptide, and measures on Pareto fronts (e.g., hypervolume) can avail or disprove these predictions.

We also discussed the concept of robustness and parameter sensitivity for the models of organelles, and in general for compartmentalized structures in the cell. While cross comparing the biological systems, we showed also a methodological variety to analyze them. We are delighted to report that the modeling of metabolic processes, which is a thriving field of research, has two immediate and important benefits: 1) a comprehensive insight into the energy balance in the cell, and 2) the improved understanding of the processes that shape health and diseases, thus providing the possibility to design new drugs.

Following our framework and extending it, in the near future we plan to design, analyze, and optimize the metabolism of systems composed of different species living and interacting in the same organism, as introduced by Cottret et al. [37] in the case of two endosymbiotic bacteria. In this paper, we also analyzed the metabolite production of organelles and identified interventions needed to overproduce metabolites of interest. We performed an in silico design that can also explore the reaction network and seek

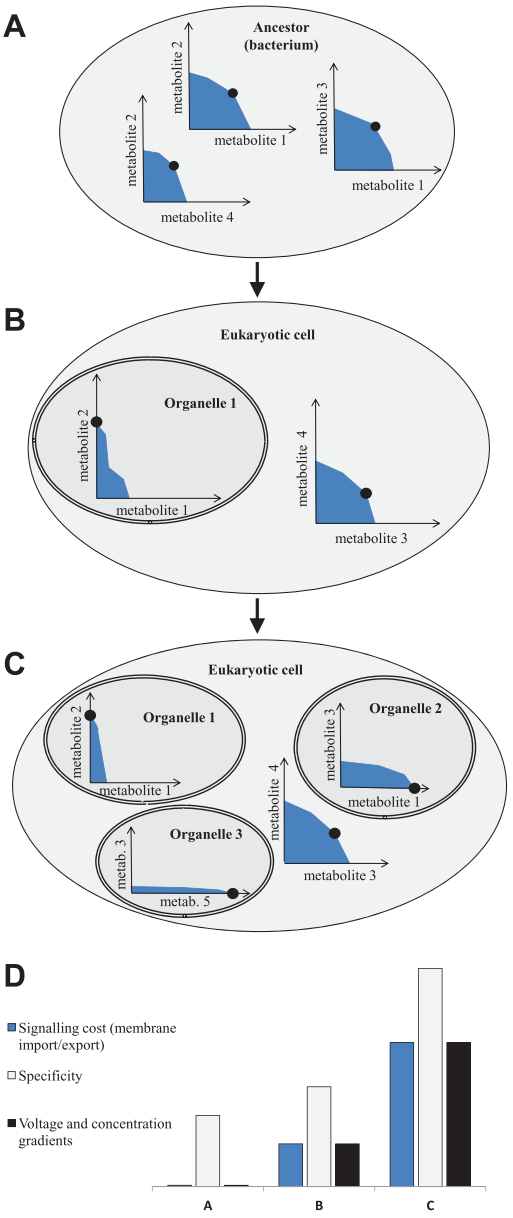


Fig. 9. Putative Pareto-front evolution associated with the engulfments in cell evolution. The double-membrane organelle is a bacterium that has been engulfed by an single-membrane eukaryote (A). After an engulfment (B), the guest organelle shows a concave Pareto front that expands along one axis, since it specializes in the production of one metabolite, thus losing versatility and decreasing the area under the front. Conversely, the hosting cytoplasm specializes in the production of all the other metabolites, thus it shows a convex Pareto front. The process is repeated for the second and third engulfment (C), after which the organelles are specialized in the production of a metabolite each. Since a slight reduction of metabolite 1 dramatically increases the production of metabolite 2, the organelle 1 can further specialize (see C) in the next evolutionary step. Organelle 3 is the most specialized, and therefore it could be thought of as an apicoplast, with Metabolite 5 representing the IPP. The histogram (D) highlights the increase of specificity, gradient, and signaling cost in the evolutionary steps.

in the search space the solutions that optimize two or more objectives simultaneously. Our technique can be extended using global and local sensitivity measures based on partial derivatives [38], which allow to obtain rankings of parameters according to their influence. This kind of analysis could easily highlight the complementarity of different metabolic networks. For instance, mitochondria and chloroplasts are (usually) both found in plants, and are part of the same functional pipeline: starting from CO_2, the photosynthesis in the chloroplast creates glucose that enters the mitochondrion to create ATP.

Although in this paper we have considered the single organelle, in cells there are usually many organelles that could differ for activity depending on their location. Therefore, we expect that applications of compartmentalization in synthetic biology (e.g., the use of organelles to optimize a cell) will increase in the near future.

Acknowledgments

Pietro Lió received funding from the EU FP7-Health-F5-2012 under grant no. 305280 (MIMOmics).

References

[1] S. Raha et al., "Mitochondria, Oxygen Free Radicals, Disease and Ageing," *Trends in Biochemical Sciences,* vol. 25, no. 10, pp. 502-508, 2000.

[2] B. Alberts et al., "Energy Conversion: Mitochondria and Chloroplasts," *Molecular Biology of the Cell,* Garland Science, 2002.

[3] P. Lió and N. Goldman, "Modeling Mitochondrial Protein Evolution Using Structural Information," *J. Molecular Evolution,* vol. 54, no. 4, pp. 519-529, 2002.

[4] T. Kuroiwa, H. Kuroiwa, A. Sakai, H. Takahashi, K. Toda, and R. Itoh, "The Division Apparatus of Plastids and Mitochondria," *Int'l Rev. of Cytology,* vol. 181, pp. 1-41, 1998.

[5] A. Shiflett and P.J. Johnson, "Mitochondrion-Related Organelles in Parasitic Eukaryotes," *Ann. Rev. of Microbiology,* vol. 64, pp. 409-429, 2010.

[6] J.I. Macrae, E. Maréchal, C. Biot, and C.Y. Botté, "The Apicoplast: A Key Target to Cure Malaria," *Current Pharmaceutical Design,* vol. 18, pp. 3490-3504, 2012.

[7] M.V.D. Giezen, "Hydrogenosomes and Mitosomes: Conservation and Evolution of Functions," *J. Eukaryotic Microbiology,* vol. 56, no. 3, pp. 221-231, 2009.

[8] W. de Souza, M. Attias, and J.C.F Rodrigues, "Particularities of Mitochondrial Structure in Parasitic Protists (Apicomplexa and Kinetoplastida)," *Int'l J. Biochemistry & Cell Biology,* vol. 41, no. 10, pp. 2069-2080, 2009.

[9] W.B.M. Paula, J.F. Allen, and M. Giezen, "Mitochondria, Hydrogenosomes and Mitosomes in Relation to the CoRR Hypothesis for Genome Function and Evolution," *Organelle Genetics,* pp. 105-119, Springer, 2012.

[10] J. Handl, D.B. Kell, and J. Knowles, "Multiobjective Optimization in Bioinformatics and Computational Biology," *IEEE/ACM Trans. Computational Biology and Bioinformatics,* vol. 4, no. 2, pp. 279-292, Apr.-June 2007.

[11] J.O.H Sendin, O. Exler, and J.R. Banga, "Multi-Objective Mixed Integer Strategy for the Optimisation of Biological Networks," *IET Systems Biology,* vol. 4, no. 3, pp. 236-248, May 2010.

[12] 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,* vol. 145, pp. 513-526, 2007.

[13] J.N. Bazil, G.T. Buzzard, and A.E. Rundell, "Modeling Mitochondrial Bioenergetics with Integrated Volume Dynamics," *PLoS Computational Biology,* vol. 6, no. 1, article e1000632, 2010.

[14] G. Stracquadanio and G. Nicosia, "Computational Energy-Based Redesign of Robust Proteins," *Computers & Chemical Eng.,* vol. 35, no. 3, pp. 464-473, 2011.

[15] C.M. Agapakis, H. Niederholtmeyer, R.R. Noche, T.D. Lieberman, S.G. Megason, J.C. Way, and P.A. Silver, "Towards a Synthetic Chloroplast," *PLoS One,* vol. 6, no. 4, article e18877, 2011.

[16] J.D. Orth, I. Thiele, and B.Ø. Palsson, "What Is Flux Balance Analysis?" *Nature Biotechnology,* vol. 28, no. 3, pp. 245-248, 2010.

[17] H. Kitano, "Towards a Theory of Biological Robustness," *Molecular Systems Biology,* vol. 3, no. 1, 2007.

[18] J. Gunawardena, "Models in Systems Biology: The Parameter Problem and the Meanings of Robustness," *Elements of Computational Systems Biology,* pp. 19-47, John Wiley & Sons, 2009.

[19] D. Gorissen, I. Couckuyt, P. Demeester, T. Dhaene, and K. Crombecq, "A Surrogate Modeling and Adaptive Sampling Toolbox for Computer Based Design," *J. Machine Learning Research,* vol. 11, pp. 2051-2055, 2010.

[20] P. Lió., "Phylogenetic and Structural Analysis of Mitochondrial Complex i Proteins," *Gene,* vol. 345, no. 1, pp. 55-64, 2005.

[21] M.D. Morris, "Factorial Sampling Plans for Preliminary Computational Experiments," *Technometrics,* vol. 33, no. 2, pp. 161-174, 1991.

[22] R. Umeton, G. Stracquadanio, A. Papini, J. Costanza, P. Liò, and G. Nicosia, "Identification of Sensitive Enzymes in the Photosynthetic Carbon Metabolism," *Advances in Systems Biology,* vol. 736, pp. 441-459, 2012.

[23] G. Stracquadanio, R. Umeton, A. Papini, P. Liò, and G. Nicosia, "Analysis and Optimization of C3 Photosynthetic Carbon Metabolism," *Proc. IEEE 10th Int'l Conf. Bioinformatics and Bioeng. (BIBE '10),* pp. 44-51, 2010.

[24] M. Rosvall and C.T. Bergstrom, "An Information-Theoretic Framework for Resolving Community Structure in Complex Networks," *Proc. Nat'l Academy of Sciences USA,* vol. 104, no. 18, pp. 7327-7331, 2007.

[25] F. Wu, F. Yang, K.C. Vinnakota, and D.A. Beard, "Computer Modeling of Mitochondrial Tricarboxylic Acid Cycle, Oxidative Phosphorylation, Metabolite Transport, and Electrophysiology," *J. Biological Chemistry,* vol. 282, no. 34, pp. 24525-24537, 2007.

[26] J.O. Ramsay, G. Hooker, D. Campbell, and J. Cao, "Parameter Estimation for Differential Equations: A Generalized Smoothing Approach," *J. the Royal Statistical Soc.: Series B (Statistical Methodology),* vol. 69, no. 5, pp. 741-796, 2007.

[27] J.A.K. Suykens and J. Vandewalle, "Least Squares Support Vector Machine Classifiers," *Neural Processing Letters,* vol. 9, no. 3, pp. 293-300, 1999.

[28] T.V. Gestel, J.A.K. Suykens, B.D. Moor, and J. Vandewalle, "Automatic Relevance Determination for Least Squares Support Vector Machine Regression," *Proc. Int'l Joint Conf. Neural Networks (IJCNN '01),* vol. 4, pp. 2416-2421, 2001.

[29] C. Higuera, A.F. Villaverde, J.R. Banga, J. Ross, and F. Morán, "Multi-Criteria Optimization of Regulation in Metabolic Networks," *PLoS one,* vol. 7, no. 7, article e41122, 2012.

[30] O. Shoval, H. Sheftel, G. Shinar, Y. Hart, O. Ramote, A. Mayo, E. Dekel, K. Kavanagh, and U. Alon, "Evolutionary Trade-Offs, Pareto Optimality, and the Geometry of Phenotype Space," *Science,* vol. 336, pp. 1157-1160, 2012.

[31] R. Umeton, G. Stracquadanio, A. Sorathiya, A. Papini, P. Liò, and G. Nicosia, "Design of Robust Metabolic Pathways," *Proc. 48th Design Automation Conf. (DAC '11),* pp. 747-752, June 2011.

[32] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II," *IEEE Trans. Evolutionary Computation,* vol. 6, no. 2, pp. 182-197, Apr. 2002.

[33] R.E. Schneider, M.T. Brown, A.M. Shiflett, S.D. Dyall, R.D. Hayes, Y. Xie, J.A. Loo, and P.J. Johnson, "The *Trichomonas vaginalis* Hydrogenosome Proteome is Highly Reduced Relative to Mitochondria, Yet Complex Compared with Mitosomes," *Int'l J. Parasitology,* vol. 41, pp. 1421-1434, 2011.

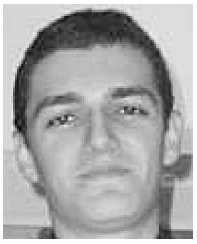
[34] M. Müller, M. Mentel, J.J. van Hellemond, K. Henze, C. Woehle, S.B. Gould, R.Y. Yu, M. van der Giezen, A.G.M. Tielens, and W.F. Martin, "Biochemistry and Evolution of Anaerobic Energy Metabolism in Eukaryotes," *Microbiology and Molecular Biology Rev.,* vol. 76, no. 2, pp. 444-495, 2012.

[35] N. Mallo, J. Lamas, and J.M. Leiro, "Hydrogenosome Metabolism Is the Key Target for Antiparasitic Activity of Resveratrol against *Trichomonas vaginalis*," *Antimicrobial Agents and Chemotherapy,* vol. 57, pp. 2476-2484, 2013.

[36] E. Zitzler, D. Brockhoff, and L. Thiele, "The Hypervolume Indicator Revisited: On the Design of Pareto-Compliant Indicators via Weighted Integration," *Proc. Fourth Int'l Conf. Evolutionary Multi-Criterion Optimization,* pp. 862-876, 2007.

[37] L. Cottret, P.V. Milreu, V. Acuña, A. Marchetti-Spaccamela, L. Stougie, H. Charles, and M.F. Sagot, "Graph-Based Analysis of the Metabolic Exchanges between Two Co-Resident Intracellular Symbionts, *baumannia cicadellinicola* and *Sulcia muelleri*, with Their Insect Host, *Homalodisca coagulata*," *PLoS Computational Biology*, vol. 6, no. 9, article e1000904, 2010.

[38] E. Balsa-Canto and J.R. Banga, "Amigo, A Toolbox for Advanced Model Identification in Systems Biology Using Global Optimization," *Bioinformatics*, vol. 27, no. 16, pp. 2311-2313, 2011.



Claudio Angione received the BSc degree in 2008 and the MSc degree in 2011 in applied mathematics from the University of Catania. He is currently working toward the PhD degree in computer science at the University of Cambridge. He received the "Anile Prize" for the best thesis from the Department of Mathematics and Computer Science of the University of Catania. He has been recently selected for the Heidelberg Laureate Forum. His current research interests include synthetic biology, computational immunology, multiobjective optimization in bacteria, organelle models, and computation with molecular machines.

Giovanni Carapezza received the graduate degree in telecommunications engineering from the University of Catania in July 2011 with a thesis entitled "Services and potentialities of Generalised Multi Protocol Label Switch." He has been a research student on electronic design automation/computer aided design and computational biology at the University of Catania since July 2011. He has taken part in the European research project that aims to increase solar power use and to deliver "Energy for a Green Society." He has coauthored four refereed conference papers with proceedings.

Jole Costanza received the BSc degree in biomedical engineering in 2007 and the MSc degree in bioengineering in 2010 from the University of Padua. She is currently working toward the PhD degree in computer science at the University of Catania, supervised by Prof. Giuseppe Nicosia. She is currently working on computational methods for systems and synthetic biology (sensitivity analysis, robustness analysis, combinatorial multiobjective optimization), reverse engineering for gene regulatory networks and artificial immune systems.

Pietro Lió has an interdisciplinary approach to research and teaching due to the fact that he received the PhD degree in complex systems and nonlinear dynamics from the School of Informatics, Department of Engineering of the University of Firenze, Italy, and the PhD degree in (theoretical) genetics from the University of Pavia, Italy. He is a senior lecturer at the Computer Laboratory, Department of Computer Science, University of Cambridge. He is a member of the Artificial Intelligence Group of the Computer Laboratory.

Giuseppe Nicosia received the PhD degree in computer science. He is an associate professor of computer science at the University of Catania, Italy. He is currently involved in the design and development of optimization algorithms for circuit design and solar cells, a joint research project supported by the Sharp and STMicroelectronics. He is a coauthor of more than 70 papers in international journals and conference proceedings, and eight coedited books. He has chaired several international conferences, summer schools, and workshops in the research areas of natural computing, optimization and electronic design automation (EDA). His primary research interests include design and analysis of optimization algorithms, sensitivity analysis, robustness, computational systems biology, and EDA.