PAPER

A design automation framework for computational bioenergetics in biological networks†

Claudio Angione,*[a] Jole Costanza,[b] Giovanni Carapezza,[a] Pietro Lió[a] and Giuseppe Nicosia[b]

The bioenergetic activity of mitochondria can be thoroughly investigated by using computational methods. In particular, in our work we focus on ATP and NADH, namely the metabolites representing the production of energy in the cell. We develop a computational framework to perform an exhaustive investigation at the level of species, reactions, genes and metabolic pathways. The framework integrates several methods implementing the state-of-the-art algorithms for many-objective optimization, sensitivity, and identifiability analysis applied to biological systems. We use this computational framework to analyze three case studies related to the human mitochondria and the algal metabolism of *Chlamydomonas reinhardtii*, formally described with algebraic differential equations or flux balance analysis. Integrating the results of our framework applied to interacting organelles would provide a general-purpose method for assessing the production of energy in a biological network.

1 Introduction

Mitochondria are the organelles of the eukaryotic cells that play a pivotal role in the bioenergetics and regulation of many signaling pathways. They are also fundamental for the evolution of complex organisms. Specifically, mitochondria are optimized symbiotic cells useful to produce energy while simultaneously being energy-saving organelles. As a result, eukaryotic cells are able to synthesize more proteins than prokaryotic cells (such as bacteria). Recent studies confirmed that mitochondria descend from bacteria, and indeed they lived outside the cell.[1] During the evolution, mitochondria entered the animal and plant cells.[2]

Mitochondria are important firstly for their energy productivity: they are the energy source of the cell, since they synthesize adenosine triphosphate (ATP), the chemical energy in the cell. Moreover, the mitochondrion is the site of carbohydrate metabolism, fatty acid oxidation and the urea cycle.

The expansion of the fields of mitochondria and other mitochondrion-like organelles is mainly due to the identification of the pivotal role that mitochondria play in human disease and ageing,[3] to the synergy showed by chloroplasts and mitochondria in energy output,[4] and to the discovery of novel factors involved in organelle division, movement, signaling and adaptation to varying environmental conditions.[5] Biomarkers for these events have been recently identified through integrative network biology.[6] In a network of organelles,[7] the study of the effects that all these conditions may have on each organelle (thought of as a submodule of the network[8]), and then on the whole network, lies in the field of differential network biology.[9] Interestingly, there are no known examples of eukaryotes lacking a mitochondrion-related organelle.[10]

In the carbohydrate metabolism, the pyruvate produced from glycolysis undergoes oxidative decarboxylation to give acetyl CoA, which is then oxidized in an eight-step process known as the tricarboxylic acid (TCA) cycle. The respiratory substrates NADH and $FADH_2$ generated through the TCA cycle are then oxidized in a process coupled with ATP synthesis. Electrons are transferred from NADH and $FADH_2$ to oxygen *via* enzyme complexes located on the inner mitochondrial membrane. Three of the electron carriers (complexes I, III and IV) are proton pumps, and couple the energy released by electron transfer with the translocation of protons from the matrix side to the external side of the inner mitochondrial membrane. The energy stored in the resulting proton gradient (*i.e.*, the proton-motive force) is used to drive the synthesis of ATP *via* the mitochondrial enzyme ATP synthetase (complex V). Under certain conditions (*e.g.* fasting), acetyl CoA molecules are converted into ketones for use as an alternative energy source (fatty acid oxidation). In the urea cycle, amino acid degradation resulting in excretion of nitrogen as urea occurs partly in the mitochondrion.

Additionally, the mitochondrion is also essential for several other processes, including the regulation of calcium homeostasis

[a] *Computer Laboratory, University of Cambridge, William Gates Building, 15 JJ Thomson Avenue, Cambridge, UK. E-mail: claudio.angione@cl.cam.ac.uk*

[b] *Department of Mathematics & Computer Science, University of Catania, Viale Andrea Doria 6, 95125, Catania, Italy*

and other inorganic ions, cellular differentiation, cell death (apoptosis), as well as for the control of the cell cycle and cell growth.[11] The number, the size and the shape of mitochondrion in a cell change depending on the tissue. For example, in the cardiac muscle, when the primary role of the heart is to pump blood, the cells have a large number of mitochondria and a large size. Mitochondria have also been detected as responsible for several human diseases, including mitochondrial disorders, cardiac dysfunction, and type 2 diabetes.[12] The mitochondrion plays a crucial role also in cancer and in neurodegenerative disorders (such as Parkinson's, Alzheimer's or ALS).

For these reasons, many researchers focused their attention on mitochondrial analysis, developing mathematical models that can simulate its activity, and in particular the oxidative phosphorylation. In the recent work by Bazil *et al.*,[13] 73 algebraic differential equations are implemented to model the mitochondrial bioenergetics, including 34 biochemical reactions. We perform an *in silico* analysis on this model in order to find the metabolites that are the most important for optimizing the energy productivity, *i.e.*, for maximizing ATP and NADH production in matrix space. We conduct five different studies with five different matrix calcium concentrations. As introduced above, mitochondria regulate calcium homeostasis, which is strictly linked to ATP and NADH production.[14]

The mitochondrial model was also set to simulate the cancer state. Specifically, we modified three features that have been found to vary between healthy and cancer conditions: (i) hexokinase activity, (ii) membrane potential differential, and (iii) concentration of hydrogen ions. The model includes kinetic parameters useful to mimic regulatory effects such as activation of enzymes by protein kinases. Therefore, it describes in a detailed way many features of the biochemical reaction and enzymatic action. On the other hand, this complex mathematical description introduces some limits: (i) the number of reactions is small, therefore the complexity of the complete network is not captured; (ii) solving the set of DAEs requires more computational effort, and solvers can only compute approximations that may not fully agree with the real behavior of the system.

For these reasons, in our work we take into account also a mitochondrial model realized with flux balance analysis (FBA), where the system is described considering a steady state for all the metabolites involved in the network.[15] The model is composed of a set of algebraic equations and does not contain kinetic parameters. This approach permits us to handle large metabolic networks (more than 2500 reactions, 2000 metabolites and 1400 genes). The FBA mitochondrial model here considered is composed of 423 reactions (including transformation reactions and transport reactions between compartments, and those between internal and external environments) and 228 metabolites. The computational time to solve the problem with FBA is highly reduced. In our analysis we include also another FBA model, namely the alga *Chlamydomonas reinhardtii*, containing also the mitochondria organelle compartment.

In this work, we propose a methodology suitable for evolutionary single- and many-objective optimization in biological models. We apply our pipeline to three case studies: the algal

FBA model, the mitochondrial FBA model, and the mitochondrial DAEs model. The systematic analysis of the Pareto surface emerging from the biological system can be performed by evaluating the sensitivity of its components, and the functional relations inferred from its constraints. To this end, we adopt sensitivity and identifiability analyses. Our methodology can be used to tackle FBA, GPR, and systems of ODEs, and also to optimize reaction fluxes and gene sets simultaneously.

2 Many-objective concurrent optimization of biological models

A trade-off between multiple tasks performed by a biological organism can be reached using a multi-objective optimization algorithm. Given r objective functions $f_1, \ldots, f_r$ to be maximized, seeking the optimal array means solving the problem $\max_x(f_1(x), f_2(x), \ldots, f_r(x))^T$, where x is the variable in the search space. (The same definition can also be adopted when minimizing a function, taking into account that minimizing f_i is equivalent to maximizing $-f_i$). The *Pareto front* is defined as the set of non-dominated points. A solution y^* is said to be non-dominated if there does not exist a point y such that $f(y)$ dominates $f(y^*)$. Formally, y^* is Pareto optimal if $\nexists y \; s.t. f_i(y) > f_i(y^*) \; \forall i = 1, \ldots, r$, where f is the vector of r objective functions that have to be maximized.

For instance, a bacterium can use the Pareto-front to maximize simultaneously two or more requirements. A point in the objective space can be regarded as the phenotype associated with a point in the variable space (the genotype). As a result, the Pareto front is the set of all the phenotypes that remain after eliminating all the feasible phenotypes dominated on all tasks.[16]

Case study 1: FBA algal mitochondria

In this first case, we take into account the flux balance analysis model of *C. reinhardtii* that contains reactions related to the mitochondrial compartment. The model includes 2188 reactions, 1706 metabolites, 718 gene sets and 112 pathways. Out of 2188 reactions, 164 are mitochondrial reactions. By using a multi-objective optimization algorithm,[17] we maximize ATP and NADH production. In order to measure ATP and NADH production, we add two reactions representing the transport of ATP and NADH from the matrix to the external environment. In this way, we can calculate their rate in the FBA framework. The aim is to find the optimal genetic strategies for increasing the algal bioenergetic yield.

Genetic strategies stand for knockout operations. A gene is knocked out if it cannot be expressed. Therefore, the gene is neither transcribed to mRNA nor translated to peptide. Accordingly, the enzymes that under natural conditions are synthesized by that gene are not present in the cell, and the corresponding reactions do not occur. Knockout strategies are represented in the model through a knockout vector y. Each element of y represents a gene set. The gene set y_l can be expressed if $y_l = 0$, otherwise $y_l = 1$. The gene sets are constituted by one or more genes related by a Boolean relationship. Each gene set is linked to the reactions in the

metabolic network by a gene-reaction mapping. For instance, the gene set "geneA OR geneB" involving the two genes *geneA* and *geneB* is linked to the reaction *reactionAB*. The reaction *reactionAB* takes place if at least one of the two genes is present and expressed in the network. Each gene set can be linked to one or more reactions.

Searching for the best genetic strategies means searching for the best knockout vector $\tilde{y}$ that maximizes ATP and NADH production. The constraint-based combinatorial optimization problem has therefore two objective functions: ATP and NADH production. Constraints include (i) thermodynamics bounds of the reactions of the metabolic networks, (ii) knockout bound, *i.e.*, the maximum number of genes that can be turned off, and more significantly (iii) a nested biomass maximization problem.[18] The biomass maximization is a constraint of the problem.

In our experiment, we set the maximum knockout number equal to 10. The result of the optimization is shown in Fig. 1. Since ATP and NADH production are conflicting objectives, we obtain a set of optimal solutions constituting the Pareto front. From the Pareto front, we can choose the best trade off design according to function values, knockout cost and robustness indices.[17] The black point in Fig. 1 corresponds to the wild type condition, *i.e.*, when all genes of the metabolic network can be expressed.

Case study 2: mitochondrial FBA model

The FBA mitochondrion model[20] contains 423 reactions and 326 metabolites. As in the previous case study, by using a multi-objective optimization algorithm we maximize ATP and NADH production. In order to measure the matrix NADH productivity, we add a reaction that represents the transport of NADH from the matrix to the external environment. In this way, we can calculate the rate of NADH production in the FBA framework. The aim is to find the optimal environment for mitochondria so as to increase their bioenergetic yield. The decision variables are the 73 uptake fluxes. We search for the best values of uptake
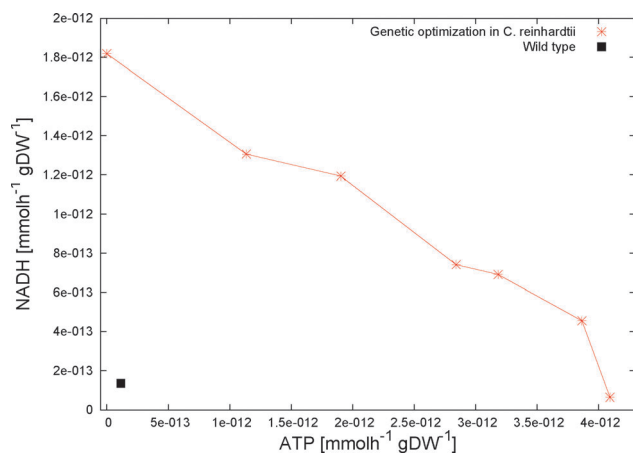

Fig. 1 Genetic strategies search in *Chlamydomonas reinhardtii*.[19] In red, the Pareto optimal points and, in black, the wild type solutions. Genetic strategies are knockout operations, useful to maximize the bioenergetic yield.
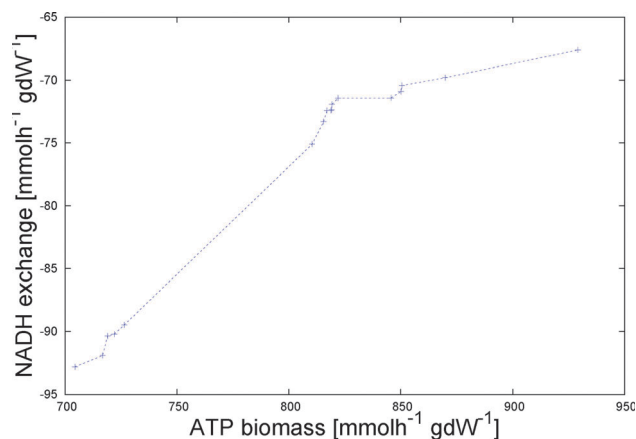

Fig. 2 Effect of the genetic algorithm on the Pareto front when optimizing ATP and NADH in the FBA mitochondrial model.[20] This Pareto front has been obtained with 100 individuals and halted at the 900th generation. We optimized the uptake rate fluxes (73 exchange fluxes) to maximize the energy. The algorithm was set in order to search for the environmental conditions to optimize the objective functions.

rate flux, whose maximum value is 1000 mmol h^{-1} gDW^{-1}. The optimization finds a single Pareto point that reaches the maximum amount of ATP (1000 mmol h^{-1} gDW^{-1}), without NADH production. In another optimization experiment, we maximize ATP production and simultaneously minimize NADH production (Fig. 2). We observe that ATP production grows more rapidly than NADH consumption.

We initialize the input fluxes of the mitochondrial model as described in the work by Smith *et al.*[20] Under these conditions (before the optimization), the ATP production is equal to 139.4264 mmol h^{-1} gDW^{-1}, while NADH is totally consumed in the metabolism, and the productivity is equal to 0. After the optimization, the maximum ATP production is 929.0940 mmol h^{-1} gDW^{-1} corresponding to an NADH consumption of 67.62 mmol h^{-1} gDW^{-1}. By considering this optimal Pareto solution and comparing the initial state (before the optimization) with the optimal state (maximum ATP production), we remark that ATP and NADH increase when the uptake rates linked to (*R*)-3-hydroxybutanoate, isocitrate, α-ᴅ-glucose, citrate and oxygen increase (Fig. S1 in ESI†). Indeed, the oxygen is the variable that changes more, from 19.8 mmol h^{-1} gDW^{-1} to 143.17 mmol h^{-1} gDW^{-1}. Therefore, the oxygen is the element that plays a major role in ATP and NADH production in mitochodria. This result is recurrent in biological experiments, and thus validates our analyses.

In this experiment, the optimization does not consider the limitation of substrates (as glucose or oxygen) in the biological environment, so we consider our approach as an asymptotic analysis for investigating the potentiality of mitochondria. Indeed, the optimization algorithm searches for the optimal environmental conditions without considering that the glucose and the other elements of the environment in a real cellular context can be limited. Each uptake flux can reach the upper bound value, *i.e.* 1000, which is usually not feasible. In a real context, such as that of a cell, glucose availability is in fact limited.

For this reason, we consider this experiment as a theoretical study that helps evaluate how the system behaves when it wants to increase ATP and NADH.

In a second experiment, in order to include in our optimization the limited availability of elements of the environment, we change the upper bound of each variable or uptake flux. In this way, we are modeling a real context, *i.e.*, an environment where glucose, lactate and other elements are present in a limited amount. In the original work,[20] the maximum uptake rate of the fluxes was limited as follows: oxygen to 19.8, arginine to 0.0068, lysine to 0.0298, proline to 0.0044, aspartate to 0.1524, α-D-glucose to 0.9000, (*R*)-3-hydroxybutanoate to 0.7000, isoleucine to 0.0039, valine to 0.0106, hexadecanoic acid to 1.0000, (*S*)-lactate to 0.5750, HCO_3^- to 0.0198. These values have been validated using experimental data.[20] In this second experiment, we only optimize the twelve fluxes cited above. Additionally, for each variable, the domain space is constrained between 0 and +33% of the maximum uptake rate used by the authors and reported above. Under these conditions, NADH production does not increase. Specifically, we observe a consumption of NADH (negative values in Fig. S2, ESI†). ATP increases and, by considering only the solutions such that NADH is positive, we find ATP = 185.4299 mmol h^{-1} gDW^{-1}. The optimization algorithm here used is the Non-dominated Sorting Genetic Algorithm II, also known as NSGA-II,[21] and described in depth in the following section.

Case study 3: mitochondrial DAEs model

Using the Non-dominated Sorting Genetic Algorithm II, also known as NSGA-II,[21] we optimize multiple energy-related objectives. The model we adopt here consists of 73 differential-algebraic equations (DAEs) to model 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. As in the previous case studies, we maximize the production of adenosine triphosphate (ATP) and nicotinamide adenine dinucleotide (NADH).

The variable space is defined as the space of feasible initial concentrations of 50 metabolites. The population is initialized in a random way and, for each individual of the population, the fitness score is computed. The fitness score is calculated according to the values of the objective functions. An individual of the population is preferred to another (has a better fitness score) if its objective functions are larger in a maximization problem or smaller in a minimization problem. The fitness score is also called *rank*. An individual is a feasible vector of concentrations of metabolites. Each individual is assigned a rank, and between two solutions with different non-domination ranks, the one with the lowest rank is preferred. After sorting the individuals according to the level of non-domination, the fitness score of each individual is computed by evaluating the objective functions associated with it.

The algorithm is based on non-domination ranking and crowding distance. The algorithm ranks the candidate solutions using the constrained non-dominated criteria. The non-dominated front is generally assigned a rank of one. Similarly, the second non-dominated front (the non-dominated front that would be detected after the removal of all the points of the non-dominated front) has a rank of two and so on. The candidate solutions with lower rank are the best candidates to be selected for the next generation. The lower the non-domination rank of a candidate solution is, the better it is. If two candidate solutions have the same non-domination rank, the algorithm prefers the candidate solution with the smaller crowding distance.

Successively, three steps are carried out iteratively: (i) 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; (ii) the algorithm selects a number of new 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; (iii) a new population of the same size of the initial population is formed by selecting only the best individuals from the parents and the offspring. By using the mutation operator, one individual of the population is selected randomly, and a random set of its elements is mutated. Instead, by using the crossover operator, two individuals of the population are chosen and their elements are crossed to form a new individual.[21] Each individual of the final population in the variable space corresponds to a point of the Pareto front in the objective space.

Before the optimization, at the fully oxidized state we obtain NADH = 1.5987×10^{-10} nmol mg^{-1} (formation) and ATP = -0.0014 nmol mg^{-1} (consumption). After the optimization, we obtain the Pareto-optimal points shown in Fig. 3.

Finally, we analyze more thoroughly two particular Pareto-optimal solutions, *i.e.*, the point with maximum ATP synthesis (and lower NADH formation) and the point with maximum NADH formation (and lower ATP synthesis). By maintaining Ca^{2+} fixed at 10^{-5} nmol mg^{-1}, after the optimization the first solution provides NADH = 6.17×10^{-15} nmol mg^{-1} and ATP = 2027.34 nmol mg^{-1}, with overproduction of SUC$_{mtx}$, SCoA$_{mtx}$, CoASH$_{mtx}$, H$_{mtx}^+$ and ATP$_{ims}$ (ims = inter-membrane space, mtx = matrix) and under-production 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 NADH = 6.07×10^{-6} nmol mg^{-1} and ATP = -3734.6 nmol mg^{-1} (consumption), over-producing the following metabolites: H$_{mtx}^+$, ISOC$_{mtx}$, SUC$_{mtx}$ and ATP$_{ims}$, whereas CIT$_{mtx}$, MAL$_{ims}$ and AMP$_{ims}$, PYR$_{ims}$, GLU$_{ims,cyt}$ and aKG$_{ims}$ are totally consumed.

If the matrix calcium content is maintained fixed or increased from 10^{-5} to 10^{-4} nmol mg^{-1}, the ATP synthesis and NADH formation stop, and both molecules are consumed by the metabolism (see Fig. 3, red signs). This experiment 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 to 1.5×10^{-5} nmol mg^{-1}, the NADH formation increases, while ATP decreases (see Fig. 3, green signs). We have also performed other two experiments, decreasing Ca^{2+} to 10^{-6} nmol mg^{-1} (see Fig. 3, blue circles) and to 10^{-5}/1.5 nmol mg^{-1} (see Fig. 3, green circles).
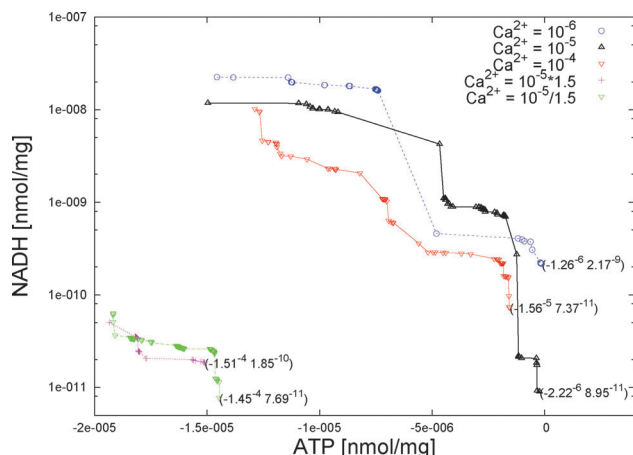
Fig. 3 ATP and NADH production maximization in the mitochondrial DAEs model.[13] We varied the initial concentrations of 55 metabolites and solved the system of DAEs. We simulated five differential states based on the concentration of calcium in the matrix: the standard concentration (10^{-5} nmol mg^{-1}), two increments of the standard concentration (10^{-4} nmol mg^{-1} and 1.5×10^{-5} nmol mg^{-1}) and two decrements of the standard concentration (10^{-6} nmol mg^{-1} and $10^{-5}/1.5$ nmol mg^{-1}).

Fig. 4 ATP and NADH production maximization and minimization in the mitochondrial DAEs model.[13] The regions define the states of mitochondria under different conditions. The red region represents the pathological state (cancer conditions), while the green zone the healthy state. In purple we show the mitochondrial pathological state with the minimization of ATP and NADH production, i.e., during the cellular apoptosis. Pareto fronts are in black.

For the cancer studies, we focused the analysis on mitochondrial activity and its role in cancer diseases. In a first step, our objective is to maximize the mitochondrial bioenergetics (ATP and NADH contents) in order to estimate the mitochondrial dynamics under cancer conditions. We simultaneously maximized ATP and NADH production by varying the initial conditions (the initial metabolites contents) used to solve the DAEs system. In particular, we varied the initial conditions for 50 metabolites (the decision variables), while maintaining at the physiological concentration the following: (1) matrix water volume, (2) inner membrane space water volume, (3) matrix free chloride, (4) total matrix ATP, (5) total matrix ADP, (6) total matrix GTP, (7) total matrix GDP, (8) total matrix NADH, (9) total matrix NAD, (10) total mitochondrial ubiquinol, (11) total mitochondrial ubiquinone, (12) total CO_2 matrix, (13) total O_2 matrix, (14) inter membrane space (IMS) free proton, (15) IMS free potassium, (16) IMS free magnesium, (17) IMS free calcium, (18) IMS free sodium, (19) total IMS cytochrome $c2+$, (20) total IMS cytochrome $c3+$. In addition, to simulate the cancer environment, we changed the initial condition for (i) the mitochondrial membrane potential $\Delta\Phi$, (ii) the total extra-mitochondrial glucose-6-phosphate content (G6P) and (iii) the extra-mitochondrial free proton content (H$^+$). We fixed $\Delta\Phi = 2$ mV, G6P $= V_{cyt} \times 10^6$ nmol mg^{-1} and H$^+ = V_{cyt} \times 10^{-5}$ nmol mg^{-1}. In the healthy state, $\Delta\Phi = 1$ mV, G6P $= V_{cyt} \times 10^{12}$ nmol mg^{-1} and H$^+ = V_{cyt} \times 10^7$ nmol mg^{-1}, where V_{cyt} is the water volume in cytosolic space.

In a second step, we used the same design to minimize ATP and NADH when the mitochondrion is under cancer conditions in order to find the variables playing a crucial role in eliminating the cancer cells. The results of the optimization are shown in Fig. 4. With respect to the state where no ATP is available in the matrix, the abuse of notation of a negative concentration in the model can be thought of as the change in ATP concentration needed to produce the corresponding NADH. Using our method
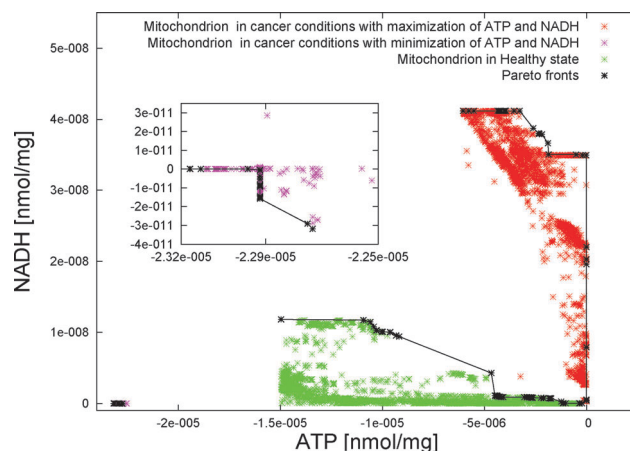
we can distinguish between healthy and pathological states. The red and green regions represent, respectively, the pathological and healthy state in mitochondria during the production of ATP and NADH. The purple region represents the apoptosis state of a cell under cancer conditions, i.e., when mitochondria are not able to produce bioenergy. The black line marks the Pareto optimal solutions. In the cancer experiments, the initial Ca^{2+} concentration used in all the simulations is 10^{-5} nmol mg^{-1}.

3 Sensitivity analysis

Sensitivity analysis allows us to rank inputs of the model according to their influence on the output. The parameters of a metabolic network can be reactions, enzymes, genes, pathways and metabolites. Ranking these elements can be very useful in that, for instance, once the most sensitive parameters are selected, one can focus its analysis on the features of the model related to these parameters.

By considering the model of the algal metabolism of *Chlamydomonas reinhardtii*,[19] and the DAEs[13] and the FBA[20] model of the mitochondrial metabolism, we calculated the sensitivity indices, respectively, for the uptake reactions, for the metabolites and for the inner matrix reactions. We used and implemented the method by Morris[22] for analyzing the uptake reactions in *C. reinhardtii* and the metabolites involved in the DAEs mitochondrial network.[13] The Morris method calculates the effect of the perturbation on a parameter s, estimating the distribution of elementary effects EE$_s$. The mean μ^* and standard deviation σ^* of the distribution indicate how the parameter should be considered important. A large (absolute) central tendency for EE$_s$ indicates an input with an important overall influence on the output. A large spread indicates an input whose influence is highly dependent on the values of the inputs.[22]

To evaluate the perturbations on the inner matrix reactions of the FBA model,[20] we implemented a new method, inspired

by Morris,[22] that perturbs each parameter and considers only the feasible solutions. Indeed, the FBA approach considers constraint-based models, and the perturbation of inner matrix reactions is not always allowed. The output considered in our analysis is the vector of output fluxes for FBA models, and the curves of integration for the DAEs model.

In Fig. 5 and Fig. S5 (ESI†), we show, respectively, the results of the sensitivity analysis for the FBA model and the DAEs model of the mitochondrion. In the first case, the *saccharopine dehydrogenase* reaction is the most sensitive, followed by *ornithine aminotransferase*, *glycine C-acetyltransferase* and *3-hydroxyacyl-CoA dehydrogenase*. In the second case, the initial concentrations of the metabolites are taken into account. Only the initial concentration variations of the extra-mitochondrial glc6phosphate, extra-mitochondrial fumarate,

inner-membrane aspartate, inner-membrane citrate, inner-membrane AMP and inner-membrane fumarate do not affect the output, unlike all the others, which have a large influence.

Fig. 5 Sensitivity analysis results for the FBA mitochondrial model.[20] We report the two sensitivity indices for each of the 135 internal reactions. The larger is the value of the two indices, the larger is their influence of the corresponding reaction on output of the model. On the key, only the most sensitive reaction IDs are reported. The labels on the key are sorted according to sensitivity ranking. For a detailed description of the key, see the work by Smith *et al.*[20]

Fig. 6 Sensitivity analysis applied to the photosynthetic algal metabolism of *C. reinhardtii.*[19] We calculate the sensitivity of the uptake rate fluxes on the output of the model. As shown in the plot, the photon uptake rate is the most sensitive flux. The labels on the key are sorted according to the sensitivity ranking.

Table 1 Identifiability analysis applied to the 100 non-dominated points of the ATP–NADH Pareto front with a standard amount of calcium. The initial concentrations of the 55 metabolites are grouped according to functional relations. The value of r^2 is an indicator of the amount of variance of the response explained by the predictors. A large ratio $cv(x) = std(x)/mean(x)$ indicates that the data are scattered, thus suggesting practical non-identifiability. ''n.a.'' stands for ''not available'', when metabolites do not play any role in the two-objective maximization of ATP and NADH. An asterisk is added when $r^2 > 0.9$ and $cv > 0.1$. Two asterisks indicate a strong interdependence between variables, *i.e.* the same functional group has been detected even if the role of response and predictors is switched

Variable	Metabolite	Metabolite groups	r^2	cv
x_1	H_m	x_1,x_{39}*	0.956	1.855
x_2	K_m	x_2,x_{53}*	0.931	0.305
x_3	Na_m	x_3*	0.975	0.407
x_4	Mg_m	x_4,x_{33}	0.941	0.094
x_5	Pi_m	x_5	0.986	0.008
x_6	CIT_m	x_6,x_{26}*	0.938	0.382
x_7	$ISOC_m$	x_7,x_{33}**	0.974	0.921
x_8	aKG_m	x_8,x_{37}*	0.937	1.868
x_9	$SCoA_m$	x_9,x_{37}*	0.942	1.181
x_{10}	SUC_m	x_{10}*	0.989	0.379
x_{11}	FUM_m	x_4,x_{11},x_{15}*	0.976	1.193
x_{12}	MAL_m	x_{12},x_{45}*	0.971	0.375
x_{13}	OAA_m	x_{13}*	0.993	1.503
x_{14}	GLU_m	x_7,x_{14}*	0.914	0.421
x_{15}	ASP_m	x_1,x_{15}*	0.967	0.166
x_{16}	PYR_m	x_1,x_{16},x_{47}*	0.969	0.673
x_{17}	$AcCoA_m$	x_{17},x_{53},x_{54}*	0.982	0.861
x_{18}	$CoASH_m$	x_{18},x_{28}**	0.957	0.176
x_{19}	H_{im}	n.a.	n.a.	n.a.
x_{20}	K_{im}	n.a.	n.a.	n.a.
x_{21}	Na_{im}	n.a.	n.a.	n.a.
x_{22}	Mg_{im}	n.a.	n.a.	n.a.
x_{23}	Ca_{im}	n.a.	n.a.	n.a.
x_{24}	ATP_{im}	x_{24},x_{52}**	0.971	0.413
x_{25}	ADP_{im}	x_{25},x_{26}*	0.979	0.917
x_{26}	AMP_{im}	x_{26},x_{43}*	0.980	0.883
x_{27}	Pi_{im}	x_{26},x_{27}*	0.975	0.868
x_{28}	PYR_{im}	x_{18},x_{28}**	0.957	1.353
x_{29}	CIT_{im}	x_{18},x_{29},x_{43}*	0.981	0.715
x_{30}	$ISOC_{im}$	x_{30}*	0.985	1.418
x_{31}	aKG_{im}	x_6,x_{17},x_{31}*	0.982	0.350
x_{32}	SUC_{im}	x_{32},x_{47}*	0.978	0.690
x_{33}	FUM_{im}	x_7,x_{33}**	0.974	0.929
x_{34}	MAL_{im}	x_{34}*	0.991	0.394
x_{35}	GLU_{im}	x_{35}*	0.990	0.545
x_{36}	ASP_{im}	x_{33},x_{36}	0.897	2.030
x_{37}	K_c	x_8,x_{26},x_{37}*	0.985	0.357
x_{38}	Na_c	x_{38},x_{45}**	0.951	0.326
x_{39}	Mg_c	x_9,x_{39}*	0.973	0.750
x_{40}	Ca_c	x_{40}*	0.980	0.368
x_{41}	ATP_c	x_{41}	0.972	0.018
x_{42}	ADP_c	x_{42}*	0.988	1.289
x_{43}	AMP_c	x_{26},x_{43},x_{53}*	0.990	0.827
x_{44}	Pi_c	x_{28},x_{44}	0.960	0.014
x_{45}	CIT_c	x_{38},x_{45}**	0.951	0.390
x_{46}	$ISOC_c$	x_{46}*	0.984	0.384
x_{47}	aKG_c	x_{24},x_{32},x_{47}*	0.986	1.112
x_{48}	SUC_c	x_{48}*	0.977	2.261
x_{49}	FUM_c	x_{14},x_{49}*	0.941	0.408
x_{50}	MAL_c	x_{25},x_{31},x_{50}*	0.983	0.957
x_{51}	GLU_c	x_2,x_{51}*	0.935	0.684
x_{52}	ASP_c	x_{24},x_{52}**	0.971	0.924
x_{53}	PYR_c	x_1,x_{53}*	0.944	0.335
x_{54}	GLC_c	x_6,x_{54}	0.940	0.092
x_{55}	Ca_m	x_{30},x_{42},x_{55}*	0.973	0.622

For the light-driven algal metabolism of *Chlamydomonas reinhardtii*, we detect that photon absorption is the most sensitive, and its influence is very relevant. The keys in the figures are sorted according to the sensitivity ranking (Fig. 6).

4 Identifiability analysis

A constraint in a biological model usually involves decision variables, thus establishing a functional relation between them. The need for inferring the model structure can be addressed with a mathematical analysis of the constraints operating in the model. When a component (*e.g.* a parameter or a variable) cannot be uniquely inferred through data measurement or experimental fitting, it is said to be *non-identifiable*. The *structural non-identifiability* occurs when an unambiguous estimation cannot be reached since non-identifiable components are functionally related with other non-identifiable components. Conversely, the *practical non-identifiability* is caused by the low amount or quality of data available for estimating the component. The identifiability analysis (IA) detects those groups of model components that comprise a functional relation by fitting the model repeatedly to experimental data.

Here we seek non-identifiable variables. Given n values for each of the m decision variables $\{x_1, \ldots, x_m\}$, let $K = [v_1, \ldots, v_m] \in \mathbb{R}^{n \times m}$ be the matrix of the repeated estimates. The column $v_i \in \mathbb{R}^n$ contains the n estimates for the ith variable. Let the variables be functionally related by unknown linear or non-linear relations linearized by the transformations α and β_j, where

$$\alpha(x_i) = \sum_{j \neq i}^{m} \beta_j(x_j) + \xi,$$

and ξ is a Gaussian noise. Using the ACE algorithm,[23] we seek the optimal transformations $\hat{\alpha}(x_i)$ and $\hat{\beta}_j(x_j), j \neq i$ such that

$$\hat{\alpha}(x_i) = \sum_{j \neq i}^{m} \hat{\beta}_j(x_j),$$

Fig. 8 Optimal transformations β (y axis) found for the three decision variables ASP, FUM and Mg (x axis) [nmol mg^{-1}] with [Ca^{2+}] = 10^{-5} nmol mg^{-1}. Although FUM has been once assigned to the same functional group of ASP and Mg, it shows a different and noisier behavior.
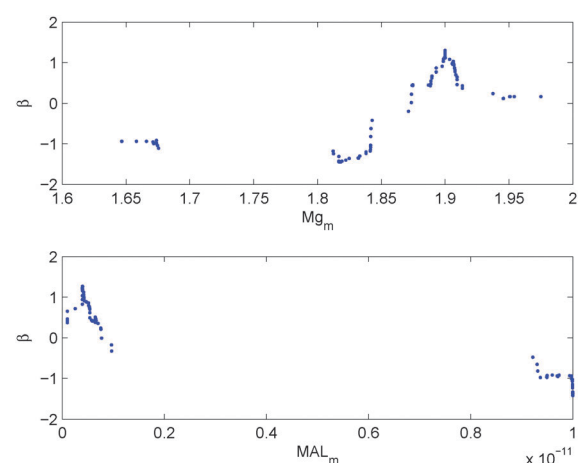
Fig. 9 Optimal transformations β (y axis) found for the two decision variables Mg and MAL (x axis) [nmol mg^{-1}] when maximizing ATP and NADH in the DAEs model of the mitochondrion simulating the cancer condition.

Fig. 7 Functional relation among the three decision variables ASP, FUM and Mg [nmol mg^{-1}] in the matrix, thus highlighting the structural non-identifiability of these variables. This group has been detected for the FUM metabolite with a standard calcium concentration [Ca^{2+}] = 10^{-5} nmol mg^{-1}.

Fig. 10 Optimal transformations β (y axis) found for the two decision variables Mg and SCoA [nmol mg^{-1}] when minimizing ATP and NADH in the DAEs model of the mitochondrion simulating the cancer condition.

where x_i is the response and the set $\{x_j\}_{j \neq i}$ is the set of the predictors.

We take into account the output of the multi-objective optimization carried out on the DAEs model of the mitochondrial metabolism by Bazil et al.[13] under different calcium conditions. Mor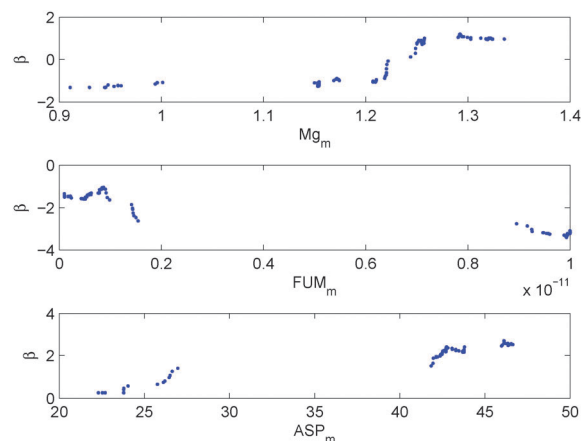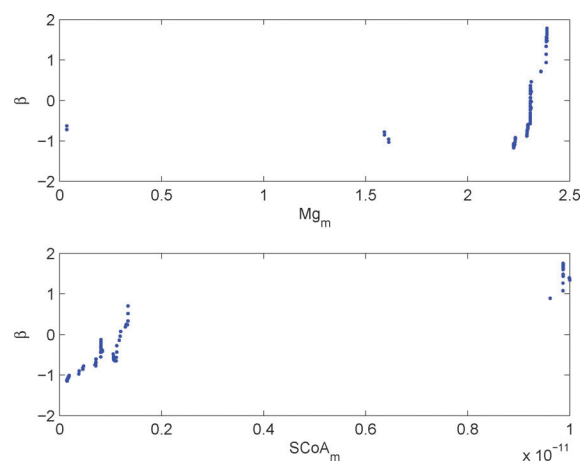e specifically, we investigate (i) the standard condition of calcium, as well as four modified conditions: (ii) a small increment (1.5 times), (iii) a large increment (10 times),

(iv) a small decrement (1.5 times) and (v) a large decrement (10 times). We take into account 55 decision variables of the model, namely the initial concentrations of its metabolites. We adopt the method proposed by Hengl et al.[24] to detect automatically structural identifiability consisting of functional relations between decision variables. These relations are detected using the alternating conditional expectation algorithm (ACE).[23]

Table 2 Functional relations under cancer conditions. We show the results of the IA applied after the multi-objective minimization and maximization of ATP and NADH

Variable	Metabolite	Min{ATP,NADH} Groups	r^2	cv	Max{ATP,NADH} Groups	r^2	cv
x_1	H_m	x_1,x_{17},x_{38}^*	0.983	0.517	x_1,x_{52}	0.962	0.024
x_2	K_m	x_2	0.994	0.044	x_2,x_9,x_{15}	0.973	0.079
x_3	Na_m	x_3,x_{52},x_{53}^*	0.990	0.671	x_3,x_{52}	0.987	0.015
x_4	Mg_m	x_4,x_9^{**}	0.982	0.237	x_4,x_{12}	0.965	0.042
x_5	Pi_m	$x_5,x_{39},x_{48},x_{51}^*$	0.995	1.455	x_5,x_{51}	0.949	0.000
x_6	CIT_m	x_6,x_{18},x_{27}	0.988	0.059	x_6,x_{38}	0.965	0.000
x_7	$ISOC_m$	x_7,x_9,x_{39}^*	0.992	1.541	x_5,x_7,x_{52}^*	0.985	0.704
x_8	aKG_m	x_8,x_{11},x_{55}^*	0.986	0.650	x_8,x_{30},x_{37}^*	0.984	0.934
x_9	$SCoA_m$	x_4,x_9^{**}	0.982	1.290	x_9,x_{52}^*	0.979	0.777
x_{10}	SUC_m	x_{10},x_{52}^*	0.972	0.686	x_{10},x_{46}^{**}	0.994	0.915
x_{11}	FUM_m	x_{11},x_{37},x_{39}^*	0.984	2.666	x_{10},x_{11}^*	0.916	1.910
x_{12}	MAL_m	x_{12}^*	0.994	2.032	x_4,x_{12}^{**}	0.965	1.118
x_{13}	OAA_m	x_{13},x_{18}^*	0.967	0.283	x_{13}^*	0.989	2.602
x_{14}	GLU_m	x_{14},x_{50},x_{55}^*	0.989	1.399	x_{14},x_{24}^*	0.976	0.709
x_{15}	ASP_m	x_{15}^*	0.988	0.208	x_4,x_8,x_{15}^*	0.987	0.197
x_{16}	PYR_m	x_{16},x_{37},x_{54}^*	0.980	1.459	x_{16}	0.985	0.000
x_{17}	$AcCoA_m$	x_{17},x_{41},x_{44}^*	0.988	1.268	x_4,x_9,x_{17}^*	0.981	0.740
x_{18}	$CoASH_m$	x_{13},x_{18},x_{37}^*	0.992	0.200	x_{18}^*	0.986	0.164
x_{19}	H_{im}		n.a.	n.a.		n.a.	n.a.
x_{20}	K_{im}		n.a.	n.a.		n.a.	n.a.
x_{21}	Na_{im}		n.a.	n.a.		n.a.	n.a.
x_{22}	Mg_{im}		n.a.	n.a.		n.a.	n.a.
x_{23}	Ca_{im}		n.a.	n.a.		n.a.	n.a.
x_{24}	ATP_{im}	x_2,x_{24},x_{49}^*	0.980	0.282	x_{14},x_{24},x_{37}^*	0.983	0.854
x_{25}	ADP_{im}	x_{25},x_{42}^*	0.971	0.239	x_3,x_{25},x_{44}^*	0.985	0.234
x_{26}	AMP_{im}	x_{26},x_{41}^*	0.958	2.195	x_{26},x_{51}^{**}	0.967	0.788
x_{27}	Pi_{im}	x_6,x_{27}^*	0.967	1.533	x_{27}	0.993	0.005
x_{28}	PYR_{im}	x_{28},x_{55}^*	0.962	0.282	x_{28}	0.980	0.004
x_{29}	CIT_{im}	x_{29},x_{53}^*	0.982	0.672	x_{12},x_{29},x_{37}^*	0.980	0.949
x_{30}	$ISOC_{im}$	x_3,x_{30}	0.977	0.006	x_8,x_{10},x_{30}^*	0.992	0.972
x_{31}	aKG_{im}	x_{31},x_{34}^{**}	0.985	1.217	x_{12},x_{31}^*	0.970	1.153
x_{32}	SUC_{im}	x_{32},x_{54}^{**}	0.966	0.946	x_{32},x_{35},x_{40}^*	0.975	0.262
x_{33}	FUM_{im}	x_{33},x_{46}^*	0.971	2.470	x_{33}^*	0.989	2.012
x_{34}	MAL_{im}	x_{31},x_{34}^{**}	0.985	0.533	x_{34},x_{48}^{**}	0.982	1.093
x_{35}	GLU_{im}	x_{32},x_{35},x_{46}^*	0.985	2.539	x_9,x_{35},x_{42}^*	0.978	1.167
x_{36}	ASP_{im}	x_{34},x_{36}^*	0.955	2.994	x_{36},x_{52}	0.935	0.008
x_{37}	K_c	x_8,x_{18},x_{37}^*	0.993	0.859	x_{24},x_{37},x_{49}^*	0.984	0.944
x_{38}	Na_c	x_{33},x_{38},x_{46}^*	0.989	0.143	x_{30},x_{38}^*	0.938	0.308
x_{39}	Mg_c	x_{39},x_{43}^*	0.974	1.342	x_{12},x_{39}^*	0.956	0.358
x_{40}	Ca_c	$x_{12},x_{18},x_{25},x_{40}^*$	0.992	0.282	x_{32},x_{40},x_{50}^*	0.973	2.272
x_{41}	ATP_c	x_4,x_7,x_{41}^*	0.988	0.865	x_9,x_{41}	0.898	0.004
x_{42}	ADP_c	x_{42}^*	0.995	0.161	x_{35},x_{42},x_{51}^*	0.980	0.752
x_{43}	AMP_c	x_{43},x_{55}^*	0.980	1.404	x_{43},x_{53}^*	0.924	0.839
x_{44}	Pi_c	x_{43},x_{44},x_{47}^*	0.986	1.241	x_{12},x_{41},x_{44}^*	0.977	0.117
x_{45}	CIT_c	x_{14},x_{45},x_{53}^*	0.991	0.687	x_{45}^*	0.989	0.246
x_{46}	$ISOC_c$	x_{38},x_{46}^*	0.981	0.231	x_{10},x_{46}^{**}	0.994	0.964
x_{47}	aKG_c	x_{24},x_{47}^*	0.969	0.557	x_{47},x_{50}^*	0.953	2.273
x_{48}	SUC_c	x_{17},x_{48}^*	0.978	0.573	x_{34},x_{48}	0.982	0.002
x_{49}	FUM_c	x_{24},x_{46},x_{49}^*	0.980	0.211	x_{10},x_{37},x_{49}^*	0.985	0.742
x_{50}	MAL_c	x_{13},x_{50},x_{52}^*	0.988	0.631	x_{10},x_{50}^*	0.971	0.856
x_{51}	GLU_c	x_{17},x_{51}^*	0.945	2.709	x_{26},x_{51}	0.967	0.003
x_{52}	ASP_c	x_3,x_{52}^*	0.964	1.326	x_3,x_7,x_{52}^*	0.989	1.143
x_{53}	PYR_c	x_3,x_{53}^*	0.982	0.686	x_{46},x_{53}	0.916	0.000
x_{54}	GLC_c	x_{32},x_{54}^{**}	0.966	0.341	x_{16},x_{46},x_{54}^*	0.958	0.674
x_{55}	Ca_m	x_{38},x_{43},x_{55}^*	0.987	1.263	x_{31},x_{55}	0.970	0.001

In our study, the connection between optimization and identifiability is due to the fact that the estimates in the fitting matrix K consist of all the non-dominated feasible points obtained in the multi-objective optimization process. In the mitochondrial DAEs model, the constraint is detected through 100 estimates of all the 55 variables (metabolites). Each estimate corresponds to a point of the Pareto front involving the ATP and NADH production. To detect the functional relations, shown in Table 1, we use the Mean Optimal Transformation Approach (MOTA).[24]

The ''groups'' column indicates the functional relations between variables. For instance, fumarate, Mg and aspartate are functionally related. In other words, the response variable x_{11} is strongly related to the predictors x_4 and x_{15}. Conversely, Na, Pi, SUC and OAA in the matrix do not have any functional relation with any other enzyme. The value of r^2 is an indicator of the amount of variance of the response explained by the predictors. A high r^2 indicates that the fixation of the predictors has a major effect on the standard deviations of the response. The cv(x) = std(x)/mean(x) helps distinguish practical identifiable from non-identifiable parameters.[24] A functional relation among k variables is tested k times considering each variable as a response variable, and gets two asterisks if it is detected at least two times.

In Fig. 7 we show the functional relations among ASP, FUM and Mg detected by the identifiability analysis applied to fumarate. Fig. 8 shows that the optimal transformation β found for x_{11} is different from the transformations found for x_{15} and x_4, although the IA applied to FUM has assigned x_{11} to the same functional group of x_4 and x_{15}. This can happen when the variables taken into account are also practically non-identifiable (which is the case of these three metabolites, since their cv is high).

In order to perform the identifiability analysis in the cases analyzed in the optimization section, we perturb the calcium concentration and detect the non-identifiabile variables for each of the cases we analyze. All the results are shown in Table S1 (ESI†). After multiplying by 1.5 the amount of calcium in the model, in Fig. S6 (ESI†) we show the functional relations among Na, Pi and AcCoA detected by the identifiability analysis applied to Na. Fig. S7 (ESI†) shows that the optimal transformation β found for AcCoA is different from the other two transformations in the functional group, thus indicating a practical non-identifiability. Fig. S8 (ESI†) shows the functional relations among Mg and CoASH when the amount of calcium in the matrix is increased from 10^{-5} to 10^{-4} nmol mg^{-1}, detected by the identifiability analysis applied to Mg. This pair has a clear functional relation but a very low cv value, indicating structural but not practical non-identifiability. Fig. S9 (ESI†) shows the optimal transformation β for Mg and CoASH.

If the amount of calcium is reduced to 10^{-5}/1.5 nmol mg^{-1}, the identifiability analysis applied to PYR detects a functional relation among H, PYR and AcCoA (see Fig. S10, ESI†). Remarkably, the optimal transformation β shown in Fig. S11 (ESI†) indicates that H may be practically non-identifiable, since its optimal transformation differs from the others in the same

Table 3 Identifiability analysis applied to the FBA model of the mitochondrion. The 73 fluxes are grouped according to functional relations

Variable	Metabolite	Metabolite groups	r^2	cv
x_1	CO_2	x_1, x_{73}	0.983	0.000
x_2	FAD	x_2, x_{40}	0.991	0.000
x_3	Oxygen	x_3, x_{24}, x_{40}	0.998	0.015
x_4	Glutathione	x_2, x_4	0.989	0.000
x_5	Arginine	x_3, x_5, x_{16}^*	0.996	0.751
x_6	Lysine	x_6, x_9^*	0.981	1.351
x_7	Glutamate	x_7, x_{33}, x_{70}^*	0.995	0.378
x_8	Proline	x_8, x_{61}^*	0.968	1.173
x_9	Fe^{2+}	x_9, x_{66}	0.992	0.000
x_{10}	Aspartate	x_{10}, x_{46}, x_{49}	0.993	0.087
x_{11}	Glycine	x_{11}, x_{46}^{**}	0.983	0.234
x_{12}	Tyrosine	x_{12}, x_{47}^{**}	0.990	0.248
x_{13}	Asparagine	$x_{13}, x_{21}, x_{31}, x_{37}, x_{38}, x_{51}$	0.998	0.000
x_{14}	Methionine	x_{14}, x_{73}^{**}	0.981	0.653
x_{15}	Phenylalanine	x_{15}, x_{28}, x_{30}^*	0.995	0.778
x_{16}	Tryptophan	x_5, x_{16}, x_{49}^*	0.994	0.655
x_{17}	Histidine	x_3, x_{17}, x_{54}^*	0.994	0.486
x_{18}	Glutamine	$x_9, x_{10}, x_{18}, x_{42}^*$	0.998	0.185
x_{19}	5-Aminolevulinate	x_2, x_{19}^*	0.973	0.317
x_{20}	Cysteine	x_{11}, x_{20}^*	0.975	0.579
x_{21}	Phosphatidylcholine	x_{21}, x_{41}	0.979	0.000
x_{22}	Phosphatidylserine	x_{13}, x_{22}, x_{51}	0.992	0.000
x_{23}	Phosphatidylethanolamine	x_1, x_9, x_{23}, x_{37}	0.998	0.000
x_{24}	α-D-Glucose	x_3, x_{24}^*	0.996	0.392
x_{25}	Glycerol	x_{25}, x_{61}	0.992	0.000
x_{26}	Propanoate	x_{24}, x_{26}^*	0.979	0.485
x_{27}	Sulfate	x_{27}, x_{38}	0.995	0.000
x_{28}	Sulfite	x_{28}, x_{42}	0.979	0.000
x_{29}	Thiosulfate	x_3, x_{29}	0.989	0.000
x_{30}	Succinate	x_{30}, x_{33}^*	0.972	0.779
x_{31}	(S)-Malate	x_2, x_{31}^*	0.982	0.381
x_{32}	Isocitrate	x_{11}, x_{32}	0.981	0.097
x_{33}	Citrate	x_3, x_7, x_{33}^*	0.997	0.243
x_{34}	Butanoicacid	x_{28}, x_{34}^*	0.972	0.192
x_{35}	Oxaloacetate	x_{35}^*	0.996	0.103
x_{36}	2-Aminoadipate	$x_7, x_{36}, x_{37}, x_{50}, x_{54}$	0.997	0.000
x_{37}	2-Oxoadipate	x_{37}, x_{51}	0.974	0.000
x_{38}	CoA	x_{38}, x_{54}	0.981	0.000
x_{39}	H_2O	x_{39}	0.998	0.000
x_{40}	(R)-3-Hydroxybutanoate	x_2, x_{40}	0.991	0.060
x_{41}	Urea	x_{41}, x_{54}	0.982	0.000
x_{42}	Orthophosphate	x_{42}, x_{54}	0.985	0.000
x_{43}	Thiamindiphosphate	$x_{13}, x_{31}, x_{43}, x_{63}$	0.997	0.000
x_{44}	Biomass	x_{32}, x_{44}, x_{64}^*	0.998	0.864
x_{45}	NAD^+	x_{45}, x_{60}	0.981	0.000
x_{46}	ADP	x_{11}, x_{46}	0.983	0.000
x_{47}	UDP	x_{12}, x_{47}	0.990	0.000
x_{48}	2-Oxoglutarate	x_3, x_{48}^*	0.984	0.307
x_{49}	GDP	x_{49}, x_{66}	0.983	0.000
x_{50}	CDP	$x_7, x_{50}, x_{54}, x_{73}$	0.997	0.000
x_{51}	dATP	x_{29}, x_{41}, x_{51}	0.994	0.000
x_{52}	dADP	x_{52}, x_{63}	0.972	0.000
x_{53}	dGTP	x_{11}, x_{53}	0.975	0.000
x_{54}	dGDP	x_{50}, x_{54}	0.986	0.000
x_{55}	dTDP	x_{55}, x_{66}	0.974	0.000
x_{56}	dCTP	x_9, x_{56}, x_{57}	0.988	0.000
x_{57}	dTTP	x_3, x_{57}	0.982	0.000
x_{58}	dCDP	$x_1, x_{30}, x_{54}, x_{58}$	0.998	0.000
x_{59}	ATP	$x_3, x_{12}, x_{52}, x_{59}^*$	0.996	3.265
x_{60}	H^+	$x_5, x_{36}, x_{45}, x_{60}^*$	0.996	0.627
x_{61}	Isoleucine	x_{25}, x_{61}^{**}	0.992	0.173
x_{62}	Leucine	x_{62}^*	0.997	0.220
x_{63}	Valine	x_9, x_{63}^*	0.983	0.522
x_{64}	Acetaldehyde	x_{16}, x_{46}, x_{64}^*	0.995	0.650
x_{65}	Hexadecanoicacid	x_{65}^*	0.994	2.486
x_{66}	(S)-Lactate	x_9, x_{66}^{**}	0.992	0.485
x_{67}	Fumarate	x_{18}, x_{67}^*	0.973	1.366
x_{68}	Octanoicacid	$x_{60}, x_{66}, x_{68}, x_{73}$	0.994	0.000
x_{69}	HCO^{3-}	x_{69}^*	0.998	1.540

Table 3 (*continued*)

Variable	Metabolite	Metabolite groups	r^2	cv
x_{70}	Putrescine	x_{32},x_{44},x_{70}	0.995	0.000
x_{71}	Alanine	x_{20},x_{71}	0.978	0.000
x_{72}	Serine	$x_{46},x_{72}{}^*$	0.977	0.677
x_{73}	Threonine	$x_{14},x_{73}{}^{**}$	0.981	0.330

functional group. Among the functional groups shown in Table S1 (ESI†), when the concentration of calcium is $10^{-5}/1.5$ nmol mg^{-1}, the one composed of Asp and CoASH is detected two times from the IA algorithm, and therefore these two metabolites are involved in a strong functional relation, shown in Fig. S12 (ESI†). The optimal transformations β are shown in Fig. S13 (ESI†). Fig. S14 (ESI†) shows the functional relations among Na and GLU when the amount of calcium is reduced to 10^{-6} nmol mg^{-1}, detected by the identifiability analysis applied to Na. Fig. S15 (ESI†) shows the optimal transformation β found for both metabolites.

In order to give some insights into the cancer condition of the mitochondrion, in Fig. S16 (ESI†), we show the functional relations among Mg and MAL detected by the identifiability analysis applied to both Mg and Malate when maximizing the production of ATP and NADH in the mitochondrial model. Indeed, this is a strong functional relation since this pair has been detected both times by the algorithm. The optimal transformation β is shown in Fig. 9.

Conversely, when minimizing ATP and NADH, Mg is detected in a strong functional relation with SCoA (see Fig. S17, ESI† and the optimal transformation β in Fig. 10). This relation has been detected two times both for Mg and for SCoA. All these conditions are summarized in Table 2.

Finally, in Table 3 we show the output of the identifiability analysis applied to the mitochondrial FBA model.

5 Conclusions

The present work shows how multi-objective optimization, sensitivity and identifiability analyses can be combined to gain an insight into computational models under a variety of conditions. In particular, we have investigated two models based on flux balance analysis (a mitochondrial model and an algal model), and a model based on differential and algebraic equations. The analysis of the Pareto optimal front, with information from sensitivity and identifiability, proves useful to tackle the complex chemical reaction networks often found in cells, consisting of many pathways with different responses to perturbations.

The sensitivity analysis evaluates the role that the variables play in the model. Specifically, the biological system is perturbed to seek the variables that mainly affect its behavior under varying conditions. The identifiability analysis finds functionally related groups of variables to reveal whether a component of a model can be uniquely determined or not. Non-identifiable variables may be fixed at an arbitrary value without modifying the model's dynamical properties.

The combination of these techniques allows us to reach the optimal configuration and carry out a systematic study on any

model consisting of ODEs, DAEs, FBA and GPR mappings. In a large model composed of different submodels, the output of these techniques applied to subnetworks can be easily integrated, allowing a convergence of different modeling techniques.[25] For instance, links between properties of each functional submodule can be established using network analysis approaches,[26] which allow us to understand the function and evolution of the entire biological process.

Interestingly, we can conclude that our approach also allows comparison of the patterns of change in metabolic processes. We are interested in providing measures so that we could eventually make statements such as "the pathway X in humans is more similar to a pathway in gorillas than mice". Our methodology points in that direction, as it allows us to reduce the metabolic network with summary statements such as "the pathway P plays a key role in the species X, whereas in the other species Y the pathways Q and R are dominating". We believe that our work could provide an interesting insight into the methodology of biological network analysis, inching towards the full understanding of the complex bioenergetics of the mitochondrion.

Acknowledgements

References

1 J. C. Thrash, A. Boyd, M. J. Huggett, J. Grote, P. Carini, R. J. Yoder, B. Robbertse, J. W. Spatafora, M. S. Rappé and S. J. Giovannoni, *Sci. Rep.*, 2011, **1**, 13.

2 P. Lió and N. Goldman, *J. Mol. Evol.*, 2002, **54**, 519–529.

3 S. Raha and B. Robinson, *et al.*, *Trends Biochem. Sci.*, 2000, **25**, 502.

4 B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter, *et al.*, *Molecular Biology of the Cell*, 4th edn, 2002, ch. 14.

5 T. Kuroiwa, H. Kuroiwa, A. Sakai, H. Takahashi, K. Toda and R. Itoh, *Int. Rev. Cytol.*, 1998, **181**, 1–41.

6 J. Dudley and A. Butte, *Pacific Symposium on Biocomputing*, 2009, p. 27.

7 C. Angione, G. Carapezza, J. Costanza, P. Lió and G. Nicosia, *EMBnet. journal*, 2012, **18**, 20.

8 T. Williams, N. Turan, A. Diab, H. Wu, C. Mackenzie, K. Bartie, O. Hrydziuszko, B. Lyons, G. Stentiford and J. Herbert, *et al.*, *PLoS Comput. Biol.*, 2011, **7**, e1002126.

9 T. Ideker and N. Krogan, *Mol. Syst. Biol.*, 2012, **8**, 565.

10 A. Shiflett and P. Johnson, *Annu. Rev. Microbiol.*, 2010, **64**, 409.

11 P. Lió, *Gene*, 2005, **345**, 55–64.

12 U. Sengupta, S. Ukil, N. Dimitrova and S. Agrawal, *PLoS One*, 2009, **4**, e8100.

13 J. Bazil, G. Buzzard and A. Rundell, *PLoS Comput. Biol.*, 2010, **6**, e1000632.

14 U. Mühlenhoff, N. Richhardt, J. Gerber and R. Lill, *J. Biol. Chem.*, 2002, **277**, 29810–29816.

15 J. D. Orth, I. Thiele and B. Palsson, *Nat. Biotechnol.*, 2010, **28**, 245–248.

16 O. Shoval, H. Sheftel, G. Shinar, Y. Hart, O. Ramote, A. Mayo, E. Dekel, K. Kavanagh and U. Alon, *Science*, 2012, 1157–1160.

17 J. Costanza, G. Carapezza, C. Angione, P. Lió and G. Nicosia, *Bioinformatics*, 2012, **28**, 3097–3104.

18 A. Burgard, P. Pharkya and C. Maranas, *Biotechnol. Bioeng.*, 2003, **84**, 647–657.

19 R. Chang, L. Ghamsari, A. Manichaikul, E. Hom, S. Balaji, W. Fu, Y. Shen, T. Hao, B. Palsson and K. Salehi-Ashtiani, *Mol. Syst. Biol.*, 2011, **7**, 518.

20 A. Smith and A. Robinson, *BMC Syst. Biol.*, 2011, **5**, 102.

21 K. Deb, A. Pratap, S. Agarwal and T. Meyarivan, *IEEE Transactions on Evolutionary Computation*, 2002, **6**, 182–197.

22 M. Morris, *Technometrics*, 1991, **33**, 161–174.

23 L. Breiman and J. Friedman, *J. Am. Stat. Assoc.*, 1985, **80**, 580–598.

24 S. Hengl, C. Kreutz, J. Timmer and T. Maiwald, *Bioinformatics*, 2007, **23**, 2612–2618.

25 M. Dumas, *Mol. BioSyst.*, 2012, 2494–2502.

26 G. Lima-Mendez and J. van Helden, *Mol. BioSyst.*, 2009, **5**, 1482–1493.