RESEARCH ARTICLE

Multi-Target Analysis and Design of Mitochondrial Metabolism

Claudio Angione[1]☯*, Jole Costanza[2]☯, Giovanni Carapezza[3], Pietro Lió[1], Giuseppe Nicosia[3]

1 Computer Laboratory-University of Cambridge, Cambridge, United Kingdom, **2** Center for Genomic Science of IIT@SEMM, Istituto Italiano di Tecnologia, Milan, Italy, **3** Department of Mathematics and Computer Science-University of Catania, Catania, Italy

☯ These authors contributed equally to this work.
* claudio.angione@cl.cam.ac.uk

Abstract

Analyzing and optimizing biological models is often identified as a research priority in biomedical engineering. An important feature of a model should be the ability to find the best condition in which an organism has to be grown in order to reach specific optimal output values chosen by the researcher. In this work, we take into account a mitochondrial model analyzed with flux-balance analysis. The optimal design and assessment of these models is achieved through single- and/or multi-objective optimization techniques driven by epsilon-dominance and identifiability analysis. Our optimization algorithm searches for the values of the flux rates that optimize multiple cellular functions simultaneously. The optimization of the fluxes of the metabolic network includes not only input fluxes, but also internal fluxes. A faster convergence process with robust candidate solutions is permitted by a relaxed Pareto dominance, regulating the granularity of the approximation of the desired Pareto front. We find that the maximum ATP production is linked to a total consumption of NADH, and reaching the maximum amount of NADH leads to an increasing request of NADH from the external environment. Furthermore, the identifiability analysis characterizes the type and the stage of three monogenic diseases. Finally, we propose a new methodology to extend any constraint-based model using protein abundances.

1 Introduction

The analysis of models and the automated design of metabolic networks and synthetic pathways are key features for investigating biochemical systems. To this end, several mathematical model approaches have been designed, based on ordinary differential expression, stochastic methods, master equations or on algebraic equations [1]. More recently, algorithms and computational methods have been implemented to perform their numerical simulation. This led to biological models that researchers used to build gene regulatory networks of cells under specific experimental conditions [2], to inferring the metabolic flux behavior in single cells for synthetic purposes [3], and to pharmacokinetics/pharmacodynamics studies [4].

In this study, we propose the BioCAD framework, which integrates different computational tools able to analyze and query biological models. In particular, it integrates the Pareto optimal principle, the ϵ-dominance analysis and the identifiability analysis, with the aim of designing robust metabolic networks that perform specific tasks. Since the models considered in BioCAD represent biological processes, the tasks here reported are biological functions, such as the energy yield of a cell. The framework is general-purpose, and the design acts on genes, reactions, enzymes and metabolites (Fig 1).

In the work by Costanza et al. [5], the authors propose a technique for robust design of microbial strains by using multi-objective optimization techniques [6]. Given two or more conflicting objectives in a system, the Pareto front is the set of trade-off points that are "non-dominated", i.e., they are such that an increase of one objective comes at the cost of a decrease in the other objective. In a pre-processing step, models are analyzed by means of sensitivity analysis methods. In a post-processing step, each Pareto front achieved by the multi-objective optimization [6] is tested by the robustness analysis [7] in order to evaluate the fragility of each Pareto point. We remark that an effective framework for analysis of metabolic networks would include the following analyses: evolutionary many-objective optimization, robustness analysis, ϵ-dominance analysis, sensitivity and identifiability analysis [8].

Our BioCAD pipeline is able to tackle biological problems in an automated way. More specifically, it is able to analyze a metabolic network and design an optimized version of the same network. In the BioCAD framework, we present here a new optimization method called OPT-BioCAD, a stochastic general-purpose optimization algorithm able to perform single and multi-objective optimization. OPTBioCAD searches for optimal (i) genetic strategies, (ii) metabolites and enzyme concentrations, and (iii) flux rates by using combinatorial/continuous optimization. The ϵ-dominance Pareto front analysis extends the search space of solutions, therefore revealing other suitable (suboptimal) points, while the identifiability analysis finds functional relations among decision variables. The source code of our pipeline is in S1 Code. The novelty of the present work lies also in the Protein-Abundance Design through Multi-objective Optimization (PADMO) algorithm, able to integrate protein abundances in genome-scale metabolic networks while ensuring flux balance and optimization. The use of gene expression and protein abundance in computer-aided-design can help to build highly functional and robust engineered strains with reduced times and costs [9].

As introduced above, the energy yield can be considered an interesting point to analyze, since the metabolism of a cell is highly correlated with the energy production and utilization. Mitochondria are organelles in eukaryotic cells and play a key role in the cell. First, they are responsible for the energy productivity, since they synthesize adenosine triphosphate (ATP), the chemical energy in the cell. Second, the mitochondrion is the site of carbohydrates metabolism, fatty acid oxidation and urea cycle. Mitochondria are implied in many other important processes, such as the regulation of calcium homeostasis and other inorganic ions [10, 11], cellular differentiation, cell death (apoptosis) [12], as well as the control of the cell cycle and cell growth [13]. Mitochondria have been also found responsible for several human diseases, including mitochondrial disorders [14], cardiac dysfunction [15], and type 2 diabetes [16]. For the reasons discussed above, we investigate two mitochondrial models [17, 18].

The first model we take into account is the mitochondrial network by Smith et al. [17], composed of 423 reactions (including transformation reactions and transport reactions between compartments and those between internal and external environment) and 228 metabolites. Here, the system is described by considering a steady state for all the metabolites, and solved through flux balance analysis (FBA) [19].

FBA is a modelling approach to simulate and investigate genome-scale metabolic networks [3]. FBA is able to handle large networks and estimate the value of the metabolic fluxes through
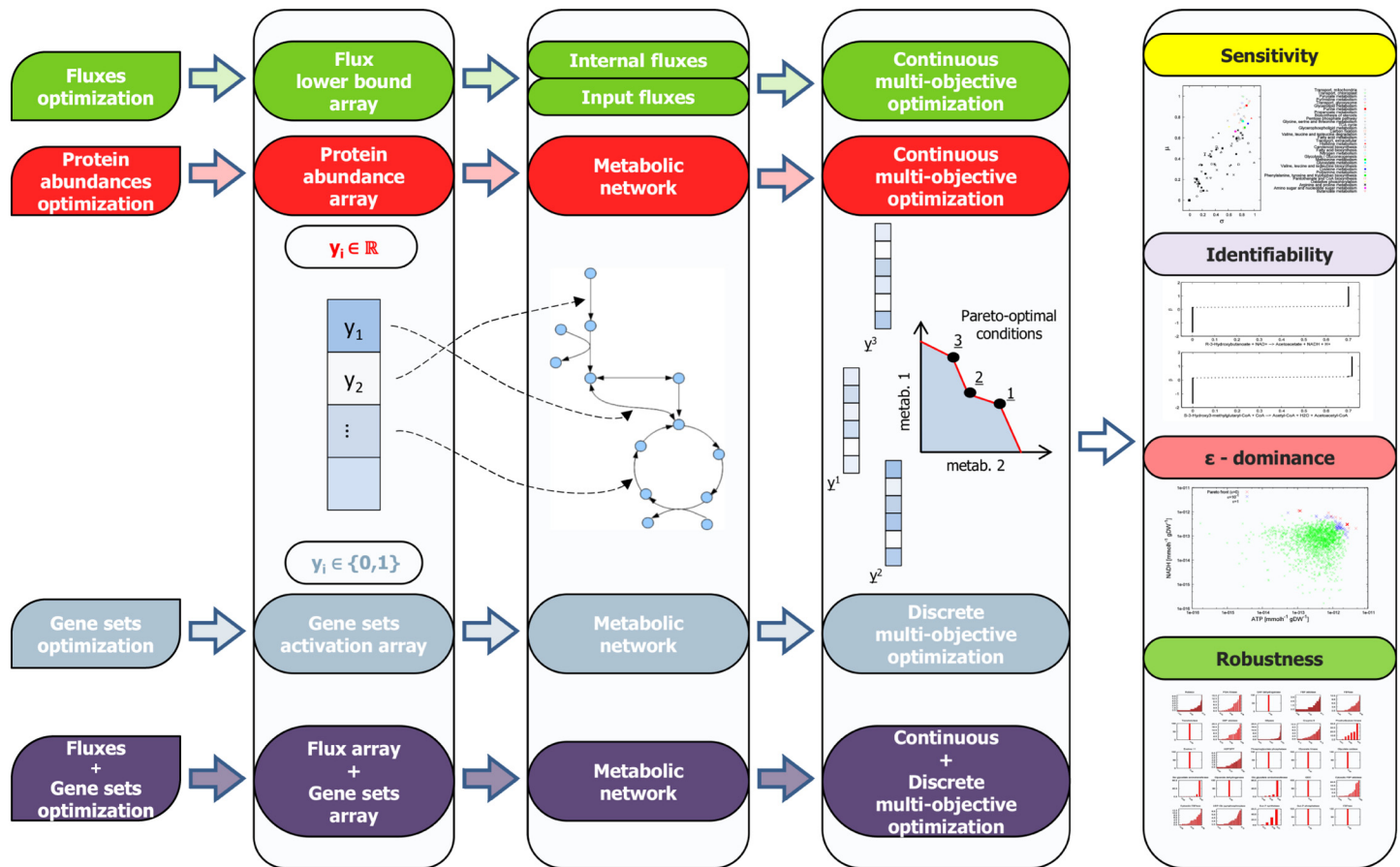
Fig 1. The B$_{IO}$CAD framework is able to perform multi-objective optimization on gene sets and protein abundances. The optimization can be applied to the Boolean arrays of gene sets (simulating an on/off condition), to the real-valued fluxes, or to the real-valued arrays representing protein abundances. In both cases, we seek for the Pareto-optimal arrays to simultaneously optimize two or more objective functions. The optimization is augmented with sensitivity, identifiability, robustness and ϵ-dominance analysis. The sensitivity analysis quantifies the importance of the input variables in the model, the identifiability analysis infers functional relations between them, the robustness is used in combination with the sensitivity and quantifies if a solution is reachable even if small perturbations are applied to the system, while the ϵ-dominance analysis identifies sub-optimal points.

the network when an objective function is defined (e.g., the growth rate of a cell or its energy yield). The network can be thought as a graph, where each edge represents a reaction and the metabolic flux through the reaction. The direction of each edge is linked to the reversibility or irreversibility of the corresponding enzymatic/transport/exchange reaction. Moreover, the vertices represent the substrates or the products of each reaction. The output of the network (the flux distribution) represents the phenotypic state of the biological system and depends on the topology of the network and on the input, i.e., the uptake rates. From a computational standpoint, the time needed to solve the network is remarkably short if compared, for instance, to methods used to solve ordinary differential equations. Furthermore, FBA does not involve kinetic parameters and is not affected by computational errors of numerical approximation. As a result, FBA allows handling large networks with a large number of components [19].

In the past, FBA models have been chosen for analysis aimed to synthetic biology. For instance, OptKnock [20] implements a bi-level programming framework, GDLS [21] proposes a local optimization algorithm, and GDMO [5] performs optimization based on genetic

algorithms, where bacterial cells are optimized and computationally designed in terms of gene knockouts.

For cross-comparison purposes, we will also consider a second model representing the genome-scale metabolic network of the *Chlamydomonas* algal cell [18]. This model includes the mitochondrial organelle; therefore, we use it to calculate the energy yield of the algal cell by searching optimal gene knockout and uptake rates in the FBA framework. The analysis performed on *Chlamydomonas* is reported in S1 File.

We search for those metabolites that are fundamental for optimizing the energy productivity, i.e., for maximizing ATP and NADH production in the matrix. Our method is able to tackle and compare different systems modeled through FBA, gene-protein-reaction mapping (GPR), systems of ODEs, and also to optimize reaction fluxes and gene sets simultaneously. Fig 1 shows the optimization steps performed by BioCAD (columns 1–4) and the additional analyses available (column 5).

The results we obtain indicate that the maximum ATP production is linked to a total consumption of NADH, probably because NADH is completely used to synthesize ATP through the electron transport chain. On the other hand, reaching the maximum amount of NADH leads also to an increasing request of NADH from the external environment. The optimization process showed also that biomass formation increases when ATP production is stopped. As expected, oxygen is more requested when ATP production is maximized, together with hydroxybutanoate, isocitrate, alpha-D-glucose and citrate metabolites. We finally report that the identifiability analysis can be used to detect and characterize the stage of a metabolic disease, as well as the type of monogenic disorder. More specifically, we highlight that functional relations among chemical reactions depend on the disease investigated and, within the same disease, on its stage.

2 Results

2.1 Optimization of the mitochondrial FBA model

Using our novel bio-inspired algorithm optBioCAD, we optimize the FBA mitochondrial model by Smith and Robinson [17]. The model contains 423 reactions and 228 metabolites. We specifically optimize energy-related objectives (ATP and NADH production). To measure the availability of NADH, we added an exchange reaction for the metabolite NADH to the external environment through the mitochondrial membrane. Given the common assumptions of flux-balance analysis, such addition does not perturb the dynamics of the metabolic network. Maximizing or minimizing the flux through this reaction represents a FBA-compatible way to require maximum of minimum overall concentration of the NADH metabolite in the mitochondrion. The aim is to find the optimal environment for mitochondria so as to increase their bioenergy yield. The decision variables are the 73 input fluxes. We therefore search for the best values of uptake rate of substrates, which are bounded by a maximum uptake rate of $1000\ \mu\text{mol min}^{-1}\ \text{gDW}^{-1}$ (DW stands for dry weight, i.e., the weight of the mitochondrion without water). The optimization finds a single Pareto point that reaches the maximum amount of ATP, without NADH production (Fig 2, left panel). In another optimization experiment, we maximize ATP production and simultaneously minimize NADH production. After 1000 generations of the optimization algorithm, we observe that ATP production grows more rapidly when NADH is consumed more. During the simulation, the algorithm finds many non-dominated points with respect to the previous generations. Instead, in the last generations (after the 900th generation), the algorithm finds only four Pareto points, which represent our final results. The results are shown in Fig 2, right panel.
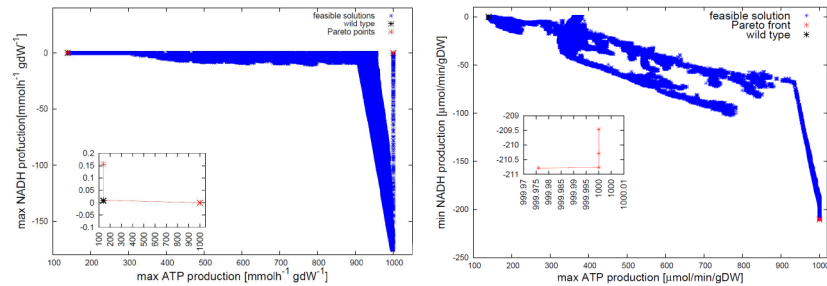
Fig 2. Optimization of the FBA mitochondrial model. (Left) Maximization of ATP and NADH production in the FBA mitochondrial model [17], carried out with 1000 individuals and halted at the 1500th generation. We optimized the uptake rate fluxes (73 exchange fluxes) to analyze the energy state of the mitochondrion. In blue the dominated feasible points, in black the wild type conditions, i.e., before optimization and in red the non-dominated Pareto points. Negative flux values represent an uptake rate, while positive values represent a production rate. (Right) Maximization of ATP production and minimization of NADH production.

doi:10.1371/journal.pone.0133825.g002

We set the input fluxes of the mitochondrial model as described by Smith and Robinson [17]. With this setting, ATP production is equal to 139.4264 μmol min^{-1} gDW^{-1}, while NADH is totally consumed, and the production is equal to 0. At the end of the optimization process, we find that the first non-dominated solution reaches NADH = -140.5484 μmol min^{-1} gDW^{-1} and ATP = 971.0874 μmol min^{-1} gDW^{-1}, and the second one reaches NADH = -139.5166 μmol min^{-1} gDW^{-1} and ATP = 971.1778 μmol min^{-1} gDW^{-1}. By comparing the initial state with the Pareto optimal state, we remark that ATP increases when the uptake rates linked to (R)-3-hydroxybutanoate, isocitrate, alpha-D-glucose, citrate and oxygen increase. Oxygen is the element that evolves more, and changes from 19.8 μmol min^{-1} gDW^{-1} to 143.17 μmol min^{-1} gDW^{-1}. In this case, the optimization does not take into account the limitation of substrates (such as glucose or oxygen) in the biological environment, so we consider this study as an asymptotic analysis for investigating the potential of mitochondria. Recently, different studies reported some experimental validations. For instance, Smith and Robinson [17] found that that when the maximum ATP production is reached (139.43 μmol min^{-1} gDW^{-1} calculated from the FBA model, and 150 μmol min^{-1} gDW^{-1} experimentally measured [22]), both oxygen and glucose uptake rates were at the maximum allowable rate [23, 24]. A negative value of NADH production indicates the NADH uptake rate. We suppose that NADH molecules are used in mitochondria to synthesize ATP molecules. When more NADH molecules are consumed, more ATP molecules are formed.

To give a more in-depth interpretation of our optimization, we choose a minimal set of decision variables, consisting of the following twelve input fluxes: oxygen, arginine, lysine, proline, aspartate, alpha-D-glucose, (R)-3-hydroxybutanoate, isoleucine, valine, hexadecanoic acid, (S)-lactate, HCO3-. Moreover, we change the maximum allowable uptake rate for each variable, that is equal to +33% of the nominal value. In this condition, NADH production does not decrease, while ATP increases and we find ATP = 185.4299 μmol min^{-1} gDW^{-1}. Also in this experiment, the oxygen is the variable that increases more (with respect to its nominal value).

Optimization of internal fluxes. The optimization of the metabolic reactions in the FBA mitochondrial model has been performed using a genetic algorithm and a novel mutation operator. The genetic algorithm is inspired to NSGA-II with a new mutation heuristic. Since perturbing metabolic reactions in FBA models may lead to unfeasible solutions, we create a mutation operator that takes into account this issue. We introduce two parameters, C and N. C is the maximum number of fluxes that can be perturbed. If the mutation results in unfeasible

solutions due to the constraints operating in the network, the procedure of mutation is repeated until a new child solution is found or a maximum number of N trials is reached. In the latter case, the current parent solution is maintained. We have performed the optimization of 229 transformation fluxes and conducted two experiments: (i) the simultaneous maximization of ATP and NADH production and (ii) the simultaneous maximization of ATP production and the minimization of NADH production. For both the experiments we have used a population of 1000 individuals. Each individual contains a vector of 229 values, and each value represents the rate of the corresponding metabolic flux. The mutation operator takes also into account the reversibility of the reactions. Our algorithm has performed the evolution of the population until the 300th generation. The results are shown in Fig S4 in S1 File. For both the experiments, the algorithm finds a set of non-dominated Pareto solutions, showed in red. Dominated and feasible solutions are shown in blue, while the point in black represents the wild type condition, i.e., the condition before the optimization.

Another multi-objective optimization that we take into account for the FBA mitochondrial model involves the production of biomass, ATP and reactive oxygen species (ROS) from complex I. Six reactions in the model represent the formation of the molecules responsible for the growth of the mitochondrion (amino acids, DNA, RNA, lipid, ATP, and heme). In an optimization to maximize ATP and biomass we define as biomass flux the sum of all the biomass fluxes except that of ATP (Fig S5 in S1 File). Conversely, in the optimization of ROS and biomass, to obtain the biomass flux we sum up all the six fluxes (Fig S6 in S1 File).

2.2 Characterization of monogenic diseases using identifiability analysis

Studying functional relations in the mitochondrial metabolic network merits further attention due to the possibility to characterize mitochondrial monogenic diseases. Here we consider the FBA model of the mitochondrion by Smith et al. [17]. The identifiability analysis reveals whether a component of a model is identifiable, i.e., uniquely determinable, thus indicating what can (and cannot) be inferred from a model. We choose ATP production as the objective function to evaluate the distribution of fluxes in the network. To reduce the non-identifiability, one should fix the non-identifiable fluxes at an arbitrary value. This does not affect the dynamical properties of the model, since all the variables related to the fixed one will change accordingly.

Our idea is that by performing the Identifiability Analysis (IA) (see Section 4) in healthy, pathological and disease conditions, we can characterize the onset of a disease by looking at the functional relations among fluxes. When taking into account a specific disease, we constrain the reaction responsible for that disease to various values in the range from 0 to the reaction flux under normal conditions [17], and we evaluate the amount of ATP and NADH as outputs of the model. We define a model condition as *disease status* if ATP production is less or equal to 33% of the production under normal conditions, and *inflammation status* if ATP production is less or equal to 66% but more than 33% of the production under normal conditions.

Fumarase deficiency. The fumarase deficiency is a monogenic disorder due to the impairment of the fumarate hydratase enzyme, caused by a mutation in the fumarate hydratase gene, which encodes the enzyme. As a result, the fumarate is not converted into malate in the TCA cycle, and therefore a marker for this disease is the presence of fumarate in the urine. Therefore, here we focus on the reaction Fumarate + $H_2O \rightarrow$ (S)-Malate. The fumarase flux is constrained to be equal to values equally distributed between 0 and 13.986 μmol min^{-1} gDW^{-1}. We first set the flux to be null, and then we increase it by 0.007 μmol min^{-1} gDW^{-1}, so as to obtain a 423×2000 matrix V containing all the fluxes corresponding to fixed fluxes of fumarate. In healthy conditions and when the objective function in FBA is the maximum ATP

production, we obtain a fumarase flux of 6.9721 μmol min^{-1} gDW^{-1}. In the fumarase deficiency conditions (about 2–3 μmol min^{-1} gDW^{-1}), ATP production is reduced until 75% of the maximum value [17]. For the fumarase deficiency, we identify these intervals of the fumarase flux: healthy stage [10,5.69), inflammation stage [5.69,3.69), pathological stage [3.69,0] μmol min^{-1} gDW^{-1} (we adopt the standard notation for half-closed intervals; [a, b) indicates that a is included and b is not included).

We first apply the IA to detect global relations between two or more variables, allowing the fumarase flux to span all the interval [0,13.986] μmol min^{-1} gDW^{-1}. In the table summarizing the results (Table A in S1 File), the "groups" column indicates the functional relations between variables. For instance, R01361MM (conversion of (R)-3-Hydroxybutanoate and NAD$^+$ into acetoacetate, NADH and H$^+$) and R01978MM (conversion of (S)-3-Hydroxy-3-methylglutaryl-CoA, and CoA into acetyl-CoA, H$_2$O and Acetoacetyl-CoA) are functionally related. In other words, the first reaction (response variable x_{47}) is strongly related to the second reaction (predictor x_{62}). (See Table A in S1 File and Table L in S1 FIle for reaction ID and stoichiometry.) In Fig 3 we plot the optimal transformations β found for these two reactions (see also Section 4). We note that the transformations are similar to each other, indicating the structural non identifiability of both variables. The functional relation between these two reactions (x_{47} and x_{62}) has been detected by the identifiability analysis applied to both reactions. This is therefore a strong relation, marked by a double asterisk in Table A in S1 File.

In Table B in S1 File, Table D in S1 File, and Table E in S1 File, we show the results of the IA applied to the metabolic network with various values of the fumarase flux. The (x_{28}, x_{50}) group is detected both in the healthy and in the pathological stage, but not in the inflammation stage. In the pathological stage only four different functional relations are detected, which means that variables are mostly unrelated to one another.

Succinate dehydrogenase deficiency. This disorders affects the mitochondrial complex II, responsible for linking the TCA cycle with the electron transport chain. It is often due to the bi-allelic inactivation of the SDHA gene. The deficiency of succinate dehydrogenase causes encephalomyopathy or tumor formation, and may affect motor and mental skills.

The behavior of ATP as function of the succinate dehydrogenase flux is equivalent to that obtained as function of the fumarase flux [17]. Therefore, we identify three intervals of the succinate dehydrogenase flux: healthy stage [10,5.69), inflammation stage [5.69,3.69), pathological stage [3.69,0] μmol min^{-1} gDW^{-1}.

In Fig 4 we show a strong functional relation between the two reactions R00713MM and R01648MM in healthy stage (full reaction stoichiometry is reported in Table L in S1 File). By cross-comparing, in the healthy stage, this functional relation under succinate dehydrogenase deficiency with the functional relation under fumarase deficiency referring to the same functional group, we report that the IA can be used to detect the type of monogenic disorder. In fact, the relation between the two variables depends on the disease investigated, although the intervals chosen are equivalent to those of the fumarase deficiency. Functional groups within succinate dehydrogenase deficiency may also include a high number of reactions (Fig 5).

α-ketoglutarate dehydrogenase deficiency. The α-ketoglutarate dehydrogenase enzyme provides protection against cyanide induced convulsions, and decreases mitochondrial damage induced by seizures caused by kainic acid. The corresponding deficiency affects the TCA cycle, and in particular the conversion of oxoglutarate into succinyl-CoA.

The intervals identified by plotting ATP as function of α-ketoglutarate dehydrogenase flux are as follows: healthy stage [0,8.31), inflammation stage [8.31,10.31), pathological stage [10.31,12] μmol min^{-1} gDW^{-1}. Interestingly, the transformations β found for the healthy, inflammation and pathological conditions (Fig 6) show that the optimal transformation in the
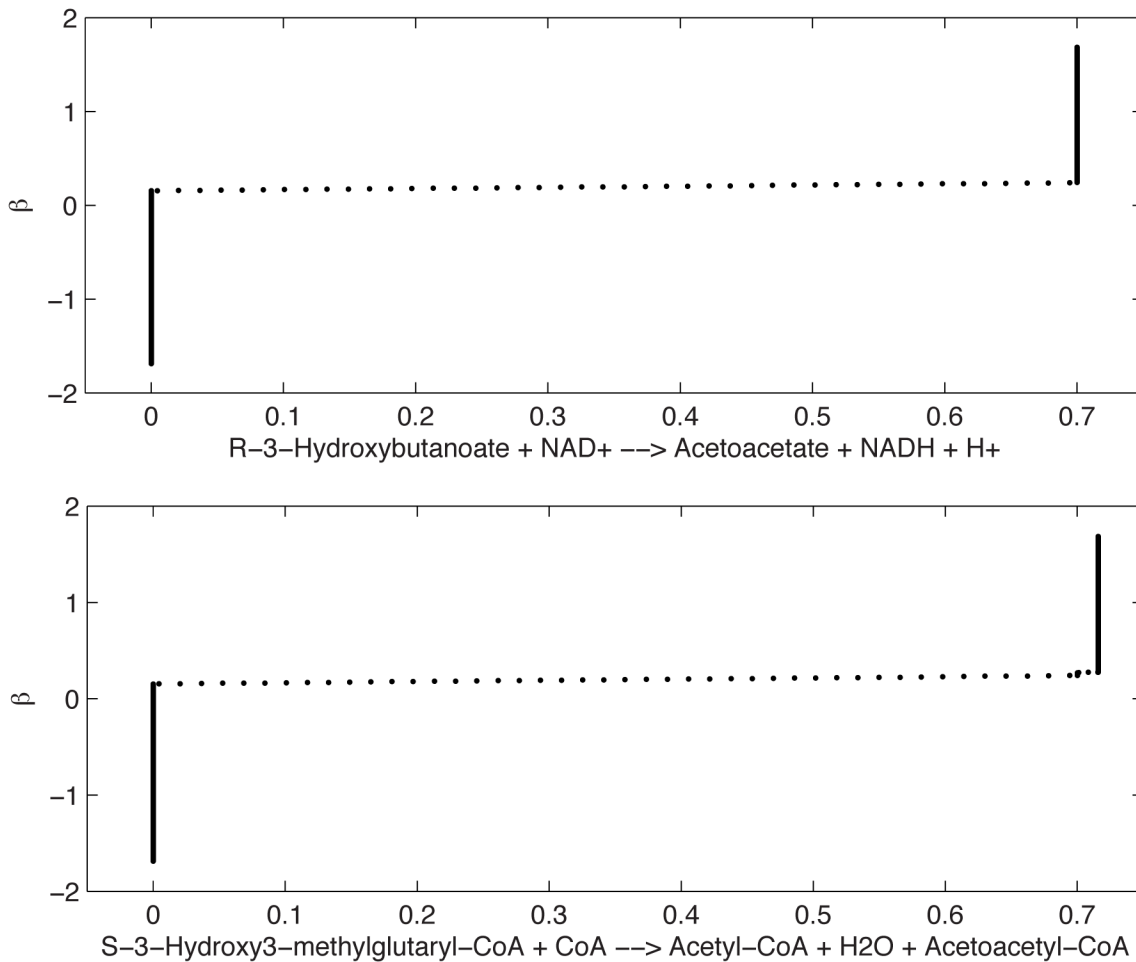
Fig 3. Optimal transformations β (y axis) found for the two fluxes R01361MM (top) and R01978MM (bottom) (x axis) [μmol min^{-1} gDW^{-1}] in the mitochondrial FBA model [17]. This plot proves that there is a strong relation between these two fluxes, with slightly different and noisier behavior in the neighborhood of 0 and 0.7.
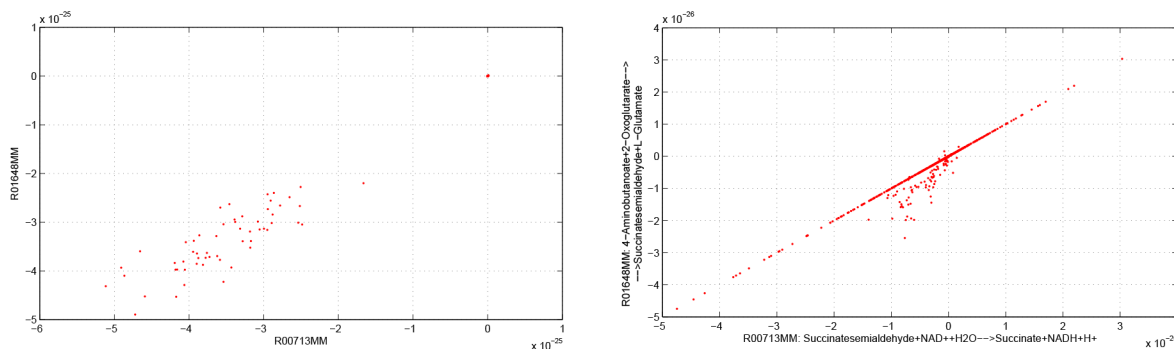
doi:10.1371/journal.pone.0133825.g003

Fig 4. Fumarase deficiency (left) and succinate dehydrogenase deficiency (right) in the healthy stage. Functional relation found for the two fluxes R00713MM and R01648MM (x axis) [μmol min^{-1} gDW^{-1}] in the mitochondrial FBA model [17]. Reaction stoichiometry is reported in Table L in S1 File. In the same stage of disease, the IA can be used to detect the type of monogenic disorder through the shape of the functional relation.
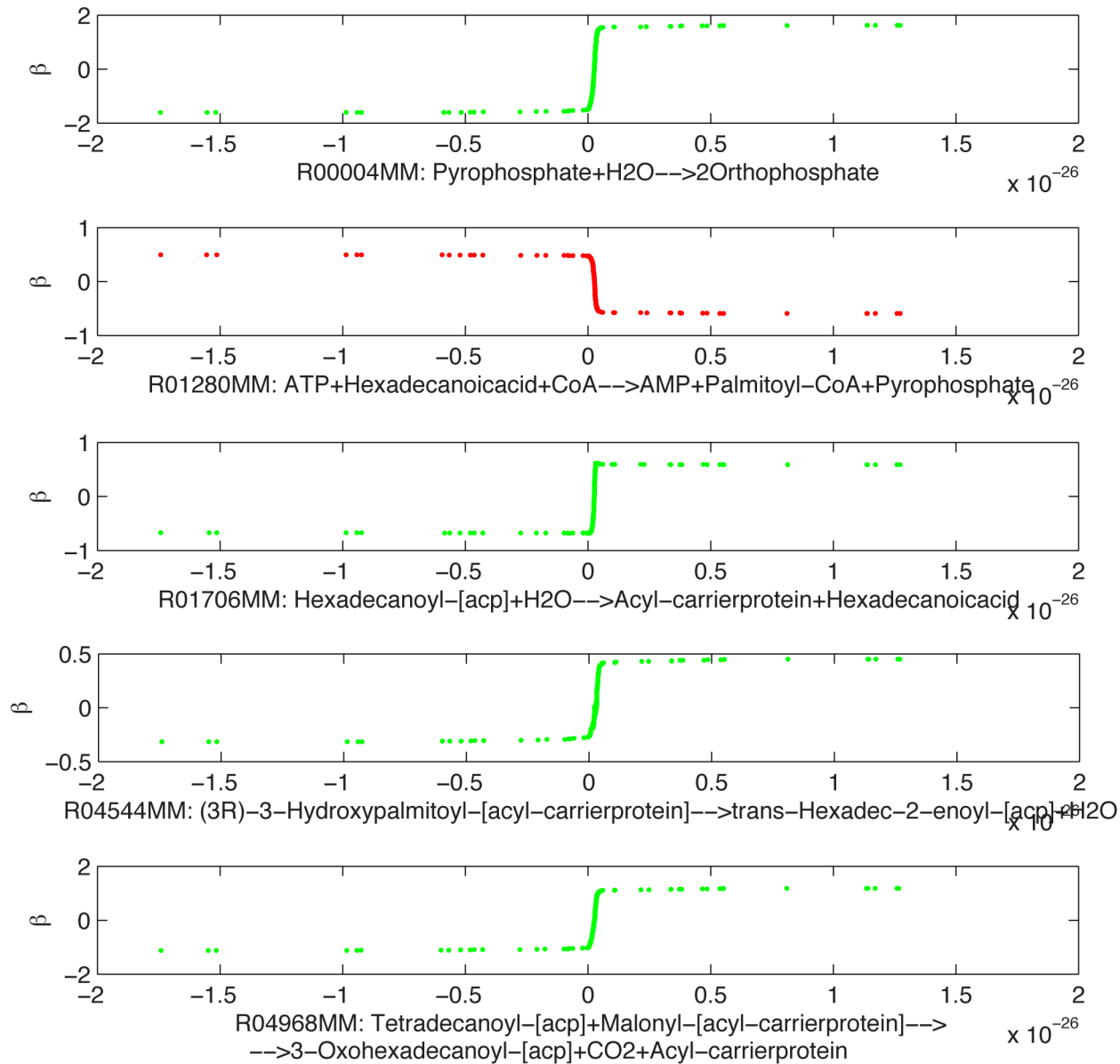
doi:10.1371/journal.pone.0133825.g004

Fig 5. Succinate dehydrogenase deficiency—inflammation stage. Optimal transformations β (y axis) found for the five fluxes R00004MM, R01280MM, R01706MM, R04544MM, and R04968MM (x axis) [μmol min^{-1} gDW^{-1}] in the mitochondrial FBA model [17].

doi:10.1371/journal.pone.0133825.g005

pathological stage has a different shape with respect to the healthy and the inflammation stages. The functional relations associated with these plots are in S1 File.

2.3 Optimization of protein abundances for FBA models

Given that protein synthesis is an outcome of expression of the genes that code for protein segments, our idea is to link the values representing protein abundances to the bounds of the flux of the reactions controlled by those proteins. We remark that the FBA models allow the variation of the bounds of each reaction flux. After the perturbation of the range of the reaction fluxes, due to the values of protein abundance given by the optimization algorithm, the actual
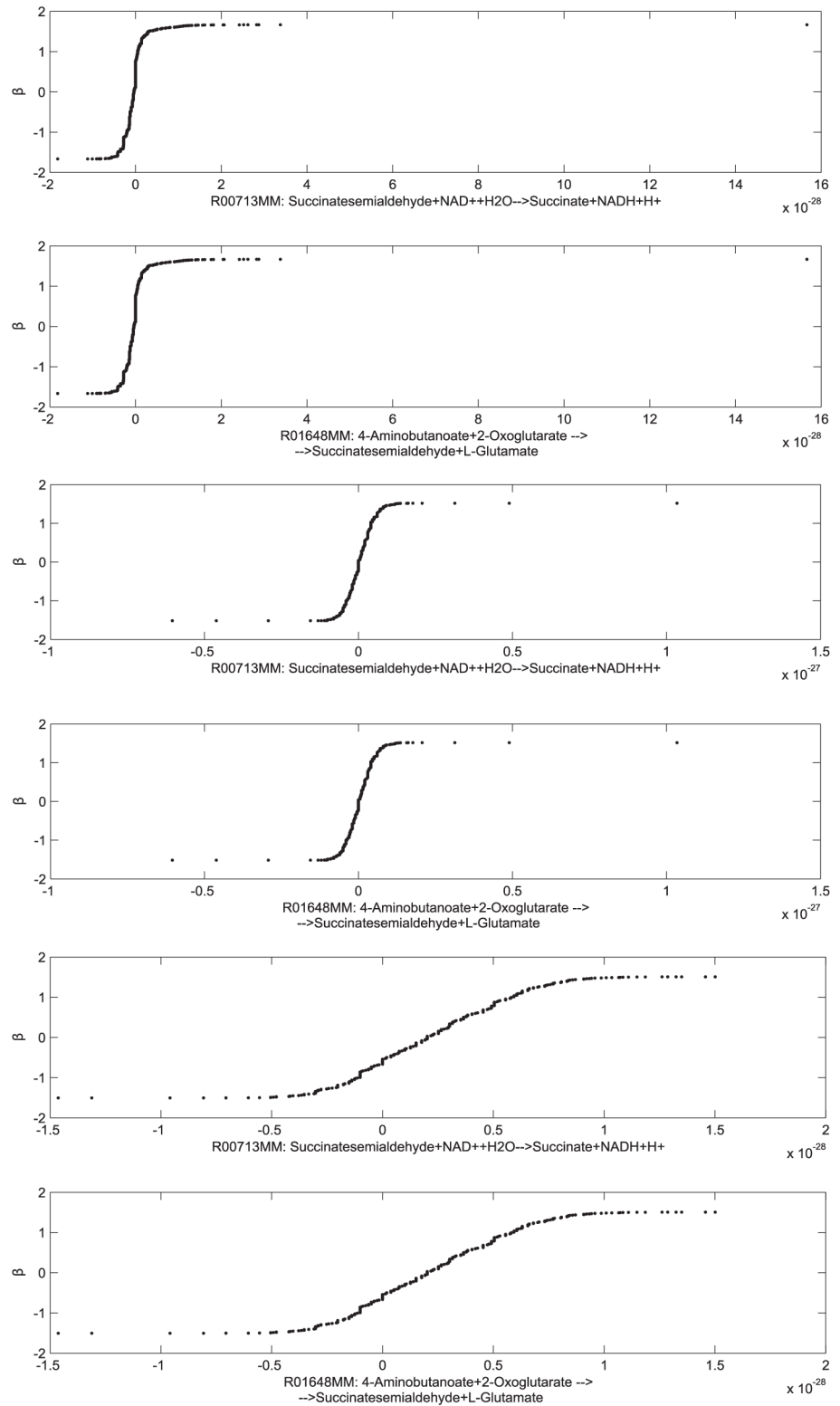
Fig 6. α-ketoglutarate dehydrogenase deficiency—healthy stage (top), inflammation stage (middle), pathological stage (bottom). The optimal transformations β (y axis) have been found for the two fluxes R00713MM and R01648MM (x axis) [μmol min^{-1} gDW^{-1}] in the mitochondrial FBA model [17].

doi:10.1371/journal.pone.0133825.g006

reaction fluxes are the results of the FBA, simulating the local variability of fluxes. As a result, by perturbing the values of protein abundance through a multi-objective optimization method, desired fluxes can be concurrently increased [25]. In other words, we perform a Protein-Abundance Design through Multi-objective Optimization (PADMO).

Let y_i be the protein abundance of the ith protein, responsible for the ith reaction of the model. In order to map the protein abundance value into a specific condition of the model, we use the following piecewise multiplicative function (Fig S11 in S1 File):

$$f(y_i) = \begin{cases} (1 + |log(y_i)|)^{(y_i-1)/|y_i-1|} & \text{if } y_i \in \mathbb{R}^+ \setminus \{1\} \\ 1 & \text{if } y_i = 1 \end{cases} \tag{1}$$

In the FBA model, we change the minimum and maximum flux of the ith reaction accordingly: $v_i^{min} = V_i^{min}f(y_i)$, $v_i^{max} = V_i^{max}f(y_i)$, where V_i^{min} and V_i^{max} are the minimum and maximum flux of the wild-type configuration of the model. The bounds of the reaction fluxes are multiplied by the logarithmic piecewise multiplicative function (Eq 1) so as to avoid that the genetic algorithm underlying the optimization is driven towards high and unfeasible values of protein abundance. Namely, the upper and lower bound of the flux in the FBA model in a specific condition are equal to the wild-type bounds multiplied by f. In $y_i = 1$, the function has a discontinuity of the third kind, removable by imposing $f(1) = 1$. The logarithm function, in a different way, has already been used in other approaches [26], so as to build new objective functions when the argument is the ratio between gene expression values. Regarding f as a $f(x)$ scaling function, the properties $f(x) \xrightarrow{x \to 0} 0$ and $f(1) = 1$ ensure that GDMO [5] becomes a particular case of PADMO. The steps of PADMO are presented in Fig 1 (red row).

The FBA model can be optimized through a multi-objective evolutionary algorithm. In GDMO [5], each "individual" of the population is represented by a binary variable set representing the knockout strategy of gene sets. Conversely, in PADMO the individuals are arrays of real values, each of which represents the abundance of a protein. Through the function $f(x)$, these abundances have a continuous effect on the FBA model, rather than only an on/off effect on reactions.

The multi-objective optimization can then be performed with OPTBIOCAD. For each generation of the algorithm, we provide the Pareto optimal solutions, in order to evaluate the evolution of the Pareto front. This loop is repeated until the solutions set does not improve, or until an individual with a desired phenotype is achieved. The number of generations and population are parameters chosen by the researcher. Using this approach, each point of the Pareto front is not merely a specific optimal model in the objective space, but also a protein abundance array representing a specific condition in the variable space. In Fig 7 we show the results of PADMO applied to the maximization of ATP and biomass in the mitochondrial FBA model. We run the model looking for the optimal arrays of protein abundances that maximize both concurrently. Since the model has five biomass reactions other than that of ATP, we define as biomass flux the sum of all the five biomass reaction fluxes.

3 Discussion

Mitochondria are the core of cellular metabolism, since they generate ATP molecules. Data from literature demonstrate that mitochondria play a crucial role in neuronal cell survival [27]. Calcium is an important ion inside mitochondria and its concentration is fundamental to regulate functions and acts at several levels during the ATP synthesis. The dysregulation of the mitochondrial Ca^{2+} homeostasis is involved in many pathologies. For example, an accumulation of Ca^{2+} ions in the mitochondrial space can lead to an increased generation of ROS
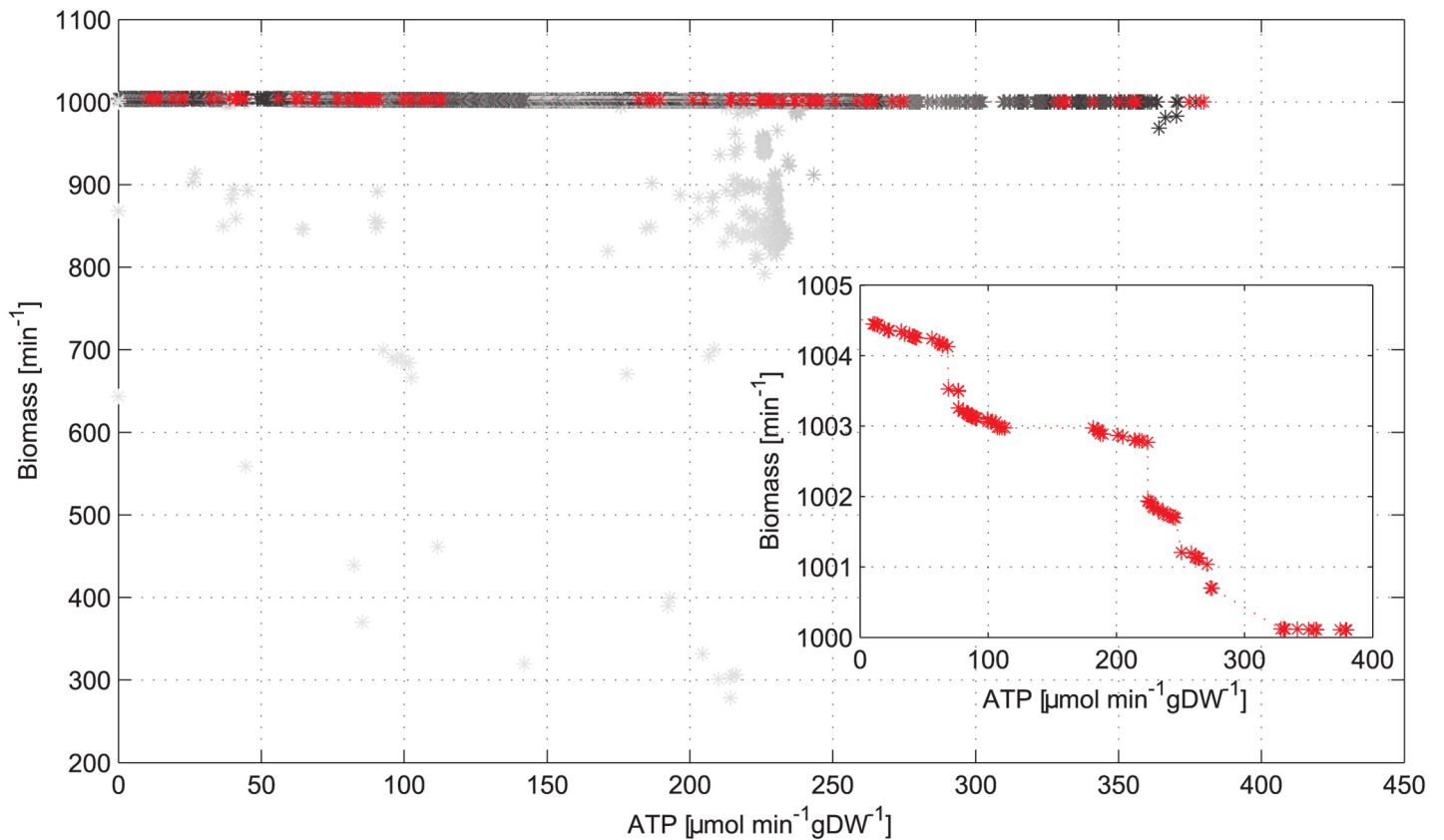
Fig 7. Maximization of ATP and biomass in the mitochondrial FBA model when taking into account the protein abundances as real-valued variables. The red points constitute the Pareto frontier, while the others represent all the feasible points. The points are color coded according to the generation to which they belong. Points from the early generations of the genetic algorithm are colored light grey, while points from the last ones are colored black.

(reactive oxygen species) that alters the permeability of the inner membrane leading the cell to apoptosis. Additionally, the dysregulation of the Ca^{2+} homeostasis is involved in neurodegenerative diseases [11]. ATP metabolism, Ca^{2+} homeostasis, NAD+, NADH and ROS are key players in the cellular mechanisms, and their alteration can lead to the cell death. Diseases like Parkinson [28], Alzheimer [29] and Amyotrophic Lateral Sclerosis [30] have a common point: mitochondrial dysfunction occurs prior to the onset of the pathology symptoms.

In this study, we analyzed how the genetic and the energy-converting pathways of mitochondria can be explored using multi-objective optimization, sensitivity, identifiability and ϵ-dominance. These techniques are framed in a unique pipeline applicable to a variety of conditions and organisms. As well as optimizing simultaneously two or more outputs of a model, the pipeline can also provide interesting insights into clusters of chemical reaction networks, which are often found in the cell and reflect the presence of different pathways with different responses to external or internal perturbations. Indeed, in addition to optimizing input and output fluxes, we defined a strategy to determine the role of the internal fluxes in the mitochondrial FBA network, evaluating the response of the network to changes of their rate.

The multi-objective optimization algorithm we propose in this manuscript, named OPTBIO-CAD, allows for a better coverage of the objective space. More specifically, it has been designed

to generate diversity among the solutions during the convergence process, and to use various perturbation operators to ensure both a local and a global search of optimal solutions. Our pipeline is suitable for comparing the structure of the metabolism through optimization and associated measures for any model taken as a "black-box" input.

For instance, by comparing Fig 2 and Fig S1 in S1 File, the analysis of the Pareto front highlights that the NADH production is flexible in the mitochondrial model by Smith et al., but not in the algal model, which instead allows higher values of ATP production. Furthermore, the algal model is able to better highlight the changes in the metabolism with respect to the wild type condition. On the other hand, the multi-objective optimization of the mitochondrial model finds many Pareto-optimal and sub-optimal points, while there are only few optimal points for each optimization run on the algal metabolism.

The sensitivity analysis has detected the variables playing the major role on the output of the model. Specifically, the PoSA algorithm assesses the sensitivity of each pathway, rather than focusing only on a single reaction or component of the model. Most importantly, PoSA is also able to handle Boolean inputs [5]. Conversely, the identifiability analysis has detected the functional groups of variables. The elements of a functional group are variables functionally related with each other, which cannot be determined unambiguously. In our work, the identifiability analysis is applied to the input variables of each model, considering only the values in the variable space that correspond to the Pareto optimal points in the objective space. The ϵ-dominance analysis is performed to investigate the neighborhood of the suitable Pareto-optimal designs (see S1 File).

The decision variables of our multi-objective optimization algorithm can also be the protein abundance values. Each point of the Pareto front represents a different strain or the same strain responding differently to different sets of environmental conditions. By refining our algorithm with the possibility of modifying real-valued protein abundances rather than binary knockout arrays, we provide a new method to optimize simultaneously two or more outputs of the model by finding the best protein abundance array. Most notably, this may permit to determine the best environmental condition in which an organism has to be grown in order to reach specific optimal output values from a range of objective functions chosen by the researcher. This method is particularly suitable for bacteria [31], when different environmental conditions cause differences in gene expression arrays and therefore in the protein abundances, whose effect can be analyzed in the FBA model using our novel PADMO algorithm. Interestingly, our approach can serve as support for the study of gene expression also in mitochondria. For instance, in studies linking gene expression data and mitochondrial disease [32], our method can be used to predict the effect of changes in gene expression due to the specific disease under investigation. In particular, using our algorithm to implement the genetic and genomic status of mitochondria can help infer biomarkers for diseases diagnoses and prognoses [33].

The interplay between these techniques in our general-purpose framework can be exploited not merely to reach the optimal and suboptimal configuration for a model in a single-objective and multi-objective fashion, but also to conduct tentative analyses on the variables and components of any model including ODEs, DAEs, FBA and GPR mapping. In fact, our framework is also suitable for general-purpose and comparative analysis, enabling to investigate and cross-compare not merely any biological pathway modeled with ordinary differential equations, differential algebraic equations, flux balance analysis and gene-protein-reaction mapping, but also models in which fluxes and gene sets need to be optimized concurrently. Striking applications of our pipeline could be in the field of metabolic engineering of a whole biological network. In particular, we propose a general and automated way to optimize biological models and assess their optimal solutions. In a network of organisms interacting with each other, where each model becomes a submodel of the whole network model, our approach can be used

for analyzing all the network. The results given by the framework applied to each submodel can be easily integrated, being output of the same pipeline, thus allowing a convergence of different modelling techniques [5, 34, 35].

In each optimization procedure we have considered a single organelle, while in the cells there are usually many compartments, each of which contains an organelle. Compartments, also referred to as submodules, may differ for their activity depending on their location in the cell. In a module of interacting organelles, most of the reactions involve more than one compartment. Indeed, any kind of circuit can be split into submodules to increase its efficiency. An appropriate approach towards whole-cell analysis would therefore be to build a Pareto front where each objective belongs to a different compartment [36, 37], linking compartments with a set of delay differential equations (DDEs) to account for events that depend on the state of the system at an earlier time (e.g., diffusion processes or maturation events). In this way, we could envisage our framework in a larger common pipeline to investigate not merely biological circuits, but also human body monitoring techniques, biosensors design, as well as a possible integration with microelectronics, e.g., CMOS biomicrosystems.

4 Methods

4.1 Searching for optimal trade-offs: the Pareto front

Standard optimization routines in computational biology are concerned with the maximization of a single objective, usually the growth rate. However, there is increasing evidence that cells need to optimize multiple, often conflicting, objectives. When two tasks are in contrast with each other, there is no phenotype that can be optimal at all of them. However, multi-objective optimization [6] allows searching for all the trade-off solutions for both tasks.

We exploit the concept of multi-objective Pareto optimality to maximize or minimize two or more desired metabolites in a model, thus obtaining new in silico synthetic strains, which are optimal in many parameters or variables simultaneously.

Let us assume to have r objective functions $f_1, \ldots, f_r$ to optimize. The problem of multi-objective optimization can be formalized as finding a solution x^* that optimizes the vector function

$$f(x) = (f_1(x), \ f_2(x), ..., f_r(x)), \tag{2}$$

where x is the variable in the search space.

The solution of a multi-objective problem is a 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* $f(y^*)$. Formally, if we consider the maximization problem, 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 to maximize in the objective space. Without loss of generality, we have assumed that all the functions have to be maximized (note that minimizing a function f_i can be thought of as maximizing $-f_i$). The objective functions f_i are usually in conflict with each other, and therefore we use the Pareto-front concept to find the set of designs that represent the best trade-off between two or more requirements. From a biological perspective, the Pareto front is the set of all the phenotypes that remain after eliminating all the feasible phenotypes dominated on all tasks [38].

4.2 OptBioCAD: a stochastic general-purpose optimization algorithm

In order to tackle multi-objective optimization problems, we design a novel algorithm, called OPTBIOCAD, which belongs to the class of the well-known evolutionary multi-objective

optimization algorithms (e.g., [39] and [40]). Each candidate solution is a vector of n values, where n is the dimension of the problem, and is assigned an age τ, initially set to 0 [40, 41]. An initial population $P^{(0)}$ of dimension d is randomly generated, and each variable is defined in the range where the search is performed. The copying phase is responsible for the production of copies of the candidate solutions. Each member of the population is copied *dup* times, thus producing a population P_{cop} of size $d \times dup$, where each copied candidate solution takes the same age of its parent; simultaneously, the age of the parent increments by one unit. Once the P_{cop} population is created, it undergoes the mutation phase in order to find better solutions; in this phase, two operators, called *local search* and *global search*, are applied to each candidate solution. Firstly, the local search operator mutates a randomly chosen variable x_i of a given candidate solution using a *self-adaptive Gaussian mutation* computed as $x_i^{new} = x_i + \sigma_i^{new} N(0,1)$, where $\sigma_i^{new} = \sigma_i e^{\gamma N(0,1) + \gamma' N_i(0,1)}$, with initial conditions $\sigma_i = 0.4(x_i^{max} - x_i^{min})/\sqrt{n}$, $\gamma = 1/\sqrt{2n}$, $\gamma' = 1/\sqrt{2\sqrt{n}}$, n being the number of decision variables.

Successively, the global search operator applies a convex perturbation to a given solution by setting $x_i^{new} = (1 - \gamma_i)x_i + \gamma_i x_k$, where x_k is a variable randomly chosen such that $x_i \neq x_k$, with $\gamma_i = N(0,1)$. These local and global search operators are controlled by a specific mutation rate α; for the local search, we define $\alpha = e^{-\rho F}$, while for the global search operator we adopted $\alpha = \frac{1}{\beta} e^{-F}$, where F is the fitness function value normalized in $[0,1]$, while ρ and β are parameters. The fitness function is defined as $F = \sum_{i=1}^{r} f_i/f_{max_i}$ where f is the vector of the r objective functions and f_{max} is the vector of the r best values (of all the populations), one for each objective function. F is normalized dividing by r.

The two operators are applied sequentially; the local search operator acts on the P_{cop} producing a new population P_{LS}. The global search operator mutates P_{LS} generating the P_{GS} population. After the local search operator, the population P_{GS} is evaluated; if a candidate solution achieves a better value of objective function, its age is set to 0, otherwise it is increased by one. Then, the *diversity enforcing* operator is applied to $P^{(t)}$ and P_{GS}; it deletes candidate solutions with an age greater than $\tau_B + 1$, where τ_B is a parameter of the algorithm. The deleted candidate solutions are saved into the archive BC_{arch}. Since the archive contains at most s_a solutions, if there is enough space, the candidate solution is put into the first available location, otherwise it is put in a random location. Finally, the selection is performed and the new population $P^{(t+1)}$ is created by picking the best individuals from the parents and the mutated candidate solutions. However, if $|P^{(t+1)}| < d$, $d - |P^{(t+1)}|$ candidate solutions are randomly picked from the archive and added to the new population.

In many real world applications, it is common to deal with constraints, which could be imposed on input and output values [42]. In general, a constraint is a function $g(x)$ that certificates if a solution for a given optimization problem is *feasible* or not. We consider constraints defined as $g(x): R^n \to R$ if $g(x) \leq \theta$, where θ is a *feasibility* threshold. The algorithm considers the constraint values during the selection procedure. Given two individuals p_1, p_2, if both are feasible, then the individual with the lowest objective function value is picked; if p_1 is feasible and p_2 is unfeasible, p_1 is chosen, otherwise if p_1 and p_2 are unfeasible the individual with the lowest constraints violation is selected.

In section 1, all the simulations are performed with the following algorithm parameters: $d = 20$, $dup = 2$, $\tau_B = 50$, $\rho = 1$, $\beta = 7$ and $s_a = 160$. Additionally, the model (input of OPTBIO-CAD) can be a FBA, FBA-GPR, ODEs or DAEs model. For each model a vector $\vec{mp}$ of parameters is defined and its elements are used as decision variables. The output of the algorithm will be $P^{(t)}$, where $t = t_{final}$, and will contain the optimal designs, $\vec{mp}_i^*$, $i = 1, \ldots, d$. Moreover, s_a additional solutions will be contained in the BC_{arch} archive and in $P^{(t)}$, where $t = 0, 1, \ldots, t_{final}$.

In section 1 and 3, the results come from these two last sets. The pseudo-code of OptBioCAD is reported in Table 1. The choice of the biomass production as a goal to optimize (e.g., allowing the maximal growth of the organism), is common for FBA models [43].

4.3 ϵ-dominance analysis

The ϵ-dominance analysis, inspired by Laumanns et al. [44], is a technique that improves the diversity of the solutions and the convergence of the optimization algorithm. In Section 1, we highlighted that a point y^* in the solution space is said to be Pareto optimal if there does not exist a point y such that $f_i(y) > f_i(y^*), \forall i = 1, \ldots, r$, where f is the vector of r objective functions to optimize in the objective space. The ϵ-dominance technique applies a "relaxed" condition of dominance. That is, a point y^* in the solution space is said to be ϵ-non-dominated if there does not exist a point y such that $f(y)$ dominates $f(y^*)$ of a value higher than ϵ. Formally, y^* is said to be ϵ-non-dominated if $\nexists\ y$ s.t. $f_i(y) \geq f_i(y^*) + \epsilon_i, \epsilon_i > 0, \forall i = 1, \ldots, r$. This "relaxed" condition captures both the "ϵ-non-dominated" solutions and the non-dominated ones (Pareto-optimal), since a Pareto-optimal solution is also ϵ-non-dominated, but the converse does not hold. This technique allows suboptimal solutions to remain in the population, therefore increasing diversity and facilitating the search for multi-modal solutions in single objective optimization problems.

In our work, we use this method to seek solutions that may have been discarded because they are dominated by a small amount ϵ that, for our purposes, can be considered negligible. After the optimization, we perform an ϵ-dominance analysis to search accurately near the edge of the Pareto-optimal region. Formally, let f be the array of the r objective functions, and suppose that all the objective functions are positive and must be maximized. Let $\epsilon > 0$ be the tolerance of our relaxed condition. We seek all points (solutions) y^* belonging to the set $\{y^* : f_i(y^*) + \epsilon_i \geq f_i(y), \forall i = 1, \ldots, r\}$, where f is the vector of the r objective functions, and y represents the non-dominated points. This set will contain both the new "ϵ-non-dominated" points and the

Table 1. OPTBioCAD Pseudo-code.

1: **optBioCAD** (*model, d, dup, τ_B, ρ, β, s_a*) /* the model can be FBA, FBA-GPR, ODEs or DAEs */
2: $t \leftarrow 0$;
3: $BC_{arch} \leftarrow Create\_Archive(s_a)$;
4: $P^{(t)} \leftarrow Initialize(d)$;
5: $Evaluate(P^{(t)}, model)$;
6: $EvaluateConstraints(P^{(t)}, model)$;
7: **while** ¬$Stop\_Condition(t)$ **do**
8: $\quad P_{cop} \leftarrow Copying(P^{(t)}, dup)$;
9: $\quad P_{LS} \leftarrow Local\_Search\_Operator(P_{cop}, \rho)$;
10: $\quad P_{GS} \leftarrow Global\_Search\_Operator(P_{LS}, \beta)$;
11: $\quad Evaluate(P_{GS}, model)$;
12: $\quad EvaluateConstraints(P_{GS}, model)$;
13: $\quad Diversity\_Enforcing(P^{(t)}, P_{GS}, \tau_B)$;
14: $\quad BC_{arch} \leftarrow (BC_{arch} \cup P^{(t)} \cup P_{GS})$;
15: $\quad P^{(t+1)} \leftarrow Selection(P^{(t)}, P_{GS}, BC_{arch})$;
16: $\quad t \leftarrow t + 1$;
17: **end while**
18: **return** $(P^{(t)})$; /* output the best d candidate solutions (metabolic networks) */

doi:10.1371/journal.pone.0133825.t001

old non-dominated ones. The "ϵ-non-dominated" solutions can be considered suboptimal solution because they are close to the Pareto-optimal region. To better understand this analysis, we report some examples in Section 2 in S1 File.

4.4 Identifiability analysis

A biological model is made up of many components (e.g., parameters, variables) estimated through fitting to experiments. The Identifiability Analysis (IA) seeks the functional relations underlying the components of a given system, and can be used after the multi-objective optimization. Coupled with the sensitivity analysis, it gives insight into the model under investigation.

A component is said to be *non-identifiable* if there is no unique solution for its estimation. The non-identifiability can be (i) *structural*, when there are relations among components and therefore they cannot be determined unambiguously; (ii) *practical*, when the low amount or quality of data available does not allow achieving a good estimate for the component. From the definitions, it follows that if a model is structurally non-identifiable, it is also practically non-identifiable. Conversely, the structural identifiability does not necessarily imply the practical identifiability. Using repeated fitting to data and estimations of components, the IA is aimed at finding the structural non-identifiable components of a model, providing hints for simplifying the model and thus avoiding redundancy, or indicating where new experimental measures are needed to ensure the identifiability of the model.

Specifically, let m be the number of decision variables $\{x_1, \ldots, x_m\}$ of the model, which are related by unknown linear or non-linear functional relations. Let n be the number of estimates available for each variable. These estimates are usually organized into a matrix $K = [v_1, \ldots, v_m] \in \mathbb{R}^{n \times m}$, where each column array $v_i \in \mathbb{R}^n$ consists of the n estimates for the ith variable.

With the aim of detecting relations among flux rates, we assign a variable x_i to each flux rate, $i = 1, \ldots, m$. Let us denote by α and β_j the true transformations that linearize the relations among variables:

$$\alpha(x_i) = \sum_{j \neq i}^{m} \beta_j(x_j) + \xi, \tag{3}$$

where ξ is a Gaussian noise. The alternating conditional expectation (ACE) algorithm [45] estimates the optimal transformations $\hat{\alpha}(x_i)$ and $\hat{\beta}_j(x_j), j \neq i$, such that

$$\hat{\alpha}(x_i) = \sum_{j \neq i}^{m} \hat{\beta}_j(x_j), \tag{4}$$

where x_i is the response, while all the other variables are the predictors. Starting from an initial guess of $\alpha(x_i) = x_i/\|x_i\|$ and $\beta_j(x_j) = 0, j \neq i$, ACE iteratively estimates α and $\{\beta_j\}_{j \neq i}$ based on the minimization of the square residuals of (Eq 4).

As a post processing step of the multi-objective optimization ATP-NADH of the FBA mitochondrial model [17], we use the IA to find all the relations among reaction fluxes. The mitochondrial FBA model is composed of 423 reactions, of which $p = 73$ are the "input fluxes", i.e., reactions transporting metabolites from the external environment into the mitochondrion; 135 are the matrix reactions, i.e., all those taking place in the matrix compartment, e.g., Krebs cycle and beta-oxidation; all the remaining reactions take place in the inner membrane space.

We take into account 135 decision variables, namely the 135 fluxes of the matrix reactions in the 2000 mitochondrial conditions obtained with 2000 different values of fumarate. To perform the IA, we infer functional relations with the ACE algorithm within the method proposed by Hengl et al. [46]. In this way, we seek structural identifiability affecting decision variables.

Each row of he matrix K, in the standard ACE approach, is obtained by estimating parameters of a system. Specifically, a row contains the estimated parameters, obtained by minimizing the square residuals between experimental data and model predictions, but with a different initial guess of the parameters in each row. In our case, this is replaced by repeating the FBA analysis for each fumarate value, thus taking into account all the configurations found by the FBA run with different fumarase flux. Therefore, the key result of our approach is the possibility to highlight stage-specific functional relations among reactions in the three monogenic diseases. Since the fitting matrix K consists of the 135 columns in V associated with matrix fluxes (see Section 3 for the definition of V), we reduce the problem of finding structural non-identifiability to the problem of detecting groups of interdependent fluxes in K. We note that the identifiability analysis depends on the constraints, since a non-identifiable constraint involving decision variables is a functional relation between them. In our approach, the constraint is detected by performing 2000 estimates of the 135 variables (matrix fluxes), with the aim of characterizing three mitochondrial diseases by looking at the functional relations among the fluxes in the model. We use the Mean Optimal Transformation Approach (MOTA) [46] and the ACE algorithm, fixing at 10 the maximum number of reactions allowed to enclose a functional relation.

In S1 File we report the tables summarizing the results of the IA applied to the three monogenic diseases. There are fluxes showing functional relations in a group of more than two elements, although in this case strong relations among variables are less likely. We remark that a variable is detected in a group depending on the contribution strength of a predictor to the response. Each variable is considered once as response variable. As a result, if a functional relation is among k variables, it is tested k times, although it is unlikely that the same functional group is really detected all the k times. The r^2 column indicates how much variance of the response can be explained by the predictors. A high amount of explained variance of the response indicates a significant effect of the fixation of the predictors on the standard deviations of the response. The $cv(x) = std(x)/mean(x)$ helps to distinguish practical identifiable from non-identifiable variables [46]. In case of practical non-identifiability, the choice of the value that needs to be fixed strongly depends on the experiments, also considering reference values in the literature.

4.5 Sensitivity analysis

The Sensitivity Analysis (SA) is a method that evaluates the importance of the input(s) in a model. The idea is to randomly perturb the input(s) of a model in order to obtain a distribution of elementary effects (on the output) due to the perturbations of each input. An elementary effect is calculated by comparing (subtracting) the output(s) of the model when the input is perturbed, with the output(s) of the model without perturbation. Plotting the average and the standard deviation of the distribution of elementary effects gives an idea of how the perturbation of the input affects the output(s) of the model. If the mean of the distribution is large, the input has an important overall influence on the output(s) [47]. If the distribution has a large standard deviation, the input has a high influence depending on the values of the other inputs.

The SA is frequently used for the in-silico design of electronic devices, and in the last decade it has been widely used in bioengineering. In electronic design automation, the input(s) of the model can be gain and tension, or resistance and conductance. Conversely, in a biological model the input(s) of the model can be: (i) nutrients of a cell, for instance the uptake rate of glucose or oxygen; (ii) gene knockouts in the genome of a bacterium, for instance the knockout of the pyruvate dehydrogenase complex; (iii) enzyme concentrations in a metabolic pathway, for instance the concentration of RuBisCO in a plant cell. According to the modelling technique and the parameters included in the model, the SA provides a ranking of selected input
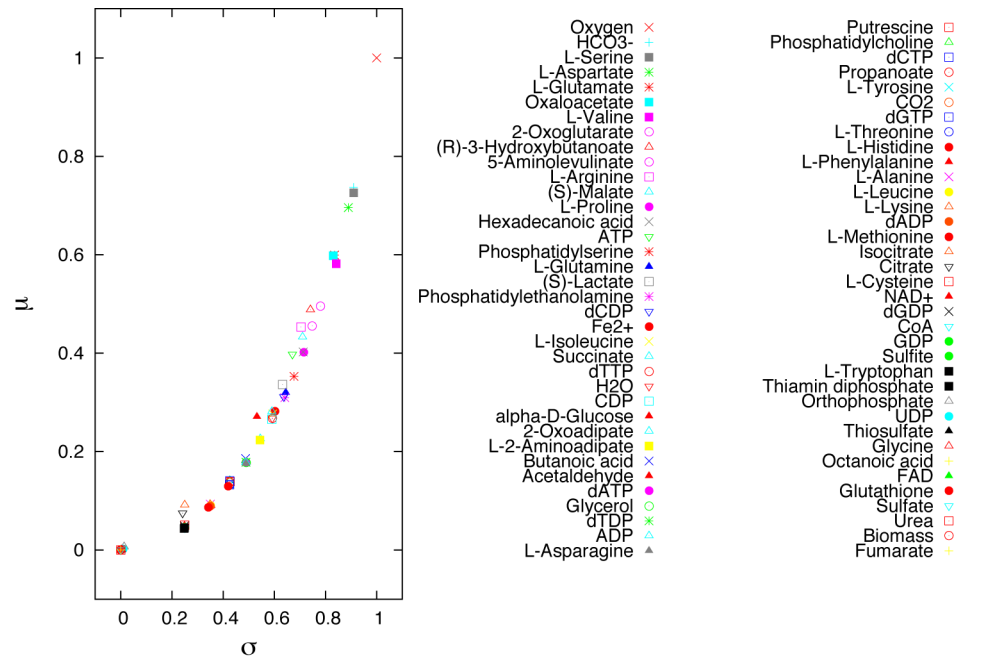
Fig 8. Sensitivity analysis on the mitochondrial FBA model. The plot shows the mean and the standard deviation of the elementary effects computed through the Morris' method applied on the upper bounds of the exchange reaction fluxes. The uptake rates are ranked according to their relative influence on the ATP production. The mitochondrial ATP production is highly sensitive to changes in the uptake rate of oxygen, HCO_3, L-serine, and L-aspartate.

(s), evaluating their importance. Our BIOCAD framework includes two sensitivity methods: Morris' [47] and Sobol's [48] methods.

In Fig 8, we show the result of the Morris method considering as input the uptake flux rates in the genome-scale metabolic network of the mitochondrion [17] discussed in Section 2. In the plot we report the mean and the standard deviation of the distribution of elementary effects for each input. The legend in the figure ranks the inputs from the most to the less important. These results were confirmed by the local robustness analysis [7], performed through perturbation of one input at a time. Finally, a pathway-oriented SA is reported in Fig S3 in S1 File for the model of *Chlamydomonas reinhardtii* [18] discussed in Section 1.

Supporting Information

S1 File. Supporting Figs. 1–32 and Tables 1–11.
(PDF)

S1 Table. Best solutions found after the optimization of the FBA algal mitochondria with 10 knockouts permitted.
(XLSX)

S2 Table. Best solutions found after the optimization of the FBA algal mitochondria with 50 knockouts permitted.
(XLSX)

S1 Code. BıoCAD source code.
(ZIP)

Acknowledgments

This work is supported by the EU MIMOmics project.

Author Contributions

Conceived and designed the experiments: GN. Performed the experiments: CA JC GC. Analyzed the data: CA JC GC PL GN. Contributed reagents/materials/analysis tools: GN PL. Wrote the paper: CA JC. Wrote the source code: CA JC.

References

1. Ingalls BP. Mathematical modeling in systems biology: an introduction. MIT Press; 2013.

2. Hecker M, Lambeck S, Toepfer S, Van Someren E, Guthke R. Gene regulatory network inference: data integration in dynamic models? A review. Biosystems. 2009; 96(1):86–103. doi: 10.1016/j.biosystems.2008.12.004 PMID: 19150482

3. Palsson BØ. Systems Biology: Constraint-Based Reconstruction and Analysis. Cambridge University Press; 2015.

4. Kirchmair J, Göller AH, Lang D, Kunze J, Testa B, Wilson ID, et al. Predicting drug metabolism: experiment and/or computation? Nature Reviews Drug Discovery. 2015;. doi: 10.1038/nrd4581 PMID: 25907346

5. Costanza J, Carapezza G, Angione C, Lió P, Nicosia G. Robust Design of Microbial Strains. Bioinformatics. 2012; 28(23):3097–3104. doi: 10.1093/bioinformatics/bts590 PMID: 23044547

6. Cutello V, Narzisi G, Nicosia G. A multi-objective evolutionary approach to the protein structure prediction problem. Journal of The Royal Society Interface. 2006; 3(6):139–151. doi: 10.1098/rsif.2005.0083

7. Stracquadanio G, Nicosia G. Computational energy-based redesign of robust proteins. Computers & chemical engineering. 2011; 35(3):464–473. doi: 10.1016/j.compchemeng.2010.04.005

8. Angione C, Costanza J, Carapezza G, Lió P, Nicosia G. A Design Automation Framework for Computational Bioenergetics in Biological Networks. Mol BioSyst. 2013;9:2554–2564. doi: 10.1039/c3mb25558a PMID: 23925151

9. Keasling JD. Synthetic biology and the development of tools for metabolic engineering. Metabolic engineering. 2012; 14(3):189–195. doi: 10.1016/j.ymben.2012.01.004 PMID: 22314049

10. Brookes PS, Yoon Y, Robotham JL, Anders M, Sheu SS. Calcium, ATP, and ROS: a mitochondrial love-hate triangle. American Journal of Physiology-Cell Physiology. 2004; 287(4):C817–C833. doi: 10.1152/ajpcell.00139.2004 PMID: 15355853

11. Flint BM. Mitochondria in Neurodegeneration: Bioenergetic Function in Cell Life and Death. Journal of Cerebral Blood Flow & Metabolism. 1999; 19:231–245.

12. Wang C, Youle RJ. The role of mitochondria in apoptosis. Annual review of genetics. 2009; 43:95–118. doi: 10.1146/annurev-genet-102108-134850 PMID: 19659442

13. Galluzzi L, Kepp O, Trojel-Hansen C, Kroemer G. Mitochondrial Control of Cellular Life, Stress, and Death. Circulation research. 2012; 111(9):1198–1207. doi: 10.1161/CIRCRESAHA.112.268946 PMID: 23065343

14. Schapira A. Mitochondrial disease. Lancet. 2006; 368(9529):70–82. doi: 10.1016/S0140-6736(06)68970-8 PMID: 16815381

15. Lesnefsky EJ, Moghaddas S, Tandler B, Kerner J, Hoppel CL. Mitochondrial Dysfunction in Cardiac Disease: Ischemia–Reperfusion, Aging, and Heart Failure. Journal of Molecular and Cellular Cardiology. 2001; 33(6):1065–1089. doi: 10.1006/jmcc.2001.1378 PMID: 11444914

16. Sengupta U, Ukil S, Dimitrova N, Agrawal S. Expression-based network biology identifies alteration in key regulatory pathways of type 2 diabetes and associated risk/complications. PloS one. 2009; 4(12): e8100. doi: 10.1371/journal.pone.0008100 PMID: 19997558

17. Smith AC, Robinson AJ. A metabolic model of the mitochondrion and its use in modelling diseases of the tricarboxylic acid cycle. BMC systems biology. 2011; 5(1):102. doi: 10.1186/1752-0509-5-102 PMID: 21714867

18. Chang RL, Ghamsari L, Manichaikul A, Hom EFY, Balaji S, Fu W, et al. Metabolic network reconstruction of Chlamydomonas offers insight into light-driven algal metabolism. Molecular systems biology. 2011; 7(1):518. doi: 10.1038/msb.2011.52 PMID: 21811229

19. Orth JD, Thiele I, Palsson BO. What is flux balance analysis? Nature Biotechnology. 2010; 28(3):245–248. doi: 10.1038/nbt.1614 PMID: 20212490

20. Burgard AP, Pharkya P, Maranas CD. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnology and bioengineering. 2003; 84(6):647–657. doi: 10.1002/bit.10803 PMID: 14595777

21. Lun DS, Rockwell G, Guido NJ, Baym M, Kelner JA, Berger B, et al. Large-scale identification of genetic design strategies using local search. molecular systems biology. 2009; 5(1). doi: 10.1038/msb.2009.57 PMID: 19690565

22. Stanley WC, Recchia FA, Lopaschuk GD. Myocardial substrate metabolism in the normal and failing heart. Physiological reviews. 2005; 85(3):1093–1129. doi: 10.1152/physrev.00006.2004 PMID: 15987803

23. Sherry A, Zhao P, Wiethoff A, Jeffrey F, Malloy C. Effects of aminooxyacetate on glutamate compartmentation and TCA cycle kinetics in rat hearts. American Journal of Physiology-Heart and Circulatory Physiology. 1998; 274(2):H591–H599.

24. Nuutila P, Koivisto V, Knuuti J, Ruotsalainen U, Teräs M, Haaparanta M, et al. Glucose-free fatty acid cycle operates in human heart and skeletal muscle in vivo. Journal of Clinical Investigation. 1992; 89(6):1767. doi: 10.1172/JCI115780 PMID: 1601987

25. Tyo KEJ, Fischer CR, Simeon F, Stephanopoulos G. Analysis of polyhydroxybutyrate flux limitations by systematic genetic and metabolic perturbations. Metabolic engineering. 2010; 12(3):187–195. doi: 10.1016/j.ymben.2009.10.005 PMID: 19879956

26. Navid A, Almaas E. Genome-level transcription data of Yersinia pestis analyzed with a New metabolic constraint-based approach. BMC Systems Biology. 2012; 6(1):150. doi: 10.1186/1752-0509-6-150 PMID: 23216785

27. Borthwick GM, Johnson MA, Ince PG, Shaw PJ, Turnbull DM. Mitochondrial enzyme activity in amyotrophic lateral sclerosis: implications for the role of mitochondria in neuronal cell death. Annals of neurology. 1999; 46(5):787–790. doi: 10.1002/1531-8249(199911)46:5%3C787::AID-ANA17%3E3.0.CO;2-8 PMID: 10553999

28. Esteves AR, Arduino DM, Silva DFF, Oliveira CR, Cardoso SM. Mitochondrial dysfunction: the road to alpha-synuclein oligomerization in PD. Parkinson's disease. 2011;. doi: 10.4061/2011/693761

29. Supnet C, Bezprozvanny I. The dysregulation of intracellular calcium in Alzheimer disease. Cell Calcium. 2010; 47(2):183–189. doi: 10.1016/j.ceca.2009.12.014 PMID: 20080301

30. Shi P, Gal J, Kwinter D, Liu X, Zhu H. Mitochondrial Dysfunction in Amyotrophic Later Sclerosis. Biochim Biophys Acta. 2010 January; 1802(1).

31. Brandes A, Lun DS, Ip K, Zucker J, Colijn C, Weiner B, et al. Inferring Carbon Sources from Gene Expression Profiles Using Metabolic Flux Models. PLoS ONE. 2012; 7(5). doi: 10.1371/journal.pone.0036947

32. Katic M, Kennedy AR, Leykin I, Norris A, McGettrick A, Gesta S, et al. Mitochondrial gene expression and increased oxidative metabolism: role in increased lifespan of fat-specific insulin receptor knock-out mice. Aging cell. 2007; 6(6):827–839. doi: 10.1111/j.1474-9726.2007.00346.x PMID: 18001293

33. Johnston DS, Su YA, Alesci S. Mitochondrial gene profiling: translational perspectives. Pharmacogenomics. 2009; 10(10):1645–1655. doi: 10.2217/pgs.09.112 PMID: 19842937

34. Dumas ME. Metabolome 2.0: quantitative genetics and network biology of metabolic phenotypes. Molecular BioSystems. 2012;. doi: 10.1039/c2mb25167a

35. Nicosia G, Stracquadanio G. Generalized pattern search algorithm for peptide structure prediction. Biophysical journal. 2008; 95(10):4988–4999. doi: 10.1529/biophysj.107.124016 PMID: 18487293

36. Carapezza G, Umeton R, Costanza J, Angione C, Stracquadanio G, Papini A, et al. Efficient Behavior of Photosynthetic Organelles via Pareto Optimality, Identifiability and Sensitivity Analysis. ACS Synthetic Biology. 2013; 2(5):274–288. doi: 10.1021/sb300102k PMID: 23654280

37. Angione C, Carapezza G, Costanza J, Lio P, Nicosia G. Pareto optimality in organelle energy metabolism analysis. Computational Biology and Bioinformatics, IEEE/ACM Transactions on. 2013; 10(4):1032–1044. doi: 10.1109/TCBB.2013.95

38. Shoval O, Sheftel H, Shinar G, Hart Y, Ramote O, Mayo A, et al. Evolutionary Trade-Offs, Pareto Optimality, and the Geometry of Phenotype Space. Science. 2012; 336(6085):1157–1160. doi: 10.1126/science.1217405 PMID: 22539553

39. Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation. 2002; 6(2):182–197. doi: 10.1109/4235.996017

40. Cutello V, Nicosia G, Pavone M, Timmis J. An immune algorithm for protein structure prediction on lattice models. Evolutionary Computation, IEEE Transactions on. 2007; 11(1):101–117. doi: 10.1109/TEVC.2006.880328

41. Cutello V, Nicosia G, Pavone M. An immune algorithm with stochastic aging and kullback entropy for the chromatic number problem. Journal of Combinatorial Optimization. 2007; 14(1):9–33. doi: 10.1007/s10878-006-9036-2

42. Pavone M, Narzisi G, Nicosia G. Clonal selection: an immunological algorithm for global optimization over continuous spaces. Journal of Global Optimization. 2012; 53(4):769–808. doi: 10.1007/s10898-011-9736-8

43. Stephanopoulos G. Metabolic fluxes and metabolic engineering. Metabolic engineering. 1999; 1(1):1–11. doi: 10.1006/mben.1998.0101 PMID: 10935750

44. Laumanns M, Thiele L, Deb K, Zitzler E. Combining convergence and diversity in evolutionary multiobjective optimization. Evol Comput. 2002 Sep; 10(3):263–282. doi: 10.1162/106365602760234108 PMID: 12227996

45. Breiman L, Friedman JH. Estimating optimal transformations for multiple regression and correlation. Journal of the American Statistical Association. 1985; 80(391):580–598. doi: 10.2307/2288477

46. Hengl S, Kreutz C, Timmer J, Maiwald T. Data-based identifiability analysis of non-linear dynamical models. Bioinformatics. 2007; 23(19):2612–2618. doi: 10.1093/bioinformatics/btm382 PMID: 17660526

47. Morris MD. Factorial sampling plans for preliminary computational experiments. Technometrics. 1991; 33(2):161–174. doi: 10.1080/00401706.1991.10484804

48. Sobol I. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation. 2001 Feb; 55(1–3):271–280. doi: 10.1016/S0378-4754(00)00270-6