Multi-objective Optimisation, Sensitivity and Robustness Analysis in FBA Modelling

Jole Costanza[1], Giovanni Carapezza[1], Claudio Angione[2], Pietro Liò[2], and Giuseppe Nicosia[1]

[1]Department of Mathematics and Computer Science, University of Catania
[2]Computer Laboratory, University of Cambridge
{costanza,carapezza,nicosia}@dmi.unict.it
{claudio.angione,pietro.lio}@cl.cam.ac.uk

Abstract. In this work, we propose a computational framework to design in silico robust bacteria able to overproduce multiple metabolites. To this end, we search the optimal genetic manipulations, in terms of knockout, which also guarantee the growth of the organism. We introduce a multi-objective optimisation algorithm, called Genetic Design through Multi-Objective (GDMO), and test it in several organisms to maximise the production of key intermediate metabolites such as succinate and acetate. We obtain a vast set of Pareto optimal solutions; each of them represents an organism strain. For each solution, we evaluate the fragility by calculating three robustness indexes and by exploring reactions and metabolite interactions. Finally, we perform the Sensitivity Analysis of the metabolic model, which finds the inputs with the highest influence on the outputs of the model. We show that our methodology provides effective vision of the achievable synthetic strain landscape and a powerful design pipeline.

Keywords: Cell Metabolism, Biological CAD, Sensitive and Fragile Pathways, Genetic Design, Multi-Objective optimisation, Flux balance analysis, Sensitivity and Robustness Analysis.

1 Introduction

Metabolic engineering is central in Biotechnology and has impact also in basic cellular biology. The aim of metabolic engineering is to direct specifically a flux through a metabolic pathway, for instance a product made during the fermentation. To this end, one needs a deep understanding not merely of the genetics of a microorganism, but also of its metabolic capacity (i.e. the amount of all the intermediates). Remarkably, through genetic manipulations (in terms of knockouts) carried out on bacteria, one can overproduce one or more metabolites of interest. A gene knockout is a genetic technique in which one gene in an organism is made inoperative through a base mutation or a deletion. Sometime the inactivation of one gene results in the inactivation of all the downstream genes of the operon. These manipulations are very useful for classical genetic studies

D. Gilbert and M. Heiner (Eds.): CMSB 2012, LNCS 7605, pp. 127–147, 2012.

as well as for modern techniques including functional genomics. Recently, many organisms have been used to analyse their metabolite production potential and to identify the metabolic interventions to produce the metabolite of interest. Indeed, strains have been systematically designed in vivo to overproduce target metabolites such as lycopene [1], ethanol [2], isobutanol [3] and many others.

Metabolic engineering requires mathematical models for accurate metabolic reconstruction of strains, as well as for seeking non-native synthesis pathways. A recent research methodology, called Flux Balance Analysis (FBA) [4], studies biochemical networks, in particular the genome-scale metabolic network reconstructions. These network reconstructions contain all of the known metabolic reactions in an organism and the genes that encode each enzyme. FBA calculates the flow of metabolites through this metabolic network, thereby making it possible to predict the growth rate of an organism or the rate of production of a biotechnologically important metabolite at steady state. Being at steady state, FBA manages large networks very quickly, since it does not require kinetic parameters. This makes it well suited to research on perturbations and genetic manipulations (knockouts) that bacteria might undergo. One of the major advantages of performing computational analysis of stoichiometric models is that the pathways are system proprieties emerging under particular genetic background and nutritional conditions. In other words, the FBA provides better treatment of metabolism than classical biochemistry drawings of metabolic pathways.

By using computational metabolic engineering methods, it is possible to explore the reaction network and search for the genetic interventions to optimise the objectives. By making inoperative the genes, the enzymes that are normally synthesised by those genes are not present anymore in the biological system. In this way, also the corresponding biochemical reactions, normally catalysed by these enzymes, do not occur. Then, the chemical species that constitute the reagents and products of these reactions do not undergo the transformations. The aim is to find the genetic manipulations that change the metabolic process in an organism, in order to increase the flow of desired metabolites, chosen according to biotechnological purposes. Additionally, changing the natural genetic function in an organism may cause the death of the growth cell. Therefore, finding genetic manipulations is a hard problem of search and optimisation.

For all the above reasons and since designing gene knockout in laboratory is very expensive and time-consuming, in the past years a variety of methods has been implemented in order to predict in silico the best knockout strategies that optimise a cellular function of interest. These methods are based on evolutionary algorithms [5], simulated annealing [6], bi-level optimisation framework [7], and mixed-integer linear programming (MILP) [8,9]. All have been tested in FBA organism models, but they require high computational efforts, since the execution times grow exponentially [8,6,5] or linearly [7] as the number of manipulations allowed in the final designs increases. Moreover, cellular metabolism is composed of a large number of reactions, thus the dimension of the solution space is very large and finding genetic manipulations is computationally expensive.

In this work, we present a novel Multi-Objective optimisation algorithm denoted by Genetic Design through Multi-Objective (GDMO), in order to search for the genetic manipulations that optimise multiple cellular functions of interest. Our idea is to use the Pareto optimality to obtain not only a wide range of Pareto optimal solutions, but also the *best trade-off design*. In this context, the multiple biological functions are represented by desired productions, e.g., vitamins, proteins, biofuel, biomass formation, antibodies, electron productivity, or the energetic yield of the organism. For this application, a Pareto solution represents a strain with a particular genetic manipulation (genotype), and that is specialised to overproduce selected metabolites (phenotype), with respect to the wild type (i.e., a strain with genes that are all operative). We test our knockout-based multi-target optimisation on the most recent metabolic data concerning *Escherichia coli*, *Geobacter* [10], *Methanosarcina barkeri* [11], and *Yersinia pestis* [12]. We report that multi-objective optimisation provides more insights than single optimisation on the capability of these organisms to adapt to the simultaneous presence of different conditions and constraints. Furthermore, our method is able to explore effectively the whole space of knockouts. We tested the performance of GDMO by maximising acetate and succinate production rates, and other multiple biological functions in *E. coli, i*AF1260, and comparing it against previous methods.

GDMO is accompanied by a robustness analysis that performs the local, global robustness and the Normalised Feasible Parameter Volume of the genetic manipulation proposed by GDMO. For each strain, we compute the robustness indexes, in order to estimate how robust is a strain obtained by GDMO when it undergoes small perturbations, external (changes in the nutrients) or internal (changes in the metabolism). This way, we are able to choose the most robust strain proposed by GDMO. Finally, the Sensitivity Analysis investigates the species solution space and determines the influence of each specie on the output of the FBA model.

2 Methods

2.1 GDMO: Genetic Design through Multi-Objective Optimisation

GDMO is a combinatorial global search method that finds the genetic manipulation strategies to simultaneously optimise multiple cellular functions (i.e., objective functions) in metabolic networks modelled with Flux Balance Analysis (FBA) and Gene-Protein-Reaction (GPR) map. The simultaneous optimisation of multiple objectives differs from the single-objective optimisation because the solution is not unique when the objectives are in conflict with each other. In a maximisation problem objectives are in conflict when the increment of an objective, causes the decrement of at least another one.

The solution of a multi-objective problem is a potentially infinite set of points, called *non-dominated* solutions or *Pareto front*. In a maximisation problem, a solution is Pareto optimal if there exist no feasible solutions that increase some objective without causing a simultaneous decrease in at least one other objective.

In our problem, the genotype of a bacterium is mathematically represented by a string of bits $y \in \{0,1\}^L$. Each bit in y is a gene set that distinguishes between single and multi-functional enzymes, isozymes, enzyme complexes and enzyme subunits; this way, it captures the complexity and diversity of the biological relationships through a Boolean approach. For example, when the genes of the l-th set are all necessary to catalyse the corresponding reactions (a single gene set can linked to more reactions), genes are related by "AND"; otherwise if it is necessary at least a gene, genes are linked by "OR". When the l-th element of y is set to 1, the corresponding gene set is inoperative. Therefore, y represents the vector of decision variables to be found, in order to obtain the higher values of objective functions, satisfying particular constraints (for instance a maximum number of gene knockouts allowable). 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 $F(y^*)$, where F is the vector of r objective functions. The variable space, (i.e., the domain of y) is defined in a discrete space.

The method we present implements a genetic algorithm inspired by NSGA-II [13] and is composed of 4 key steps. We start with the *initialisation* of the population *Pop* and the computation of the fitness score. The population *Pop* is a set of individuals, i.e., a set of feasible solutions. *Pop* is represented by a $I \times (L + r + 2)$ matrix, where I is the number of individuals, L is the number of the decision variables and r is the number of the objective functions, obtained solving the problem (2). The last two columns are used to store two parameters of the algorithm linked to each individual and useful to evaluate the quality of the solution. Each individual is composed of the proposed knockout strategy $\tilde{y}$ and the corresponding objective function values. Each generation select the individuals that are maximal with respect to the product ordering. The individuals of the initial population can be initialised in different ways: either randomly, assigning present status to all genes or selecting a set of knocked out genes.

Successively, three steps are iteratively carried out. In a *binary tournament selection* process, two individuals are selected at random, and their fitness is compared. The individual with the best fitness is selected as a parent. The algorithm selects a number of parents (i.e. the best individuals) equal to the half of the population. Parents are mutated using a *combinatorial mutation operator* convenient to create an offspring population. Mutation represents a switch, from 0 to 1, or from 1 to 0 for the l-th gene set. The process is randomly executed and for each parent individual ten offspring have been formed and only the best is chosen. Mutation can achieve the maximum knockouts number equal to the parameter C (fixed to 50 by default). A novel population of size *Pop* is formed selecting the best individuals from the parents of the previously generation and the current offspring. The new population undergoes a new round of evaluation. For each generation of the algorithm, Pareto optimal solutions are provided. Finally, a *selection* operator is performed in order to reach the last front.

This cycle is repeated until the solution set does not improve, or until an individual with a desired phenotype is achieved or when the number of generation is bounded out. The number of generations D and individuals of the

population I are parameters chosen by the user. The time-complexity of the genetic algorithm is $O(2DI^r)$, where D is the number of generations, I is the population size and r is the number of the objectives. GDMO finds a set of Pareto optimal solutions (non-dominated solutions) for a combinatorial multi-objective optimisation problem, which is also a NP-complete problem.

Pareto Optimality is very useful for the analysis of metabolism, as reported in the previous works [14,15,16], where authors used multi-objective approaches to evaluate the fluxes distributions and genetic manipulations in metabolic networks. In our work, we remark the usefulness of Pareto optimality and adopt an effective and state-of-the-art algorithm to investigate the knockout space. Additionally, for the first time, we used the ϵ-dominance optimality, to do an accurate search in a neighbourhood of the edge of the Pareto region.

2.2 Pareto ϵ-Dominance

Another analysis that we perform is inspired by the idea described in [17]. They use a condition of approximated dominance for their evolutionary multi-objective algorithm with the aim of improving the diversity of solutions and convergence. We, however, use this idea to perform a post-processing analysis in order to calculate an approximated Pareto front. This calculation is designed to search for new solutions and, in particular, solutions that may have been discarded, but they are dominated by an amount that, for our purposes, can be considered negligible. Therefore, once the optimisation routine has been carried out, all the sampled points are revisited. Then, a new set of solutions is built, called "ϵ-non-dominated" set, by applying a "relaxed" condition of dominance, called ϵ-dominance. Formally, assuming that all the objective functions must be maximised, given $\epsilon > 0$, we seek all the points (solutions) belonging to the set: $\{w : w_z + \epsilon \geq u_z,\ \forall\ z = 1, ..., r\}$. Remarkably, this set contains both the new "ϵ-non-dominated" solutions and the previous non-dominated ones.

2.3 FBA Modelling and the Combinatorial Optimisation Problem

FBA is a modelling framework used for studying biochemical networks and in particular the m metabolites and n reactions that are involved (e.g., their formation and degradation, transport and cellular utilisation). For each metabolite X_i, $i = 1, \ldots, m$ a material balance is $\frac{dX_i}{dt} = \sum_{j=1}^{n} S_{ij} v_j$, where S_{ij} is the stoichiometric coefficient associated with each flux v_j, $j = 1, \ldots, n$. At steady state, $\sum_{j=1}^{n} S_{ij} v_j = 0$ holds. This balance equation can be written in matrix form $Sv = 0$, where S is the stoichiometric matrix of m rows and n columns, and v is the vector of fluxes (metabolic and transport fluxes). The matrix S is not square and $n > m$, so we have a plurality of solutions. Each solution is a flux distribution representing a particular metabolic state, depending on the genotype and the transport fluxes. The FBA approach finds the metabolic state in order to optimise a particular objective function, given as a linear combination of fluxes (e.g., growth rate, ATP production). Consequently, the problem can be formulated as a linear programming problem:

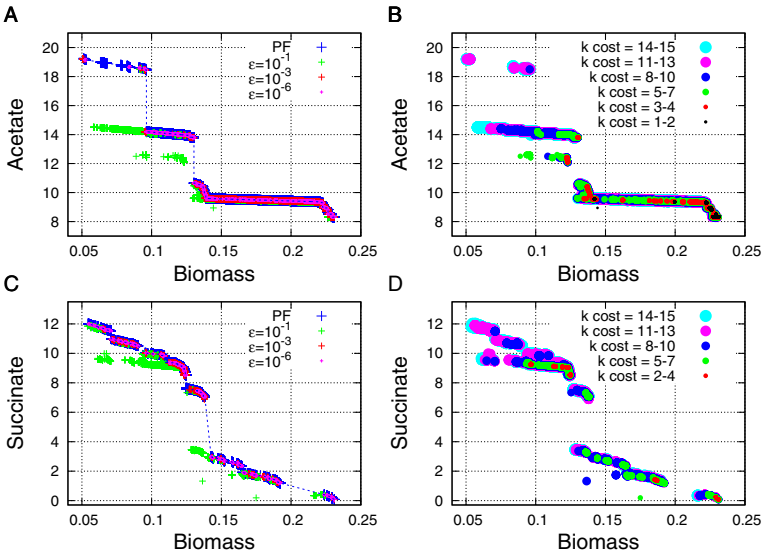

Fig. 1. ϵ-dominance analysis results in *E. coli* network for acetate (A) and succinate (C) multi-objective optimisation. Figures B and D report the knockout cost associated with the solutions reported respectively in Figures A and C, and the dimension of circles reflects the knockout cost associated with the solution point.

$$\begin{aligned} &\text{maximise (or minimise)} && f'v \\ &\text{such that} && Sv = 0 \\ & && v_j^L \leq\ v_j \leq\ v_j^U, j = 1, \ldots, n, \end{aligned} \tag{1}$$

where f is a vector of weights (n dimensional). All the elements in f are either 0 or 1. In our work, f_i is equal to 1 if v_i is the biomass core, but it is possible any combination of fluxes as an objective functions. v_j^L and v_j^U are the lower and upper bound values (thermodynamic constraints) of the generic flux v_j (in our analysis, we consider $v_j^U = 100$ and $v_j^L = -100$ for the fluxes that represent reversible reactions). The output of FBA is a particular distribution of fluxes, denoted by v, which optimises the objective function. Remarkably, FBA does not describe how a certain flux distribution is realised (by kinetics or enzyme regulation), but which flux distribution is optimal for the cell.

We propose the gene-protein-reaction (GPR) mappings to allow our algorithm to work at the genetic level. GPR mappings provide links between each gene and all the reactions v_j depending on it, and define how certain genetic manipulations affect reactions in the metabolic network. For a set of L genetic manipulations, the GPR mappings are represented by a $L \times n$ matrix G, where the *(l,j)*-th element is 1 if the *l*-th genetic manipulation maps onto the reaction j, and is 0 otherwise.

Our approach is based on the technique adopted in OptKnock [8], which finds the fluxes distribution in the metabolic network in order to reproduce the desired productions (synthetic objectives) and achieve the maximal growth.

Unlike Optknock, we are able to optimise more than one objective. The bi-level problem [8] is represented by the following formulation:

$$\begin{array}{lll} \max & g'v & \\ \text{such that} & \sum_{l=1}^{L} y_l \leq C & \\ & y_l \in \{0,1\} & \\ & \max & f'v \\ & \text{such that} & Sv = 0 \\ & & (1-y)'G_j v_j^L \leq \ v_j \leq \ (1-y)'G_j v_j^U, \\ & & j = 1, \ldots, n, \end{array} \tag{2}$$

where g is a vector of weights (n dimensional) associated with the synthetic objectives, and g' is its transpose. For example, when the synthetic objectives v_j and v_h have to be maximised, the weights g_j and g_h are equal to 1. y is the knockout vector (L dimensional). If there are no impaired reactions in the metabolic network, y contains only zeros. Conversely, when $y_l = 1$, the gene set involved in the manipulation l is turned off, and the corresponding reactions are in the absent status (the lower and upper bounds are set to zero, resulting in a modified metabolic network). C is an integer representing the maximum number of knockouts allowed. The bi-level problem can be converted to a MILP problem as described in [8] (for a detailed description, see the original work [8]). We implemented and solved the problem using the GLPK solver.

2.4 Sensitivity Analysis

In modelling, Sensitivity Analysis (SA) is a method used to discover which inputs play a key role on the output of the model. In the last years, scientists used SA indexes in systems biology interrogating the reactions space (RoSA - Reactions oriented Sensitivity Analysis) [18], [19] and species space (SoSA - Species oriented Sensitivity Analysis) to find their influence on the outputs of the system [20]. We perform SA to find the most sensitive inputs in FBA model of *E. coli* using the Matlab SensSB Toolbox [21].

The *E. coli* model analysed in this work contains 2382 fluxes, 299 of which represent exchange fluxes, 2082 represent inner metabolic reactions, and 1 the growth rate or biomass.

The n_{ex} exchange reactions ($n_{ex} < n$) described by the vector $v_{ex} \subset v$, allowing nutrients to enter and leave the system, are unconstrained in the forward direction (v_{ex}^U, upper bound vector), while are constrained in reverse directions (v_{ex}^L, lover bound vector) to zeros when uptake rates is not allowed. Moreover, the *"EX glc"* is an exchange reaction for glucose and has a lower bound of "-10" indicating a potential glucose uptake rate of 10 mmol gDW^{-1} h^{-1}.

We performed the SA method considering as inputs of the model the v_{ex}^L lower bound vector of exchange fluxes. For each of n_{ex} exchange fluxes we varied each

element of v_{ex}^L in the interval [-100, 0] of the region of interest Ω, n_{ex}-dimensional unit hypercube. We adopted the Morris [22] method in order to identify the uptake rates whose tuning results in a major system response. SA is based on the calculation of the *elementary effect* due to the variation of each input. For a given value of v_{ex}^L, we define the elementary effect of the h-th input as:

$$d_h(v_{ex}^L) = \frac{F(v_{ex}^L(1), \ldots, v_{ex}^L(h-1), v_{ex}^L(h) + \Delta, v_{ex}^L(h+1), \ldots, v_{ex}^L(n_{ex})) - F(v_{ex}^L)}{\Delta}. \tag{3}$$

We considered the vector of fluxes v as output $F(v_{ex}^L)$ for *E. coli* model, calculated by solving the problem (1). For each of the n_{ex} exchange fluxes, the attention is restricted to a region of experimentation ω, n_{ex}-dimensional k-level grid, where each v_h^{ex} may take a value from $\omega = \left\{-100, -100\frac{k-2}{k-1}, \ldots, -100\frac{2}{k-1}, -100\frac{1}{k-1}, 0\right\}$. Δ is a predetermined multiple of $\frac{1}{(k-1)}$ and represents the *perturbation* of the input v_h^{ex}. The distribution of elementary effects EE_h for the input v_h^{ex} is obtained by randomly sampling Q points from ω. The estimation of the mean μ^* and standard deviation σ^* of those distributions EE_h will be used as indicator of which inputs should be considered

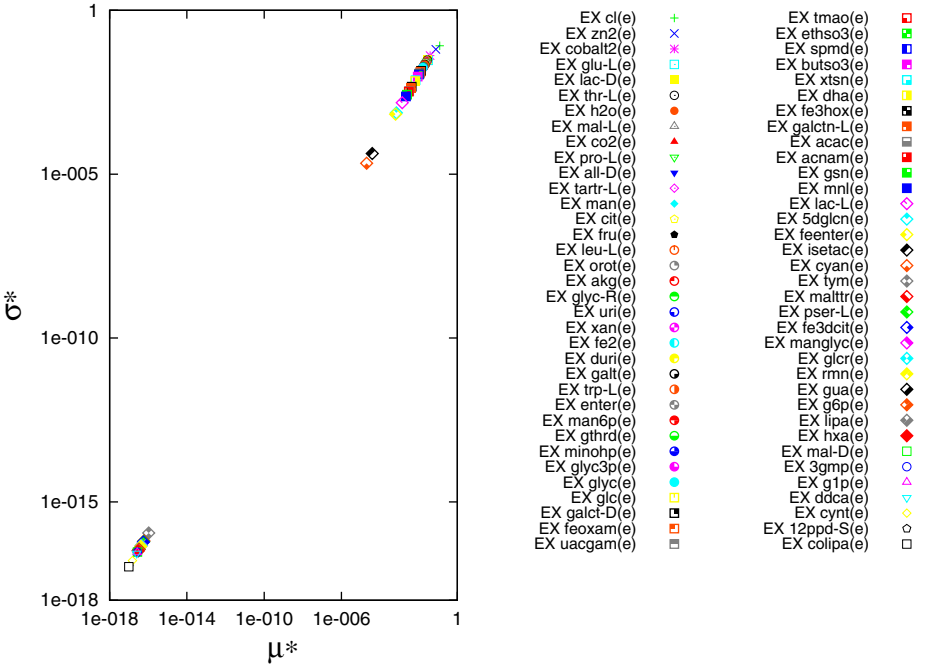

Fig. 2. Uptake Rate-oriented Sensitivity Analysis for the *E. coli* model iAF1260. In this analysis we investigate the input fluxes of the model (299 nutrients) and evaluate their sensitivity with respect to all fluxes of the model. We find that only 70 fluxes (reported in the key) out of 299 are influent, the other ones have sensitivity indexes equal to zero. Results have been obtained averaging over 3000 evaluates of function F.

important. A high μ^* mean indicates an input with an important "overall" influence on the output. A large measure of σ^* variance indicates an input whose influence is highly dependent on the values of the inputs.

2.5 Robustness Analysis

The ability of a system to adapt to perturbations due to internal or external agents, aging, temperature, environmental changes and, in our case, also due to molecular noise and mutation is one of evolutionism guidelines and should also be a fundamental design principle. To optimise the production of a specific metabolite (and simultaneously the formation of biomass, which is necessary to maintain the survival of the bacteria), we used GDMO that obtain a *strain* that maximises the feature required by us. At this point, the validity of the biological strain, designed in-silico, must be tested as regards robustness and sensitivity to endogenous and exogenous perturbations, and this is done by the *robustness analysis*. In this way, we also know the ability of a strain to adapt to small perturbations that can occur at any stage of the biochemical processes within the bacterium, or caused by the environment in which it reproduces. As we shall see, by the term "adaptive capacity" we mean the ability to maintain "acceptable" the performances relative to the metabolite production and biomass formation previously optimised.

There are numerous methods that can be used to fulfil this task. Among these, in [23] the authors consider a big network (in this case, however, considered the Internet network) and use the *theory of percolation on random graphs* to test the robustness of the network in case of random or targeted node deletion, or in case of random link deletion. They associate occupation of nodes or links with their functioning, and for occupation probability they mean the probability of operation of them. They consider that this probability is uniform or depends on the degree of each node (that is the number of connections at that node) distribution. So they analyse the robustness of the network connectivity as the occupation probability is varied. Through this analysis, they highlight that a network with these characteristics is robust to random removal of nodes or links, but not if they are targeted nodes with highest degree. In another work [24], the relationship between the general characteristics of a chemical reaction network and the sensitivity of his equilibrium is investigated according to changes in the overall supply of reagents. The authors define *the sensitivity of a species* as the variation of it with respect to the element concentration one, and they find a lower bound to such sensitivity that depends on the network structure alone. In particular, they argue that a strong robustness of the equilibrium against element variations is likely only if the various species are constructed from building block highly gregarious (i.e. each one binds with many others) or present in some species with high multiplicity. Finally, in [25] the authors use a combined approach of global and local robustness that they call *Glocal Robustness*. The global analysis investigates the parameter space with the aim of finding where a circuit cell shows experimental observed features (global), while the local one determines the robustness of parameter sets sampled during the previous phase.

Similar work making use of the robustness analysis for parameter estimation are also present in [26] and in [27]. In our work, however, we use very simple robustness analysis that shows a high degree of transversality because easily applicable in other fields, as was done in [28] and in [29].

The basic principle of this analysis is as follows. Firstly, we define the perturbation as a function $\tau = \gamma(\Psi, \sigma)$ where γ applies a stochastic noise σ to the system Ψ and generates a trial sample τ. The γ-function is called γ-perturbation. Without loss of generality, we assume that the noise is defined by a random distribution. In order to make statistically meaningful the calculation of robustness, we generate a set T of trial samples τ. Each element τ of the set T is considered robust to the perturbation, due to stochastic noise σ, for a given property (or metric) ϕ if the following condition is verified:

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

where Ψ is the *reference system*, ϕ is a *metric* (or property), τ is a *trial sample* of the set T and ϵ is a *robustness threshold*. The definition of this condition makes no assumptions about the function ϕ. It can be anything (not necessarily related to properties or characteristics of the system); however, it is implicitly assumed that it is quantifiable. The robustness of a system Ψ is the number of robust trials of T, with respect to the property ϕ, over the total number of trials ($|T|$). In formal terms:

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

where Γ is a dimensionless quantity that states, in general, the robustness of a system and, in this case, of a strain.

Robustness index is a function of ϵ, so the choice of this parameter is crucial and not a trivial task. Since we are interested in the behaviour of strain when subjected to small perturbations and because the behaviour is acceptable when the deviations from the original value is as small as possible, we choose the values of epsilon equal to 1% of the metric and sigma equal 1% of the perturbed variable.

Based on this principle, we evaluate two values of robustness, the *Global Robustness* value (GR) and the *Local Robustness* value (LR). Also we evaluated the *Normalised Feasible parameter Volume* ([25]) to give a comparison between these results and the GR/LR values. The first two values only differ in the perturbation kind, in particular, chosen σ, it will differ the set of variables that will be perturbed.

Global Robustness. As regards the Global Robustness of a strain, we perturbed the upper and lower bounds of each metabolic flux. Hence, a trial τ is created by perturbing all the upper v_j^U and lower bounds v_j^L, $j = 1, \ldots, n$ of the metabolic flux. We create a set T_τ of trials, and for each of them we perturb all the bounds and evaluate the property $\phi(\tau)$ (by flux balance analysis), which in our case can be the value of acetate, succinate, biomass or a combination of them; and then, we calculate the function ρ. Once a value of ρ is obtained for

each of the trials, we compute the value of robustness (Equation 5), which in this case we call *Global Robustness* because all the parameters are perturbed.

Local Robustness. In this case, we perturb again the upper v_j^U and lower bounds v_j^L, $j = 1, \ldots, n$, of a metabolic flux, but we create a sample trial perturbing a single flux, we evaluate the property $\phi(\tau)$ and we calculate the function ρ. After creating a set T_τ of trials, we calculate the robustness (Equation 5), which in this case we call *Local Robustness*. Hence, we calculate a LR value for each metabolic flux.

Normalised Feasible parameter Volume. We also implemented the analysis described in [25] to compare the results obtained by GR and LR. In this analysis, the authors implement a procedure that calculates the volume occupied by those parameters such that the system maintains the desired characteristics. The volume is computed in the 2n-dimensional parameter space. In our case, the volume is such that Equation 4 holds. Since this research requires a huge computational effort, given the high number of dimensions (R^{2n}, where 2n is the number of parameters), it is guided by an iterative procedure that involves the Principal Component Analysis (PCA). In the second part, they calculate local coefficients, and from these they derive which parameters are influential on the robustness (by Spearman's partial correlation coefficient).

In particular, the first part requires two steps. The first is a Monte Carlo sampling obtained with 2n-dimensional Gaussian random variations centred around a parameter vector (known in advance). In our case this vector is represented by the 2n parameters: v_j^U and v_j^L. Then a set $T_\tau^{(1)}$, $2n \times K$ is created, that contains K parameter vectors. Among these, only a fraction will satisfy the Equation 4, the set comprising this fraction is the set of the feasible parameter vectors $V^{(1)}$. The second step begins with a principal component analysis of $V^{(1)}$; this analysis allows to identify statistical linear structures within high-dimension data sets. Here, instead, it is used to guide the sampling of the parameter vectors in subsequent iterations. In particular, $T_\tau^{(2)}$ and the subsequent sets $T_\tau^{(h)}$ are generated from $V^{(1)}$ and, in general, from $V^{(h-1)}$, where $h = 1, \ldots, H$ are the iterations number. In particular the generic element $\tau_{j,k}$ of $T_\tau^{(h)}$ is generated as:

$$\tau_{j,k} = \frac{\sum_{t^*=1}^{T^*} V_{j,t^*}^{(h-1)}}{|T^*|} + \lambda^{(h-1)} \cdot \xi_{j,k}, \tag{6}$$

where $j = 1, \ldots, 2n$, since the columns of $T_\tau^{(h)}$ contain the perturbed values of the parameters v_j^U and v_j^L, $j = 1, \ldots, n$; $k = 1, \ldots, K$ is the cardinality of $T_\tau^{(h)}$; the first term, on the right side, is the average of the elements for each perturbed parameter (that is the average for each row) of the set $V^{(h-1)}$ obtained in the previous iteration; $\xi_{j,k}$ is a Gaussian noise with zero mean and standard deviation equals to the $(j,k)^{th}-$element of the covariance matrix $\Sigma^{(h-1)}$, i.e. the pair wise covariance calculated for all vectors τ_a and τ_b of $V^{(h-1)}$ (the eigenvectors of this

matrix are the principal axes of the $V^{(h-1)}$ set by PCA); finally, the real value $\lambda^{(h-1)}$ guides the h^{th} Gaussian process by scaling the standard deviations of the distribution along the PCA directions. The purpose of Equation 6 is to avoid unnecessary sampling in a parameter space region where there are no probably feasible vectors. At the end of this procedure, a hyper-box B is constructed in the parameter space, whose axes are parallel to the PCA axes of the last iteration. The bounds of this box, for each direction, are given by the more extreme elements in the set V^H of the last iteration. Then B is uniformly sampled constructing the final set T_τ; a subset V of T_τ will verify the Equation 4. Finally, the feasible parameter volume will be calculated as $R^{2n} = (|V| \setminus |T_\tau|) * Vol(B)$, where $|.|$ determines the cardinality. The logic of this measure is that as the value of R^{2n} increases as the likelihood that perturbing a parameter vector, another feasible parameter vector is generated increases. Finally, for comparing systems with different number of parameters the *normalised feasible parameter volume* R is defined as $R = \sqrt[2n]{R^{2n}}$. R can be considered as the *permissible average variation per-parameter* that leaves intact the system performance.

The second part of this analysis is connected to the global part. The authors take into account the final set of the feasible parameter vectors V and for each parameter vector produces Q sample trials perturbing the $2n$ parameters by Gaussian noise with zero mean and sigma equal to 0.2; then, they calculate the fraction of robust trials; after that, they repeat the calculations for all vectors. Finally, for the 2n-parameters, they calculate the Spearman partial correlation coefficient with respect to the robust trial fraction values and the different values assumed by the parameters $\delta_j(V(j), X)$, where $j = 1, \ldots, 2n$; $V(j)$ is the j^{th} row of V (containing the observations of the j^{th}-parameter) and X is a vector (containing the values of the robust trial fractions).

3 Results

We tested the performance of GDMO to maximise the production of acetate and succinate in the recent model of *E. coli* K-12 MG1655, iAF1260 [30], and we compared it with GDLS [7], OptFlux [6] and OptGene [5]. In Table 1, we report the productions in wild type, and the results obtained by previous methods and in particular the greater level of acetate and succinate we reach. Pareto and ϵ-optimality present several suitable solutions, which are reported in details in Table 3. Our method reaches interesting results in terms of acetate, succinate, biomass and, mostly, in knockout cost. The knockout cost is defined according to the Boolean relationship between genes. For example, if a gene set is composed by two genes linked by an "AND" relation, the cost to ensure the catalysis of the corresponding reactions is 2, since both genes are necessary to turn on the reactions. Instead, the cost to ensure the turning off of the corresponding reactions (knockout cost) is 1. In our optimisation, indeed, we select as third objective the minimisation of the knockout cost, since in vivo knockout is an expensive and a difficult biotechnological procedure. In all the simulations we initialised the network with an empty set of knockouts, in order to compare our results with GDLS.

Table 1. Comparison between GDMO and previously genetic design methods. We compare OptFlux [6], OptGene [5], GDLS [7], OptKnock [8] and our multi-objective optimisation algorithm (GDMO) for maximising acetate (Ac) and succinate (Suc) production [mmolh^{-1} gdW^{-1}]. The third and fourth rows provide the biomass (Bm) [h^{-1}] and the knockout cost (kc). We report two candidate solutions for acetate optimisation: the first strain, named A_5 (Table 3), provides a low kc equal to 3, and the second one (A_2) reaches an elevated value of acetate, +130.7%, outperforming the previous methods. For succinate production, we obtain +13659% with respect to wild type, deleting only 8 genes (B_3). The last three rows provide the robustness indexes. R values [25] and GR values are global robustness indexes. For LR we report the minimum value found that is associated with the less robust flux (glucose uptake rate). "n.a." stands for *not applicable*.

	Wilde Type	OptFlux	OptGene	GDLS	GDLS	OptKnock	OptKnock	**GDMO**	**GDMO**	**GDMO**
Ac	8.30	15.129	15.138	15.914	n.a.	n.a.	12.565	13.791	19.150	n.a.
		(+82.3%)	(+82.4%)	(+91.7%)	n.a.	n.a.	(+51.4%)	***(+66.13%)***	***(+130.7%)***	n.a.
Suc	0.077	10.007	9.874	n.a.	9.727	9.069	n.a.	n.a.	n.a.	10.610
		(+12877%)	(+12704%)	n.a.	(+12514%)	(+12362%)	n.a.	n.a.	n.a.	***(+13659%)***
Bm	0.23	n.a.	n.a.	0.0500	0.0500	0.1181	0.1165	0.130	0.053	0.087
		n.a.	n.a.	(-78.4%)	(-78.4%)	(-77.9%)	(-49.6%)	(-43.72%)	(-77.10%)	(-62%)
kc	n.a.	n.a.	n.a.	14	26	54	53	***3***	10	***8***
GR	54.76%	n.a.	n.a.	13.76%	16.6%	43.24%	43.08%	45.32%	27.6%	40.40%
LR	54.0%	n.a.	n.a.	16.0%	21.33%	40.0%	40.60%	39.33%	24.0%	46.0%
R	1.30	n.a.	n.a.	1.45	1.45	1.18	1.02	0.78	0.44	1.32

For each solution, we also calculate the Robustness indexes. In particular the Global Robustness index (GR) can be seen as an index to discriminate a strain from each other to choose the best, as regard the robustness with respect to the selected metrics. Moreover, if GR is high, the likelihood of the strain to maintain the performance increases, even if subjected to perturbations. R values indicate the *permissible average variation per-parameter* that leaves intact the system performance. Therefore, also these values can be seen as an index to discriminate a strain from each other. If we consider the GR and R values, we can see a similar behaviour in most cases. *Local Robustness* (LR) index represents the absolute and relative minimum of the results obtained for each strain. Only for the flux related to *D-glucose exchange (Ex glc)*, we obtain LR values less than 100%. Finally, we evaluated Spearman partial correlation coefficients. Since this procedure requires a considerable computational effort, the results of this analysis are calculated for only one strain ($A4$, Table 2). The results indicate that the highest value is $\delta i^* = -0.24$ (the other values are smaller at least by one order of magnitude) that corresponds to the *D-glucose exchange (EX glc)* reaction. It indicates that the Robustness of the strain is more correlated to this reaction. The result is identical to that we obtained with the Local Robustness analysis. Also in this case, the *D-glucose exchange (EX glc)* is the fragile reaction of the strain. Results are shown in Table 2.

The study of genes and reactions of *E. coli* has involved inferring several Pareto trade-offs in anaerobic and aerobic conditions (Figure 3 and Figure 4). The experiments reported in Figure 3 show that GDMO overcomes all the above mentioned methods, and in particular GDLS. The latter performs a single-objective optimisation maximising the synthetic objective function acetate

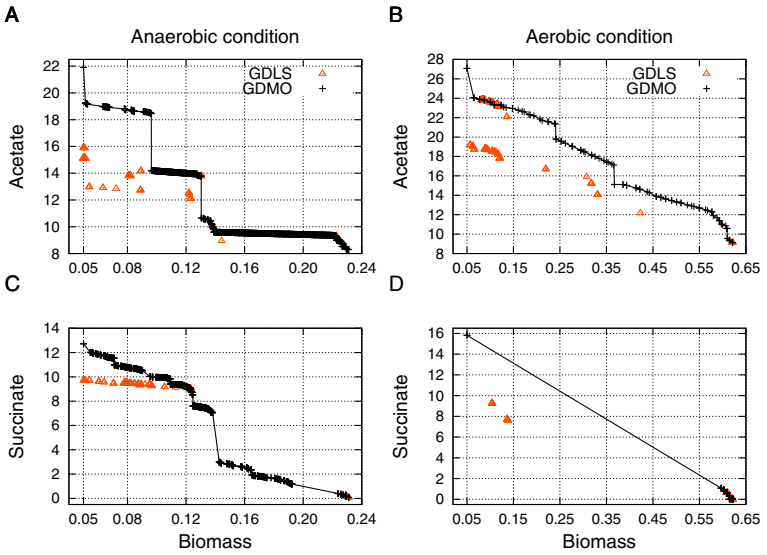

Fig. 3. Maximisation of biomass and acetate production in anaerobic and aerobic conditions (A,B), and maximisation of biomass formation and succinate production in anaerobic and aerobic Condition (C,D), with glucose uptake rate 10 mmolh^{-1} gDW^{-1} in iAF1260. In black the Pareto solutions obtained by GDMO, and in red the optimal results obtained by GDLS [7].

production with knockout cost equal to 14, or optimising succinate production with knockout cost 26, as shown in Table 1 and Figure 3.

The Pareto front strategy is useful to investigate the biological and statistical complexity in several organisms. Figure 5 reports four Pareto curves obtained optimising the acetate/succinate production and the biomass formation in different organisms: the *E. coli* [30], the *Methanosarcina barkeri* [11], *Geobacter sulfurreducens* [10] and *Yersinia pestis* [12]. *M. barkeri* is an archaea able to live in anaerobic condition and produce methane using three known metabolic pathways for methanogenesis. *Geobacter* species are of ecological importance due to bioremediation capabilities. The organism can metabolise uranium, has the ability to generate electricity and can decompose petroleum contaminants in polluted groundwater. For all the organisms, glucose uptake rate is fixed at a maximum of 10 $mmolh^{-1}gDW^{-1}$ as a carbon source (except for *M. barkery*, which does not features the glucose exchange flux in its metabolic network). Some external metabolites (e.g., calcium, ammonia, sulfate, phosphate, oxygen, water, proton, iron (II-III), potassium, sodium, copper, chloride and carbon dioxide) are allowed to both enter and leave the system, while the others are allowed only to leave the system. GDMO highlights the response of different systems and the ability of the organisms to produce the desired metabolite. In the same conditions, *M. barkeri* and *G. sulfurreducens* reach higher levels of acetate than *E. coli*.

Table 2. Robustness analysis results. For each strain we report: the *Global Robustness Value* (GR), the *normalised feasible parameter volume* (R) and the *Local Robustness* (LR) values. For the LR values are shown the minimum associated with the glucose uptake rate. For all other fluxes, we obtained 100% of local robustness.

Strain	**GR**(%)	**R**	**LR**(%)
A_1	28.72	0.39	26.67
A_2	27.60	0.44	24.00
A_3	40.72	1.27	35.33
A_4	41.52	1.74	36.0
A_5	45.32	0.78	39.33
B_1	44.60	0.15	44.67
B_2	43.48	0.92	42.0
B_3	40.40	1.32	46.0
B_4	44.64	1.30	44.0

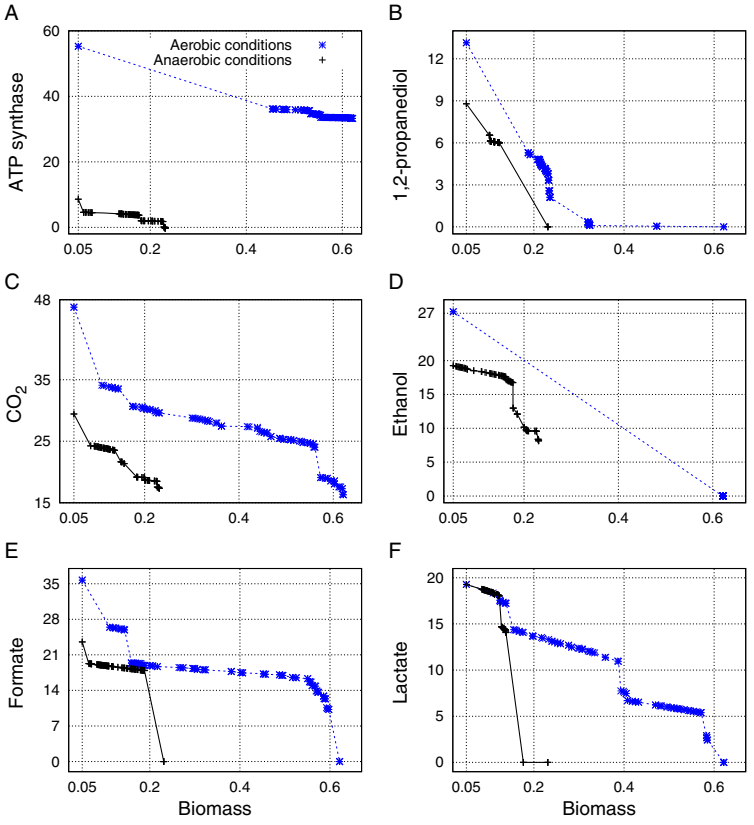

Fig. 4. Pareto fronts for six optimisation problems. We simultaneously maximise biomass formation [h^{-1}] and A) ATP synthase rate, B) 1,2-propanediol, C) CO_2, D) ethanol, E) formate , F) lactate production rates [mmolh^{-1} gDW^{-1}]. In blue we simulate aerobic conditions with O_2=10 mmolh^{-1} gDW^{-1}, in black anaerobic conditions.

Table 3. We report some of the proposed solutions obtained by GDMO to maximise acetate and succinate productions [mmolh^{-1} gDW^{-1}] in *E. coli* network. For each strategy, we report the biomass formation [h^{-1}], the knockout cost (k cost) and the corresponding genes and reactions switched off. The variation of acetate, succinate and biomass in comparison with the wild type is enclosed in brackets.

Strain	Acetate	Biomass	k cost	Knocked out Genes	Deleted Reactions
A_1	19.198	0.052	12	(b0351) OR (b1241)	acetaldehyde dehydrogenase (acetylating)
	(131.26%)	(-77.38%)		(b0910)	cytidylate kinase (CMP)
					cytidylate kinase (dCMP)
				(b2975) OR (b3603)	D-lactate transport via proton symport
					glycolate transport via proton symport, reversible
					L-lactate reversible transport via proton symport
				(b4381)	deoxyribose-phosphate aldolase
				(FdhF and Hyd4) or (FdhF and HycB)*	Formate-hydrogen lyase
				(b0243)	glutamate-5-semialdehyde dehydrogenase
				(b3617)	glycine C-acetyltransferase
				(b0963)	methylglyoxal synthase
				Nuo*	NADH dehydrogenase
A_2	19.150	0.053	10	(b0351) OR (b1241)	acetaldehyde dehydrogenase (acetylating)
	(130.7%)	(-77.10%)		(b3945)	aldose reductase (acetol)
					Glycerol dehydrogenase
					D-Lactaldehyde:NAD+ 1-oxidoreductase
				(b4381)	deoxyribose-phosphate aldolase
				(FdhF and Hyd4) or (FdhF and HycB)*	Formate-hydrogen lyase
				(b3617)	glycine C-acetyltransferase
				(b1380) OR (b2133)	D-lactate dehydrogenase
				(b3236)	malate dehydrogenase
A_3	18.532	0.096	9	(b0351) OR (b1241)	acetaldehyde dehydrogenase (acetylating)
	(123.2%)	(-58.6%)		(b0910)	cytidylate kinase (CMP)
					cytidylate kinase (dCMP)
				(b2975) OR (b3603)	D-lactate transport via proton symport
					glycolate transport via proton symport, reversible
					L-lactate reversible transport via proton symport
				(b4381)	deoxyribose-phosphate aldolase
				(b3617)	glycine C-acetyltransferase
				(b0963)	methylglyoxal synthase
				Nuo	NADH dehydrogenase
A_4	14.046	0.104	5	(b0351) OR (b1241)	acetaldehyde dehydrogenase (acetylating)
	(69.20%)	(-55.14%)		(b3617)	glycine C-acetyltransferase
				(b4025)	glucose-6-phosphate isomerase
				(b3708)	Tryptophanase (L-tryptophan)
A_5	13.791	0.130	3	(b0351) OR (b1241)	acetaldehyde dehydrogenase (acetylating)
	(66.13%)	(-43.72%)		(b1539)	L-allo-threonine dehydrogenase
					D-serine dehydrogenase
					L-serine dehydrogenase
Strain	**Succinate**	**Biomass**	**k cost**	**Knocked out Genes**	**Deleted Reactions**
B_1	12.012	0.055	15	(b0351) OR (b1241)	acetaldehyde dehydrogenase (acetylating)
	(15476%)	(-76.33%)		(b2587)	2-oxoglutarate reversible transport via symport
				(b0870) OR (b2551)	D-alanine transaminase
					alanine transaminase
					L-allo-Threonine Aldolase
					Threonine aldolase
				(b1852)	glucose 6-phosphate dehydrogenase
				(b1849)	GAR transformylase-T
				(b1380) OR (b2133)	D-lactate dehydrogenase
				(b2463)	malic enzyme (NADP)
				(b0963)	methylglyoxal synthase
				(b4388)	phosphoserine phosphatase (L-serine)
				(b2661)	succinate-semialdehyde dehydrogenase (NADP)
				(b1602 AND b1603)	NAD(P) transhydrogenase (periplasm)
				(b3708)	Tryptophanase (L-tryptophan)
B_2	11.530	0.070	10	(b0351) OR (b1241)	acetaldehyde dehydrogenase (acetylating)
	(14875%)	(-69.3%)		(b2587)	2-oxoglutarate transport via symport
				(b3945)	aldose reductase (acetol)
					Glycerol dehydrogenase
					D-Lactaldehyde:NAD+ 1-oxidoreductase
				(b1852)	glucose 6-phosphate dehydrogenase
				(b1380) OR (b2133)	D-lactate dehydrogenase
				(b2463)	malic enzyme (NADP)
				(b2661)	succinate-semialdehyde dehydrogenase (NADP)
				(b1602 AND b1603)	NAD(P) transhydrogenase
B_3	10.610	0.087	8	((b0351)OR(b1241))	acetaldehyde dehydrogenase (acetylating)
	(13659%)	(-62%)		((b3945))	aldose reductase (acetol)
					Glycerol dehydrogenase
					D-Lactaldehyde:NAD+ 1-oxidoreductase
				((b1380)OR(b2133))	D-lactate dehydrogenase
				(b2463)	malic enzyme (NADP)
				(b0767)	6-phosphogluconolactonase
				((b1602ANDb1603))	NAD(P) transhydrogenase
B_4	9.284	0.093	5	((b0356) OR (b1241) OR (b1478))	alcohol dehydrogenase (ethanol)
	(11939%)	(-59.55%)		(b4025)	glucose-6-phosphate isomerase
				(b2501)	polyphosphate kinase
					polyphosphate kinase

*We report the protein 1) "Nuo" associated to the gene set:(b2276 AND b2277 AND b2278 AND b2279 AND b2280 AND b2281 AND b2282 AND b2283 AND b2284 AND b2285 AND b2286 AND b2287 AND b2288), and 2) "(FdhF and Hyd4) or (FdhF and HycB)" associated to (b4079 AND (b2481 AND b2482 AND b2483 AND b2484 AND b2485 AND b2486 AND b2487 AND b2488 AND b2489 AND b2490) or (b4079 AND (b2719 AND b2720 AND b2721 AND b2722 AND b2723 AND b2724)))

In order to study the favourable environmental conditions, i.e. nutrients for *E. coli*, we performed the simultaneous optimisation of acetate, succinate and biomass on the complete network, i.e. without knockouts. We consider the anaerobic and aerobic condition (O_2 uptake rate = 10 $mmolh^{-1}gDW^{-1}$) by keeping fixed the glucose uptake rate to 10 $mmolh^{-1}gDW^{-1}$. We use the Non-Dominated Sorting Genetic Algorithm II [13] to perform optimisation by exploring the continuous space of exchange fluxes. The algorithm implements the *Simulated Binary Crossover* operator for crossover and the *polynomial mutation*. In our analysis, the decision variable vector is the lower bound vector of the flux values that constitute the 297 exchange fluxes (glucose and oxygen are kept constant) in the FBA model of *E. coli*. The decision variables are real values from 0 to -100 (0 when the uptake is not allowed, -100 when the potential uptake rate is 100 mmol gDW^{-1} h^{-1}). The algorithm parameters are the population size (set as 100 individuals) and the generation number (set at 500).

Our method reaches the maximum value of acetate (+100 mmol gDW^{-1} h^{-1}), and highlights conflictive behaviour of biomass and succinate (see their maximisation in the Pareto fronts of Figure 6). In anaerobic condition, we found 100 $mmolh^{-1}$ gDW^{-1} h^{-1} of acetate, 42.918 $mmolh^{-1}$ gDW^{-1} of succinate and 3.6204 h^{-1} of biomass (the trade-off). In this condition, we individuated a significant increment in the L-Aspartate, Citrate, Lactose, Fumarate and Malate uptake rates. Instead, in aerobic condition, we found 100 $mmolh^{-1}$ gDW^{-1} h^{-1} of acetate, 21.889 $mmolh^{-1}$ gDW^{-1} of succinate and 4.16 h^{-1} of biomass and a significant increment in the L-Asparagine, 1, 4-alpha-D-glucan, Fe(III)dicitrate, 2-Oxoglutarate uptake rates. In our analysis, we perturbed simultaneously almost

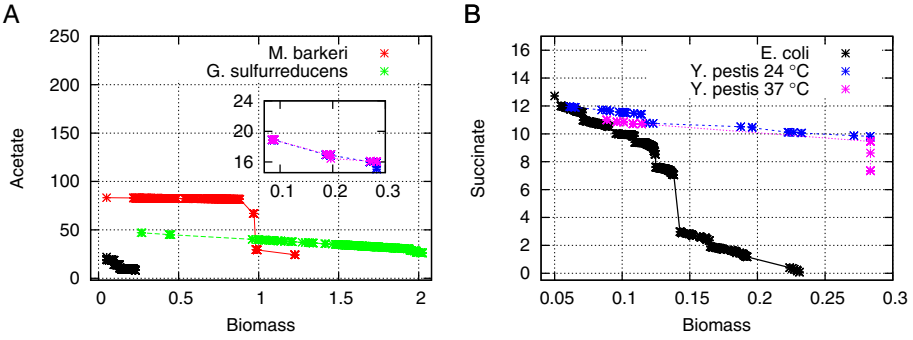

Fig. 5. Pareto fronts obtained by optimising the acetate production [$mmolh^{-1}$ gDW^{-1}] (A), succinate production (B) and the biomass formation [h^{-1}] using GDMO algorithm in four organisms models: *E. coli*, *M. barkeri*, *G. sulfurreducens* and *Y. Pestis*. For *Y. Pestis* we consider two biomass compositions: at 24-28 ℃ and 37 ℃. The significance of these two temperatures stems from the two types of hosts that *Y. Pestis* infects in the natural environment, namely insect vectors at ambient temperature and mammalian hosts with regulated body temperatures of about 37℃. In *M. barkeri*, *G. sulfurreducens* and *Y. Pestis*, the yield of acetate and biomass is larger than *E. coli* due to the lower number of reactions in the metabolic reconstructions.

all the exchange fluxes, but it is possible to select a smaller set of nutrients to study according to experimental requirement.

Sensitivity analysis results are shown in Figure 2, revealing that only 70 out of 299 are influent in the output of the model, i.e., the remaining do not change significantly the metabolic network. In particular, *Chloride, Zinc, Co2+, L-Glutamate exchanges* are the most sensitive (the complete list is reported in Figure 2).

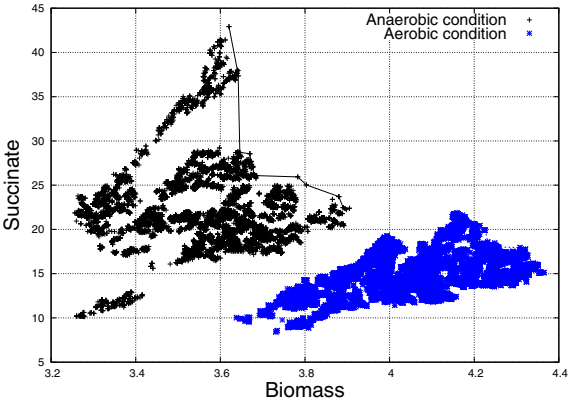

Fig. 6. Feasible regions for acetate production, succinate production (y axis) and biomass formation (x axis). We consider the wild-type bacteria (i.e. knockout zero) and perform the optimisation in aerobic (O2 = 10 mmolh^{-1} gDW^{-1}, in blue) and anaerobic (black) conditions on a basis of 10 mmolh^{-1} gDW^{-1} glucose fed to identify favourable nutrients. In both conditions, the algorithm reaches the maximum production of acetate (100 mmolh^{-1} gDW^{-1}).

4 Conclusions

This paper highlights that the Pareto front has a close link with the biotechnology productivity. For the biosynthesis, Pareto optimality is important to obtain not only a wide range of Pareto optimal solutions, but also the *best trade-off design*. Pareto front provides not merely the visualisation of the optimisation process, but also significant information in metabolic design automation. For instance, the size of non-dominated solutions, the first derivative and the area under the Pareto curve could play a key role for the best design within the same organism or between different organisms. Remarkably, the reduced size of the Pareto front could indicate the incompleteness of the model in terms of the number of reactions modelled; in this case, the Pareto optimality could be thought of as a parameter describing the improvement of a model for a bacterium with respect to a previous model for the same bacterium.

Exploratory analysis and comparative metabolic models suggest that the area underlying the Pareto provides an estimate of the number of intermediates which may be exploited for biotechnological purposes (optimisation of an additional objective) or to build synthetic pathways (synthetic biology). Given two bacteria

or two conditions for the same bacterium, a larger area under the Pareto front probably represents the best conditions for adding or optimising pathways leading to new biotechnological products. The slope of the Pareto front reflects the progressive lack of pathways able to sustain the production of one component when we are optimising the metabolism to maximise the other. The anaerobic Pareto front has also many more jumps (quick decreases) than the aerobic one. Jumps mark the sudden loss of pathways due to the critical unavailability of an enzymatic step. In other words they correspond to sudden decreases in the availability of entire pathways and subnetworks when a crucial hub is eliminated (e.g., the Krebs cycle). The region of the Pareto front nearby a jump suggests that slight changes of conditions, or a handful of genetic mutations, may result in a large change in the amount of product. Hence, the first derivative, and in particular its discontinuity, indicates the preferable conditions for the metabolites production as highlighted in the Figure 3-C-D regarding succinate and biomass optimisation. In fact, Pareto front in aerobic condition presents a wide jump, confirming that anaerobic condition is favourable for succinate fermentation as given in literature, while in aerobic condition succinate is used as intermediate to produce energy and is totally consumed. GDMO scales effectively as the size of the metabolic system and the number of genetic manipulations increase. Moreover, our results show that the multi-objective approach is very suitable for the genetic design strategies (GDS) discovering. We believe that the algorithm could be further extended and tuned to specific cases.

In the framework we propose in our work, the robustness analysis allows currently to discriminate the strains based on GR or R value: the higher these values, the greater the possibility that bacteria, reproduced in laboratory, maintain the desired performance. In future works, the local robustness analysis and other statistical connected analysis will enable us to reach a better understanding of the metabolic network fragility and this could help the GDMO algorithm to find more robust strains.

References

1. Alper, H., Miyaoku, K., Stephanopoulos, G.: Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets. Nature Biotechnology 23(5), 612–616 (2005)
2. Jarboe, L.R., Zhang, X., Wang, X., Moore, J.C., Shanmugam, K.T., Ingram, L.O.: Metabolic engineering for production of biorenewable fuels and chemicals: Contributions of synthetic biology. Journal of Biomedicine and Biotechnology (2010)
3. Atsumi, S., Wu, T.Y., Eckl, E.M., Hawkins, S.D., Buelter, T., Liao, J.C.: Engineering the isobutanol biosynthetic pathway in Escherichia coli by comparison of three aldehyde reductase/alcohol dehydrogenase genes.. Applied Microbiology and Biotechnology 85(3), 651–657 (2010)
4. Orth, J.D., Thiele, I., Palsson, B.Ø.: What is flux balance analysis? Nature Biotechnology 28(3), 245–248 (2010)
5. Patil, K.R., Rocha, I., Förster, J., Nielsen, J.: Evolutionary programming as a platform for in silico metabolic engineering. BMC Bioinformatics 6(1), 308 (2005)

6. Rocha, M., Maia, P., Mendes, R., Pinto, J.P., Ferreira, E.C., Nielsen, J., Patil, K.R., Rocha, I.: Natural computation meta-heuristics for the in silico optimization of microbial strains. BMC Bioinformatics 9(1), 499 (2008)
7. Lun, S.D., Rockwell, G., Guido, N.J., Baym, M., Kelner, J.A., Berger, B., Galagan, J.E., Church, G.M.: Large-scale identification of genetic design strategies using local search. Mol. Syst. Biol. 5(296) (2009)
8. Burgard, A.P., Pharkya, P., Maranas, C.D.: Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnology and Bioengineering 84(6), 647–657 (2003)
9. Pharkya, P., Maranas, C.: An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. Metabolic Engineering 8(1), 1–13 (2006)
10. Sun, J., Sayyar, B., Butler, J.E., Pharkya, P., Fahland, T.R., Famili, I., Schilling, C.H., Lovley, D.R., Mahadevan, R.: Genome-scale constraint-based modeling of *Geobacter metallireducens*. BMC Systems Biology 3(1), 15+ (2009)
11. Feist, A.M., Scholten, J.C.M., Palsson, B.Ø., Brockman, F.J., Ideker, T.: Modeling methanogenesis with a genome-scale metabolic reconstruction of Methanosarcina barkeri. Mol. Syst. Biol. 2 (January 2006)
12. Charusanti, P., Chauhan, S., McAteer, K., Lerman, J.A., Hyduke, D.R., Motin, V.L., Ansong, C., Adkins, J.N., Palsoon, B.Ø.: An experimentally-supported genome-scale metabolic network reconstruction for *Yersinia pestis* co92
13. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation 6(2), 182–197 (2002)
14. Sendin, J.O., Alonso, A., Banga, J.: Multi-objective optimization of biological networks for prediction of intracellular fluxes. In: Corchado, J., De Paz, J., Rocha, M., Rocha, M., Fernández Riverola, F. (eds.) 2nd International Workshop on Practical Applications of Computational Biology and Bioinformatics (IWPACBB 2008). AISC, vol. 49, pp. 197–205. Springer, Heidelberg (2009)
15. Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M., Sauer, U.: Multidimensional optimality of microbial metabolism. Science 336(6081), 601–604 (2012)
16. Xu, M., Bhat, S., Smith, R., Stephens, G., Sadhukhan, J.: Multi-objective optimisation of metabolic productivity and thermodynamic performance. Computers & Chemical Engineering 33(9), 1438–1450 (2009)
17. Laumanns, M., Thiele, L., Deb, K., Zitzler, E.: Combining convergence and diversity in evolutionary multiobjective optimization. Evol. Comput. 10(3), 263–282 (2002)
18. Stracquadanio, G., Umeton, R., Papini, A., Liò, P., Nicosia, G.: Analysis and optimization of c3 photosynthetic carbon metabolism. In: Rigoutsos, I., Floudas, C.A. (eds.) Proceedings of 10th IEEE International Conference on Bioinformatics and Bioengineering (IEEE BIBE), Philadelphia, PA, USA, May 31-June 3, pp. 44–51. IEEE Computer Society (2010)
19. Umeton, R., Stracquadanio, G., Papini, A., Costanza, J., Lio, P., Nicosia, G.: Identification of sensitive enzymes in the photosynthetic carbon metabolism. Advances in Experimental Medicine and Biology 736, 441–459 (2012)
20. Zhang, H.X., Goutsias, J.: A comparison of approximation techniques for variance-based sensitivity analysis of biochemical reaction systems. BMC Bioinformatics 11(246) (2010)
21. Rodriguez-Fernandez, M., Banga, J.R.: Senssb: a software toolbox for the development and sensitivity analysis of systems biology models. Bioinformatics 26(13), 1675–1676 (2010)

22. Morris, M.D.: Factorial sampling plans for preliminary computational experiments. Technometrics 33(2), 161–175 (1991)
23. Callaway, D.S., Newman, M.E.J., Strogatz, S.H., Watts, D.J.: Network robustness and fragility: Percolation on random graphs. Physical Review Letters 85, 5468–5471 (2000)
24. Shinar, G., Alon, U., Feinberg, M.: Sensitivity and robustness in chemical reaction networks. SIAM Journal of Applied Mathematics 69(4), 977–998 (2009)
25. Hafner, M., Koeppl, H., Hasler, M., Wagner, A.: Glocal robustness analysis and model discrimination for circadian oscillators. PLoS Comput. Biol. 5(10) (2009)
26. Donaldson, R., Gilbert, D.: A Model Checking Approach to the Parameter Estimation of Biochemical Pathways. In: Heiner, M., Uhrmacher, A.M. (eds.) CMSB 2008. LNCS (LNBI), vol. 5307, pp. 269–287. Springer, Heidelberg (2008)
27. Lodhi, H., Gilbert, D.: Bootstrapping Parameter Estimation in Dynamic Systems. In: Elomaa, T., Hollmén, J., Mannila, H. (eds.) DS 2011. LNCS, vol. 6926, pp. 194–208. Springer, Heidelberg (2011)
28. Umeton, R., Stracquadanio, G., Sorathiya, A., Papini, A., Lio, P., Nicosia, G.: Design of robust metabolic pathways. In: Design Automation Conference (DAC), 2011 48th ACM/EDAC/IEEE, pp. 747–752 (June 2011)
29. Nicosia, G., Rinaudo, S., Sciacca, E.: An evolutionary algorithm-based approach to robust analog circuit design using constrained multi-objective optimization. Knowledge-Based Systems 21(3), 175 (2008), The 27th SGAI International Conference on Artificial Intelligence
30. Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., Palsson, B.Ø.: A genome-scale metabolic reconstruction for escherichia coli k-12 mg1655 that accounts for 1260 orfs and thermodynamic information. Mol. Syst. Biol. 3(121), 291–301 (2007)