Evolutionary Computation for Predicting Optimal Reaction Knockouts and Enzyme Modulation Strategies

Pedro Evangelista[*][†], Miguel Rocha[†], Isabel Rocha[*]
[*]IBB-Institute for Biotechnology and Bioengineering, Centre of Biological Engineering,
Universidade do Minho, 4710-057 Campus de Gualtar, Braga, Portugal
Email:{ptiago, irocha}@deb.uminho.pt
[†]CCTC - Computer Science and Technology Center,
Universidade do Minho, 4710-057 Campus de Gualtar, Braga, Portugal
Email:mrocha@di.uminho.pt

***Abstract*—One of the main purposes of Metabolic Engineering is the quantitative prediction of cell behaviour under selected genetic modifications. These methods can then be used to support adequate strain optimization algorithms in a outer layer. The purpose of the present study is to explore methods in which dynamical models provide for phenotype simulation methods, that will be used as a basis for strain optimization algorithms to indicate enzyme under/over expression or deletion of a few reactions as to maximize the production of compounds with industrial interest. This work details the developed optimization algorithms, based on Evolutionary Computation approaches, to enhance the production of a target metabolite by finding an adequate set of reaction deletions or by changing the levels of expression of a set of enzymes. To properly evaluate the strains, the ratio of the flux value associated with the target metabolite divided by the wild-type counterpart was employed as a fitness function. The devised algorithms were applied to the maximization of Serine production by *Escherichia coli*, using a dynamic kinetic model of the central carbon metabolism. In this case study, the proposed algorithms reached a set of solutions with higher quality, as compared to the ones described in the literature using distinct optimization techniques.**

I. Introduction

Progress in molecular biology technologies permitted uncovering new molecular interactions aiding in the better characterization of cells. Modeling a cell based on the understanding of the interplay of its constituents, in connection with information from different omics, is the purpose of Systems Biology (SB) as advocated by Kitano [1].

The application of engineering concepts to SB provides valuable insights, helping to consolidate ongoing efforts in Biotechnology. Of particular interest, in the scope of this work, is Metabolic Engineering (ME). This discipline is concerned with the understanding and use of metabolic pathway modifications, using biological models under an engineering perspective to attain a specific industrial objective [2].

There has been a trend in industry to replace chemical synthesis techniques by biotechnological processes, due to environmental and sustainability concerns. Optimization of microbial strains has an important role in this scenario, due to increases in bioprocess productivity and, consequently, in profitability. Generally, the metabolism of wild-type microorganisms is geared to its survival and reproduction, without engaging in the production of compounds outside this scope. Thus, the metabolism has to be modified in order to meet the desired industrial outcome, typically the overproduction of a target compound.

Until recently, in bioprocess engineering, cells were modeled as black box entities responsible for consuming substrates and producing certain compounds, ignoring the underlying biological mechanisms. The genetic improvement of microorganisms has been driven by selective pressure based on empirical principles to obtain organisms with desired characteristics.

More recently, rational approaches for ME have been proposed, where researchers attempt to build mechanistic whole cell models to elucidate and provide tools for studying metabolic responses under different environments and perturbations. However, these still face hurdles such as the lack of knowledge about the reaction kinetics and the cellular responses to specific external perturbations. Nonetheless, ME has paved the way to induce cells to over-synthesize target products, to engineer new metabolic pathways and to control the production of a set of metabolites of interest.

The prediction of metabolic states has been accomplished mainly by the use of genome-scale stoichiometric models and constraint based phenotype simulation methods. These are developed based on a microorganism's specific biochemical network, using mass balances and reaction flux constraints derived from biophysical principles or empirical observations.

Several stoichiometric genome scale models have been published in the literature for microorganisms such as *Escherichia coli* [3] and *Saccharomyces cerevisiae* [4]. Typically, these models do not contain kinetic and/or regulatory information. Even so, it is possible to predict cellular behavior under certain assumptions (e.g. pseudo steady-state). From a ME point of view, these models allow to investigate the response of a metabolic network to specific genetic manipulations and/or environmental conditions.

There are several simulation methods that can be employed to estimate the microorganism flux distribution using stoichiometric models such as Flux Balance Analysis (FBA) [5], mini-

978-1-4799-0454-9/13/$31.00 ©2013 IEEE

mization of metabolic adjustment (MOMA) [6] and Regulatory on/off minimization of metabolic flux changes(ROOM) [7]. Each of these methods returns a unique optimal solution from the solution space, but in many cases several optimal solutions may exist and there is no information concerning which of those is indeed used by the cell. Thus, it is hard to identify the cell's true state [8]. Also, the employed objective functions may not represent the biological reality and other objectives for the cell can be considered instead [9].

Despite the described limitations, these methods can provide useful insights for ME. Several tools have been developed in the last years to calculate the best set of reaction (or gene) deletions or levels of expression of enzyme sets to attain a specific objective. Within this context, the problem of finding a gene/ reaction knockout set belongs to the class of combinatorial optimization [10], while the reaction down/up regulation task is included in the numerical optimization class. It is not feasible to test all gene/ reaction deletion combinations or enzyme expression level values using a brute force approach in a reasonable amount of time.

OptKnock [11] provides an alternative for the reaction deletion task, based on a MILP formulation, finding an optimal set of reactions to delete. However, it is constrained to linear objective functions and it cannot be applied with large networks due to the NP complexity of the problem [12].

OptGene [10] tackles this problem using Evolutionary Algorithms (EAs), in conjunction with FBA to estimate the effect of certain sets of reaction deletions. This method gives no guarantees of finding the best global reaction deletion set, but often provides (near) optimal solutions in a reasonable time being also more flexible in terms of the definition of the fitness functions. In recent work, other variants of Evolutionary Computation (EC) approaches such as set-based representation EAs and Simulated Annealing have been proposed and evaluated [13]. Also, methods that try to estimate the best under/over expression levels for a set of enzymes have been proposed, namely OptReg [14] that is based on a MILP formulation and more recently EAs [15].

An important shortcoming of all these methods based on constraint-based approaches is the absence of dynamic features concerning the metabolic state, not allowing to cope with enzyme kinetics and regulation. Therefore, the obtained results do not portray these effects and are bound to be incomplete.

One approach to overcome these hurdles is to use dynamic models. These models are usually based in ordinary differential equations and produce a more detailed description of cellular systems by capturing transient behavior. This type of models mimic better the phenomena observed *in vivo* in microbial strains than its purely stoichiometric counterparts. These mathematical abstractions also allow obtaining a specific steady state from an initial set of conditions wihtout further assumptions.

On the down side, they require detailed enzyme kinetic information that is often incomplete and spread across several information sources. This gives rise to inconsistencies due to the unavailability of experimental data and methodology standardization concerning the estimation of kinetic parameters. Another hurdle is the imprecise knowledge of the mechanistic rate laws underlying several reactions. It is important to bear in mind that it is not usually possible to measure all cellular compounds precisely in order to build the respective kinetics. On the whole, these models account for a small part of the metabolism. These obstacles can be attenuated by utilizing kinetic law approximations [16].

Despite the limitations, and due to the use of kinetic information, dynamic models are able to represent enzyme interactions not possible with steady-state models, such as metabolic inhibition. These models are also better suited to simulate the effects of enzyme expression level changes.

Therefore, and in spite of the lack of information to build large-scale dynamic models, a few attempts have been made regarding their use in ME applications. In [17] a Mixed Integer Non-Linear Progamming (MINLP) method for finding optimal modulation strategies was developed. The main limitations of this method are computational tractability[18].

In [19] the problem of finding the best set of enzyme expression levels modifications and reaction knockouts using a dynamic model of central carbon metabolism of *Eschericia coli* [20] was addressed. A MILP formulation and a generalized linearization of the kinetic model were used to find a ME strategy. However, like in Optknock [11] the effort to solve a MILP problem increases exponentially with the size of the problem at hand. This method also assumes flux and concentration bounds around the reference state, to control the error of the linerarized model regarding the original model.

In [21], the problem of finding the best set of changes in enzyme expression levels using the aforementioned model was addressed. Simulated Annealing [22] was used to search the enzyme set space, while a sequential quadratic programming method estimated the respective enzyme expression levels, forcing the objective function and the constraints to be continuous in the considered ranges and of class C^2. This method assumes a value for the overall maximum allowed metabolite changes at steady-state and also that overall system enzyme levels remain constant within a constant value proportional to the number of modifications. In this work this constraint will not be used due to its specificity and lack of experimental data to corroborate it in *Escherichia coli*. This restriction may also limit the algorithm generalization capabilities.

A. Aims and overview of the approach

This work entails the development of Evolutionary Computation (EC) approaches to find a set of metabolic modifications, such as reactions knockouts and reaction up/down regulation that will optimize the production of a metabolite with an industrial interest, utilizing as a basis for simulation dynamic models composed of ordinary differential equations (ODEs).

One aim of this method is to provide a proof of concept of a scalable approach able to deal with larger scale dynamic metabolic models than the ones that currently exist. The existing methods are not able to cope with the definition of ME strategies in larger dynamic metabolic models, due to the combinatorial increase in the number of possible strategies. Also, they do not take into consideration invalid solutions that may contain valid building blocks for the optimal solution. It is important to bear in mind that it is impossible to scan the whole state space, when it has a high dimensionality. A brute force approach is not feasible for the enzyme level modulation task and it becomes unpractical in the reaction deletion scenario as the number of modifications increases.

Also, most of the current methods deal with the parallel optimization of enzyme and knockout expressions by employing Mixed Integer Non-Linear Programming methods [17] that are unable to solve problems with hundreds of equations, or rely on the approximation of the non-linear dynamic model around a reference state (usually a steady state) and a *posteriori* use a MILP formulation [19]. The approximation of the non-linear dynamic model around a reference state also enforces the use of reaction and metabolite ranges around the reference state that may exclude valid solutions of interest.

The present approach deals with these shortfalls by using the original non-linear model without doing any approximation and by searching ME strategies by means of EAs that adapt the solution size. Thus, this method does not need to assume a range of flux and metabolite values where solutions are considered valid.

In this situation, dynamic models are used to generate a single steady-state solution without the need of specifying further assumptions such as in the cases of FBA, MOMA, or ROOM for stoichiometric models. Another advantage of using these models in ME applications is the straightforward implementation of over/under expression of enzymes as ME strategies.

Two tasks are used to test the devised optimization techniques, whose purpose is to maximize the production of a metabolite at steady-state: (i) *reaction deletion* - the objective is to discover the best set of reaction deletions (knockouts). The ideal number of reactions to remove is also determined simultaneously; and, (ii) *reaction up/down regulation* - the main goal is to find the best set of enzymes to tweak and the respective level of expression concerning the base values present in the original model.

In this work, a novel encoding scheme is proposed that will adress both tasks, allowing the algorithm to choose the best ME strategy given a permitted set of constraints, as well as the number of modifications. The solution decoding affects the simulation of the dynamic model by multiplying the v_{max} parameter of the reaction rate law by the decoded enzyme modulation level contained in the solution's genome. This corresponds to a change in the total enzyme concentration assuming that v_{max} is directly proportional to it. In the reaction deletion case, the v_{max} parameter is multiplied by zero, therefore constraining the reaction's flux to 0.

The design of the algorithm also allows the discretization of the enzyme modulation value into a set of pre-defined ranges. In a wet lab setting it is often not possible to fine tune the exact enzyme expression levels as returned by the algorithm. Thus, this discretization may allow a more flexible representation of what may be achieved *in vitro*. This representation provides for the simultaneous optimization of discrete and continuous enzyme levels. The developed method also allows to incorporate non-modifiable reactions (reactions that cannot be tweaked by the algorithm or have to respect specific constraints, like for example directionality or flux intervals).

As a basis for phenotype simulation, a metabolic dynamical model of selected pathways of *Eschericia coli* will be used, based on ordinary differential equations, namely, the mechanistic model of the central carbon metabolism [20] consisting of mass balance equations for glycolysis and for the pentose-phosphate pathway.

The aforementioned tasks are used in a case study related to the maximization of Serine production, allowing to contrast the obtained results to the ones published in [19]. Co-metabolite concentrations were assumed to be constant as in the previous study. Nowadays, Serine plays a major role in several industrial applications. Serine is used in cosmetic and food industries and is produced by fermentative routes[23]. In this case study, it is not apparent how to find the best set of genetic modifications to enhance the production of Serine due to the high number of interacting reactions.

II. Methods

A. Mechanistic model of the central carbon metabolism

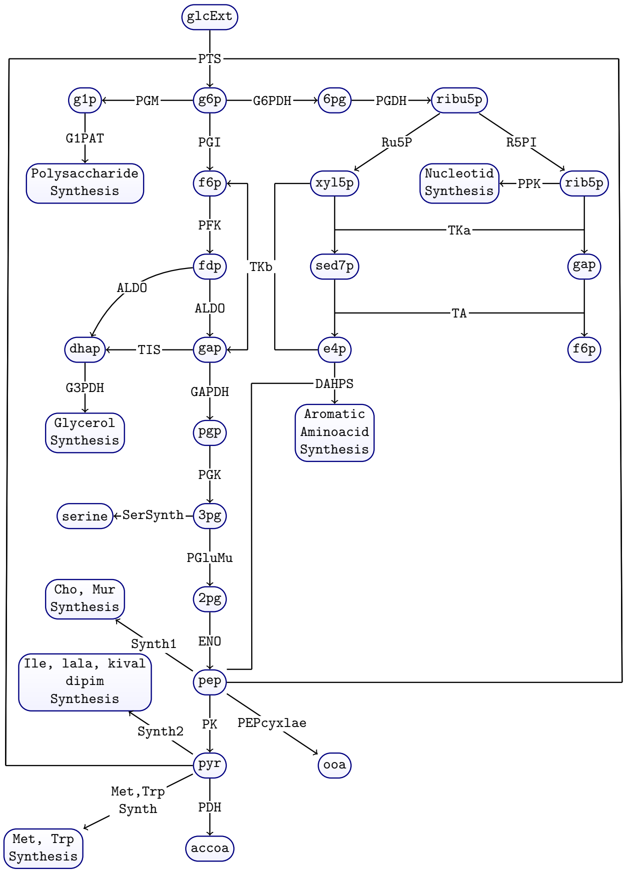

Fig. 1. *Escherichia coli* central carbon metabolism network.

The mechanistic model of the central carbon metabolism [20] encompasses the phosphotransferase system, glycolysis and the pentose-phosphate pathway. This model is curated and available from Biomodels [24]. In Figure 1, a schematic representation of the reaction network is shown. The mass balances take the following form:

$$\frac{dX}{dt} = SV - \mu X \tag{1}$$

where X represents the vector of metabolite concentrations, μ is the specific growth rate, S is the stoichiometric coefficient matrix and V is the reaction rate vector. The equation for extra-cellular glucose has the following form:

$$\frac{d[GlcExt]}{dt} = D([GlcFeed] - [GlcExt]) + fPulse - \frac{[bio]v_{PTS}}{\rho_{bio}} \tag{2}$$

where $[GlcExt]$ represents the external glucose concentration, $[GlcFeed]$ is the concentration of glucose in the feed, $fPulse$ is a function allowing to introduce glucose pulses, $[bio]$ is the biomass concentration, ρ_{bio} is the biomass density and v_{PTS} is the flux through the phosphotransferase system reaction.

The reaction fluxes at steady-state are described by:

$$v_i^0 = v_{iMax} f_i(X_i^0, P_i^0) \tag{3}$$

where the superscript ^{0} denotes a variable at steady-state, v_i^0 is the rate of reaction i at steady-state, v_{iMax} is the maximum reaction rate and $f_i(X_i^0, P_i^0)$ is a function of X_i^0 metabolite concentrations at steady-state that participate in the reaction in conjunction with a set of parameters P_i^0. Thus, the v_{iMax} for each reaction is computed by the following equation (as described in [20]):

$$v_{iMax} = \frac{v_i^0}{f_i(X_i^0, P_i^0)} \tag{4}$$

B. Objective function formulation

The Reaction deletion and Enzyme over/under expression problems can be stated as the maximization of $\frac{vMutant_j^0}{vWildType_j^0}$, where $vMutant_j^0$ and $vWildType_j^0$ represent the flux for the target reaction j at steady-state in the mutant and in the wild-type strains, respectively.

C. Solution evaluation

To assign a fitness value for each solution suggested by the evolutionary method, the following algorithm was used:

1) Perform the model modifications by decoding the solution being evaluated, described in the next section;
2) Simulate the modified model, by adding the constraints from the solution decoded and performing the numerical integration of the ODEs in the model in a given time range;
3) Verify whether metabolite concentrations do not change significantly in a given time range encompassing the end of the simulation. If this condition is met, the system is considered to be in steady-state.
4) If the previous step is completed with success, the ratio of the solution target flux by the wild-type strain target flux value is returned. Otherwise, zero is returned.

D. Solution encoding

In this work, a novel variable size representation for inferring enzyme expression levels was developed. This representation allows searching simultaneously for the set of enzymes to modify and the respective expression level. The expression level can be a real number in a specific interval or a set of discrete values defined by the user. In the proposed representation, solutions are quite simple, being represented as vectors of real numbers with values between 0 and 1.

When considering enzyme expression levels optimization, the values in an even position are mapped to a reaction index, while the values in the following odd position encode the enzyme expression level for that reaction, in a continuous or discrete interval. Each reaction may have different expression level modulation ranges.

In Figure 2, the solution decoding process is illustrated with an example considering a very simple model with three reactions R1, R2 and R3. In a), it is possible to observe that each consecutive pair of elements encodes a reaction index and its enzyme level modulation. The reaction in position six is discarded because it does not possess an enzyme modulation part.

In b), the decoding process of the first two reactions is shown. The reaction index encoded at position zero is mapped to reaction R1, as follows: the interval $[0, 1]$ is divided into three equally spaced sub-intervals (the number of reactions in the model) being each interval mapped to a reaction. The interval that contains the encoded value maps to the specific model reaction. In the example, the value 0.2 is contained in the interval $[0, \frac{1}{3}]$, that is mapped to reaction R1. Position one encodes the enzyme level for that reaction (R1). This reaction was defined as varying in the continuous interval $[0, 2]$. In this scenario, the expression level coding value 0.4 is multiplied by the interval length 2 giving the expression level equal to 0.8. Note that in this case, the lower limit of both intervals is zero and therefore the mapping is easier; in general, a mapping to an interval $[a, b]$ is obtained by multiplying the encoded value by $b - a$ and adding a.

The reaction index of the next modulation in position 2 is calculated as in the previous case, corresponding to the mapping of 0.5 to reaction R2. In this case, it is assumed that reaction R2 modulation can only vary in a discrete set of values $\{0, 0.5, 2.5\}$. In this case, the mapping of the enzyme modulation level occurs in an analogous way to the reaction index mapping and, therefore, the value 0.9 at position 3 is mapped to a modulation of 2.5. If a solution has several occurrences of the same index, only the last one is considered.

The same representation can be used for representing reaction deletions (knockouts). In this case, all expression level coding positions encode a discrete set with the value zero for the modulation level.

E. Reproduction operators

For reproduction purposes within the EA, the following operators are used:

- *Random mutation*: replaces an element of the vector by another, randomly generated in the allowed range.

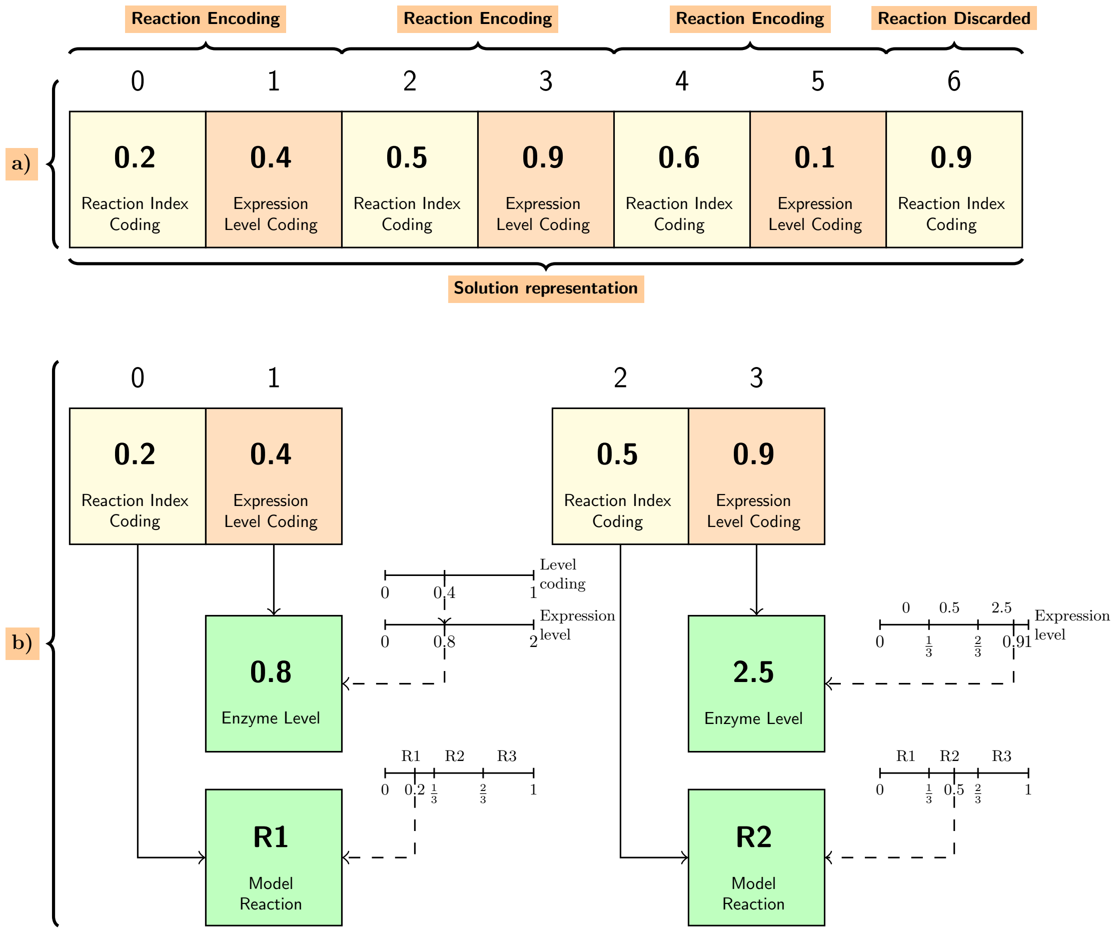

Fig. 2. Enzyme expression level solution decoding example. In a) a solution encoding for a model with reaction set $\{R1, R2, R3\}$ is shown. In b) the decoding process for the first two reactions is illustrated.

- *Cut and splice crossover*: A distinct crossover point is selected in both parents and the genes before and after that point are swapped giving rise to two new individuals. This operator has the capacity to modify the length of the resulting offspring.

In the proposed EA the operators have the following probabilities of being selected to generate new solutions from the selected parents: the mutation operator has 10% probability of being chosen while the crossover operator has 90% probability.

F. Experimental setup

In the first step of the evaluation function, the time course simulation is computed for the time interval $[0, 1E6]$ seconds. The system is considered in steady-state if the metabolite concentration change is inferior to 5% in the interval $[1E4, 1E6]$ seconds.

The enzyme up/down regulation allows reaction fluxes to vary by a multiple in the linear interval $[0, 2]$. The upper bound value was chosen based on the values employed by [19] with the linearized models. This reaction modulation range needs to be imposed in order to model the experimental capacity and the physiological reality inside the cell.

The algorithms were executed with an incremental number of restricted modifications from one up to six. In the case study, the algorithms for each problem are executed 30 times. In the knockout task the algorithm is run for 250 iterations, while in the enzyme over/under expression the algorithm is executed for 500 iterations, values that allow the convergence of the EA. Both algorithms employ the following configuration:

- *Population size*: 100 individuals;
- *Population initialization*: individuals are generated randomly with size varying between 1 and 100;
- *Elitism value*: 1 individual (the best) is always kept;
- *Number of selected individuals for reproduction*: 50 individuals;
- *Number of reinserted individuals in the population*: 49 individuals;
- *Selection operator*: Tournament selection with three individuals randomly selected, where the fittest is selected.

G. Implementation

Regarding the implementation, the software for the proposed tasks was developed using the Java, Scala, and Matlab languages. The following libraries were utilized: JECoLi, a library for EAs developed by the authors [25] and JSBML [26] a java library allowing to parse SBML encoded files. Differential equations were simulated using the solver ODE15s from Matlab. The source code is released under the GPLv3 license and is available from http://darwin.di.uminho.pt/Software/EADynamic.

III. Results and Discussion

The best solutions obtained with the proposed EA are displayed in Tables I and II. These solutions are contrasted to the ones found in the literature [19], namely the ones resulting from a linearized approximation of the non-linear model of central carbon metabolism of *Escherichia coli* around a steady-state. These solutions are also constrained by flux and concentration bounds to reduce the likelihood of not portraying the behavior of the original model. All the fitness values (v_j/v^0_{j0}) concern the non-linearized version of the model.

In both tables, EAK and EAE represent the data regarding the solutions found with devised EA (for knockout and enzyme level optimization, respectively), while VLS and VLL are related to the application of the method developed in [19]. In VLS, the enzyme expression levels (e^0_i) at steady state in the linearized model are restricted by the inequality $0.5e^0_i \leq e^0_i \leq 2e^0_i$, while metabolite concentrations at steady-state (x^0_i) are constrained by $0.5x^0_i \leq x^0_i \leq 1.5x^0_i$. In VLL, the enzyme modulation levels in the linearized model are constrained by $0.5e^0_i \leq e^0_i \leq 2e^0_i$ and the metabolite concentrations by $0.5x^0_i \leq x^0_i \leq 10x^0_i$.

The developed EA overcomes these restrictions by integrating the non-linear model and by assuming that the model depicts adequately the subjacent reality. Nonetheless, it is important to note that even the original model may not be valid in all range of metabolite concentrations due to the absence of data regarding those states when the model was fitted.

Comparing the results obtained by the proposed EA with the ones in [19], it is possible to check that they show equal or, in most cases, higher fitness values. Also, it is easy to check that, although being a stochastic method, EAs are capable of locating good quality results with a low variability.

In all the studied scenarios the first proposed transformation is also part of the underlying proposed modifications. The first modification is the one that implies a larger flux gain in the Serine synthesis flux. However, reaction knockouts are geared towards reactions leading to an increase in the concentration of compounds that contribute to Serine formation, whereas in the enzyme modulation, the algorithm tends to maximize the flux that leads directly to the production of Serine (in this case the serineSynth reaction). Notwithstanding, as the number of reaction modification increases, the marginal serine synthesis flux gain usually tends to decrease. This fact can be observed in the present case studies as well as in [19].

In addition, even without the constraints that restrict most the variation of fluxes and concentrations, solutions tend to knockout reactions directly related with drains, as it can be observed in Table I. In VLS and VLL knockout solutions, the first two modifications will normally imply reactions that drain compounds out of the network. These reactions are selected because they imply a smaller change in the overall metabolite concentrations, with an increase of the Serine production. Contrarily to what would be empirically expected, the EA only deletes reactions not directly related with drain reactions (PK, PGI) with 6 modifications. This fact is owed to the absence of constraints in the EA algorithm regarding flux and metabolite concentration constraints that may limit what are considered valid solutions.

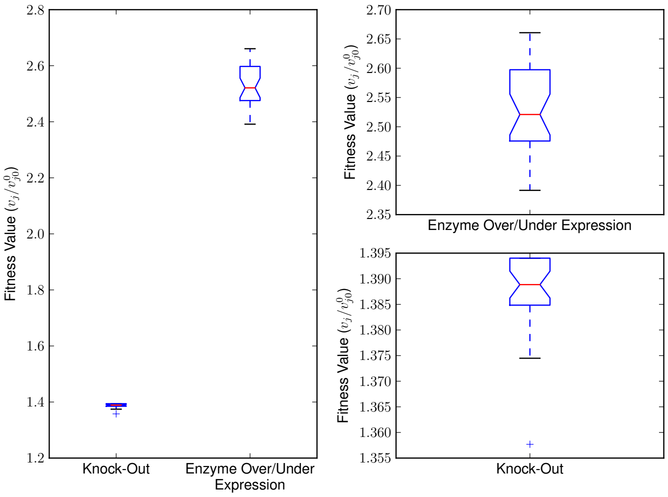

Fig. 3. Boxplots concerning the best solutions with six modifications found in the knock-out and enzyme modulation tasks, regarding the Serine maximization case study.

In all knockout algorithms, as the number of allowed modifications increases, some of the previously utilized reactions are swapped by a not apparent set of reactions that cause an increase in the target flux. This fact can be observed in table I. For instance, with five knockouts, the best solution found by the EA is composed by the PEPC, DHAPS, MURS, PGM and PPK reactions, while with six modifications the PGM and MURS reactions are changed by Syn1, PGI and PK. This swap of reaction produces an approximated 0.35% increase in the target flux.

In the devised EA knockout algorithm up until five modifications, the reactions of the solution tend to converge to the same solution in all runs. With three modifications there is a twelve-fold change in variability regarding the previous cases. This increase in variability can be explained by the increase in the search space.

In Figure 3, it is possible to observe that the knock-out algorithm tends to converge to a set of solutions with lower variability than the enzyme over/under expression counter part. This fact may be explained by the larger search space of the enzyme modulation task. It is also noticeable that the enzyme over/under expression task requires more iterations to reduce the variability in the best solutions, as can be checked from an analysis of the convergence plots provided by Figure 4. These results cannot be extrapolated to other models or scenarios and depend on the objective function, constraints and the underlying metabolic model.

TABLE I. KNOCKOUT TASK - BEST SOLUTIONS

#Modifications	Algorithm	Modifications	Fitness (v_j/v_{j0}^0)	Mean Fitness ±95% Confidence Interval
1	EAK	PEPC	**1.149**	$1.149 \pm 8.082 \times 10^{-17}$
	VLS	DHAPS	1.057	–
	VLL	PEPC	**1.149**	–
2	EAK	PEPC DHAPS	**1.254**	$1.254 \pm 8.082 \times 10^{-17}$
	VLS	DHAPS G1PAT	1.073	–
	VLL	PEPC PK	1.226	–
3	EAK	PEPC DHAPS PGM	**1.352**	$1.352 \pm 1.765 \times 10^{-5}$
	VLS	DHAPS Syn1 Syn2	1.092	–
	VLL	PEPC PK Syn1	1.250	–
4	EAK	PEPC DHAPS PGM PPK	**1.387**	1.380 ± 0.00449
	VLS	DHAPS G1PAT PK PGI	1.124	–
	VLL	PEPC PK G1PAT Syn1	1.273	–
5	EAK	PEPC DHAPS MURS PGM PPK	**1.389**	1.386 ± 0.00242
	VLS	DHAPS G1PAT PK G3PDH PGI	1.124	–
	VLL	PEPC PK Syn1 PPK TRPS	1.262	–
6	EAK	PEPC DHAPS Syn1 PK PPK PGI	**1.394**	1.387 ± 0.00280
	VLS	DHAPS G1PAT PK G3PDH PGI METS	1.124	–
	VLL	PEPC PK Syn1 PPK TRPS METS	1.262	–

TABLE II. ENZYME MODULATION TASK - BEST SOLUTIONS

#Modifications	Algorithm	Modifications	Fitness (v_j/v_{j0}^0)	Mean Fitness ±95% Confidence Interval
1	EAE	(2.0)Sersynth	**1.876**	$1.8020.0761 \pm 0.0761$
	VLS	(2.0)Sersynth	**1.876**	–
	VLL	(2.0)Sersynth	**1.876**	–
2	EAE	(1.99)Sersynth (0.0033)PGluMu	**2.413**	2.189 ± 0.0119
	VLS	(2.0)Sersynth (0)PK	2.115	–
	VLL	(2.0)Sersynth (2.0)PTS	1.876	–
3	EAE	(1.99)Sersynth (1.92)GAPDH (0.0032)PGluMu	**2.582**	2.385 ± 0.0251
	VLS	(2.0)Sersynth (1.94)GAPDH (1.57)PFK	2.001	–
	VLL	(2.0)Sersynth (0)PEPC (1.84)PTS	2.191	–
4	EAE	(1.99)Sersynth (0.043)TKA (0.0032)PGluMu	**2.639**	2.475 ± 0.0222
	VLL	(2.0)Sersynth (2.0)PTS (0)PEPC (2.0)GAPDH	2.369	–
5	EAE	(1.99)Sersynth (0.015)R5PI (0.015)PEPC (1.99)GAPDH (0.015)PK	**2.661**	2.529 ± 0.0254
	VLL	(2.0)Sersynth (1.94)PTS (0)PEPC (2.0)GAPDH (0)PK	2.532	–
6	EAE	(1.99)Sersynth (0.0035)PEPC (1.86)GAPDH (0.0035)PGluMu (0.0035)R5PI (1.79)TRPS	**2.705**	2.553 ± 0.0251
	VLL	(2.0)Sersynth (1.38)PTS (0)PEPC (1.90)GAPDH (0)PK (0)DHAPS	2.671	–

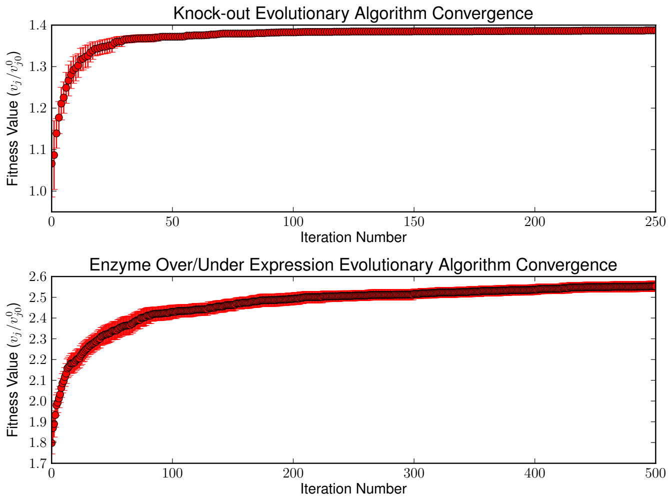

Fig. 4. Knock-out and enzyme modulation evolutionary algorithms convergence with 95% confidence bounds, concerning the best solutions found with six modifications during the 30 runs of the algorithms in the Serine maximization case study.

IV. CONCLUSION

This work encompassed the development of algorithms to design *in silico* improved microbial strains for the production of industrial relevant compounds. These algorithms achieved these modifications by finding the near/best set of reaction deletions to remove from a model and/or to infer the optimum expression levels for the enzymes in the model (or a predefined subset). A dynamic ordinary differential model describing the central carbon metabolism of *Escherichia coli* was used as basis for the simulation of the devised ME strategies. These models are capable of describing regulatory effects in the metabolism not possible to represent with steady-state models.

The best solutions returned by the devised method outperformed the ones in [19] due to the fact that no approximations of the model were needed. Solutions were computed allowing the metabolite concentrations and the fluxes to vary with no bound restrictions. During the execution of the algorithms a set of reaction modifications that might lead to an invalid steady-state were not immediately discarded. Thus, a subset of these reactions could serve as building blocks for better and valid solutions.

In future work, the remaining issues to be tackled are the validation of the work with other real-world case studies and also the integration of the developed software in a user-friendly software platform such as Optflux [27]. The utilization of multi-objective optimization algorithms [28] is also an expected extension to the current methods.

ACKNOWLEDGEMENTS

This work is funded by National Funds through the FCT - Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project PEst-OE/EEI/UI0752/2011. The work is also partially funded by ERDF - European Regional Development

Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT within project ref. COMPETE FCOMP-01-0124-FEDER-015079. PE's work is supported by a PhD grant FCT SFRH/BD/51016/2010 from the Portuguese FCT.

NOMENCLATURE

Reaction		**Metabolites**	
PTS	phosphotransferase system	glcExt	glucose
PGI	glucose-6-phosphate isomerase	g6p	glucose-6-phosphate
PFK	phosphofructo-kinase	f6p	fructose-6-phosphate
ALDO	aldolase	fdp	fructose-1,6-bisphosphate
TIS	triosephosphate isomerase	gap	glyceraldehydes-3-phosphate
GAPDH	glyceraldehyde-3-phosphate dehydrogenase	dhap	dihydroxyacetonephosphate
PGK	phosphoglycerate kinase	pgp	1,3-diphosphoglycerate
PGM	phosphoglucomutase	3pg	3-phospho-glycerate
G1PAT	glucose-1-phosphate adenyltransferase	2pg	2-phospho-glycerate
PPK	ribose-phosphate pyrophosphokinase	pep	phosphoenolpyruvate
G3PDH	glycerol-3-phosphate dehydrogenase	pyr	pyruvate
SerSynth	serine synthesis	6pg	6-phosphogluconate
Syn1	synthesis1	ribu5p	ribulose-5-phosphate
Syn2	synthesis2	xyl5p	xylulose-5-phosphate
DAHPS	DAHP synthases	sed7p	sedoheptulose-7-phosphate
G6PDH	glucose-6-phosphate dehydrogenase	rib5p	ribose-5-phosphate
PGDH	6-phosphogluconate dehydrogenase	e4p	erythrose-4-phosphate
RU5P	ribulose-phosphate epimerase	g1p	glucose-1-phosphate
R5PI	ribose-phosphate isomerase	accoa	acetyl-coenzyme A
TKA	transketolase A		
TKB	transketolase B		
TA	transaldolase		
MURS	murine synthesis		
TRPS	tryptophan synthesis		
MetSynth	methionine synthesis		

REFERENCES

[1] H. Kitano, "Systems biology: a brief overview," *Science*, vol. 295, no. 5560, p. 1662, 2002.

[2] G. Stephanopoulos, A. Aristidou, J. Nielsen, and J. Nielsen, *Metabolic engineering: principles and methodologies*. Academic Pr, 1998.

[3] J. Reed, T. Vo, C. Schilling, and B. Palsson, "An expanded genome-scale model of escherichia coli k-12 (ijr904 gsm/gpr)," *Genome Biol*, vol. 4, no. 9, p. R54, 2003.

[4] J. Forster, I. Famili, P. Fu, B. Palsson, and J. Nielsen, "Genome-scale reconstruction of the saccharomyces cerevisiae metabolic network," *Genome research*, vol. 13, no. 2, p. 244, 2003.

[5] K. Kauffman, P. Prakash, and J. Edwards, "Advances in flux balance analysis," *Current opinion in biotechnology*, vol. 14, no. 5, pp. 491–496, 2003.

[6] D. Segre, D. Vitkup, and G. Church, "Analysis of optimality in natural and perturbed metabolic networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 23, p. 15112, 2002.

[7] T. Shlomi, O. Berkman, and E. Ruppin, "Regulatory on/off minimization of metabolic flux changes after genetic perturbations," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 21, p. 7695, 2005.

[8] J. Reed and B. Palsson, "Thirteen years of building constraint-based in silico models of escherichia coli," *Journal of Bacteriology*, vol. 185, no. 9, p. 2692, 2003.

[9] R. Schuetz, L. Kuepfer, and U. Sauer, "Systematic evaluation of objective functions for predicting intracellular fluxes in escherichia coli," *Molecular systems biology*, vol. 3, no. 1, 2007.

[10] K. Patil, I. Rocha, J. F
"orster, and J. Nielsen, "Evolutionary programming as a platform for in silico metabolic engineering," *BMC bioinformatics*, vol. 6, no. 1, p. 308, 2005.

[11] A. Burgard, P. Pharkya, and C. Maranas, "Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization," *Biotechnology and bioengineering*, vol. 84, no. 6, pp. 647–657, 2003.

[12] B. Papp and E. Simeonidis, "Flux balance analysis and its applications," *BMC Systems Biology*, vol. 1, no. Suppl 1, p. P77, 2007.

[13] M. Rocha, P. Maia, R. Mendes, J. Pinto, E. Ferreira, J. Nielsen, K. Patil, and I. Rocha, "Natural computation meta-heuristics for the in silico optimization of microbial strains," *BMC bioinformatics*, vol. 9, no. 1, p. 499, 2008.

[14] P. Pharkya and C. Maranas, "An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems," *Metabolic engineering*, vol. 8, no. 1, pp. 1–13, 2006.

[15] E. Gonçalves, R. Pereira, I. Rocha, and M. Rocha, "Optimization approaches for the in silico discovery of optimal targets for gene over/underexpression," *Journal of Computational Biology*, vol. 19, no. 2, pp. 102–114, 2012.

[16] F. Wang, C. Ko, and E. Voit, "Kinetic modeling using s-systems and lin-log approaches," *Biochemical engineering journal*, vol. 33, no. 3, pp. 238–247, 2007.

[17] J. Dean and G. Dervakos, "Design of process-compatible biological agents," *Computers & chemical engineering*, vol. 20, pp. S67–S72, 1996.

[18] ——, "Redesigning metabolic networks using mathematical programming," *Biotechnology and bioengineering*, vol. 58, no. 2-3, pp. 267–271, 2000.

[19] F. Vital-Lopez, A. Armaou, E. Nikolaev, and C. Maranas, "A computational procedure for optimal engineering interventions using kinetic models of metabolism," *Biotechnology progress*, vol. 22, no. 6, pp. 1507–1517, 2006.

[20] C. Chassagnole, N. Noisommit-Rizzi, J. Schmid, K. Mauch, and M. Reuss, "Dynamic modeling of the central carbon," *Biotechnol Bioeng*, vol. 79, pp. 53–73, 2002.

[21] E. Nikolaev, "The elucidation of metabolic pathways and their improvements using stable optimization of large-scale kinetic models of cellular systems," *Metabolic engineering*, vol. 12, no. 1, pp. 26–38, 2010.

[22] S. Kirkpatrick, C. Gelatt, and M. Vecchi, "Optimization by simulated annealing," *science*, vol. 220, no. 4598, p. 671, 1983.

[23] M. Sadovnikova and V. M. Belikov, "Industrial applications of aminoacids," *Russian Chemical Reviews*, vol. 47, no. 2, p. 199, 2007.

[24] N. Le Novere, B. Bornstein, A. Broicher, M. Courtot, M. Donizelli, H. Dharuri, L. Li, H. Sauro, M. Schilstra, B. Shapiro *et al.*, "Biomodels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems," *Nucleic acids research*, vol. 34, no. suppl 1, pp. D689–D691, 2006.

[25] P. Evangelista, P. Maia, and M. Rocha, "Implementing metaheuristic optimization algorithms with jecoli," in *Intelligent Systems Design and Applications, 2009. ISDA'09. Ninth International Conference on*. IEEE, 2009, pp. 505–510.

[26] A. Dräger, N. Rodriguez, M. Dumousseau, A. Dörr, C. Wrzodek, N. Le Novère, A. Zell, and M. Hucka, "Jsbml: a flexible java library for working with sbml," *Bioinformatics*, vol. 27, no. 15, pp. 2167–2168, 2011.

[27] I. Rocha, P. Maia, P. Evangelista, P. Vilaça, S. Soares, J. P. Pinto, J. Nielsen, K. R. Patil, E. C. Ferreira, and M. Rocha, "Optflux: an open-source software platform for in silico metabolic engineering," *BMC systems biology*, vol. 4, no. 1, p. 45, 2010.

[28] P. Maia, I. Rocha, E. C. Ferreira, and M. Rocha, "Evaluating evolutionary multiobjective algorithms for the in silico optimization of mutant strains," in *BioInformatics and BioEngineering, 2008. BIBE 2008. 8th IEEE International Conference on*. IEEE, 2008, pp. 1–6.