Matteo N. Amaradio ORCID iD: 0000-0003-3942-4460

Varun Ojha ORCID iD: 0000-0002-9256-1192

Giorgio Jansen ORCID iD: 0000-0001-7320-4053

Massimo Gulisano ORCID-iD: 0000-0001-8654-1745

Jole Costanza ORCID iD: 0000-0002-0290-3970

Giuseppe Nicosia ORCID iD: 0000-0002-0650-3157

Pareto Optimal Metabolic Engineering for the Growth-coupled Overproduction of Sustainable Chemicals

Matteo N. Amaradio[1,*], Varun Ojha[2,*], Giorgio Jansen[1,3,*], Massimo Gulisano[4], Jole Costanza[5], Giuseppe Nicosia[1,3]

[1]Department of Biomedical & Biotechnological Sciences, University of Catania, Catania Italy
[2]Department of Computer Science, University of Reading, Reading, United Kingdom
[3]Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom
[4]Department of Drug Science, University of Catania, Catania, Italy
[5]National Institute of Molecular Genetics, Milan, Italy

* First Authors

Corresponding Authors:

Varun Ojha, Department of Computer Sciences, University of Reading, Reading, RG6 6DH, Reading, United Kingdom, Telephone: +44 (0)118 378 8609 Email address: v.k.ojha@reading.ac.uk

Giuseppe Nicosia, Department of Biomedical & Biotechnological Sciences, University of Catania, Catania Italy, Telephone: + 39 095 478 1289 Email address: giuseppe.nicosia@unict.it

Abstract — Our research aims to help industrial biotechnology develop a sustainable economy using green technology based on microorganisms and synthetic biology through two case studies that improve metabolic capacity in yeast models Yarrowia lipolytica (Y. lipolytica) and Saccharomyces cerevisiae (S. cerevisiae). We aim to increase the production capacity of beta-carotene (β-carotene) and succinic acid, which are among the highest market demands due to their versatile use in numerous consumer products. We performed simulations to identify *in silico* ranking of strains based on multiple objectives: the growth rate of yeast microorganisms, the number of used chromosomes, and the production capability of β-carotene (for Y. lipolytica) and succinate (for S. cerevisiae). Our multiobjective optimization methodology identified notable gene deletions by searching a vast solution-space to highlight near-optimal strains on Pareto Fronts, balancing the above-cited three objectives. Moreover, preserving the metabolic constraints and the essential genes, this work produced robust results: 7 significant strains of Y. lipolytica and 7 strains for S. cerevisiae. We examined gene knockout to study the function of genes and pathways. In fact, by studying the frequently silenced genes, we found that when the *GPH1* gene is knocked out in S. cerevisiae, the isocitrate lyase enzyme is activated, which converts the isocitrate into succinate. Our goals are to simplify and facilitate the in vitro processes. Hence, we present strains with the least possible number of knockout genes and solutions in which the genes are turned off on the same chromosome. Therefore, we present results where the constraints mentioned above are met, like the strains where only two genes are switched off and other strains where half of the knockout genes are on the same chromosome. This research offers solutions for developing an efficient *in vitro* mutagenesis for microorganisms and demonstrates the efficiency of multiobjective optimization in automatizing metabolic engineering processes.

Keywords—Genome-scale metabolic models; Metabolic engineering; β-carotene production; Succinate production; *Y. lipolytica; S. cerevisiae;* Pareto fronts, Yield maximization; Productivity maximization

1. Introduction

Synthetic biology tools offer a cost-efficient and eco-compatible alternative to the traditional high energy rate manufacturing processes for producing renewable feedstocks (King et al., 2015b; Nielsen and Keasling, 2016). In this field, mathematical modeling can help find optimal genetic manipulations and speed up the

identification of better-performing strains to produce specific metabolites of interest that can replace their petrochemical-derived equivalents. The yeast models are the bioreactors of metabolic precursors, which are converted into a wide range of consumer products, e.g., *beta-carotene* (β-carotene*)* and *Succinic acid* (or succinate).

The β-carotene has nutraceutical and antioxidant properties. These properties make β-carotene a highly desired product in agriculture, food, pharmaceutical, and industries alike. In 2018, it had an estimated $1.4 billion market demand (Larroude et al., 2018; Abdel-Mawgouda et al., 2018). Furthermore, the high versatility of *succinic acid* in multiple industrial chemical applications and consumer products makes its market demand grow steadily at a compound annual growth rate of around 27.4% to reach $1.8 billion (768 million MT at $2.3/kg) by 2025 (Nghiem et al., 2017).

The β-carotene and succinic acid are mainly produced by specific host yeasts: the *oleaginous yeast Yarrowia lipolytica (Y. lipolytica)* is the preferred host to produce carotenoids (β-carotene) because of its naturally high supply of carotenoids precursor cytosolic as acetyl-CoA and redox cofactor: *nicotinamide adenine dinucleotide phosphate* (*NADPH*), which *reduces nicotinamide adenine dinucleotide* (Kildegaard et al., 2017). On the other hand, the preferred succinic acid producer yeast is *Saccharomyces cerevisiae (S. cerevisiae)*, known for its ability to grow under acidic conditions and for its well-characterized role in wine acidity (Cao et al., 2013; Franco-Duarte et al., 2017; Vilela, 2019).

Our study proposes a robust methodology based on *a multiobjective evolutionary algorithm* (MOEA) for *in silico* identification of competitive genetically manipulated strains of *Y. lipolytica* and *S. cerevisiae* metabolism (Section 2). This is through selecting effective gene deletions (knockout) in the respective genome-scale metabolic models. In addition, the algorithm attempts to knock out genes on the same chromosome to simplify the in vitro process. We describe the results of the two case studies in Section 3. The results show that our methodology is an effective strategy to optimize the design of competitive strains to produce organic compounds in large-scale fermentation that leads to a more targeted *in vitro* mutagenesis (King et al., 2015b; Nielsen and Keasling, 2016; Patané et al., 2015; Patané et al., 2019).

The β-carotene is the main source of provitamin A, this is its main nutritional function. Vitamin A is also known as retinol, which is used as a visual pigment chromophore in the eyes. It is also implicated in the growth and reproductive efficiency of the epithelial tissue. Thus, retinoids have been used in dermatological treatments, such as for acne. The antioxidant property of carotenoids is linked to their capacity to bind with singlet oxygen by conjugated double bonds system. Moreover, β-carotene is used as a food colorant and as a nutritional supplement of vitamin A. In fact, 100% of β-carotene can be converted into vitamin A. The international trade of β-carotene is dominated by private companies such as Roche and BASF.

These companies produced β-carotene through a synthetic approach. Both companies started the synthesis using the same molecule, namely β-ionone but used different methods. Roche used a synthesis for polyenic aldehydes in the form of enol-ether condensation. BASF used the Witting condensation for the production of β-carotene. The aspect that enhances the interest in the natural β-carotene is that it contains several other carotenoids in low concentrations, which provides further health benefits. Only a low percentage of the total β-carotene produced worldwide is natural, and it is a really interesting purpose to increase the natural production of this chemical and make the process more environmentally friendly (Riberio et al. 2011). It was illustrated that globally β-carotene production via herbal sources comprises 2%. Carotenoids are isolated from components of flowers, plants, and fruits. They can be found in vegetables and in fruits, in which orange carrots are the most common source of β-carotene. As said above, industrial carotenoids are produced by extraction and chemical synthesis; chemical synthesis produces hazardous wastes, which are harmful to the environment (Gupta et al., 2022).

The succinic acid is traditionally made from fossil resources used in different ways, such as a chemical intermediate in medicine, in the manufacture of lacquers, and perfume. An interesting application field is that it can be used as an intermediate for producing biodegradable polymers (Rex et al., 2017). In fact, succinic acid was selected as one of the top bio-based chemicals. Moreover, it can be converted into other valuable chemicals such as 1,4-butanediol and tetrahydrofuran. In 2015 the total market for succinic acid was hovering around 30,000 - 50,000 tons per year, but this value has grown, probably because the knowledge of technology for bio-based production has increased. Four companies (Reverdia, Succinity, Bioamber, and

Myriant) commercialized the processes for the bio-based production of succinic acid. Also, it works on the conversion of succinic acid to various derivatives such as Polybutylene succinate (PBS) by BioAmber, Sinopec. The versatility of this molecule has allowed the development of an increasing interest in the synthesis, extraction, and purification of this chemical, with processes that will increase the yield and especially make the process more environmentally friendly (Choi et al., 2016).

2. In silico engineered strain and case studies for yeasts

2.1 Automated in silico strain design of genome-scale models

The genome-scale models are among the most effective tools for *in silico* representation and analyses to tackle strain design and optimization tasks (Palsson, 2015; King et al., 2015b; Nielsen and Keasling, 2016; Lu et al., 2019). Such genome-scale models can include a *metabolic network* of pathways or organisms that offers a representation of microorganisms as realistic as possible (Palsson, 2015). The metabolic network reconstruction is based on the *stoichiometry* of metabolic chemical reactions and metabolites involved as products or reagents.

In the stoichiometry of metabolic chemical reactions and metabolites, the stoichiometric coefficients are grouped in a sparse matrix S of size $m \times n$, where m is the number of metabolites in the model (each row is a unique compound) and n is the number of simulated reactions (each column is a reaction). The matrix S permits the definition of a system of m mathematical constraints, given by a mass balance equation: $Sv = 0$, where v is a vector of variables representing flux through all reactions in a steady-state. Solving $Sv = 0$ is known as flux balance analysis (FBA) (Palsson, 2015).

The constraints $Sv = 0$ ensure that the balance of mass for each metabolite in a model holds. Moreover, in all reconstructions, $m < n$, due to more reactions than the metabolites). However, such a system of constraints defines an *unconstrained* solution space. Therefore, a *feasible solution space* is usually defined by applying a lower bound and upper bound lb_i and ub_i on each flux v_i. These bounds are applied to internal

reactions and external exchange reactions simulating the uptake and secretions of chemicals to-and-from the extracellular regions. These bounds mainly represent the environmental conditions in which a cell is located and its metabolic foot printing.

Once a feasible solution space is defined, it is possible to select a specific point v that is optimal for a specific objective function v_{bio}, which is defined as linear combination vectors v and c, where c is a vector of weights indicating the quantity of each reaction (flux intensity) that contributes to the objective function. The objective function usually maximizes flux through a single artificial reaction simulating biomass production or the *growth rate.* The definition of this objective function is crucial for the precision of fluxes. This stoichiometric definition of objective is often complicated in both its composition and values of coefficients, even though it only refers to the exponential growth phase of a cell's cycle. The optimal point v can be found by optimizing systems by solving $Sv = 0$ and by using linear programming as

$$maximize \sum_j^n \quad c_j v_j = c^T v = v_{bio} \tag{1}$$

$$subjected\ to\ Sv = 0 \text{ and } lb_i \leq v_i \leq ub_i, for\ all\ i = 1, \ldots, n\,.$$

When the optimal value v If the growth rate is established, a subsequent optimization called *flux variability analysis* (FVA) (Palsson, 2015) is performed to explore a more *constrained feasible solutions space* where this new constraint is defined by hyperedges of the polytope. FVA allows the definition of flux ranges for a single reaction in the hyperedge.

The range (lower and upper bounds) is critical for evaluating the robustness of single strain prediction. For example, if we consider the production of a compound, a small range could mean that, despite the variance that the predictions might have along the edge-link in the reaction graph, there is still a minimum production always predicted regardless of other fluxes, and this prediction is comparable to the theoretical maximum under stress growth conditions. The parsimonious flux balance analysis (pFBA) is another approach that could

lead to more reliable prediction, specifically for the fluxes of internal metabolic reactions occurring in the internal compartments, different from the extracellular region (Palsson, 2015).

Similar to FVA, the pFBA approach calculates the optimal growth rate from FBA optimization as a constraint and then minimizes the sum of absolute values of fluxes through all reactions in a network. Thus, this new optimization returns a parsimonious distribution of fluxes (optimal array of fluxes) throughout the network, avoiding numerical predictions that do not have any biological justification, a common issue in FBA (Palsson, 2015). For an optimal array of fluxes, we define two quantities to evaluate the results of chemical production: the $yield$ and $productivity$ of specific chemicals. These quantities are respectively defined as:

$$yield = \frac{v_{bio}}{v_{cs}} \tag{3}$$

and

$$productivity = h^{-1} = \frac{v_{bio}}{v_{cs}} \cdot v_{obj} . \tag{4}$$

After normalizing $yield \leq 1$ using the corresponding molar masses v_{cs}, the $yield$ becomes a scalar quantity, expressing the fraction of chemicals produced over the quantity of glucose v_{glc} (in general, the quantity of carbon source v_{cs}) that the cell has used. The theoretical upper bound for yield equals 1, i.e., when all the glucose is converted to the specific chemical without waste. The productivity, instead, is measured as $[h^{-1}]$, which gives the rate of speed at which a product can be obtained, where v_{obj} is the growth rate. Moreover, Using the extreme ranges obtained from the FVA for each reaction, the ranges for $yield$ and $productivity$ can be obtained, and this will give us the energy and carbon source quantity that each reaction uses.

The predictions on the distribution of reaction fluxes obtained in *wild-type* (WT) and *simulated conditions* come from evaluating the metabolic network. The next question is then how the linear programming problem must be changed to reproduce the metabolic engineering techniques, namely, in our case, to simulate the gene deletions. We use a *multiobjective evolutionary algorithm* (MOEA) to address this issue.

2.2 Metabolic engineering frameworks

We created our metabolic engineering framework (Figure 1) by applying MOEA on genome-scale models of yeasts, where fitness functions of the models were computed using FBA. As explained in Figure 1, the genes are present in the genome-scale models through gene protein reaction (GPR) relationships that link the genes with the coded enzymes and coded enzymes with the corresponding catalyzed chemical reaction.

Using a series of logical rules defined by atomic propositions referring to single genes of the simulated genome and logical connectors AND OR, the genes are simulated in the model as a prerequisite for each reaction to be active. The logical connectors help represent complex relationships, such as the isoenzymes and subunits. The rules for each reaction can be evaluated from the atomic true and false values, and the reactions with satisfied propositions (that refer to genes) are considered in the metabolic network. A reaction that must be excluded from the model following one or more gene deletions is simulated, posing the lower and upper bounds equal to 0 and forcing the corresponding flux variable to assume a null value.

Multiobjective optimization algorithms are used for this class of problems since an exhaustive evaluation of all the possible gene deletion combinations is practically infeasible, even using a simple approach such as FBA. The number of different configurations to be considered for such a study is equal to 2^g for g genes. Even when considering a fixed maximum number of deletions, it would be equal to a sum of $\sum d(g)$ for d gene deletions (knockout). Hence, a multiobjective optimization framework is useful for efficiently exploring the large solution space of gene deletions. We propose using a modified version of the multiobjective metabolic engineering algorithm (Patané et al., 2019) to explore the possible sets of deletions and their impact on the growth rate and chemical production (see Supplementary Section A). It is an ad-hoc evolutionary algorithm that follows the principles of natural selection (Kenneth, 2016). That means it starts from an initial population of candidate strains (logical arrays in which each value represents a gene), and the population evolves over a fixed maximum number of generations. In each generation, genetic operators such as mutation and crossover attempt to switch genes on and off (i.e., randomly switch binary components of the array). Then the selection operator selects the candidates with the most promising and best-performing value of one or more *fitness functions (objectives)*. The fitness functions, in our case, are *growth rate* and

production capacity. Such an iterative process evolves a population in the next generation fitter than the previous generation and helps obtain the near-optimal solutions in the final populations.

The optimization of two or more objectives, which in most cases compete for resources available for the cell from the exchange reactions, defines a classical *multiobjective optimization problem,* in which it is possible to define a set of optimal points, called *Pareto Front* (Patanè et al., 2015). Each point in a Front cannot be improved in all the considered objective functions simultaneously. An optimization technique for these problems aims to find such a Front or a good approximation of the problem.

3. Results

3.1 Beta-carotene production in engineered *Yarrowia lipolytica*

3.1.1 Optimal strains of beta-carotene production by Y. lipolytica

The production of β-carotene in *Y. lipolytica* is simulated and evaluated using the iYL619_PCP genome-scale model (Pan and Hua 2012). In addition, based on our literature review, to promote β-carotene production in *Y. lipolytica,* we added the heterologous metabolic pathway of three knock-in genes: geranylgeranyl diphosphate synthase (*GGS1* from *Y. lipolytica*), phytoene synthase/lycopene cyclase, and phytoene dehydrogenase (*carPR* and *carB* from *Mucor circinelloides*) (Larroude et al., 2018; Celinska et al. 2017; Gao et al., 2014). Consequently, the model was modified by adding the corresponding arch (a new pathway) to the metabolic network. After introducing three knock-in genes, the next step is to maximize the expression of these genes within the yeast. This can be implemented by engineering the yeast or by biotechnology strategies like Celińska et al. (2017), where they found that β-carotene production was enhanced by increasing lipogenesis and gene copy number and by identifying the best combination of promoters and genes. For this, they performed a promoter shuffling strategy by using a golden gate toolbox for *Y. lipolytica.*

There are several other examples of the strategy to optimize Y. Lipolytica. For example, Zhang et al. (2020) enhanced the β-carotene production by increasing copies of carB (three copies) and carRP (two copies) genes and overexpressing the genes (GGS1, ERG13, and HMG) correlated with Mevalonate (MVA) pathway. This pathway contributes to the production of carotenoid precursors: isopentenyl pyrophosphate (IPP) and

dimethylallyl pyrophosphate (DMAPP). The DMAPP metabolite is converted into Farnesyl diphosphate (FPP) in multiple-step reactions. The overexpression of the GGS1 enzyme makes the whole of FPP convert into geranylgeranyl pyrophosphate (GGPP) rather than entering into the squalene pathway. From GGPP, two knock-in genes carRP and carB convert GGPP into Phytoene and Lycopene, and in the final step of the pathway, carRP converts lycopene into β-carotene. Moreover, Zhang et al. (2020) successfully analyzed eleven sites for CRISPR/Cas9-mediated heterologous gene knock-in Y. lipolytica and found that four sites are involved in β-oxidation (POX2, POX3, POX4, POX6), six sites belonged to nonfunctional pseudogenes due to frameshift (E1, A1, B1, A2, F1, E2), and the last site LIP1 is engaged in lipid metabolism.

Liu et al. (2021) suggested a modern strategy to optimize the β-carotene production by constructing codon-adapted genes and minimizing the intermediate accumulation, which plays an important role in metabolic balance. The metabolic balance means no accumulation of intermediates at the connecting node when combining upstream and downstream pathways. This methodology inserts the β-carotene biosynthesis pathway consisting of knock-in genes carRA and carB from B. trispora, where these two genes were codon-adapted for a better expression. Here metabolic balance is an important factor. In the biosynthesis of β-carotene, there are four enzymes that limit the rate of the process: tHMGR, GGS1, carRA, and carB. The metabolites (intermediates) that are converted by these enzymes are HMG-CoA, FPP, GGPP, Lycopene, and phytoene. Therefore, an inadequate expression of these enzymes will lead to a high accumulation of these intermediates, which results in a small production of β-carotene. Hence, they overexpressed the genes tHmgR, Ggs1, carRA, and CarB with Snf, Lip1, Pox3, and Pox4 as the target sites, which caused the deletion of these genes that led to increases in lipid body formation that allowed more storage space for β-carotene.

Yang et al. (2021) proposed a new approach that focuses on a new feature called *DID2* genes. This gene is a subunit of the ESCRT (endosomal sorting complex required for transport). This complex is made up of cytosolic protein complexes known as ESCRT-0, ESCRT-I, ESCRT-II, ESCRT-III that, together with other accessory proteins, enable a remarkable way of membrane remodeling. The *DID2* gene in *Y. lipolytica* was amplified and inserted into pJN44, leading to pJN44-Did2. As a subunit of the ESCRT protein complex, the *DID2* improves β-carotene production (increased by 260%) and does not cause metabolic stress for the host

cell. In order to understand why this gene improved β-carotene production in *Y. lipolytica*, they studied the mRNA, protein, and precursor that are part of the β-carotene pathway. From the study of mRNA of *Thmg, Ggs1, carRA, and carB* genes, they realized that due to the introduction of the *DID2* gene, the mRNA levels of β-carotene pathway genes were high. In their study, *DID2* elevated the mRNA level of the β-carotene synthesis pathway genes in *Y. lipolytica*. They found that the *DID2* also increased glucose consumption during the exponential growth phase and stationary phase, which is an important feature in metabolic engineering.

In our study, for an in-silico simulation, we set the growth conditions of yeast and configured two main objectives: maximization of β-carotene production and maintenance of biomass as close as possible to that of the wild type. We employ MOEA for retrieving gene deletions leading to suboptimal strains. MOEA starts from a population of wild-type strains with the highest growth rate prediction but a null β-carotene production and is located at the bottom of the Pareto Front (see Figure 2). Then MOEA explores possible deletions and phenotypes of the resulting strains. The phenotypes tend to cluster in various separate regions. We discovered that these regions are related to specific gene deletions that characteristically change the predictions. Hence the Fronts in Figure 2 appear to be divided into steps. However, the clusters vary only slightly from each other. The MOEA finds the importance of a small set of genes that share a similarity in predictions through its evolutionary optimization procedure. The presence of these clusters highlights the algorithm's efforts in finding other useful deletions. This is particularly evident at the top region of Figure 2, where the algorithm considered several points. Still, none led to other Pareto optimal points, despite the sensible reduction in the growth rate.

Similarly, we analyzed the productivity of the yeast. The productivity values are in the order of magnitude of -4 because both growth rate and production have low values. Furthermore, the glucose level varies among the points, especially those with low growth where the metabolic network does not use all the carbon sources. These differences also change the distribution of the explored points and the Pareto Front; many points have various productivity levels that do not correspond to the productions. Notably, not all the points of the Pareto Front (in Figure 2) are optimal when considering either productivity or growth rate. Instead,

there is a clear trade-off. Thus, the best way of comparing the strains can vary depending on the desired phenotypes and the specificity of applications.

Figure 3 represents a three-dimensional graph to compare minimum productivity, maximum productivity, and growth rates to determine the characteristics of strains. These three parameters are essential to determine which strains are best suited for use in the laboratory. The parameters maximum and minimum productivity measure the quantity of β-carotene production under optimal and non-optimal conditions. Indeed, it is necessary to find a high value of the maximum and minimum productivity (see x-axis and z-axis in Figure 3) to achieve high production of β-carotene. Finding a high value of minimum productivity is more significant because it is challenging to maintain stable optimal growth conditions in both *in vitro* and industrial bioreactors. Hence, a strain where the value of minimum productivity is high ensures a high β-carotene production is possible even in sub-optimal conditions. This gives us a measure of how resistant an obtained strain is. In Figure 3, the most significant strains are at the top right of the graph because they represent the best compromise between yeast growth and β-carotene productivity.

3.1.2 Carbon Source analysis

The relationship between the production of the chemical and the amount of used carbon source measures the yield. For example, our study used glucose as a carbon source for yeast growth and to produce β-carotene. However, different carbon sources can also be used, for example, glycerol.

Larroude et al. (2018) used glycerol (GLY) as a *cheap carbon source* with two different media, rich Yeast Extract- Peptone- Dextrose medium (YPD) and synthetic Yeast Nitrogen Base medium (YNB), and with different carbon source concentrations (10, 20, 30, and 60 g/L), keeping the amount of nitrogen constant. They selected culture media YPD10, YPD60, YNB20, YNB30, YNB60, and YNBGLY60 for β-carotene production and found that the production varied significantly based on culture media usage. Additionally, Braunwald et al. (2013) showed that both the carbon-nitrogen ratio and the applied initial carbon and nitrogen contents influenced the parameters, i.e., high carbon-nitrogen ratio promotes lipid production and other carbon-

based molecules such as carotenoids. However, they found that lipid yield was not affected by ammonium contents, while the carotenoid production decreased significantly at low and high ammonium supply levels.

Therefore, we can suggest that carbon source has little influence on production since glucose and glycerol have similar titer and yields. In addition, a clear correlation between the increase in the initial glucose content and the production titer was found in both rich and synthetic media (Larroude et al., 2018). Noticeably, the production yields for all the YNB based media were similar and were independent of the amount or kind of carbon source used. However, this was not the case for rich media (YPD), and it offered a trade-off between production titer and yield. For example, the best β-carotene titer was 1.5 g/L in YPD60, while the best yield was 0.048 g/g in YPD10. Moreover, in all cases, YPD had higher titers and yields than YNB. (Larroude et al. 2018). Therefore, they selected rich media to further optimize the culture conditions in a controlled fermentation in a bioreactor.

3.1.3 Analysis of strains and relative genes knockout

We analyzed the results of the MOEA simulation to identify the most suitable strains for *in vitro* testing. Table 1 offers growth values of β-carotene production, ATP, NADH, NADPH, and especially FAD(H2). From Table 1, we identify that strain number 7 (row 7 in Table 1) has a significant exponential increase in its energy value parameters, i.e., the values of ATP, NADH, NADPH, and FADH have notably increased (see columns in Table 1). Specifically, the production of ATP rises by 108.72%, NADH by 116.60 %, NADPH by +146.05%, and FAD(H2) by +603.74% for strain number 7. This exponential growth can bring either positive or negative results, which means that a significant amount of energy is required to grow yeasts, but such a substantial concentration of molecules can produce toxic substances as well.

Additionally, Table 1 shows that the strains for the β-carotene production fluctuate only marginally. This is due to the metabolic pathway obtained by adding three knock-in genes, which was common to all the strains, and they only differ for their knockout genes. Table 2 identifies the number and type of genes removed (knocked out) from the genome of the selected strains. We observed that there were between 1 to 7 genes switched off. Hence, to analyze knockout genes and the involved pathway, we used the *Kyoto Encyclopedia*

of Genes and Genomes (KEGG) (Kanehisa et al., 2000). Moreover, we studied the relationship between yield and the number of knockout genes. This relation is shown in Figure 4(left plot), which indicates that the increase in yield is directly proportional to the number of knockouts.

Table 2 allows biotechnological considerations to be made. Namely, as mentioned above, the following is given: the number of genes deleted in each strain, the name of each gene, and the chromosomal location of each. The chromosome belonging to each gene is clarified by the name. Specifically, the chromosome is indicated in the sixth character of the identifier of the corresponding gene. Knowledge of the associated chromosome becomes useful in the transition from in silico to in vitro. This is because the deletion of genes belonging to the same chromosome is greatly facilitated by knockouts.

We examined the strains from Table 2 in which most of the knockout genes were located on the same chromosome. For example, in strain number 2, we notice three knockout genes (YALI0A05379g, YALI0F11935g, YALI0F17996g); as we can see from the name, the first gene is located on chromosome A and the other two on chromosome F. In this strain, we have a β-carotene production of 0.22634 $[mmol \cdot gDW^{-1} \cdot h^{-1}]$. We further study the frequently silenced genes *YALI0F17996g* and *YALI0A05379g* in Table 2. The gene *YALI0F17996g* has a length of 4527 bp, GC%= 53.15 %, and is a part of Chromosome F. This gene translates to a protein that catalyzes the reaction of ergosterol transport. The gene *YALI0A05379g* has a length of 2361 bp, GC%= 51.16 %, and is part of Chromosome A. This gene translates an enzyme, chorismate: L-glutamine aminotransferase, for para-aminobenzoate (PabA) synthase ABZ1. This enzyme is composed of two parts, PabA and PabB. In the absence of PabA and glutamine, PabB converts ammonia and chorismate into 4-amino-4-deoxychorismate (in the presence of Mg2+). On the other hand, the PabA converts glutamine into glutamate only in the presence of stoichiometric amounts of PabB. Additionally, this enzyme is coupled with EC 4.1.3.38, aminodeoxychorismate lyase, to form 4-aminobenzoate. Thus, the reaction catalyzed by this enzyme is chorismate + L-glutamine → 4-amino-4-deoxychorismate + L-glutamate, and the standard Gibbs Free Energy (ΔrG'°) for this reaction is -2.0558853 kcal/mol, which is an exergonic reaction. Therefore, this indicates a spontaneous reaction.

In order to establish the role of *YALI0F17996g* and *YALI0A05379g*, we set up an additional simulation, where we knocked in the heterologous genes (*GGS1, carPR, and carB*) and knocked out *YALI0F17996g* and *YALI0A05379g*. We obtained an increase in β-carotene productivity and a lower D-glucose exchange than a wild-type strain (see Table 3).

3.2 Succinic acid production in Saccharomyces cerevisiae

3.2.1 Optimal strains of succinate production by S. cerevisiae

We changed the MOEA framework to tackle critical points that could affect the precision and reliability of the results for our case study on succinic acid production in *S. cerevisiae.* The analysis of the succinic acid production in *S. cerevisiae* is conducted using the genome-scale metabolic (GEM) yeast model, v. 8.3.1 (Lu et al., 2019). In this genome-scale model, no changes were needed as the pathway for succinate production is already included in the model. Furthermore, for these simulations, we set the bounds of the external exchange reactions of the models to simulate the growth in a rich medium, i.e., the synthetic defined medium for the S. cerevisiae model.

Firstly, we improved the framework based on the knowledge obtained from β-carotene production. We found that some obtained points offered a low growth rate, which signifies an impeded cell metabolism. Therefore, we limit the tolerance of the MOEA to a reduced growth rate compared to the wild type. We set a bound of 10% on this growth rate.

Secondly, using the fluxes predicted by the pFBA, we induced a similar bound on the sum of the fluxes through the reactions in the network that produce the metabolites ATP, GTP, NADH, NADPH, and FADH2. These constraints aimed to improve the quality of the results obtained by the algorithm, ensuring that the strains do not differ excessively from the wild type at every step. In other words, we force the algorithm to explore more extensively a narrow region to produce better results.

Thirdly, we included some restrictions on the gene deletions by not allowing MOEA to delete some of the essential genes and allowing the deletions of genes involved in the synthetic double deletions. For this, we use the database of Heavner and Price, (2015). This restriction helped the algorithm avoid unfeasible mutated

strains that are not always correctly predicted by FBA (Heavner and Price, 2015). Finally, we considered the maximization of productivity. The results of the algorithm using this setting are summarized in Figure 5 (left plot), where we observe a slight reduction in allowable growth. Nonetheless, the algorithm increased the productivity from the initial null point. Thus, a step-like clustering behavior is still present but less prominent than the one obtained for β-carotene shown in Figure 2.

Contrary to the previous simulation, Figure 5 (left plot), instead of growth rate, our final simulation in Figure 5(right) included two extremes: *min* and *max* productivity values as per FVA. The resulting strains of this simulation are shown in Figure 5(right). The number of points in this simulation is in only half of the phenotypic space. The points lying on the highlighted line (dashed line in Figure 5(right)) correspond to strains for which the range of productivity collides to a single value. We speculate that these strains are more robust as the *in-silico* productivity is always ensured when a higher value of min productivity is obtained.

For example, the only point of the Pareto Front on the dashed line in Figure 5 (right plot) is a strain with five gene deletions. Compared to the other points of the front, the points (strains on the dashed line) have lower maximum productivity but have the highest minimum productivity predicted. Thus, the points (strains) with the highest min productivity values should be preferable to the other points to ensure a less competitive minimum prediction, even though other parameters such as the metabolic foot printing should also be considered along with min productivity.

3.2.2 Analysis of strain in succinate production by S. cerevisiae

The analysis of the knockout genes through FBA answers the questions: how knockout of a specific gene influences the production of a specific metabolite within a cell and how it changes the metabolism of yeast. Mathematically, this process is described by a GPR map (Orth et al., 2010). In GPR, the organism's genes are grouped using "Boolean" relationships, which associate each gene to a group based on common reactions catalyzed by their respective associated proteins. From the analysis of S. cerevisiae, we obtained 483 strains derived from gene deletions, which resulted from exploring possible sites of gene deletion by MOEA.

MOEA results provided important information about the choice of strains for *in vitro* testing and the actual succinate production. For example, Figure 4 (right plot) shows the relationship between the maximum Succinate yield changes and the number of knockout genes. Here, we observe that the value of succinate yield increases when the number of knockout increases. This is obvious because the silencing of genes spares energy that strains might use to produce succinate.

For our research, we used a set of parameters to select strains. First, we select 7 significant strains (shown in Table 4). After this selection, we focused on analyzing genes. Mainly, we focus on identifying knockout genes of each strain and their position and length within the genome of S. cerevisiae. In addition, between 3 to 9 genes were silenced on average for each strain, as shown in Table 4. Table 5 reports silenced genes of each strain, and a roman number in the second column in Table 5 explicitly locates chromosomes a silenced gene belonged to.

MOEA, importantly, knockout a varied number of genes for strains, leading to heterogeneous results. For example, Table 4 shows that up to 9 genes were knocked out, and despite a large number of knockouts, the algorithm was able to simulate the life of yeast and, crucially, had an increased Succinate production. Thus, intuitively, knockout genes result from an optimal compromise between the cost of knockout and the production rate of succinate. The analysis of genes was carried out through Genemania (Warde-Farley et al., 2010) software for predicting functions and involved pathways. Similar to our analysis of knockout genes in β-carotene, we characterize the genes that were silenced with higher frequency in succinate production. Table 5 identifies frequently silenced (knockout) genes of S. cerevisiae. These are *GLT1, ALD6, and GPH1*.

The knockout gene *GLT1* of S. cerevisiae encodes for a glutamate synthase (GOGAT) which is essential in central nitrogen metabolism (CNM). CNM contains two pathways [glutaminases (GDA) and GOGAT, which is NADH-dependent, converts one molecule of glutamine and one molecule of α-Ketoglutarate into two glutamate molecules (Guillamon et al., 2001)] for glutamate biosynthesis using glutamine as the sole source of nitrogen. The presence of two pathways makes it harder to choose the most significant routes for the biosynthesis of the end product.

Although the pathway GDA (glutaminases)-encoding genes are unknown, these glutaminases may exist because mutants grow well on glutamine even without the GOGAT enzyme. Some authors (e.g., Tempest et al., 1970) have suggested that the role of the GOGAT pathway, with the concerted action of the glutamine synthetase (GS), is to assimilate ammonium and synthesize glutamate even under shortage of ammonium. However, NADPH-dependent glutamate dehydrogenase (NADPH-GDH) is used to incorporate ammonia during a shortage or excess of nitrogen in other microorganisms. This hypothesis suggests that NADPH-GDH is the main pathway for glutamate biosynthesis. Therefore, physiological studies have been reported that in GOGAT or NADPH-GDH activities, both wild-type and mutant strains are impaired. These show that GOGAT has different roles in different microorganisms (Valenzuela et al., 1998; Barel and MacDonald, 1993), but its function in *S. cerevisiae* is still unclear. Although the clear reason for *GLT1* knockout could not be established, the unclear role of GOAGT in *S. cerevisiae* may conclude that *GLT1* may have less influence on succinate production and the growth of the yeast.

The ALD6 gene of *S. cerevisiae* encodes the cytosolic Mg2+-activated NADP-dependent ACDH and exhibits 60% and 30% activity of wild-type activated acetaldehyde dehydrogenase (ACDH) (Remize et al. 2000). The main cytosolic Mg2+-activated ACDH isoform preferentially uses NADP, and this isoform plays an important role in both ethanol (deletion of ALD6 gene disables the organism to use ethanol as a carbon source) and glucose. Since the deletion of a mutant is feasible on glucose, the enzyme encoded by *ALD6* is not solely responsible for producing cytosolic acetyl-CoA (Meaden et al., 1997) and thus not solely responsible for succinate production.

The *GPH1* gene of *S. cerevisiae* translates to glycogen phosphorylase enzyme, which is an essential allosteric enzyme in carbohydrate metabolism, but not essential for the life of the yeast. Hence, this gene draws our attention because we need to answer our question: ‘why does the genetic algorithm knock out this gene frequently?’ To do this, using Escher (King et al. 2015), we analyzed the strains in which we knocked out one by one all of these genes. In the case of the *GPH1* gene, we noticed a particular rearrangement of the metabolism. We found that the yeast’s response to gene knockout was to activate the transcription of the isocitrate lyase enzyme. This enzyme catalyzes the conversion of isocitrate into succinate. Therefore, we

found evidence that the MOEA algorithm can find new information for the production of specific chemicals (see Supporting Information in Section C).

We also analyzed the second strain of Table 4, in which 4 genes have been knocked out (PDB1, GLT1, ALD6, and GPH1), with a model that includes Expression and Thermodynamics FLux (ETFL), which efficiently integrates RNA and protein synthesis with traditional genome-scale metabolic models. To adapt this model for *Saccharomyces cerevisiae*, Oftadeh et al. (2021) developed yETFL, in which they increase the original formulation with supplementary considerations for biomass composition, the compartmentalized cellular expression system, and the energetic costs of biological processes. The results of this analysis are that the strain with 4 Knockout genes loses 1.52% of the growth rate compared to the wild type.

4. Conclusions

We developed an automated tool for in silico implementation of genetic deletions and a precision modulation of the phenotype of the host yeasts to select optimized strains. We implemented a multiobjective evolutionary algorithm and its refinements to optimize strains to obtain results that can realize a sustainable synthesis of metabolic precursors used in large-scale manufacturing processes. Our approach optimizes two yeasts *Y. lipolytica* and *S. cerevisiae*. This provides many strains with varied features for selecting the best strains based on the production capacity of chemicals (β-carotene and succinate) with the lowest biomass losses. By examining knockout genes, we characterize pathways influenced by the knockout. We found 7 strains of *Y. lipolytica* and 7 strains of *S. cerevisiae* capable of producing a high amount of β-carotene and succinate. Such in silico processes of strains creation save costs, leading to a high bio-sustainability.

References

Abdel-Mawgoud, A. M., Markham, K. A., Palmer, et al. (2018) Metabolic engineering in the host Yarrowia lipolytica. Metabolic Engineering, 50, 192-208.

Agren R, Otero JM, Nielsen J. (2013) Genome-scale modeling enables metabolic engineering of Saccharomyces cerevisiae for succinic acid production. J Ind Microbiol Biotechnol 40(7):735–47.

Barel, I., and MacDonald, D. W. (1993) Enzyme defects in glutamate-requiring strains of Schizosaccharomyces pombe. FEMS Microbiology Letters, 113(3), 267-272.

Braunwald, T., Schwemmlein, L., Graeff-Hönninger et al. (2013) Effect of different C/N ratios on carotenoid and lipid production by Rhodotorula glutinis. *Applied Microbiology and Biotechnology*, *97*(14), 6581-6588.

Cao, Y., Zhang, R., Sun, C. et al. (2013) Fermentative succinate production: an emerging technology to replace the traditional petrochemical processes. *BioMed Research International*, *Article ID 723412*.

Celińska, E., Ledesma-Amaro, R., Larroude, M. et al.(2017). Golden Gate Assembly system dedicated to complex pathway manipulation in Yarrowia lipolytica. *Microbial Biotechnology*, *10*(2), 450-455.

Chen Y, Xiao W, Wang Y. et al. (2016). Lycopene overproduction in Saccharomyces cerevisiae through combining pathway engineering with host engineering. Microb Cell Fact. 15:113.

Choi S, Song H, Lim SW et al. (2016) Highly selective production of succinic acid by metabolically engineered Mannheimia succiniciproducens and its efficient purification. Biotechnol Bioeng 113(10):2168–2177.

Ciccazzo A., Conca P., Nicosia G. et al. (2008) An advanced clonal selection algorithm with ad-Hoc network-based hypermutation operators for synthesis of topology and sizing of analog electrical circuits. In the 7th International conference on artificial immune systems – ICARIS, 10th–13th August 2008, Phuket, Thailand. Springer, LNCS, 5132, pp. 60–70.

Clarke, K G (2013) The oxygen transfer rate and overall volumetric oxygen transfer coefficient. *Bioprocess Engineering*. p147-170.

Coussement P, Bauwens D, Maertens J et al. (2017) Direct combinatorial pathway optimization. ACS Synth Biol. 6:224–32.

Deb, K. (2001) *Multiobjective optimization using evolutionary algorithms*. New York: Wiley.

Franco-Duarte, R. Bessa, D.et al. (2017) Genomic and transcriptomic analysis of Saccharomyces cerevisiae isolates with focus in succinic acid production. *FEMS Yeast Research*, *17*(6).

Gao, S., Han, L., Zhu, L. et al. (2014) One-step integration of multiple genes into the oleaginous yeast Yarrowia lipolytica. *Biotechnology Letters*, *36*(12), 2523-2528.

Gao, S., Tong, Y., Zhu et al. (2017) Iterative integration of multiple-copy pathway genes in Yarrowia lipolytica for heterologous β-carotene production. Metabolic Engineering, 41, 192-201.

Guillamón, J. M., van Riel, N. A. Giuseppin et al. (2001) The glutamate synthase (GOGAT) of Saccharomyces cerevisiae plays an important role in central nitrogen metabolism. FEMS Yeast Research, 1(3), 169-175.

Gupta I., Sveda Nashvia A., B. P. Panda, Mohd Muieeb. (2022) β-Carotene—production methods, biosynthesis from Phaffia rhodozyma, factors affecting its production during fermentation, pharmacological properties: A review

Heavner, B. D., and Price, N. D. (2015) Comparative analysis of yeast metabolic network models highlights progress, opportunities for metabolic reconstruction. PLoS Computational Biology, 11(11), e1004530.

Henke NA, Wiebe D, Perez-Garcia F et al. (2018) Coproduction of cell-bound and secreted value-added compounds: simultaneous production of carotenoids and amino acids by *Corynebacterium glutamicum*. Bioresour Technol. 247:744–52.

Kanehisa M. and Goto S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28, 27-30.

Kenneth A. De Jong. (2016) Evolutionary computation: A Unified Approach. MIT press.

Kildegaard, K. R., Adiego-Pérez, B., Belda et al. (2017) Engineering of Yarrowia lipolytica for production of astaxanthin. *Synthetic and Systems Biotechnology*, *2*(4), 287-294.

King, Z. A., Dräger, A., Ebrahim, A et al. (2015a) Escher: a web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLoS Computational Biology*, 11(8), e1004321.

King, Z. A., Lloyd, C. J., Feist et al. (2015b) Next-generation genome-scale models for metabolic engineering. *Current Opinion in Biotechnology*, *35*, 23-29.

Larroude, M., Celinska, E. Back, A. et al.(2018) A synthetic biology approach to transform Yarrowia lipolytica into a competitive biotechnological producer of β-carotene. *Biotechnology and Bioengineering*, *115*(2), 464-472.

Lee JW, Yi J, Kim TY et al. (2016) Homo-succinic acid production by metabolically engineered Mannheimia succiniciproducens. Metab Eng 38:409–417.

Lee SJ, Lee DY, Kim TY et al.(2005) Metabolic engineering of Escherichia coli for enhanced production of succinic acid, based on genome comparison and in silico gene knockout simulation. Appl Environ Microbiol 71(12):7880–7887.

Lee SY, Hong SH, Moon SY. (2002) In silico metabolic pathway analysis and design: succinic acid production by metabolically engineered Escherichia coli as an example. Genome Inf 13:214–223.

Liu Liang, Qu Yu Ling, Dong Gui Ru et al. (2021) Elevated β-Carotene Production Using Codon-Adapted CarRA&B and Metabolic Balance in Engineered Yarrowia lipolytica. Frontiers in Microbiology 12: 627150.

Lu, H., Li, F., Sánchez et al.(2019) A consensus S. cerevisiae metabolic model Yeast and its ecosystem for comprehensively probing cellular metabolism. Nature Communications, 10(1), 1-13.

Mantzouridou F, Roukas T, Achatz B. (2005) Effect of oxygen rate on β-carotene production from synthetic medium by Blakeslea trispora in shake flask culture. Enzyme Microb Technol. ;37:687–94.

Mantzouridou FT, Naziri E. (2017) Scale translation from shaken to diffused bubble aerated systems for lycopene production by Blakeslea trispora under stimulated conditions. Appl Microbiol Biotechnol. 101:1845–56.

Meaden, P. G., Dickinson, F. M., Mifsud A. et al. (1997) The ALD6 gene of Saccharomyces cerevisiae encodes a cytosolic, Mg2+-activated acetaldehyde dehydrogenase. Yeast, 13(14), 1319-1327.

Meng, J., Wang, B., Liu, D. et al. (2016) High-yield anaerobic succinate production by strategically regulating multiple metabolic pathways based on stoichiometric maximum in Escherichia coli . Microb Cell Fact 15, 141.

Nghiem, N. P., Kleff, S., & Schwegmann, S. (2017) Succinic acid: technology development and commercialization. Fermentation, 3(2), 26.

Nielsen, J., & Keasling, J. D. (2016) Engineering cellular metabolism. Cell, 164(6), 1185-1197.

Oftadeh O., Salvy P., Masid, M. *et al. (2021)* A genome-scale metabolic model of *Saccharomyces cerevisiae* that integrates expression constraints and reaction thermodynamics. *Nat Commun* 12**,** 4790.

Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010) What is flux balance analysis? *Nature Biotechnology*, *28*(3), 245-248.

Otero, J. M., Cimini D., Patil K. R. et al.(2013) Industrial systems biology of Saccharomyces cerevisiae enables novel succinic acid cell factory. *PloS One*, *8*(1), e54144.

Palsson, B. Ø. (2015) *Systems Biology: Constraint-based Reconstruction and Analysis*. Cambridge University Press.

Pan, P., & Hua, Q. (2012) Reconstruction and in silico analysis of metabolic network for an oleaginous yeast, Yarrowia lipolytica. *PloS One*, *7*(12), e51535.

Patané A., Jansen G., Conca P. et al.(2019) Multiobjective optimization of genome-scale metabolic models: the case of ethanol production. *Annals of Operations Research*, *276*(1), 211-227.

Patane, A., Santoro, A., Costanza, J., et al. (2015) Pareto optimal design for synthetic biology. *IEEE Transactions on Biomedical Circuits and Systems*, *9*(4), 555-571.

Patil, K. R., Rocha, I., Förster, J., & Nielsen, J. (2005) Evolutionary programming as a platform for in silico metabolic engineering. *BMC bioinformatics*, *6*(1), 1-12.

Remize, F., Andrieu, E., & Dequin, S. (2000) Engineering of the pyruvate dehydrogenase bypass in Saccharomyces cerevisiae: role of the cytosolic Mg2+ and mitochondrial K+ acetaldehyde dehydrogenases Ald6p and Ald4p in acetate formation during alcoholic fermentation. *Applied and Environmental Microbiology*, *66*(8), 3151-3159.

Rex E., Rosander, E., Røyne, F., et al. (2017) A systems perspective on chemical production from mixed food waste: The case of bio-succinate in Sweden, Resources, Conservation and Recycling, Volume 125, Pages 86-97.

Ribeiro, B.D., Barreto, D.W. & Coelho, M.A.Z. (2011) Technological Aspects of β-Carotene Production. Food Bioprocess Technol 4, 693–701.

Singh A, Soh KC, Hatzimanikatis V et al. (2011) Manipulating redox and ATP balancing for improved production of succinate in E. coli. Metab Eng 13(1):76–81

Sol Choi, Chan Woo Song, Jae Ho Shin et al. (2015) Biorefineries for the production of top building block chemicals and their derivatives, Metabolic Engineering, Volume 28, Pages 223-239.

Su A, Chi S, Li Y et al. (2018) Metabolic redesign of Rhodobacter sphaeroides for lycopene production. J Agric Food Chem. 66:5879–85.

Tempest DW, Meers JL, Brown CM. (1970) Synthesis of glutamate in Aerobacter aerogenes by a hitherto unknown route. Biochem J. 117(2):405-7.

Valenzuela, L., Ballario, P., Aranda, C. et al. (1998) Regulation of expression of GLT1, the gene encoding glutamate synthase in Saccharomyces cerevisiae. *Journal of Bacteriology*, *180*(14), 3533-3540.

Valenzuela, L., Guzman-León, S., Coria et al. (1995) A NADP-glutamate dehydrogenase mutant of the petit-negative yeast Kluyveromyces lactis uses the glutamine synthetase-glutamate synthase pathway for glutamate biosynthesis. *Microbiology*, *141*(10), 2443-2447.

Vilela, A. (2019) Use of nonconventional yeasts for modulating wine acidity. *Fermentation*, *5*(1), 27.

Wang Q, Chen X, Yang Y et al. (2006) Genome-scale in silico aided metabolic analysis and flux comparisons of Escherichia coli to improve succinate production. Appl Microbiol Biotechnol. 73(4):887-94.

Warde-Farley, D., Donaldson, S. L., Comes et al. (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Research, 38, W214-W220.

Yang F, Liu L, Qiang S et al. (2021) Enhanced β-carotene production by overexpressing the DID2 gene, a subunit of ESCRT complex, in engineered Yarrowia lipolytica. Biotechnol Lett. :1799-1807.

Yang J, Wang Z, Zhu N et al. (2014) Metabolic engineering of Escherichia coli and in silico comparing of carboxylation pathways for high succinate productivity under aerobic conditions. Microbiol Res 169(5–6):432–440.

Zhang, X. K., Wang, D. N., Chen, J. et al.(2020) Metabolic engineering of β-carotene biosynthesis in Yarrowia lipolytica. Biotechnology Letters, 42(6), 945–956.

Zhao J, Li Q, Sun T et al. (2013) Engineering central metabolic modules of Escherichia coli for improving b-carotene production. Metab Eng. 17:42–50.

List of Figures

Figure 5. Succinic acid production in S. cerevisiae. Pareto Fronts are shown in red and feasible solutions are in blue dots. ***Left plot*:** The trade-off between the competitive objectives (Succinate production versus growth rate) constitutes the observed Pareto Front. ***Right plot*:** The trade-off between the competitive objectives (Min and Max Productivity) constitutes the observed Pareto Front. Minimum productivity is computed to highlight the most robust strains.

Figure 1

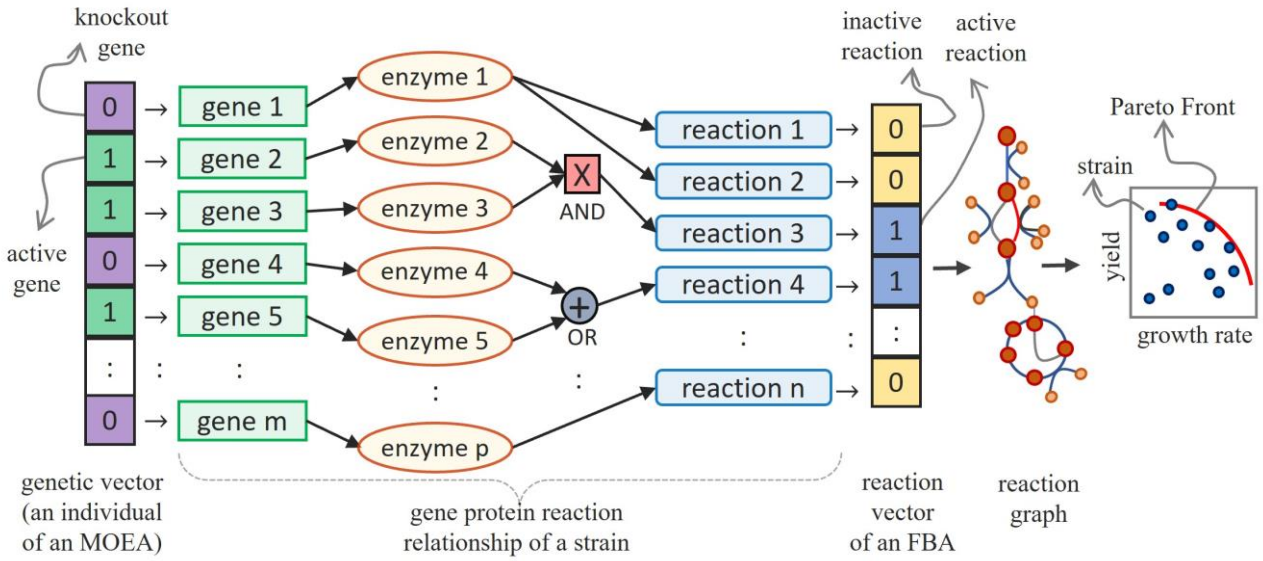


Figure 2

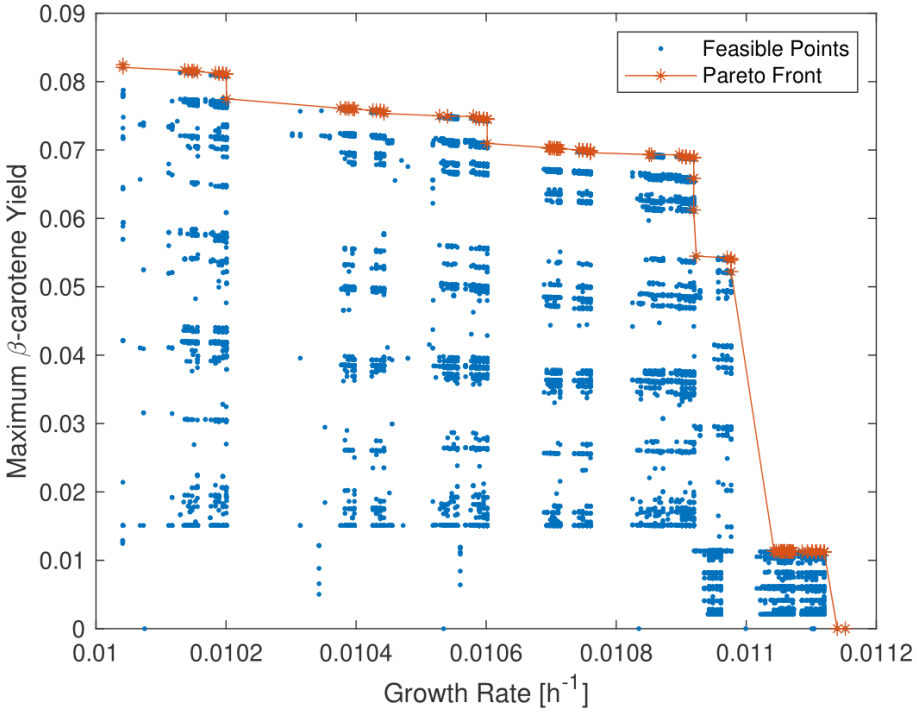


Figure 3

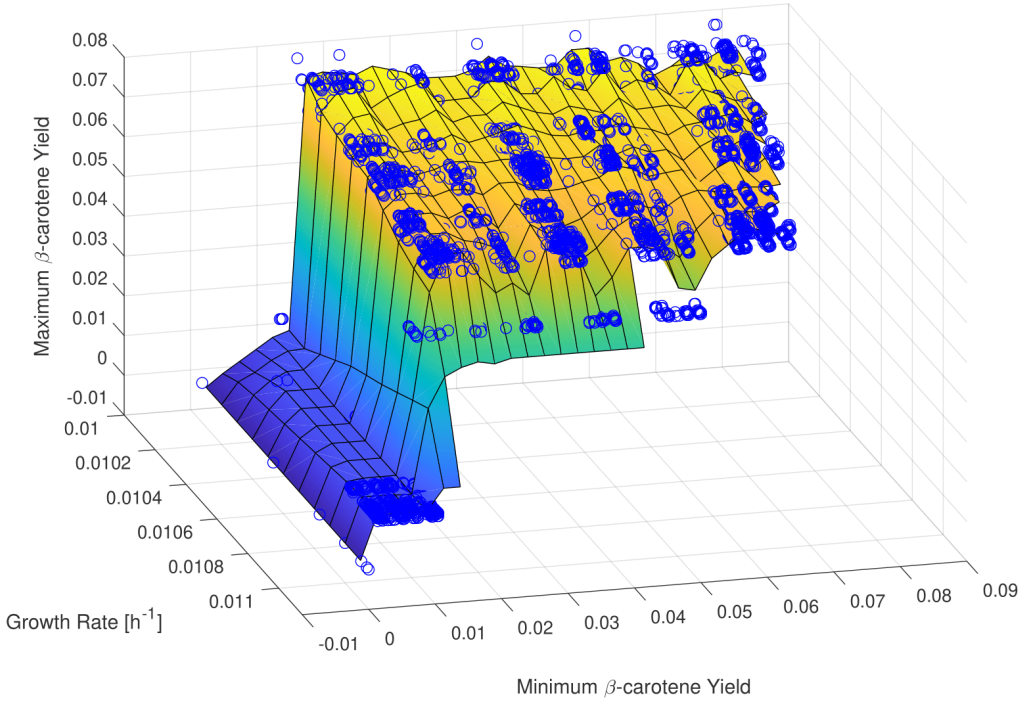
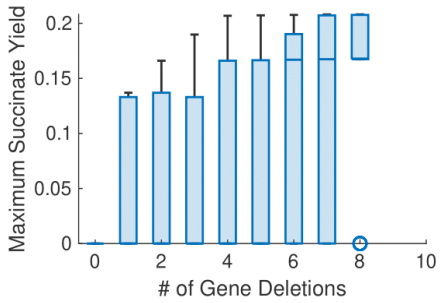


Figure 4



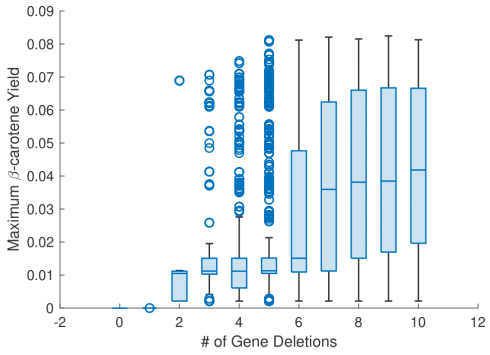
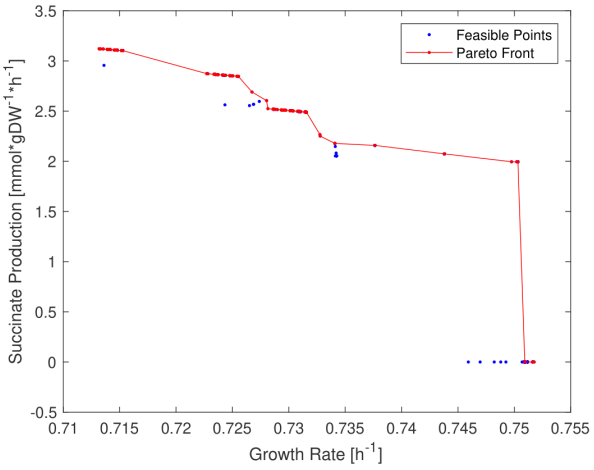
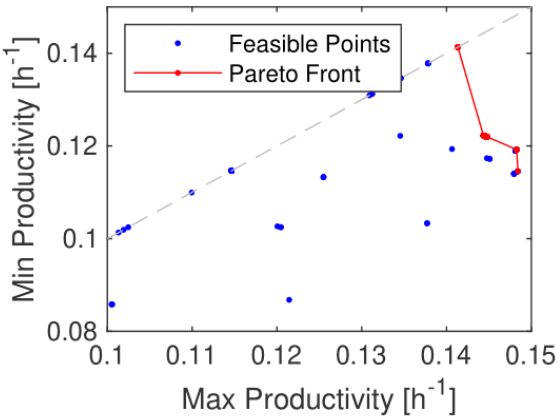


Figure 5





List of Tables

Table 1. The 7 selected significant strains of Y. lipolytica from the Pareto Front. We select 20 strains based on the parameters shown in columns. The first row shows results on wild-type (WT) strain. Column 1 is the numbering of strains. The second column is the biomass (the total mass of all living material in a specific area, habitat, or region). Other columns from left to right are β-carotene production (the quantity of β-carotene produced by strains), adenosine triphosphate (ATP) production, NAD(H) production, NADP(H) production, FAD(H2) production. The last column is the number of knockout genes of strains. We focused primarily on two parameters: biomass and β-carotene production. Thus, we selected the strains with smaller biomass loss and higher β-carotene production. The highest values are indicated in bold.

Table 2. Knockout (KO) genes of Y. lipolytica**.** Results identify the genes that were removed from the genome of Y. lipolytica of each strain and explain this removal. Strains in rows are arranged in the ascending order of the number of their KO genes. The frequently occurring KO genes are YALI0A05379g and YALI0F17996g.

Table 3. Comparison between wild type and strains obtained by deletion of YALI0A5379g and YALI0f17996g genes, which were frequently silenced during multiobjective optimization.

Table 4. The 7 selected significant strains of S. cerevisiae from the observed Pareto Front. The parameters from left to right related strains are Max Productivity, Min Productivity, Max Yield, Min Yield, Succinate Production, Biomass, and Knockout (KO) genes. The values of parameters Succinate Production and Biomass are used to select 7 strains. The strains in rows are arranged in the ascending order of the number of their KO genes. Row 1 indicates wild-type (WT) strain. The highest values are indicated in bold.

Table 5. Knockout (KO) genes of S. cerevisiae**.** For each strain in Rows, the number of silenced (KO) genes are reported in column 1, names of KO genes are reported in column 2, and the chromosomes they belong to are indicated in column 3 by a roman number (note that the budding yeast S. cerevisiae has a 16-chromosome organization). In column 2, from left to right, the acronym provides a description of the genes, where Y indicates yeast's unknown sequence, the second letter represents the chromosome, the third letter indicates the left or right arm of the chromosome, the number indicates the sequence of the open reading frame (ORF), and the last letter W or C represents Watson (5'→3') or Crick strand respectively. Column 3 shows that the standard name of a gene is composed of three letters followed by a number and a roman number written in brackets indicating the chromosome it belongs to, the final letter if an uppercase character indicates a dominant gene, while if it is a lowercase character, it indicates a recessive gene.

Table 1

No. of Strain	Biomass (WT variation)	β-Carotene production $[mmol \cdot gDW^{-1} \cdot h^{-1}[mmol \cdot gDW^{-1} \cdot h^{-1}]]$	ATP Production (WT var. %) $[mmol \cdot gDW^{-1} \cdot h^{-1}[mmol \cdot gDW^{-1} \cdot h^{-1}]]$	NAD(H) Production (WT var. %) $[mmol \cdot gDW^{-1} \cdot h^{-1}[mmol \cdot gDW^{-1} \cdot h^{-1}]]$	NADP(H) Production (WT var. %) $[mmol \cdot gDW^{-1} \cdot h^{-1}[mmol \cdot gDW^{-1} \cdot h^{-1}]]$	FAD(H2) Production (WT var. %) $[mmol \cdot gDW^{-1} \cdot h^{-1}[mmol \cdot gDW^{-1} \cdot h^{-1}]]$	No. of KO
WT	0.011152362	0	62.68911709	11.83086467	29.84459341	0.128944987	
1	**0.01112 (-0.29%)**	0.031725	62.7378 (+0.08%)	11.5504 (0.19)	29.7883 (-0.19%)	0.12895 (0%)	1
2	0.010907 (-2.20%)	0.22634	73.8947 (+17.87%)	14.7743 (+24.88%)	32.5591 (+9.10%)	0.90737 (+603.69%)	3
3	0.010897 (-2.29 %)	0.22736	69.2036 (+10.39%)	14.7437 (+27.40%)	32.5388 (+9.03%)	**0.91146 (+606.86%)**	4
4	0.010897 (-2.29%)	0.22736	69.2036 (+10.39%)	14.7437 (+27.40%)	32.5388 (+9.03%)	**0.91146 (+606.86%)**	4
5	0.010885 (-2.40%)	0.22736	73.8978 (+17.88%)	14.776 (+24.89%)	32.5414 (+9.04%)	**0.91146 (+606.86%)**	5
6	0.010838 (-2.82%)	**0.22737**	73.7797 (+17.69%)	14.7484 (+24.66%)	32.4141 (+8.61%)	**0.91146 (+606.86%)**	6
7	0.010619 (-4.78%)	0.22637	**130.8433 (+108.72%)**	**25.6254 (+116.60%)**	**73.4325 (+146.05%)**	0.90744 (+603.74%)	6

Table 2

No. of Strain	No. of KO	Genes KO
1	1	YALI0F17996g
2	3	YALI0A05379g, YALI0F11935g, YALI0F17996g
3	4	YALI0A5379g, YALI0E22649g, YALI0F15587g, YALI0F17996g
4	4	YALI0A5379g, YALI0B15598g, YALI0F15587g, YALI0F17996g
5	5	YALI0A04983g, YALI0A05379g, YALI0E22649g, YALI0F15587g, YALI0F17996g
6	6	YALI0A04983g, YALI0A05379g, YALI0C23408g, YALI0E22649g, YALI0F15587g, YALI0F17996g
7	6	YALI0A05379g, YALI0C04433g, YALI0D06325g, YALI0E16643g, YALI0E26004g, YALI0F17996g

Table 3

Reactions	Wild Type	Deletion of YALI0A5379g and YALI0f17996g
Growth Rate [h^{-1}] (WT variation %)	0.011152362	0.010919286 (-2.16%)
Max Productivity [h^{-1}]	0	0.062331504
Min Productivity [h^{-1}]	0	0.059657794
D-glucose exchange (WT variation %) [$mmol \cdot gDW^{-1} \cdot h^{-1}$]	-3.034654555	-3.631218038 (-16.42%)

Table 4

Max Productivity $[h^{-1}]$	**Min Productivity** $[h^{-1}]$	**Max Yield**	**Min Yield**	**Succinate Production** $[mmol \cdot gDW^{-1} \cdot h^{-1}]$	**Biomass** (WT variation)	**KO**
0	0	0	0	0	0.751751629	WT
0.14064	0.10696	**0.2395**	0.19708	2,9562	0.71361 (-5.07%)	4
0.14795	0.114	0.2295	**0.20685**	3.1027	**0.71528 (-4.85%)**	4
0.14795	0.114	0.2295	**0.20685**	3.1027	**0.71528 (-4.85%)**	4
0.14795	0.114	0.2295	**0.20685**	3.1027	**0.71528 (-4.85%)**	4
0.14795	0.114	0.2295	**0.20685**	3.1027	**0.71528 (-4.85%)**	4
0.14795	0.114	0.2295	**0.20685**	3.1027	**0.71528 (-4.85%)**	4
0.14841	**0.11455**	0.2282	0.2081	**3.1215**	0.71319 (-5.13%)	9

Table 5

KO	**Genes Knockout**	**Genes KO (standard name) and chromosomic locations**
4	YDL171C, YPL061W, YPR160W, YPR127W	GLT1(IV), ALD6 (XVI), GPH1 (XVI), YPR127W
4	YBR221C, YDL171C, YPL061W, YPR160W	PDB1(II), GLT1(IV),ALD6 (XVI), GPH1(XVI)
4	YDL171C, YGR193C, YPL061W, YPR160W	GLT1(IV), PDX1(VIII), ALD6(XVI), GPH1(XVI)
4	YDL171C, YNL071W, YPL061W, YPR160W	GLT1(IV), LAT1(XIV), ALD6 (XVI), GPH1 (XVI)
4	YDL171C, YER178W, YPL061W, YPR160W	GLT1(IV), PDA1(V), ALD6(XVI) , GPH1(XVI)
4	YDL171C, YHR002W, YPL061W, YPR160W	GLT1(IV), LEU5(VIII), ALD6(XVI) , GPH1(XVI)
9	YDL171C, YHR144C, YJR105W, YLR209C, YNL071W, YNL169C, YOR175C, YPL061W, YPR160W	GLT1(IV), DCD1(XII), ADO1(X), PNP1(XII), LAT1 (XIV), PSD1(XIV), ALE1(XV), ALD6 (XVI), GPH1 (XVI)

Supplementary Information

Pareto Optimal Metabolic Engineering for the Growth-coupled Overproduction of Sustainable Chemicals

Matteo N. Amaradio[1,*], Varun Ojha[2,*], Giorgio Jansen[3,*], Massimo Gulisano[4], Jole Costanza[5], Giuseppe Nicosia[1,3]

[1]Department of Biomedical & Biotechnological Sciences, University of Catania, Catania Italy
[2]Department of Computer Science, University of Reading, Reading, United Kingdom
[3]Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom
[4]Department of Drug Science, University of Catania, Catania, Italy
[5]National Institute of Molecular Genetics, Milan, Italy

Summary:

This supplementary Section is divided into six sections. Section A describes the 'pseudocode' used in our work. Section B presents a comparison table of different strategies used for the production of carotenoids, where we specifically focus on β-*Carotene production.* Subsequently, Section C compares different studies based on succinate production and different strategies used for its fabrication. Section D describes the analyses conducted by silencing the three genes (GLT1, ALD6, and GPH1) of interest for S. cerevisiae. Finally, Section E supplements a list of abbreviations.

A. Pseudocode

The pseudocode presented in Algorithm 1 is based on the NSGA-II algorithm (Deb 2001). The NSGA-II algorithm works by sampling from the optimization problem input domain an initial set of candidate solutions to the optimization problem, i.e., a population, and it iteratively attempts to optimize the problem objective function by applying to the population a set of *evolutionary operators* (Deb 2001). The parameters of the algorithm are (1) *pop,* the size of the population; (2) *maxGen,* the maximum number of *generations* to be performed; (3) *dup,* the strength of the *cloning operator* (Ciccazzo et al. 2008); and (4) *uKC,* the maximum knockout cost allowed to be taken into account by the algorithm (Patanè et al. 2019). Therefore, we have an initial population that is randomly initialized by *Initpop,* which samples the domain of the problem by applying a few random mutations to the wild-type strain. Then, we apply FBA to each strain in the population of candidate strain $P^{(gen)}$ in order to evaluate the production rate of metabolites and the corresponding growth rate; we measure rank and crowding distance for each member of the population. Rank ensures the Pareto-orientation of our procedure, redirecting the search towards the problem Pareto Front. The *crowding distance is* a rough estimation of the population density near each candidate solution.

During the main loop of the optimization, candidate solutions in *unexplored* regions of the objective space (thus having small values of crowding distance) are preferred to those which lie in "crowded" regions of the objective space. This has the purpose of obtaining good approximations of the actual Pareto Front of the problem. We then initialize the generation counter and enter the main loop, which is performed *maxGen* times. At the beginning of each generation, the *Selection* procedure generates a mating pool $Pool^{(gen)}$ by selecting individuals from the current population $P^{(gen)}$. This is done following a *binary tournament selection* approach. Namely, tournaments are performed until there are [*pop*/2] individuals (*parents*) in the mating pool. Each tournament consists of randomly choosing two individuals from $P^{(gen)}$ and putting the best of the two individuals (in terms of rank and crowding distance) into $Pool^{(gen)}$. *Children are* thus generated from the *parents* by using binary *mutation*. Namely, we randomly generate *dup* different children from each parent, generating the $Q_{dup}{}^{(gen)}$. Then, we keep only the best solution of these *dup* children for each parent, hence

defining the actual offspring set $Q^{(gen)}$. This is because many of the mutations allowed in an FBA model are *lethal* mutations, i.e., they severely compromise the bacteria's growth. Of course, a greater value for *dup* implies that feasible mutations are more likely to be found, whereas smaller values reduce the computational burden of the optimization. In order to achieve this, we first ensure that each individual of $Q^{(gen)}$ is feasible with respect *dup* to our optimization problem (i.e., it has less than *uKC* knockouts). Namely, if a child is not in the allowed region, we randomly knockin genes until it is forced back to the feasible region.

We, therefore, evaluate the biomass and metabolites production of each new individual, and the algorithm computes new values of rank and crowding distance for each individual. Procedure *BestOutOfDup* selects from each of the *dup* children of each parent the best one and puts it in the $Q^{(gen)}$ set. Finally, procedure *Best* generates a new population of *pop* individuals, considering the current best individuals and children. The output of the optimization algorithm is the union of the populations of all the generations. We then analyze the optimization results by means of Pareto analysis, hence computing the *observed* Pareto fronts, i.e., the set of $U_{gen}\ P^{(gen)}$ elements which are not dominated by any other element in $U_{gen}\ P^{(gen)}$ (Notice that $U_{gen}\ P^{(gen)}$ covers only a portion of the feasible region. Hence, we talk about observed Pareto optimality) (Patané et al. 2019).

Algorithm 1. Pareto Optimal Metabolic Engineering (POME) Algorithm

POME Algorithm (*Genome-scale metabolic model, medium, carbon source,* Oxygen level, *pop*, *maxGen*, *dup*, *uKC*)

objective_function_1 ← Biomass /* maximization of the biomass */

objective_function_2 ← β-Carotene production (or Succinate Production, Yield and Productivity) /* maximization of the chemical or maximization of the Yield/Productivity */

objective_function_3 ← Number_of_Used_Chromosomes /* Minimization of the used chromosomes */

gen ← 0

$P^{(gen)}$ ← *InitPop*(*pop*)

FBA($P^{(gen)}$)

Rank_and_crowding_distance($P^{(gen)}$)

while (*gen* < *maxGen)* **do**

$Pool^{(gen)}$ ← *Selection*($P^{(gen)}$,[*pop/2]*)

$Q_{dup}^{(gen)}$ ← *GenOffspring*($Pool^{(gen)}$, *dup*)

$Q_{dup}^{(gen)}$ ← *Force_to_feasible*($Q^{(gen)}$,*uKC*)

FBA($Q_{dup}^{(gen)}$)

Rank_and_crowding_distance($Q_{dup}^{(gen)}$)

($Q^{(gen)}$) ← *BestOutOfDup*($Q_{dup}^{(gen)}$, *dup*)

$P^{(gen+1)}$ ← *Best*($P^{(gen)} \cup Q^{(gen)}$, *pop*)

gen ← *gen* + 1

end_while

return (U_{gen} $P^{(gen)}$)

B. Comparison of algorithms and methods for β-Carotene production

In this section, we present a list where we integrate information about our work (we added our best strains) and other research to provide an exhaustive summary of methods and strategies used to produce β-Carotene (see Table B). We debated not only Y. lipolytica but other microorganisms: *Blakeslea trispora, Escherichia coli, Saccharomyces cerevisiae, Corynebacterium glutamicum, Rhodobacter sphaeroides, and Xanthophyllomyces dendrorhous.* The engineering strategies utilized are different. For example, in the case of *Blakeslea trispora,* Mantzouridou et al. (2005) used the control of Oxygen transfer rate (OTR). This methodology scales the OTR under steady-state conditions where the dissolved oxygen concentration remains constant. Under this condition, the OUR (Oxygen uptake rate) of the microorganism equals the OTR. The OUR is calculated by means of an oxygen-mass balance around the reactor, and since OUR = OTR, OTR is quantified as a result. An oxygen-mass balance around the reactor is conceptually defined as: Rate of oxygen entering- the rate of oxygen exiting- the rate of oxygen used equally to the rate of accumulation of oxygen in the system (Clarke 2013). This was done to test the effect of OTR on β-Carotene production. The results indicated that the concentration of β-carotene (704.1 mg/l) was the highest in a culture grown at a maximum OTR of 20.5 mmol/(l h) (Mantzouridou et al. 2005).

Another engineering strategy used by Mantzouridou et al. (2017) to produce lycopene is to optimize the fermentation with lycopene cyclase inhibitor. The addition of this inhibitor involves a lower production of the carotenoid β-and derived, and consequently a major production of the carotenoid Ψ-end derived. Instead, an approach used to increase β-carotene production in E. coli consists in the engineering of the methylerythritol phosphate (MEP) pathway to improve the production of IPP (The inositol pyrophosphate) and DMAPP (Dimethylallyl Diphosphate). Both are precursors of β-carotene then improving their synthesis, we will have a major production of β-carotene. Moreover, the engineering of the TCA pathway is used as a strategy to refine the production of β-carotene because, in this way, they increase the production of ATP and NADH, so it will be easier to produce the chemical (Zhao et al. 2013). We know that Acetyl-coA is a precursor of carotenoid synthesis. So if we wanted to increase the production of lycopene or β-carotene, in theory, we

should intensify the concentration of Acetyl-coA. Chen et al. (2016) have worked on this through a process of host engineering in which YPL062W, a distant genetic locus in S. cerevisiae CEN.PK2, was deleted. In this way, little acetate was accumulated, and an approximately 100 % increase in cytosolic acetyl-CoA pool was achieved relative to that in the parental strain (Chen et al. 2016).

Larroude et al. (2018) used an interesting strategy to improve the production of β-carotene. Their strategy is slightly dissimilar from the others because they developed a combinatorial synthetic biology approach based on Golden Gate DNA assembly to screen the optimum promoter-gene pairs for each transcriptional unit expressed. The major conclusion of their results is that the cassette with the three genes controlled by TEFp is the optimum producer. So, they constructed a new car-cassette, where the three genes (GGS1, carB, and carPR) are under the control of the TEF1 promoter (Larroude et al. 2018). Therefore, we looked at some of the strategies that have been used in different studies in order to give a broader and more complete view of the topic and to make it clear how synthetic biology can offer countless possibilities in the field of research.

Table B. In this table is shown the results derived from some research in which we can see that the strategies adopted are different and various. We can see as strategies, for example, control of oxygen transfer rate, optimization of fermentation with lycopene cyclase inhibitor, Regulation of lycopene synthesis pathway expression, Engineering MEP pathway for IPP and DMAPP supply and central pathway (TCA, PPP) for carbon flux, increase of acetyl-CoA pool and optimization of the lycopene synthesis pathway

Host strain	**Descriptions**	**Products and titers**	**Engineering strategies**	**References**
Blakeslea trispora	Native producer of carotenoids	β-Carotene 704.1 mg/L	Control of oxygen transfer rate	Mantzouridou et al. 2005
Blakeslea trispora	Native producer of carotenoids	Lycopene, 256 mg/L	Optimization of fermentation with lycopene cyclase inhibitor	Mantzouridou et al. 2017
Escherichia coli	Genetically tractable, non-native producer	Lycopene, 0.5 g/g DCW	Regulation of lycopene synthesis pathway expression	Coussement et al. 2017
Escherichia coli	Genetically tractable, non-native producer	β-Carotene, 2.1 g/L	Engineering MEP pathway for IPP and DMAPP supply and central pathway (TCA, PPP) for carbon flux	Zhao et al. 2013
Saccharomyces cerevisiae	Genetically tractable, non-native producer	Lycopene, 56 mg/g DCW	Increase of acetyl-CoA pool and optimization of lycopene synthesis pathway via genome manipulation	Chen et al. 2016
Corynebacterium glutamicum	Native producer of C50 Carotenoid	β-Carotene, 7 mg/L	Deletion of *crtR* and integration of *crt* pathway genes	Henke et al. 2018
Rhodobacter sphaeroides	Phototroph with carotenogenic genes	Lycopene, 10 mg/g DCW	Replacement of *crtI*, augmentation of MEP pathway, and	Su et al. 2018

			block of PPP pathway	
Yarrowia lipolytica	Genetically tractable, non-native producer	β-Carotene, 6.5 g/L	Optimization of promoter-gene pairs of heterologous *crt* pathway	Larroude et al. 2018
Yarrowia lipolytica	Genetically tractable, non-native producer	β-Carotene, 4 g/L	Iterative integration of multiple-copy pathway genes	Gao et al. 2017
Yarrowia lipolytica	Genetically tractable, non-native producer	β-Carotene 0.22636 [$mmol \cdot gDW^{-1} \cdot h^{-1}[mmol \cdot gDW^{-1} \cdot h^{-1}]$]	Iterative integration of three genes (GGS1, carPR, carP)	**Our work** (strain with 3 Knockouts)
Yarrowia lipolytica	Genetically tractable, non-native producer	β-Carotene 0.22634 [$mmol \cdot gDW^{-1} \cdot h^{-1}[mmol \cdot gDW^{-1} \cdot h^{-1}]$]	Iterative integration of three genes (GGS1, carPR, carP)	**Our work** (strain with 2 Knockouts)
Yarrowia lipolytica	Genetically tractable, non-native producer	β-Carotene 0.22635 [$mmol \cdot gDW^{-1} \cdot h^{-1}[mmol \cdot gDW^{-1} \cdot h^{-1}]$]	Iterative integration of three genes (GGS1, carPR, carP)	**Our work** (strain with 4 Knockouts)

C. Comparison of algorithms and methods for succinate production

Table C shows a comparison between our research work and other algorithms mentioned, where the objective is to maximize the production of succinate. From left to right, we reported: Approach/ Algorithm, year of publication, organism, main carbon source, medium, Genetic modification, Results, and References.

Table C. In this table, we have selected the best strains (in terms of succinate production), comparing our work with some articles in the literature. As we can see, each strain has a variable number of knockout genes. In the table from left to right, we reported: Approach/ Algorithm, year of publication, organism, main carbon source, medium, Genetic modification, Results, and References. In this table, only studies that use computational methods to guide the corresponding experimental engineering strategy are considered. Even though FBA and TFBA are not strained design algorithms as such and, consequently, do not allow the direct identification of cellular targets, they have been widely used as a valuable tool in rational strain design. FBA flux balance analysis, CDM Chemically defined medium, TFBA thermodynamics-based flux balance analysis.

Approach/ Algorithm	**Organism**	**Main Carbon Source**	**Medium**	**Genetic Modification**	**Result**	**References**
FBA	*E. coli*	Sorbitol	Complex + anaerob	*DldhA, Dpfl + sfcA overexpression*	38% improvement in succinate productivity	Lee et al., 2002
FBA	*E. coli*	Glucose	Complex + anaerob	*DptsG, DpykF and DpykA*	235% increase in titer	Lee et al., 2005
FBA	*S. cerevisiae*	Glucose	aerobic glucose-limited conditions	SDH-complex, ZWF1, PDC6, U133, U221	0.39 mg/(g glucose x h)	Patil et al., 2005
FBA	*E. coli*	Glucose	Complex + anaerob	*DptsG, DiclR + pyc*	760% increase in	Wang et al., 2006

				overexpression	succinate yield	
FBA	*S. cerevisiae*	Glucose	Not available	*sdh3, ser3, ser33*	0.90 g of succinate/L	Otero et al., 2013
TFBA	*E. coli*	Glucose	Complex + anaerob	*DldhA, Dpflb, DptsG, Dppc + PEPCK from Actinobacillus succinogenes*	60% improvement in succinate titer	Singh et al., 2011
FBA	*S. cerevisiae*	Glucose	Minimal supplemented with vitamins and amino acids + anaerob	*Ddic1*	0.02 (C-mol/C-mol) yield	Agren et al., 2013
OptGene	*S. cerevisiae*	Glucose	Minimal supplemented with vitamins + aerob	*Dsdh3, Dser3, Dser33 + icl1 overexpression*	30-fold improvement in succinate titer	Otero et al., 2013
CASOP	*E. coli*	Glucose	Complex + aerob	*DsdhA, DackA-pta, DpoxB, DmgsA, DiclR + pyc overexpression*	52% improvement in specific productivity and 58% in yield	Yang et al., 2014
FBA	*E. coli*	Glucose	Minimal + anaerob	*DackA-pta + pgl, tktA, talB, sthA, dcuB, dcuC overexpression + pepck from Actinobacillus succinogenes + pyc from C. glutamicum + mutated zwf243 and gnd361 from Corynebacterium glutamicum*	52% improvement in yield	Meng J et al. 2016
FBA	*M. succiniproducens*	Glucose	Complex + anaerob	*DackA-pta and*	35% improvement in maximum productivity	Choi et al., 2016

				DldhA		
FBA	*M. succiniproducens*	Sucrose + Glycerol	CDM + anaerob	*DackA-pta and DldhA*	34% improvement in overall productivity and 21% improvement in yield	Lee et al., 2016
FBA	*S. cerevisiae*	Glucose	Rich medium	***GLT1****(IV), DCD1[XII], ADO1(X), PNP1[XII], LAT1[XIV], PSD1[XIV], ALE1(XV),****ALD6****[XVI] , GPH1[XVI]*	3.1215 $[mmol \cdot gDW^{-1} \cdot h^{-1}]$ Max Productivity: 0.14841 $[h^{-1}]$ Max Yield: 0.2282	**Our work** (strain with maximal Succinate Production)
FBA	*S. cerevisiae*	Glucose	Rich Medium	*PDB1(II), GLT1(IV),ALD6[XVI] , GPH1[XVI]*	3.1027 $[mmol \cdot gDW^{-1} \cdot h^{-1}]$ Max Productivity: 0.14795 $[h^{-1}]$ Max Yield: 0.2295	**Our work** (strain with a minimal number of gene Knockouts)
FBA	*S. cerevisiae*	Glucose	Rich medium	*GLT1(IV), ALD6 (XVI), GPH1 (XVI), YPR127W*	2,9562 [$mmol \cdot gDW^{-1} \cdot h^{-1}]$ Max Productivity: 0.14064 $[h^{-1}]$ Max Yield: 0.2395	**Our work** (strain with maximal yield)

D. Analysis of the role of genes of S. cerevisiae

We present an additional analysis to determine the role of frequently silenced genes, GLT1, ALD6, and GPH1 in S. cerevisiae by our multiobjective evolutionary algorithm framework. In this additional analysis, genes were first switched off individually, then switched off pair-wise, and lastly switched off altogether to understand the impact that silencing these genes has on the microorganism. These analyses were performed using MATLAB and then represented as graphs produced by Escher (King et al., 2015a).

We notice a higher flow leading to an increase of the isocitrate concentration in the mitochondrial compartment (isocitrate is a key organic acid of the TCA cycle) when the ALD6 gene was knocked out (see Figure D1). The reaction involving the *citrate transport* in Figure D1 that catalyzes the transport of isocitrate from mitochondria compartment to cytoplasm shows the value of this flux is lower in the strain (where ALD6 is knocked out) than the wild type. In fact, next to the name of the enzyme that catalyzes the reaction, there are three values, the first represents the flux of the strain obtained by silencing of ALD6, the second represents a wild type, and the third represents the difference between these two values (more reddish the edges are higher the difference and more bluish the edges are lower this difference in Figure D1). The difference is equal to 2.28, which means that in the strain, there is a major concentration of isocitrate in the mitochondria. Moreover, there is a lower activity of *isocitrate dehydrogenase* (NAD+) in the strain than in the wild type. This enzyme catalyzes the conversion of isocitrate in 2-oxoglutarate. It is known that the isocitrate can be transformed into succinate by *isocitrate lyase*. Therefore, by switching off the ALD6 gene, there is a higher isocitrate concentration given by a lower activity of the enzyme *citrate transport* and *isocitrate dehydrogenase* (NAD+) (Figure D1). This could lead to increased production of succinic acid. Hence, looking at the TCA cycle, the flows are channeled towards a higher accumulation of isocitrate at the mitochondrion level. Similarly, in the case of the strain obtained by silencing the GLT1 gene, we find a redirection of the fluxes leads to a higher accumulation of isocitrate at the mitochondrial level, although it was less than the strain in which the ALD6 gene was knocked out (Figure D2).

Figure D3 presents the analyses for silencing the GPH1 gene. This analysis shows that the flow of ATP: D-glucose 6-phosphotransferase is increased highly. This enzyme catalyzes the conversion reaction of D-glucose to D-glucose 6- phosphate. This results in much higher production of D-glucose 6-phosphate cytoplasmic and consequently increased use of glucose. Therefore, this could lead to a higher yield of succinate. We observed that in the case of the strain obtained by knocking out of the GPH1 gene, the flux of isocitrate lyase reaction (that catalyzes the conversion of isocitrate into succinate) increases from zero in the wild type to 0.604 in the strain. An increase in this flux value suggests that the knockout of the GPH1 gene increases the synthesis of the enzyme isocitrate lyase and therefore increases the production of succinate in the microorganism. The increase in the flux values when the GPH1 gene was more prominent than knocking out GLT1 and ALD6 confirms the significance of GPH1 knockout. We can conclusively demonstrate that knocking genes ALD6 and GLT1 genes increase the concentration of isocitrate in the mitochondrial compartments, leading to higher succinate production. Additionally, knocking out GPH1 genes help the growth of Isocitrate lyase, which converts isocitrate into succinate, leading to higher succinate production. These analyses also validate the finding of our multiobjective evolutionary algorithms optimization of strains for automatizing metabolic engineering.

In our final analysis, we first combined the knockout of two genes one at a time to perform four different analyses and finally turned off all three genes simultaneously. We observed that the metabolism of the strain obtained by knocking out all three genes together includes the characteristics of the three strains obtained by knocking out each gene separately (see Figures D1, D2, and D3). This means there is a higher accumulation of isocitrate in the mitochondrion, D-glucose 6-phosphate cytoplasm, and especially the activation of the isocitrate lyase enzyme.

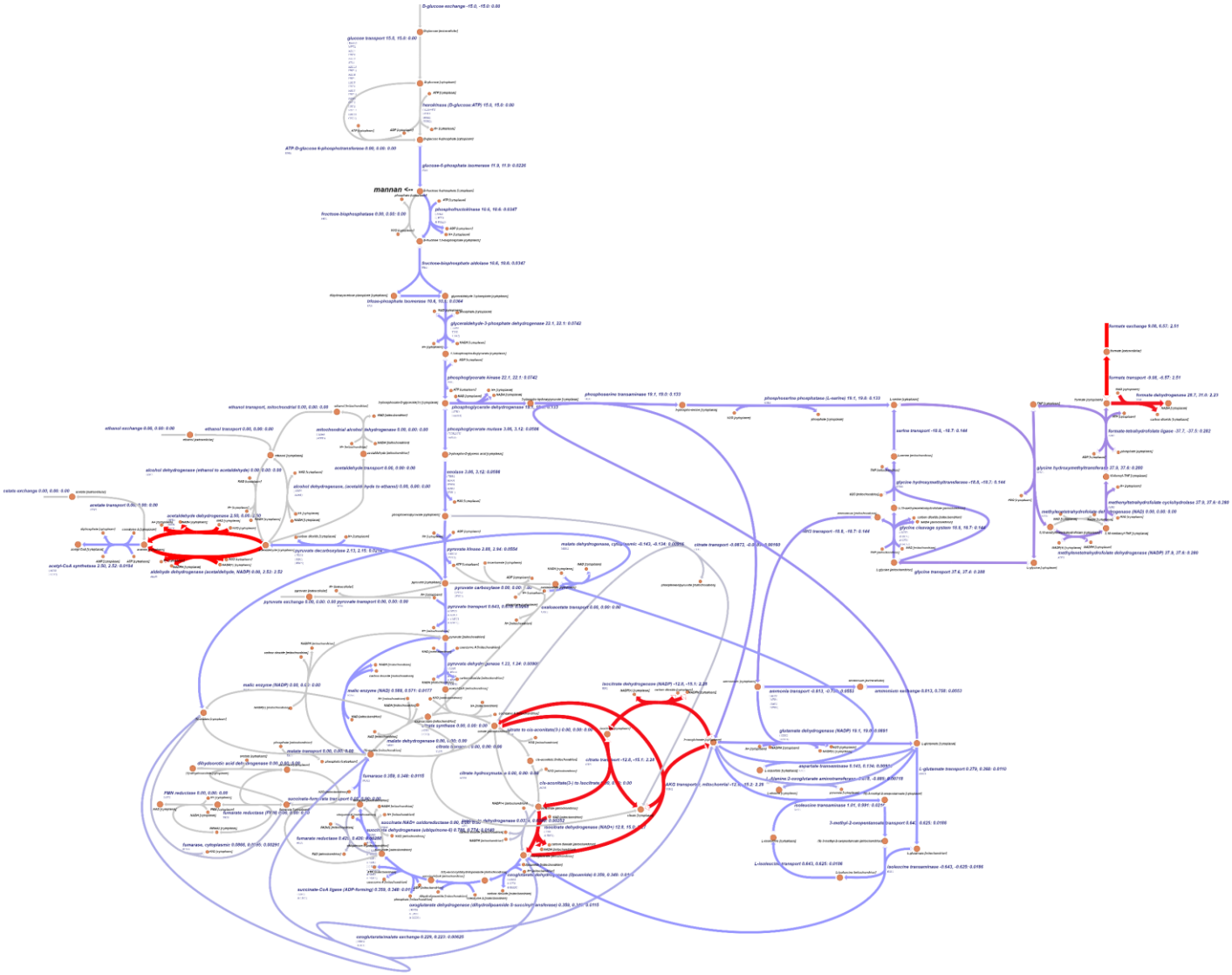


Figure D1. Confront of the Krebs cycle of S. cerevisiae wild type and the strain with ALD6 gene KO. The pathways are shown in blue and red colors, and some pathways are not colored. These colors represent the difference between wild-type flux and the flux of the strain obtained by silencing the ALD6 gene. The pathways colored in red represent a higher difference of value than the pathways colored in blue. The zoom-in versions of Figure D1 are shown in Figure D1-A, D1-B, and D1-C.

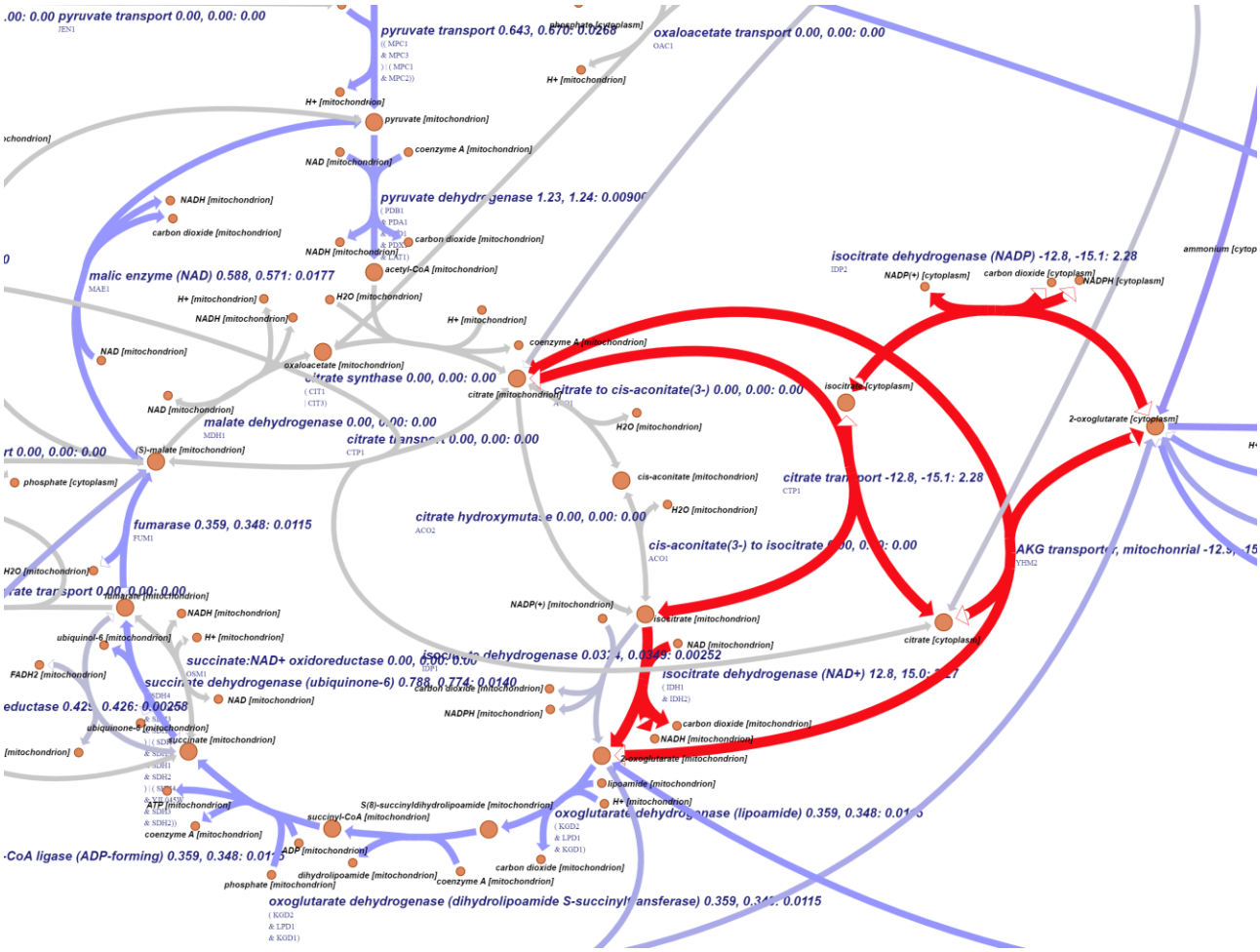


Figure D1-A. Flux changes at the levels of the tricarboxylic acid cycle part of Figure D1. This part focuses on the fluxes that are affected by the Knockout of the ALD6 gene. In red we have highlighted the fluxes that are mainly changed by the Knockout of ALD6. One of these enzymes is the citrate transport, which catalyzes the transport of isocitrate from the mitochondrial compartment to the cytoplasm. The value of the flux of this reaction is lower in the strain (where ALD6 is knocked out) than in the wild type. Therefore, through the Knockout of the gene ALD6, we obtained a major concentration of isocitrate in the mitochondrial compartment.

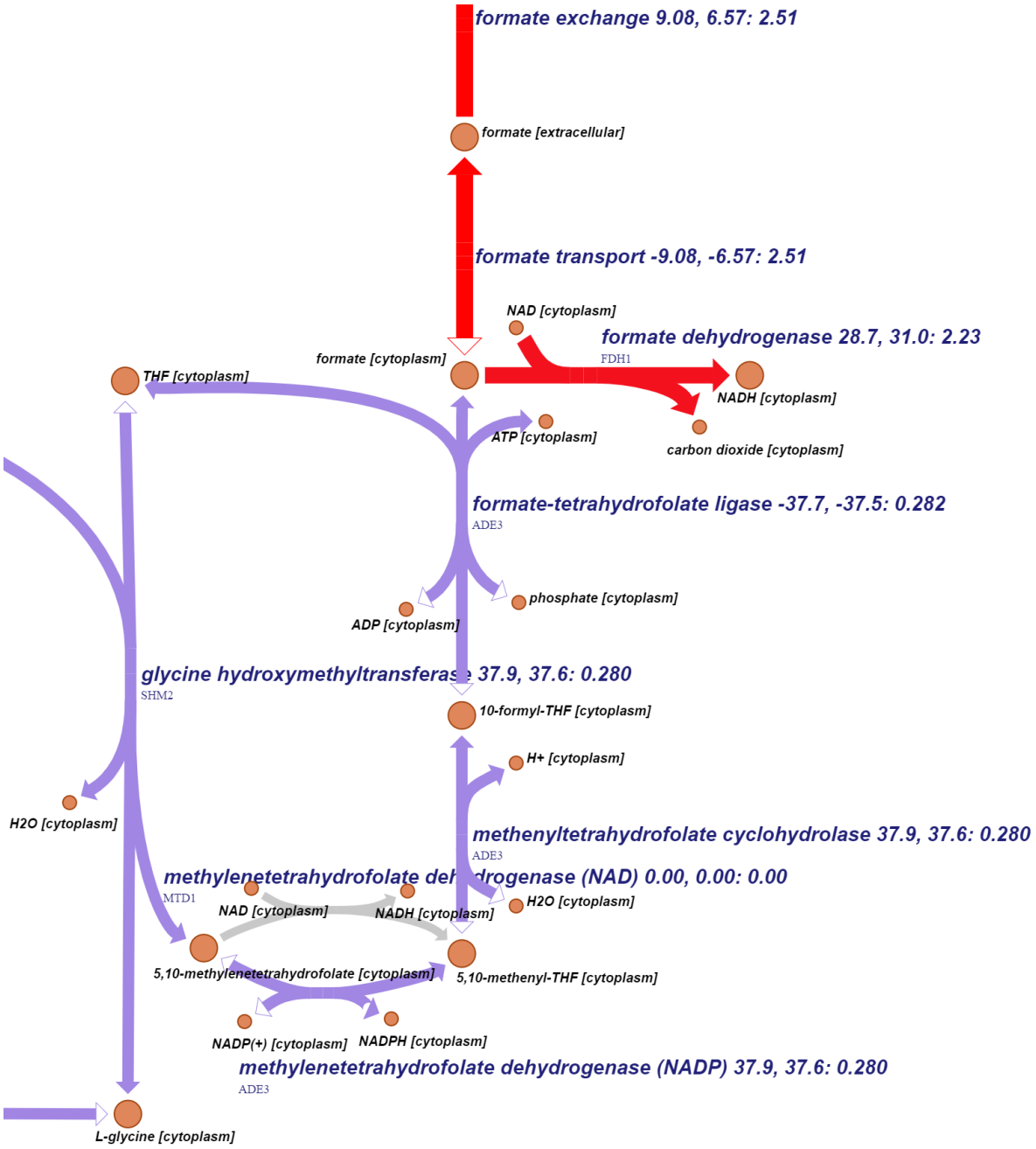


Figure D1-B. This part of Figure D1 shows the change that derives from the Knockout of the ALD6 gene lies in the variation of the formate concentration. Formate is a monocarboxylic acid anion that is the conjugate base of formic acid. It has a role as a metabolite in the S. cerevisiae metabolism. This part shows that knocking out the ALD6 gene has a minor activity of enzyme 'formate dehydrogenase'. This implies a major concentration of formate in the cytoplasm and consequently in a minor production of NADH.

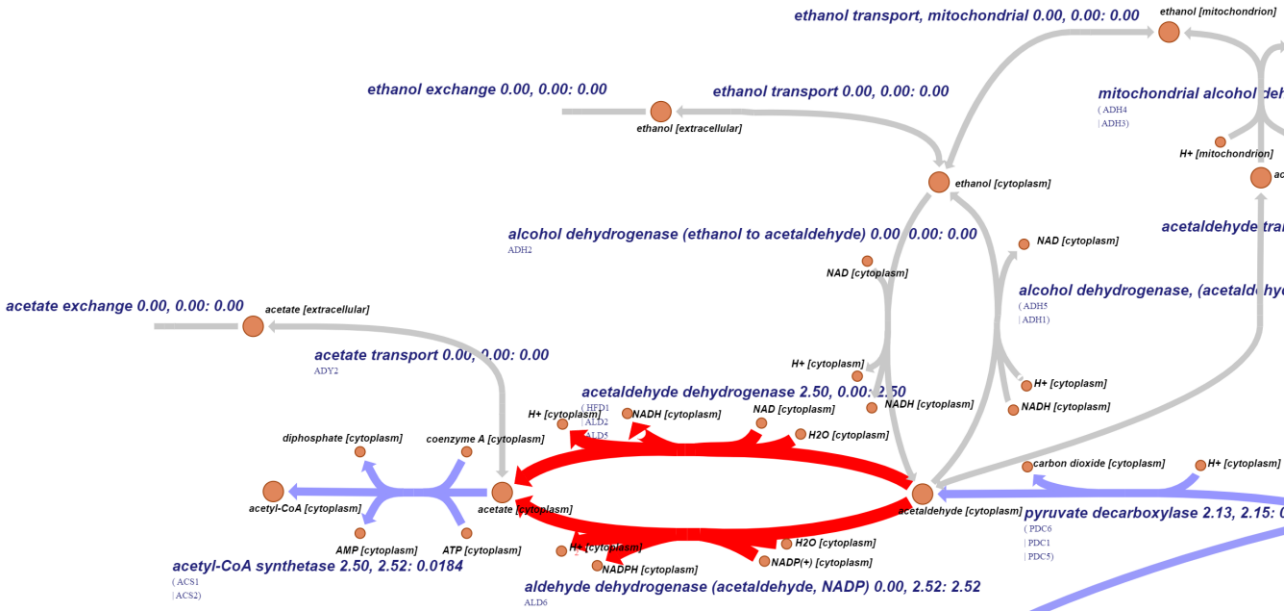


Figure D1-C. This part of Figure D1 shows the effect of the knockout of the ALD6 gene that leads to the activation of a specific isoenzyme and simultaneously to the inactivation of the other corresponding enzyme. This part indicates that the enzyme ‘aldehyde dehydrogenase (acetaldehyde, NADP)’ in the wild type is fully active, while the enzyme ‘acetaldehyde dehydrogenase’ has a flow of zero. Knocking out the ALD6 gene has a contrary situation, that is, the enzyme ‘acetaldehyde dehydrogenase’ increases its flow, while the enzyme ‘aldehyde dehydrogenase (acetaldehyde, NADP)’ resets it.

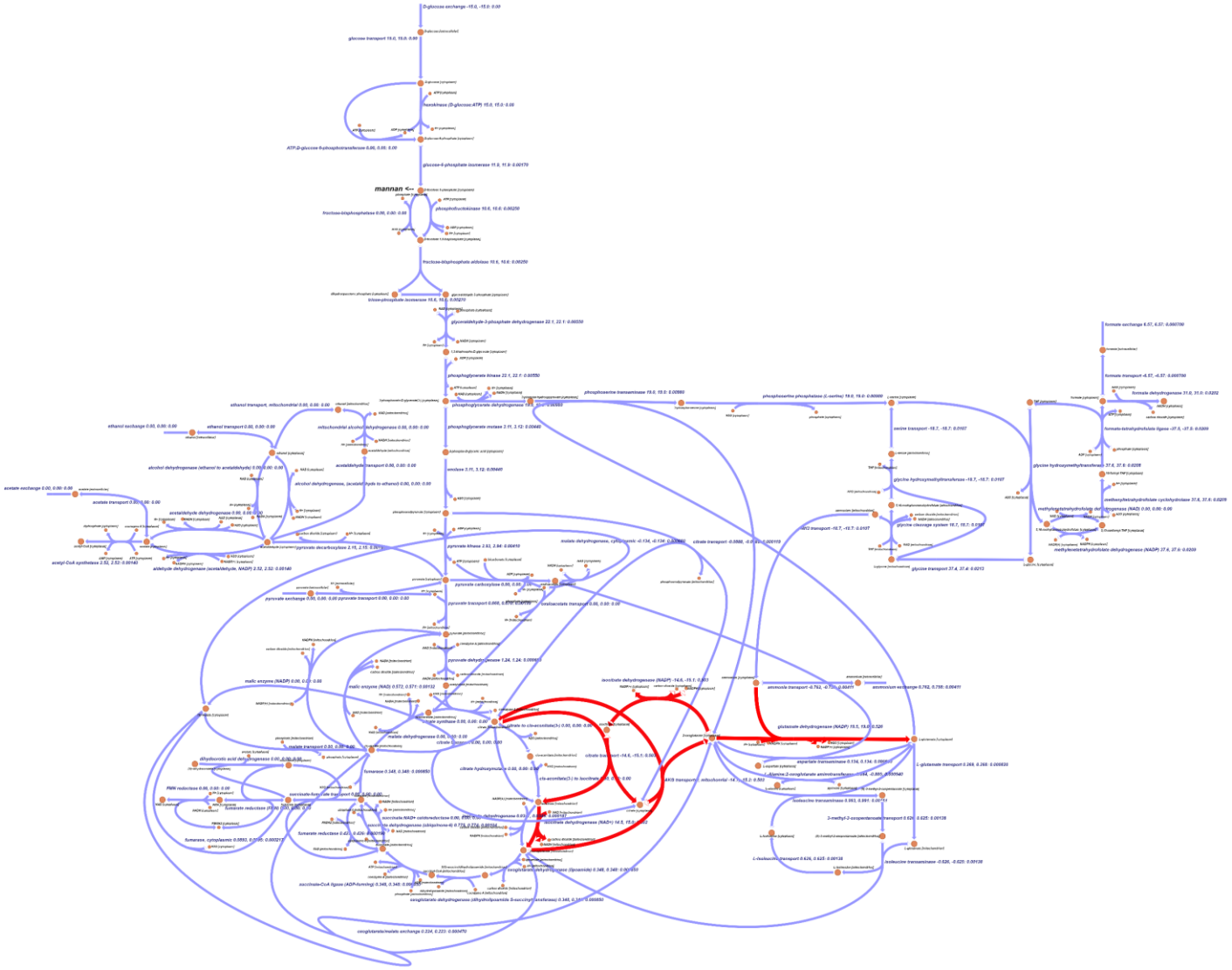


Figure D2. Confront of Krebs cycle of S. cerevisiae wild type and the strain with GLT1 gene knockout. Red and blue colors represent the difference between wild-type flux and the flux of the strain obtained by silencing the GLT1 gene. The pathways that are colored in red represent a higher difference of value than pathways that are colored in blue.

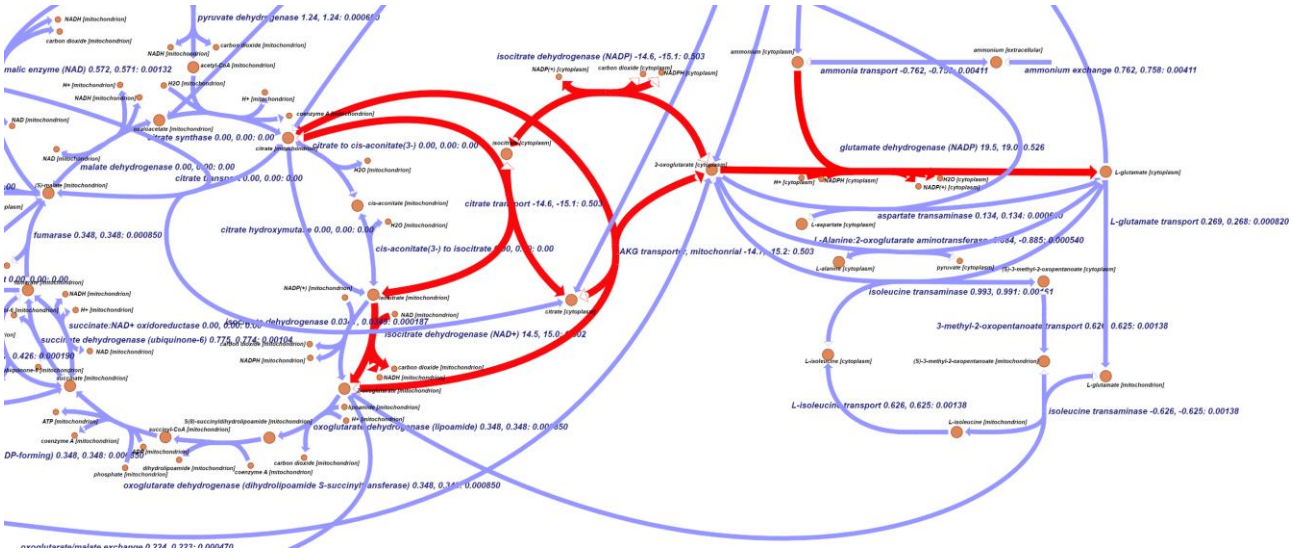


Figure D2-A. Similar to Figure D1-A, in the case of the strain obtained by silencing the GLT1 gene, we find a redirection of the fluxes leads to a higher accumulation of isocitrate at the mitochondrial level, although it was less than the strain in which the ALD6 gene was knocked out (Figure D1-A).

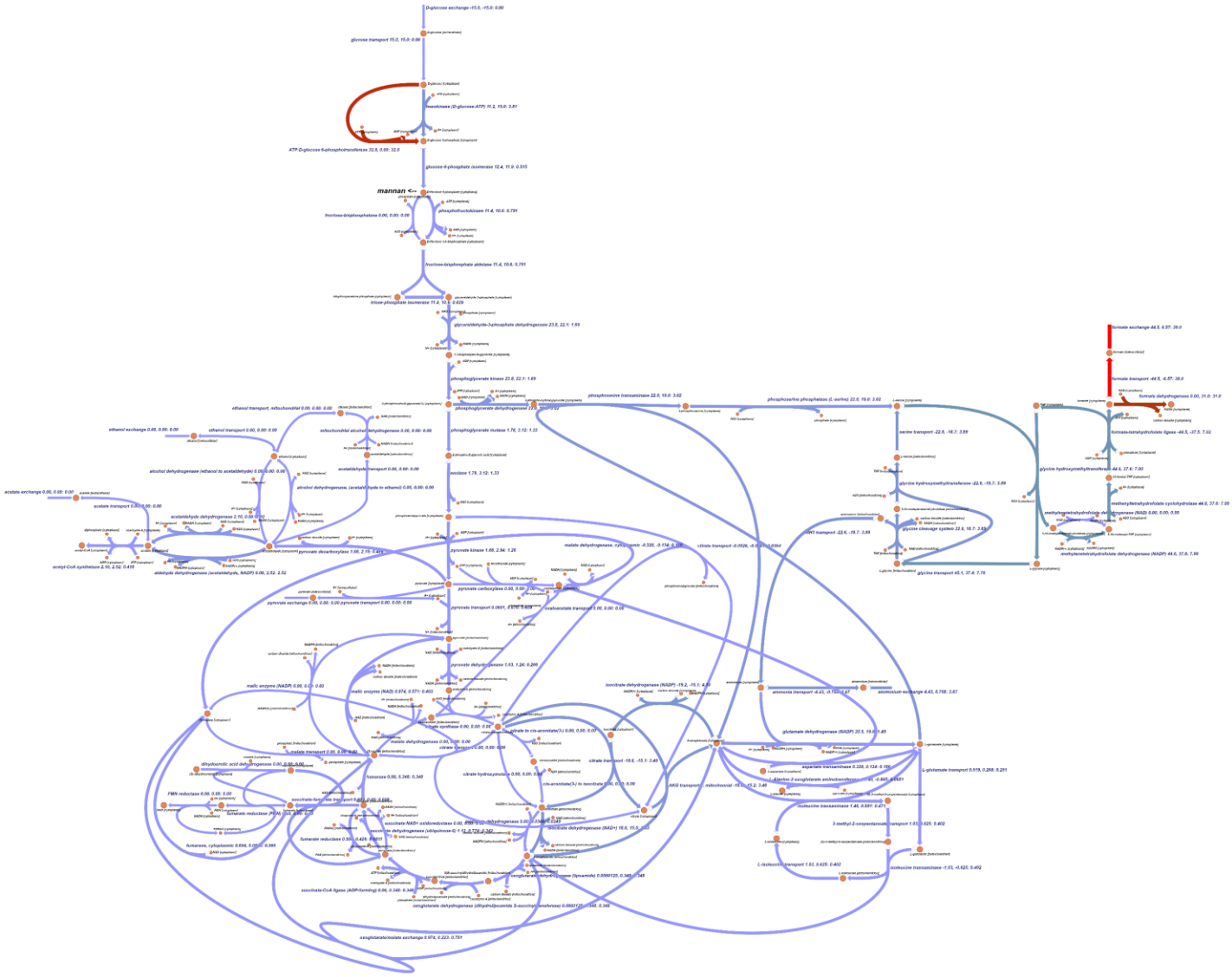


Figure D3. Confront of Krebs cycle of S. cerevisiae wild type and the strain with GPH1 gene knockout. Red and blue colors represent the difference between wild-type flux and the flux of the strain obtained by silencing the GPH1 gene. The pathways that are colored in red represent a higher difference of value than pathways that are colored in blue.

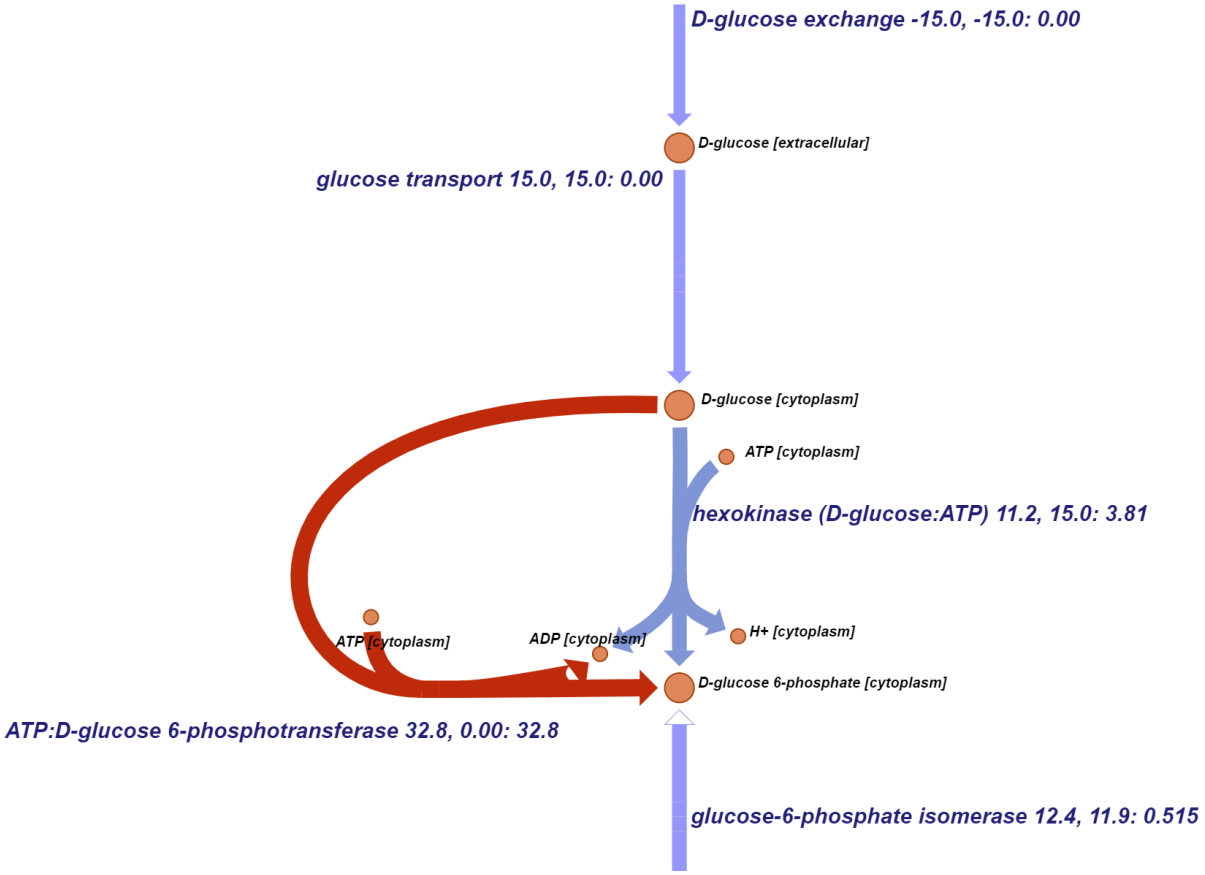


Figure D3-A. This part of Figure D3 shows that the flow of ATP: D-glucose 6-phosphotransferase is increased highly. This enzyme catalyzes the conversion reaction of D-glucose to D-glucose 6- phosphate. This results in much higher production of D-glucose 6-phosphate cytoplasm and consequently increased the use of glucose. Therefore, this may have led to a higher yield of succinate. We also notice a substantial increase in this flux, in fact, it passed from 0.00 in the wild type to 32.8 in the strain with the GPH1 gene KO.

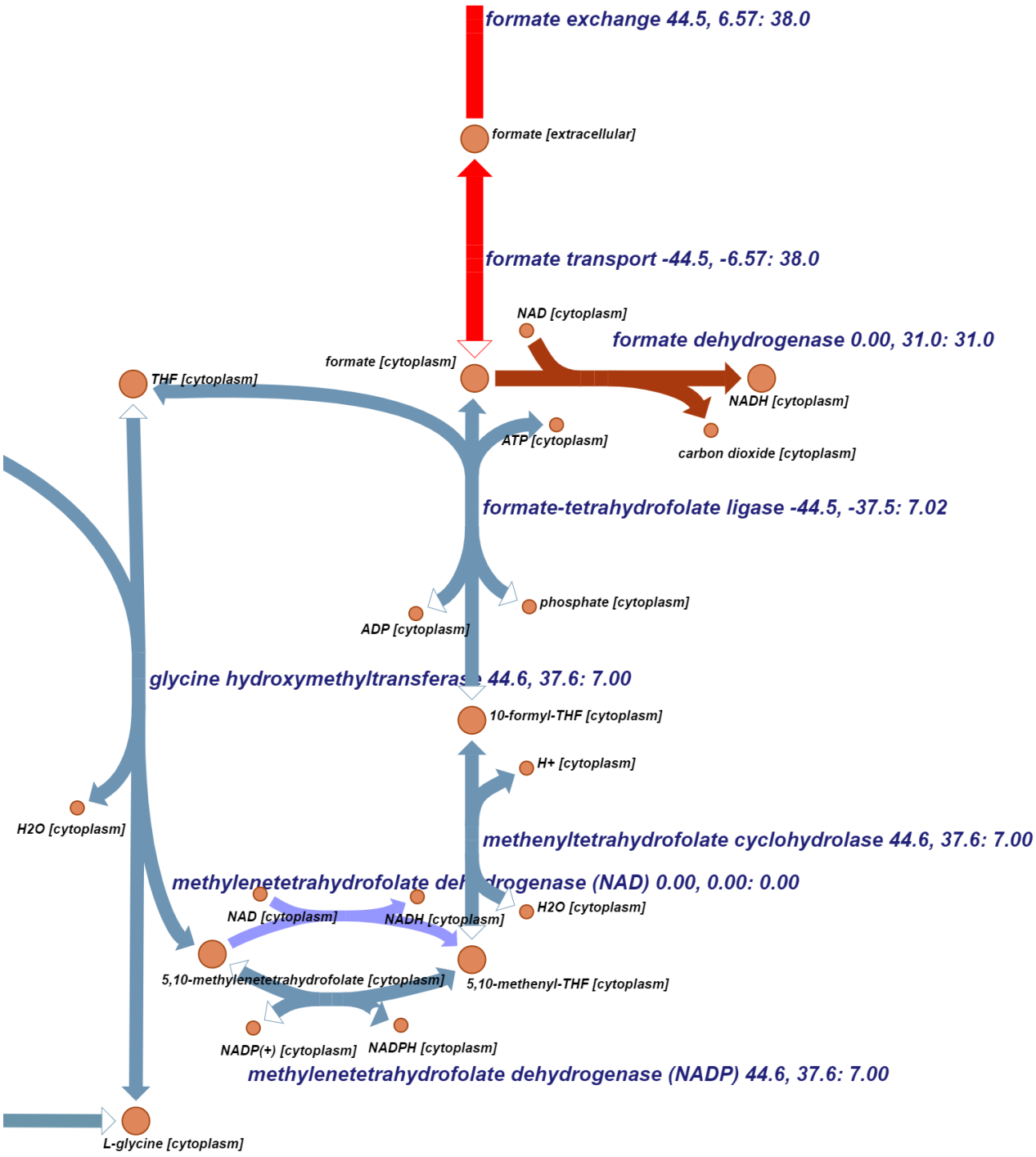


Figure D3-B. This part of Figure D3 shows that knocking out of the GPH1 gene brings the 'formate dehydrogenase' reaction flow to zero. The GPH1 gene affects this reaction more than the ALD6 gene. In fact, as per Figure D1-A, where the knockout of the ALD6 gene led to a decrease in the flow of 'formate dehydrogenase,' in the case of knocking out the GPH1 gene the similar flow is brought to zero. This will produce a much larger increase in formate concentration, compared to a noticeable decrease in cytoplasmic NADH concentration.

E. List of Abbreviations

Ac-ac- CoA	Acetyl- acetyl coA
ACDH	Activated acetaldehyde dehydrogenase
ACS	acetyl-CoA synthetase
AGP3	glutamate permease
ATP	Adenosine triphosphate
BPCY	Biomass Product Coupled Yield
carB	phytoene dehydrogenase
carPR	phytoene synthase/ lycopene cyclase
CNM	Central Nitrogen Metabolism
CoA	Coenzyme A
FAD(H2)	Dihydroxyflavone-adenine dinucleotide
FBA	Flux Balance Analysis
FPP	Farnesyl diphosphate
FVA	Flux Variability Analysis
GGS1	geranylgeranyl diphosphate synthase
GOGAT	Glutamate synthase
GPR	Gene- Protein- Reaction
HMG-coA	3- Hydroxy- 3- methylglutaryl coenzyme A
HMGS	Hydroxymethylglutaryl- CoA- synthase
KO	Knockout
MOEA	Multiobjective Evolutionary Algorithm
POME	Pareto Optimal Metabolic Engineering algorithm
MVA	Mevalonate
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
PABA	para-amino benzoate

PDC	Pyruvate decarboxylase
PDC6	pyruvate decarboxylase
PDH	Pyruvate dehydrogenase
pFBA	parsimonious Flux Balance Analysis
SDH- complex	Succinate dehydrogenase complex
TCA	Tricarboxylic Acid
THR1	homoserine kinase
YNB	Yeast Nitrogen Base medium
YPD	Yeast Extract- Peptone- Dextrose medium
WT	Wild Type strain
sgRNA	Single strain RNA