OptEnvelope: a target point guided method for growth-coupled production using knockouts

Ehsan Motamedian, Kristaps Berzins, Reinis Muiznieks, Egils Stalidzans*
*Corresponding author
Institute of Microbiology and Biotechnology, University of Latvia, Jelgavas Street 1, Riga LV1004, Latvia

Abstract
Finding the best knockout strategy for coupling biomass growth and production of a target metabolite using a metabolic model is a challenge in biotechnology. In this research, a three-step method named OptEnvelope is developed based on finding minimal active reactions for a target point in the feasible solution space using a mixed-integer linear programming formula. The method initially finds the reduced desirable solution space (envelope) in the product versus biomass plot by removing all inactive reactions, and then, with reinsertion of the deleted reactions, OptEnvelope attempts to reduce the number of knockouts so that the production envelope is preserved. Additionally, OptEnvelope searches for envelopes with higher minimum product yields or fewer knockouts for different target points within the desired solution space. A number of knockouts can be set and evaluated using OptEnvelope in order to determine how it affects the intended envelope. The method was implemented on metabolic models of *E. coli* and *S. cerevisiae* to benchmark the capability of these industrial microbes for overproduction of acetate and glycerol under aerobic conditions and succinate and ethanol under anaerobic conditions. The results indicate that *E. coli* is more appropriate to produce acetate and succinate while *S. cerevisiae* is a better host for glycerol. The positive effect of deleting some genes responsible for the proposed reactions for knocking out was previously confirmed by reported experimental data. Both organisms are suitable for ethanol production, however, more knockouts for the adaptation of *E. coli* are required. OptEnvelope is available at https://github.com/lv-csbg/optEnvelope.

Keywords: Coupling, target point, minimal active reactions (MAR), host selection, genome-scale metabolic model, knockout strategy for overproduction, mixed-integer linear programming (MILP).

1. Introduction

The organism strains used in biotechnology define the opportunities and limitations of biotechnological production, therefore, their selection and modification are important ways to improve the productivity features [1] and sustainability of biotechnology [2]. Coupling is one of the methods to ensure the expected behavior of the strain [3]. Coupling in metabolic modelling means that two independent metabolites become linked with each other: one metabolite can not be produced without producing the other one. The coupling is useful especially when the secretion of a product is coupled with growth because optimization of the strain for higher growth rates using adaptive laboratory evolution and serial passages evolves the strain towards higher target production rates by coupling [4]. This growth-coupling

relationship can be projected on the product versus biomass plot creating a production envelope that represents all feasible steady states. Due to the dominance of the fastest-growing strains in a medium, the most probable point in the growth-production space described by the envelope is the one with the highest growth rate. Therefore, metabolic engineering efforts are frequently devoted to the limitation of the metabolic production envelope [4] in a way that the point of the fastest possible growth brings also sufficient product generation rate.

Several methods have been published for growth-coupled strain design and the developers choose different platforms to create the method including Matlab (e.g., optKnock, robustKnock, and modCell2) and Python (e.g., optCouple and strainDesign) packages. OptKnock[5], the very first method for growth coupling, is a bilevel programming framework for identifying reaction knockout strategies. The first level of the framework focuses on maximizing the yield of the product while the second level optimizes the growth. This bilevel problem can be expressed as a two-dimensional matrix and solved using mixed integer linear programming (MILP). Results obtained using this method are often overly optimistic [6] and have weak growth coupling. To solve these problems, RobustKnock [6] was developed. This approach searches for a set of knockouts that guarantees a minimal production of target metabolite considering the entire feasible solution space. RobustKnock guarantees that the target metabolite is "an obligatory by-product of growth". The authors achieve this by using bilevel max-min optimization that finds knockouts which guarantee non-zero flux of product at maximum growth rate. Another improvement of the optKnock method was achieved with optCouple [7] addressing the possibility of simultaneous knockouts, knockins, and modification to the growth medium. This method also uses a bilevel programming framework to find control values that provide the highest growth-coupling potential. Postprocessing workflow is used to identify modifications that do not contribute to growth-coupling to decrease calculation times. Another approach to the growth-coupling problem is modCell2 [8] which uses modular cell design to solve multiobjective optimization. The creators of this method used evolutionary game theory and evolutionary algorithms to manipulate the network to achieve overproduction. This method is more complicated to use because users have to create manually specific files for each model. The fifth method – strainDesign [9] – is a Python package created to integrate four of the most popular MILP-based strain design approaches – optKnock, robustKnock, optCouple, and MCS [10] - to improve calculation times and reach better results. User input is minimized (e.g., definition of the objective function), and MILP construction is automated. This method also allows user to add desired and undesired regions for production envelope. The previously developed powerful network compression techniques are applied to reduce the size of the problem and thus improve the calculation time.

The earlier described methods have one common disadvantage: they only present one envelope after each run. In contrast, the method developed in this research named OptEnvelope attempts to find all envelopes located in the desired space for coupling (the solution space with higher product yield and lower biomass yield). Thus, a comparison of the envelopes based on various criteria such as fewer knockouts, higher minimum and maximum production rates, etc. is possible in a simple manner. OptEnvelope can find multiple envelopes in one run because it defines a target point for coupling and by moving the target point in the desired space, various valuable envelopes are discovered. While the strategy of the other methods is to find the best knockouts, OptEnvelope finds minimum active reactions (MAR) for a target point and then,

with reinsertion of reactions attempts to reduce the required number of knockouts. The strategy of defining a target point combined with the low CPU time of OptEnvelope makes it easy to switch the target point to the middle feasible space and find other favourable envelopes. OptEnvelope is also a good tool for fast benchmarking of different organisms for production of different metabolites due to its fast operation. The appropriateness of an organism for the production of a particular metabolite is determined by the shape of the envelope and the number of necessary deletions.

2. Material and methods

2.1. Genome-scale metabolic models and simulated growth medium

In this study, the genome-scale metabolic models iJR904 [11] and iMM904 [12] were applied for *E. coli* and *S. cerevisiae*, respectively. The models were used in the irreversible form to simplify the solution. Growth on glucose minimal media was simulated by considering upper bounds of 10 mmol/gDCW/h for glucose uptake rate and 1000 mmol/gDCW/h for NH_4, H_2O, SO_4, O_2, H^+, phosphate, and inorganic ion consumption reactions under aerobic conditions. Oxygen uptake was bound to zero in iJR904 for the anaerobic growth simulation of *E. coli*. For *S. cerevisiae*, cytochrome c oxidase was removed from iMM904 to implement anaerobic conditions as very small amounts of oxygen are essential for biomass[13]. The uptake rate of other metabolites was set to zero. MATLAB (R2020a) was used for modelling using the COBRA Toolbox [14] and Gurobi was applied to solve linear programming (LP) and mixed-integer linear programming (MILP) problems. A computer with an i7 Intel 4 GHz processor and 32 GB RAM was used to run the method.

2.2. Method

OptEnvelope (see https://github.com/lv-csbg/optEnvelope) includes three steps: 1) selecting a target point in the feasible region of a production envelope, 2) determining the minimal active reactions (MAR) at the selected point, and 3) reinsertion of the deleted reactions at step 2 to specify minimal knockouts. For selecting a point in the feasible region of the product-biomass plot, Flux Balance Analysis (FBA) was applied according to the LP problem presented in Eq. 1.

$$\begin{aligned} &\text{Maximize } Z = cv \\ &S \times v = 0 \\ &0 \leq v \leq v_{max} \end{aligned} \tag{1}$$

where S is an m×n stoichiometric matrix where m and n indicate the number of metabolites and reactions, respectively. c is the vector of the objective function coefficients, and vector v_{max} denotes the upper bounds of reaction fluxes.

The maximum and minimum growth rate on the glucose minimal medium was initially calculated considering biomass reaction as the objective function. Then, the space between the minimum and maximum growth values was divided into 100 points including the endpoints. The objective function was changed to the exchange reaction of the desired product and the minimum and maximum production rates at each point of the growth rate were calculated. The final plot can be presented based on the rate of production or yield on the substrate. In the code, if the molecular weight of a substrate is presented in the input data by the user, the plot of product yield versus biomass yield is drawn, otherwise, production rate versus growth rate is applied. Product versus biomass yield was plotted and the feasible region was specified. A

typical envelopes of product excretion rate versus growth rate for a wild-type strain and growth-coupled strains are presented in Figure 1.

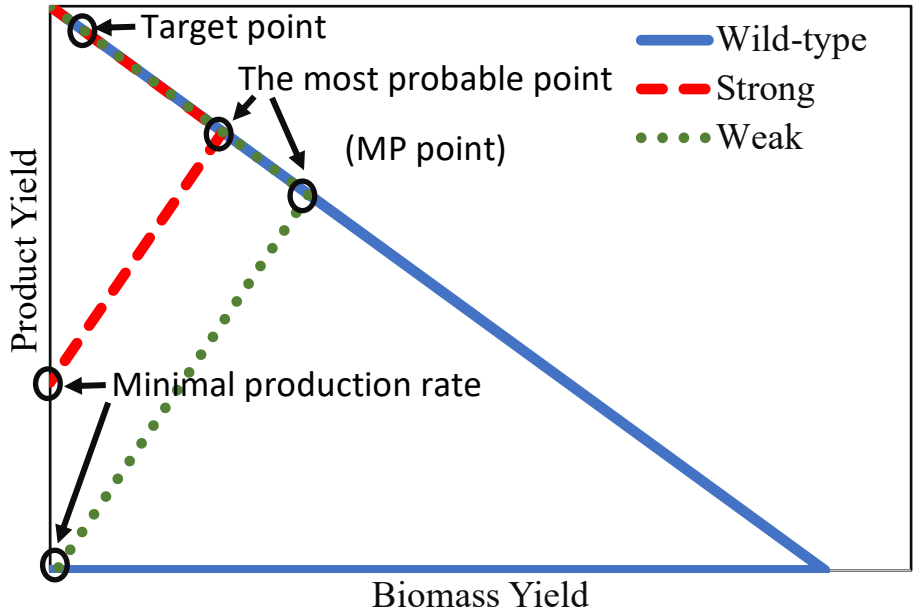

Figure 1. A typical envelope of product excretion yield versus growth yield for wild-type strain and strong and weak couplings. The target point for coupling has a high production rate and a low and non-zero growth yield. The minimal production yield is non-zero for strong coupling. The points with the highest growth yield in the coupled envelopes are the most probable (MP) points.

The plots in Figure 1 indicate the maximum and minimum excretion yields of the desired product at each growth yield. Selecting a point on the boundary of maximum excretion yields is desired because the secretion yields of by-products are minimal on this boundary. Furthermore, more excretion yield of the desired product along with a low but non-zero growth yield is mostly desired. So, the target point labelled in Figure 1 (located on the upper edge) was selected as a point for which MAR are determined, however, any other point in the feasible region can be selected to find its MAR in step 2. It should be mentioned that the envelope which is presented in step 2 may change when altering the target point for coupling in step 1.
For finding MAR of each target point, it has to be mentioned that the reactions can be categorized by flux variability range into four groups at each feasible point in Figure 1: 1) inactive reactions with no variability at zero rate, 2) reactions with flux variability including zero rates, 3) active reactions with no variability at non-zero rate and 4) active reactions with flux variability range that does not include zero rate. So, to determine MAR, the reactions belonging to group 1 can be easily removed from the model, however, all reactions of group 2 cannot be eliminated simultaneously. So, Eq. 2 was added to the LP problem to convert it into a MILP problem for finding MAR at a target point.

$$v_j - v_{j,max} \times w_j \leq 0 \quad , \begin{cases} v_j = 0 \rightarrow z_j = 0,1 \\ v_j \neq 0 \rightarrow z_j = 1 \end{cases} \tag{2}$$

The inequality shown in Eq. 2 was defined for each metabolic reaction (v_j); the variable z_j is a binary variable that can only take the values of zero and one; and $v_{j,max}$ is the upper bound of

reaction j. Eq. 2 was not written for exchange reactions as they are pseudo-reactions and the diffusion-based protein-free transport reactions of metabolites, such as carbon dioxide. After bounding rates of biomass and desired product to the values at the target point, to calculate MAR in the second step of OptEnvelope, the minimization of the sum of z_j according to Eq. 3 was used as the objective function of the MILP problem.

$$\text{Minimize} = \sum z_j \quad (3)$$

As the metabolic model is in the irreversible form, the metabolic model might activate the forward and backward fluxes of reversible reactions at the same time and make cycles to unrealistically increase the number of genes with z_j=0. So, to avoid the simultaneous use of forward and backward directions in reversible reactions, Eq. 4 was added for each forward and backward flux,

$$z_{j,f} + z_{j,b} \leq 1 \quad (4)$$

where $z_{j,f}$ and $z_{j,b}$ are binary variables for forward and backward fluxes of reversible reaction j, respectively. Thus, if one direction of the reaction has a non-zero flux, Eq. 4 forces the other one to be zero.

After solving the MILP problem, all fluxes were bound to zero except reactions belonging to the set of MAR, and then, the product-biomass plot was drawn. This envelope is generated because of the flux variability of MAR at the target point. The coupling is strong if the minimum secretion rate of the desired product in lack of growth as shown in Figure 1 is more than zero, otherwise, it is considered a weak coupling [15]. It is easy to determine if the strong coupling is possible using OptEnvelope and this method was used to evaluate the best possibility of coupling a product with biomass for different hosts. As the adaptive laboratory evolution can be resulted in the highest growth rate based on Darwin's natural selection, the point with the maximal growth rate in the envelope can be considered the most probable (MP) point (Figure 1).

After finding the optimal envelope and MAR, sequential reinsertion of the deleted reactions was done to find minimal reaction knockouts needed to support the minimum secretion rate of the desired product if the coupling is strong. In fact, the minimum secretion rate was preserved in the reinsertion step to be sure that the coupling is strong and the lack of the secretion of the desired product is infeasible.

As MILP can have alternative equally optimal solutions and the solvers may not return feasible solutions due to the way solvers approach the NP-complete problems, at least calculations for 8 points including the points 3 to 10 of those 100 points for the envelope of wild-type product versus growth were done. The number of points can be set as input by the user in the code. The final envelope is the same for all points but the number of knockouts could be different because of different sets of MARs. So, the result with minimum knockouts is selected because fewer gene knockouts are desirable. A time limitation of 10 minutes was considered for each MILP problem.

2.3. Switching the target point to the middle space

Sometimes switching the target point to the middle space is desirable because it results in an envelope with comparable maximum product yield but minimum product yield is significantly increased with fewer knockouts. So, evaluation of the envelopes for middle target points and

comparison with the primary envelope obtained for the target point located on the upper edge is necessary. Thus, in this research, after finding the primary envelope for the primary target point, the line space between the production yield of the target and MP points as shown in Figure 2 is divided into a specific number of identical distances. The points lower than the MP point are not desirable because the chance of generating envelopes better than the primary envelope is almost zero. In the code, the number of new target points in the middle area can be set by the user. The ratio of the production rate per maximum production rate for the primary envelope at zero growth rate is presented after running the code to determine the location of each target point.

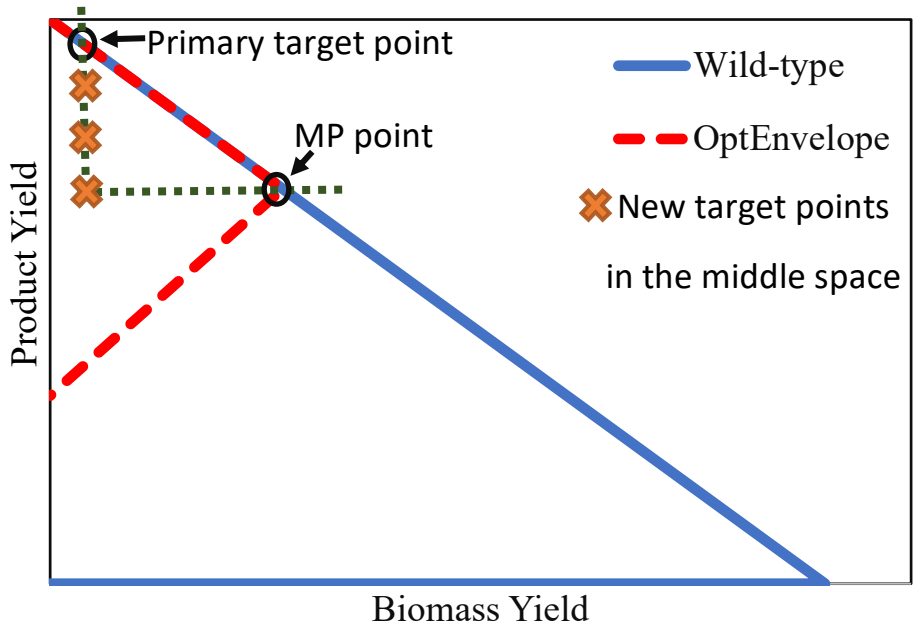

Figure 2. Determination of new target points in the middle space using the primary target and most probable (MP) points of the primary envelope obtained for the primary target point. The number of new target points can be determined by the user.

For all target points, the envelope and its important parameters (minimum and and maximum production rate, production and growth rates at MP point and the number of knockouts) are determined and compared to select the best envelope. If minimum production rate for an envelope is lower than that for the primary envelope or number of knockouts for an envelope is more than that for the primary envelope, the results are not reported for that envelope. For the selection of the best envelope, various criteria have to be considered. Higher values of maximum and minimum production yields as well as more production yield at the MP point of the envelope are desirable. It should be mentioned that the maximum and minimum production yields are calculated when growth rate is assumed to be zero. Furthermore, the required number of knockouts to obtain the envelope is an important criterion for the decision. After running the code, all the data are presented in a table on the MATLAB's command and the user can select the most appropriate envelope depending on the specificity of the biotechnological task.

2.4. Limitation of the number of knockouts for each envelope

After selecting the best envelope, the number of knockouts can still be high for the user and hence, the effect of the reduction of the number of knockouts is required to be studied. A

procedure for finding the dual for an inner LP problem used by the Robustknock [6] is embedded in OptEnvelope so that the number of knockouts can be set as an input in the code. The OptEnvelope transformation for inner problem is defined as:

$$A = \begin{bmatrix} S \\ -S \\ I_{not} \\ -I_{not} \\ I \\ -I \end{bmatrix}, A_y = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -I \cdot ub \\ I \cdot lb \end{bmatrix}, B = \begin{bmatrix} 0 \\ 0 \\ ub \\ -lb \\ 0 \\ 0 \end{bmatrix}$$

Where S is a stoichiometric matrix of a model, I_{not} is an eye matrix of reactions that are protected while I is an eye matrix of reactions that can be deleted. Ub and lb are upper and lower bounds of reactions.

Similar to RobustKnock, another transformation is performed to achieve max-min problem. It is defined as:

$$A2 = \begin{bmatrix} -C^T & -C_w^T \\ A & 0 \\ 0 & A_w \end{bmatrix}, A_y2 = \begin{bmatrix} 0 \\ A_y \\ A_{y_w} \end{bmatrix}, B2 = \begin{bmatrix} 0 \\ B \\ B_w \end{bmatrix}$$

Where elements with subscript w are corresponding parameters of dual LP problem and finally, a third transformation is performed to include minimal number of reaction deletions:

$$A3 = \begin{bmatrix} A & 0 & A_y \\ 0 & A2_w & A_y2_w \\ 0 & 0 & -1 \end{bmatrix}, B3 = \begin{bmatrix} B \\ B2_w \\ -numDel \end{bmatrix}$$

The main difference with RobustKnock is that MAR for the selected envelope is added to the list of reactions that are not allowed to be knocked out. Thus, if the number of knockouts is set to be less than the determined value by OptEnvelope, an expansion in the area of the predicted envelope by OptEnvelope is observed after finding the envelope by the modified RobustKnock.

2.5. Selecting an appropriate host for a product using OptEnvelope

OptEnvelope can be implemented simply and quickly for various hosts, and thus can easily evaluate the impact of host change. Similar to the process of selecting the best envelope for a model, various criteria including higher values of maximum and minimum production yields, more production yield at the MP point of the envelope and fewer knockouts can be considered for selection of the host. The coupling of acetate, glycerol, ethanol and succinate in *S. cerevisiae* and *E. coli* was investigated using OptEnvelope. The primary envelopes of two hosts were compared in this research.

3. Results

3.1. Implementing OptEnvelope on *E. coli* for acetate production as a case study

The characteristics of envelopes including the number of knockouts, name of reactions and the rates for all envelopes found by OptEnvelope on *E. coli* for acetate production under aerobic condition with middle points are presented in Table 1 and Figure 3. It can be seen that maximum and minimum production rate and the production rate at MP point is 27.49, 4.47, and 21.39 mmol/gDCW/h, respectively, for the primary envelope shown with red colour. Maximum growth rate is reduced from 0.91 1/h for wild-type to 0.24 1/h and eight reaction knockouts are needed to have this envelope. The range between 21.39 and 27.49 mmol/gDCW/h were divided to 15 target points and calculations for the middle points were carried out. The eleven envelopes having higher minimum production rate compared the primary envelope and needing reaction knockouts less than 10 are presented in Table 1.

Table 1. Characteristics of the found envelopes by OptEnvelope for acetate production in *E. coli*. Units for production rates and growth rate are mmol/gDCW/h and 1/h, respectively. The first envelope is the primary envelope found for the first target point located on the upper edge. The abbreviated names of knockout reactions are presented in the table.

Envelope No.	No. of knockouts	Minimum production rate	Maximum production rate	Production rate at MP point	Growth rate at MP point	Name of knockout reactions
1	8	4.47	27.49	21.391	0.243	GLCDe, MTHFD_1, F6PA, PGL, PFK, SUCCt2r, SUCD4, AKGt2r
2	6	9.50	27.07	22.681	0.143	ATPS4r, GNK, F6PA, PGL, PFK, SUCOAS
3	8	6.91	26.17	22.962	0.140	GLCDe, MTHFD_1, F6PA, PGL, PFK, SUCCt2r, SUCD4, PPM
4	10	8.94	25.87	23.203	0.111	CYTBO3, GLCDe, GLYCL, F6PA, PGL, PDH, PFK, SUCCt2r, SUCD4, PPM
5	8	6.91	26.17	22.962	0.140	GLCDe, MTHFD_1, F6PA, PGL, PFK, SUCCt2r, SUCD4, PPM
6	8	6.33	26.95	22.393	0.180	GLCDe, MTHFD_1, NADH6, F6PA, PGL, PFK, SUCCt2r, SUCD4
7	9	4.90	27.35	21.685	0.223	GLCDe, MTHFD_1, F6PA, PGL, PDH, PFK, SUCCt2r, SUCD4, AKGt2r
8	3	7.60	25.60	17.400	0.146	ATPS4r, PGM, SUCOAS
9	8	12.67	25.51	23.953	0.065	GLCDe, GLYCL, NADH6, F6PA, PGL, PFK, SUCD4, PPM
10	9	19.00	25.18	24.645	0.022	GLCDe, GLYCL, NADH6, F6PA, PFL, PGL, PFK, SUCD4, PPM

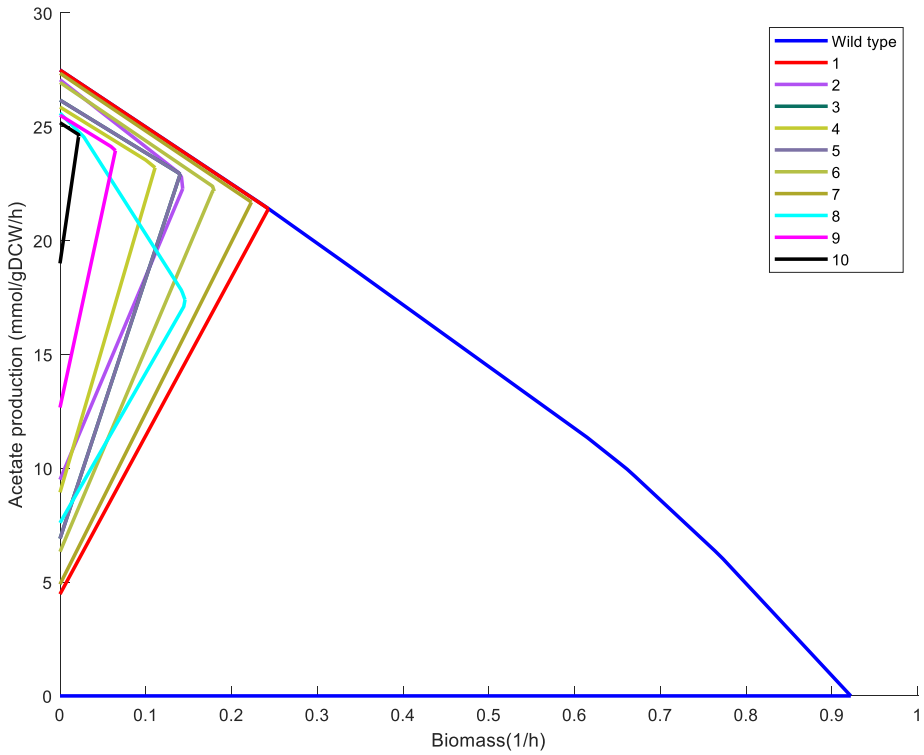

Figure 3. The wild-type envelope (blue) and the ten envelopes predicted by OptEnvelope for production of acetate using *E. coli* under aerobic condition.

The envelopes indicate that diverse results are presented for the middle points and it has to be mentioned that some of the envelopes are found multiple times. Minimum number of knockouts is three (ATP synthase, succinyl-CoA synthetase and phosphoglycerate mutase) that belongs to an envelope with maximum growth rate of 0.146 1/h at production rate of 17.4 mmol/gDCW/h and minimum and maximum production rates of 7.6 and 25.6 mmol/gDCW/h, respectively. While this envelope represents minimum number of knockouts, its production rate at MP point is much less than the others even less than the primary envelope. An envelope with six knockouts is also recommended having maximum growth rate of 0.143 1/h at production rate of 22.68 mmol/gDCW/h and minimum and maximum production rate of 9.5 and 27.1 mmol/gDCW/h, respectively. Minimum production rate is increased 2.1 times while a slight reduction in the maximum production is observed. Production rate at MP point is a little more than that for the primary envelope that is valuable. So, all criteria for this envelope are better compared to the primary envelope and this demonstrates the importance of switching the target point to the middle space. It is interesting that knockout of ATP synthase is only suggested for envelopes with 3 and 6 knockouts. For the other envelopes with more knockouts removal of this important reaction is not needed.

The highest minimum production rate and the production rate at MP point is 19.0 and 24.65 mmol/gDCW/h, respectively, that belongs to an envelope with nine knockouts. This envelope has the minimum value of maximum growth rate (0.022 1/h). Its maximum production rate is reduced to 25.18 mmol/gDCW/h which is about 8 percent reduction compared to the primary

envelope. So, if the higher number of knockouts is not important, this envelope can be selected as the candidate envelope for designing because of its specific characteristics.

3.2. Effect of reduction of the number of knockouts on an envelope

After selecting an envelope, its MAR can be used to determine the effect of reduction of number of knockouts using the modified RobustKnock method. An expansion of the envelope for the values less than that proposed by OptEnvelope is expected. To indicate the effect of the reduction, the expansion of the primary envelope for acetate (presented in the previous section) versus number of knockouts is presented in Figure 4. It can be seen that the knockout envelope presented with red colour become bigger with decrease of the number of knockouts. The envelope for one and two knockouts is the same and a little smaller than that for the wild-type.

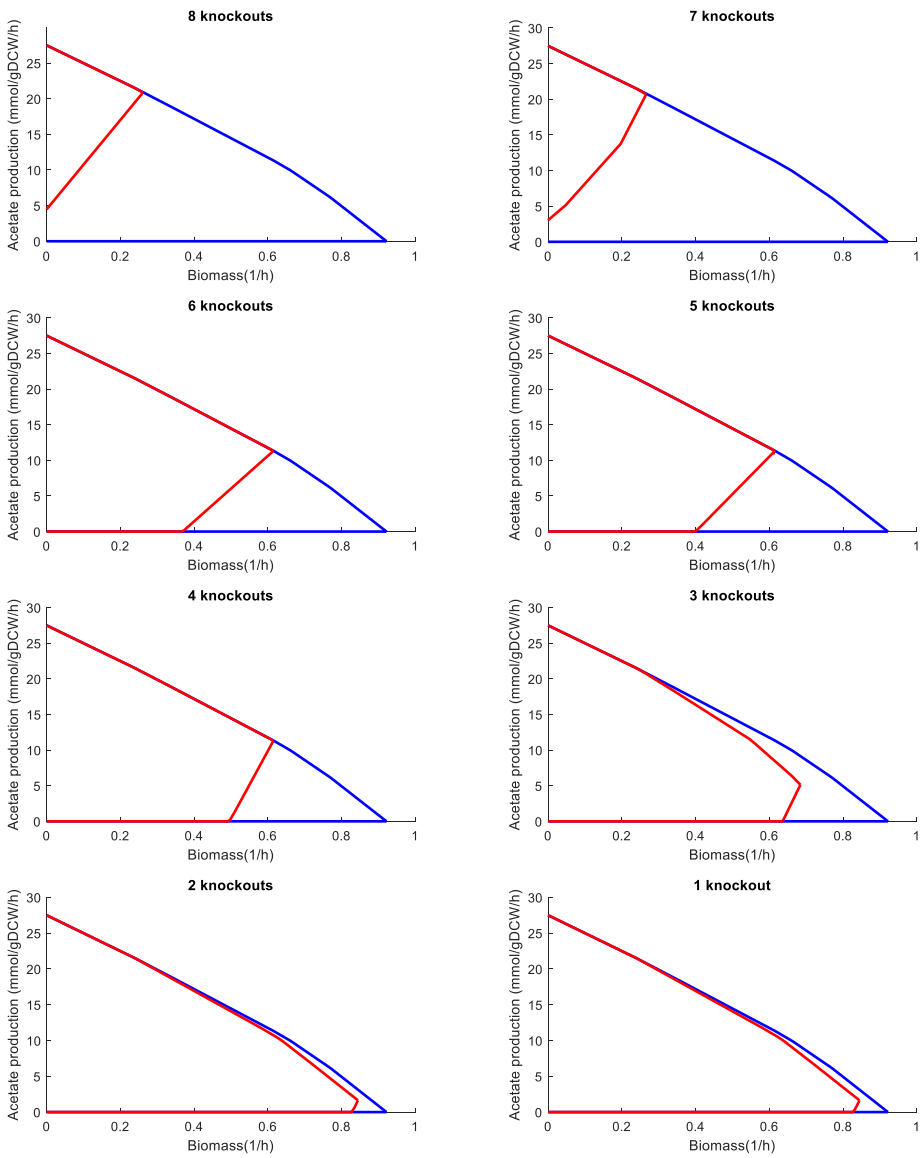

Figure 4. The effect of number of knockouts on the envelope for acetate production using *E. coli*.

3.3. Finding optimal envelope and minimal knockouts

OptEnvelope was used to study the production of acetate and glycerol under aerobic growth and succinate and ethanol under anaerobic conditions in two model prokaryotic and eukaryotic microbes *E. coli* and *S. cerevisiae* as shown in Figure 5. The results are presented in the unit g/g for better comparison with the reported experimental data in the literature.

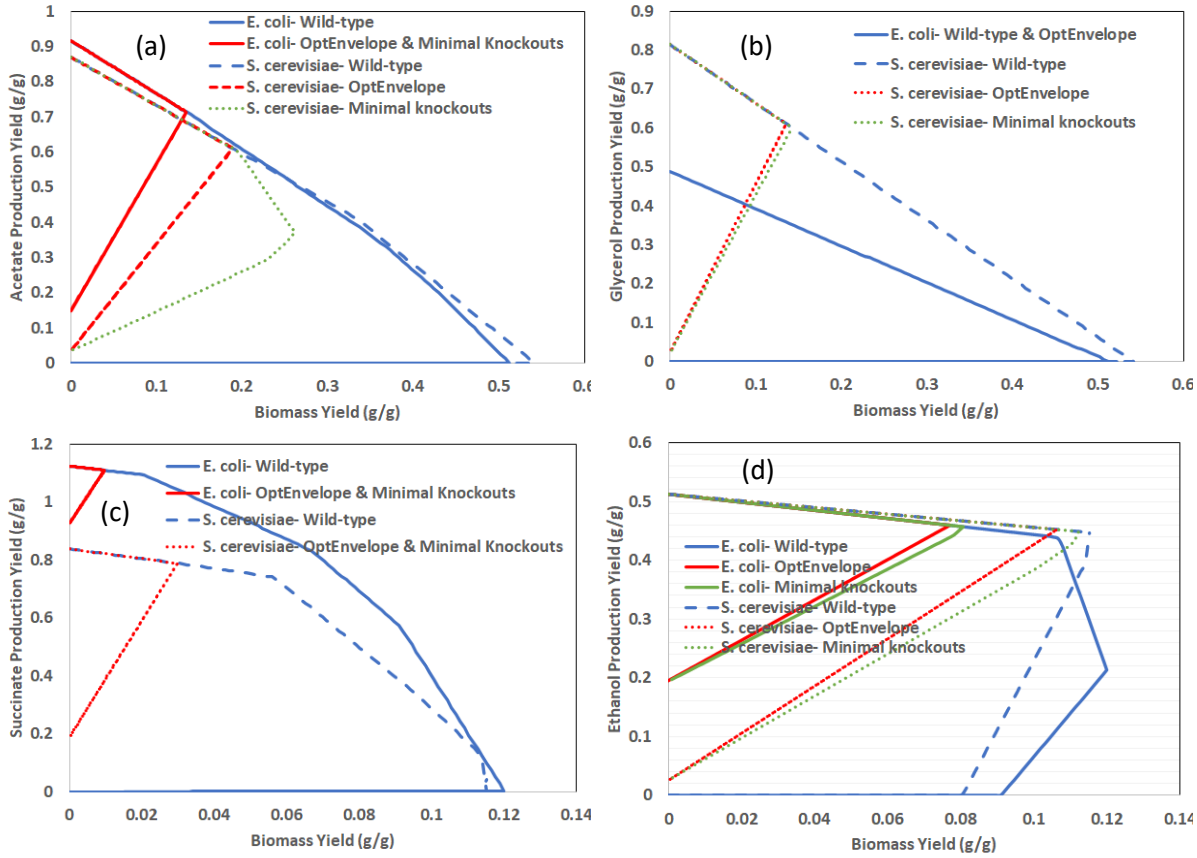

Figure 5. The wild-type and optimal envelopes of production yield versus biomass yield for growth of *E. coli* and *S. cerevisiae* on a glucose minimal medium. Aerobic conditions for a) acetate and b) glycerol and anaerobic conditions for c) succinate and d) ethanol were applied. Coupling glycerol with biomass is impossible for *E. coli*.

3.3.1. Case 1: Acetate production

Figure 5a indicates that the found optimal envelope by OptEnvelope for *E. coli* is better than that for *S. cerevisiae*. The maximum acetate production yield is 0.91 and 0.87 g/g and the minimum yield is 0.15 and 0.003 g/g for *E. coli* and *S. cerevisiae*, respectively. The maximal growth yield of the bacterium has decreased from about 0.51 g/g in the wild-type envelope to 0.14 g/g in the optimal envelope while this reduction is from about 0.54 to 0.19 g/g for yeast. Acetate production yield at the most possible point is 0.71 and 0.6 g/g (0.78 and 0.69% of maximum theoretical yield) for *E. coli* and *S. cerevisiae*, respectively. Furthermore, the envelope of *S. cerevisiae* after reinserting the reactions is worse than its optimal envelope (Figure 5a); the maximal growth yield is elevated to 0.26 g/g with a reduction of the acetate yield to 0.36 g/g (41% of maximum theoretical yield) while there is no change for *E. coli*'s envelope after reinsertion. The envelope for *E. coli* was the same as the optimal envelope with

the identical value of maximal growth yield after reinserting reactions. In the third step, 8 and 48 knockouts are needed for *E. coli* and *S. cerevisiae*, respectively.

All the results indicate that *E. coli* is a better host for acetate production and it is in accordance with the characteristic of *E. coli* as a natural producer of acetate. Acetate is a natural overflow metabolism product in *E. coli*. Due to its capability of inhibiting growth even in small concentrations and diverting carbon flow from other desired products such as recombinant proteins, acetate production in *E. coli* is usually tried to be minimized [16, 17]. Experimental data demonstrate that the maximum theoretical yield for aerobic acetate production in *E. coli* is 0.67 g/g glucose and *E. coli* has been reported to achieve 86% of the maximum theoretical yield [18]. This has been achieved using a strain with deletions of genes focA pflB frdBC ldhA atp adhE sucA.

Although small amounts of acetate are produced during ethanol fermentation in *S. cerevisiae*, growth on glucose is severely inhibited with even small acetate concentrations, because of glycolysis repression when acetate is present [19]. This means, that maximizing acetate production in *S. cerevisiae* is ineffective and the possible production of this metabolite can't be studied effectively. This seems to suggest, that even if *S. cerevisiae* could be engineered to have a higher acetate tolerance, it still wouldn't be a preferred host for acetate production, strictly because of the structure of its metabolic network requiring 48 knockouts.

The 8 reactions that must be knocked out to preserve the minimal acetate production of the *E. coli* optimal envelope include AKGt2r, F6PA, GLCDe, MTHFD, PFK, PGL, SUCCt2b, and SUCD4.

3.3.2. Case 2: Glycerol overproduction

Figure 5b indicates that the maximum theoretical yield of glycerol production is 0.48 and 0.81 g/g using *E. coli* and *S. cerevisiae*, respectively. This initial result demonstrates the higher capability of the metabolism of *S. cerevisiae* for glycerol production. iJR904 predicts that optimal production of glycerol without acetate secretion is impossible for *E. coli* and the carbon dioxide secretion rate is about two times higher than this value for iMM904 when glycerol production is the objective function. Furthermore, OptEnvelope could not find a better envelope than that for wild type when using the *E. coli* model while it resulted in an optimal envelope with a positive minimum glycerol yield of 0.03 g/g for yeast. The maximal growth yield is reduced to about 0.14 g/g with a glycerol yield of 0.6 g/g at the MP point. The optimal envelope and the envelope with minimal knockouts including 10 reaction knockouts are almost the same.

S. cerevisiae is a well-known producer of glycerol and its better results illustrate that its metabolism has evolved to be a good glycerol producer. The experimental results show that *S. cerevisiae* uses glycerol as a sink for excess NADH and glycerol production increases with increased cytosolic NADH concentrations. Hence, the deletion of NADH2_u6cm (NADH dehydrogenase) suggested by OptEnvelope is logical. In addition, the deletion of gene responsible for TPI is previously suggested for the overproduction of glycerol and by deleting the TPI1 gene, which is responsible for the isomerization of glyceraldehyde-3-phosphate and DHAP, a glycerol yield of 0.51 g/g glucose can be achieved [20, 21]. By deleting TPI1 and ADH1 genes and overexpressing GPD1 and ALD3, Cordier et al. [21] achieved a glycerol yield of 0.46 g/g glucose.

The experimental data also agree with the prediction of OptEnvelope for *E. coli* and previous reports confirm that *E. coli* is not a natural producer of glycerol and its fermentation has been achieved in *E. coli* using added pathways from other organisms [22]. A glycerol yield of 1 mol/mol glucose has been reported, using a strain evolved with the glycerol pathway from *S. cerevisiae* with deletions of the *tpiA, mgsA* and *edd* genes [23].

3.3.3. Case 3: Succinate overproduction

Unlike with glycerol, *E. coli* is a better succinate producer than *S. cerevisiae.* The maximum theoretical yield of succinate for *E.coli* under anaerobic growth is 1.12 g/g which is higher than this value for *S. cerevisiae* (0.84 g/g) as presented in Figure 5c. OptEnvelope presents envelopes for both models with the same envelope after reaction reinsertion. The envelope is excellent with a minimum succinate production yield of 0.93 g/g and maximal growth yield of 0.01 g/g at a succinate yield of 1.1 g/g. This envelope can be obtained with nine knockouts while for achieving the envelope of *S. cerevisiae*, at least 46 knockouts are required. The succinate yields presented in two review papers [24, 25] confirm that *E. coli* is an ideal host, and the yields for *S. cerevisiae* are much lower. It is also mentioned that achieving the highest yield of 1.172 mol/mol glucose (1.12 g/g) is possible using *E. coli* [24]. The experimental results [24-26] also prove the positive effect of deletion of genes ldhA, pta, citF and ptsFGH supporting reactions LDH_D, PTAr, CITL and GLCpts proposed by OptEnvelope for overproduction.

3.3.4. Case 4: Ethanol overproduction

While for three earlier cases, one of the two hosts was clearly superior, both microbes present acceptable results for ethanol production that indicate *S. cerevisiae* and *E. coli* can be promising ethanol producers (Figure 5d). Both microbes have the same maximum theoretical yield of 0.51 g/g and wild-type strains produce ethanol under optimal anaerobic growth. However, ethanol is the main fermentation product of *S. cerevisiae*, and the wild-type envelope of this microbe secrets ethanol with a yield of 0.45 g/g glucose (88% of maximum theoretical yield) at a maximal anaerobic growth yield of 0.115 g/g according to Figure 5d. This ethanol yield is close to the maximum theoretical yield and the overproduction proposed by OptEnvelope with five reaction knockouts slightly increases the yield. This indicates that the metabolism of this microbe is designed for ethanol production and more knockout is not needed. There are experimental reports that prove even wild-type *S. cerevisiae* strains can achieve ethanol yields above 90% of the maximum theoretical, with some production processes in the Brazilian fuel production industry reporting 0.469 g ethanol per g glucose [27]. In addition, considering other important factors such as the tolerance to ethanol concentration and the possibility to produce ethanol without metabolic engineering interventions, *S. cerevisiae* seems to be the better ethanol producer.

OptEnvelope is highly effective in the overproduction of *E. coli* and the ethanol yield at maximal anaerobic growth is improved from 0.12 to 0.466 g/g (91% of maximum theoretical yield) with eleven reaction knockouts. These results demonstrate that more knockouts are required to have an ethanol producer of *E. coli* compared to *S. cerevisiae*. The experimental data confirm that achieving high values of ethanol yield is possible even for *E. coli* with multiple gene knockouts. For example, by deleting eight genes from the *E. coli* MG1655, Trinh et al. [28] produced the strain TCS083 (MG1655 *Δzwf Δndh ΔsfcA ΔmaeB ΔldhA ΔfrdA ΔpoxB Δpta*), which could produce ethanol with a yield of

0.36 g/g glucose in anaerobic conditions. Furthermore, by adding pdc and adhB genes from *Zymomonas mobilis* to the TCS083 strain, the yield improved to 0.49 g/g. Woodruff et al. [29] also developed an ethanol-producing strain LW06 which can reach up to 80% of the theoretical yield on glucose.

4. Discussion

While the previous methods of coupling attempt to directly find the best knockouts, OptEnvelope first uses a MILP formula to find minimal active reaction (MAR) at an intended point located on the upper edge of the desired space of the wild-type envelope. In this research, the first target point was selected to be located on the upper edge with small growth yield because by-product secretion on the upper edge of the envelope is minimal and the maximal shift of resources from biomass to the desired product occurs. The space having a higher production rate and lower growth rate is desired because more product yield can be achieved after coupling. After finding MAR and removing the reactions that are not included in this subset, an envelope with a smaller area than the wild-type envelope is acquired, because of the variability of some active reactions at the point. So, the smallest envelope including the target point is obtained and then, in the third step, OptEnvelope tries to preserve one property of this envelope while the number of removed reactions is reduced because deletion of fewer reactions is often desired and is more feasible for the implementation. The strategy of initially finding the envelope and then finding knockouts makes OptEnvelope able to suggest a strong coupling, if it exists.

Considering that sometimes the envelopes located in the middle area are more desirable because of some of their features, OptEnvelope tries to find all desired envelopes based on moving the target point towards the middle of the feasible space. As OptEnvelope is simple and fast to run, different target points can be easily evaluated and exploited. Then, the user can compare the envelopes based on the various criteria (minimal production rate, MP production rate, number of knockouts, biomass production at MP point) and select the best envelope for strain design. The user can decide about the number of knockouts and whether moving to the upper edges or staying in the middle of the feasible solution is better if that can be reached by a lower number of knockouts. It is also possible to determine the number of allowed knockouts as an input of OptEnvelope. After selecting the best envelope, if the number of knockouts is not acceptable, it can easily be reduced to a favourable value by the user and the effect of the reduction of the number of knockouts on the shape of the envelope is presented by OptEnvelope. Increasing the number of the middle points can increase the chance of finding good envelopes at the cost of increasing the computation time.

The size of the envelope and the number of deletions depend on the peculiarities of the metabolism of the producer organism. OptEnvelope is a perfect tool for finding the most appropriate organism for particular product having better productivity at maximal growth and smaller number of deletions as a criterion. This method ranked correctly the appropriateness of the analysed hosts indicating natural adaptation of microbes for a specific product. A host is not appropriate for a product if the optimal envelope is not strong or if the number of knockouts is high. In such case changing the host is can be considered instead of shifting to a worse envelope of existing host with a lower number of knockouts. Better result for a host is an important indication that the metabolism of a microbe has naturally evolved to secrete a product of interest.

OptEnvelope can serve also for the assessment of sustainability features of an organism for production of metabolite of interest due to easy switch between the models of organisms and the possibility to implement sustainability objective function [2] as a criterion in constraint-based stoichiometric modelling carried out by COBRA toolbox.

Acknowledgments

This research has been supported by the European Regional Development Fund within the project No. 1.1.1.1/20/A/137 ‘‘genome scale metabolic modelling linked bioreactor control system (GenCon)”.

References

1. Ko, Y.-S., et al., *Tools and strategies of systems metabolic engineering for the development of microbial cell factories for chemical production.* Chemical Society Reviews, 2020. **49**(14): p. 4615-4636.
2. Stalidzans, E. and E. Dace, *Sustainable metabolic engineering for sustainability optimisation of industrial biotechnology.* Computational and Structural Biotechnology Journal, 2021. **19**: p. 4770-4776.
3. Schneider, P., R. Mahadevan, and S. Klamt, *Systematizing the different notions of growth-coupled product synthesis and a single framework for computing corresponding strain designs.* Biotechnology Journal, 2021. **16**(12): p. 2100236.
4. Alter, T.B. and B.E. Ebert, *Determination of growth-coupling strategies and their underlying principles.* BMC bioinformatics, 2019. **20**(1): p. 1-17.
5. Burgard, A.P., P. Pharkya, and C.D. Maranas, *Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization.* Biotechnology and bioengineering, 2003. **84**(6): p. 647-657.
6. Tepper, N. and T. Shlomi, *Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways.* Bioinformatics, 2010. **26**(4): p. 536-543.
7. Jensen, K., et al., *OptCouple: joint simulation of gene knockouts, insertions and medium modifications for prediction of growth-coupled strain designs.* Metabolic engineering communications, 2019. **8**: p. e00087.
8. Garcia, S. and C.T. Trinh, *Multiobjective strain design: a framework for modular cell engineering.* Metabolic engineering, 2019. **51**: p. 110-120.
9. Schneider, P., et al., *StrainDesign: a comprehensive Python package for computational design of metabolic networks.* Bioinformatics, 2022.
10. Schneider, P., A. von Kamp, and S. Klamt, *An extended and generalized framework for the calculation of metabolic intervention strategies based on minimal cut sets.* PLoS computational biology, 2020. **16**(7): p. e1008110.
11. Reed, J.L., et al., *An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR).* Genome biology, 2003. **4**(9): p. 1-12.
12. Mo, M.L., B.Ø. Palsson, and M.J. Herrgård, *Connecting extracellular metabolomic measurements to intracellular flux states in yeast.* BMC systems biology, 2009. **3**(1): p. 1-17.
13. von Kamp, A. and S. Klamt, *Growth-coupled overproduction is feasible for almost all metabolites in five major production organisms.* Nature communications, 2017. **8**(1): p. 1-10.
14. Becker, S.A., et al., *Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox.* Nature protocols, 2007. **2**(3): p. 727-738.

15. Klamt, S. and R. Mahadevan, *On the feasibility of growth-coupled product synthesis in microbial strains.* Metabolic engineering, 2015. **30**: p. 166-178.
16. Eiteman, M.A. and E. Altman, *Overcoming acetate in Escherichia coli recombinant protein fermentations.* Trends in biotechnology, 2006. **24**(11): p. 530-536.
17. De Mey, M., et al., *Minimizing acetate formation in E. coli fermentations.* Journal of Industrial Microbiology and Biotechnology, 2007. **34**(11): p. 689-700.
18. Causey, T., et al., *Engineering the metabolism of Escherichia coli W3110 for the conversion of sugar to redox-neutral and oxidized products: homoacetate production.* Proceedings of the National Academy of Sciences, 2003. **100**(3): p. 825-832.
19. Giannattasio, S., et al., *Molecular mechanisms of Saccharomyces cerevisiae stress adaptation and programmed cell death in response to acetic acid.* Frontiers in microbiology, 2013. **4**: p. 33.
20. Overkamp, K.M., et al., *Metabolic engineering of glycerol production in Saccharomyces cerevisiae.* Applied and Environmental Microbiology, 2002. **68**(6): p. 2814-2821.
21. Cordier, H., et al., *A metabolic and genomic study of engineered Saccharomyces cerevisiae strains for high glycerol production.* Metabolic engineering, 2007. **9**(4): p. 364-378.
22. Lim, C., A.L. Lin, and H. Zhao, *Metabolic strategies for microbial glycerol overproduction.* Journal of Chemical Technology & Biotechnology, 2018. **93**(3): p. 624-628.
23. Salles, I.M., et al., *Evolution of a Saccharomyces cerevisiae metabolic pathway in Escherichia coli.* Metabolic Engineering, 2007. **9**(2): p. 152-159.
24. Cheng, K.-K., et al., *Improved succinate production by metabolic engineering.* BioMed research international, 2013. **2013**.
25. Cao, Y., et al., *Fermentative succinate production: an emerging technology to replace the traditional petrochemical processes.* BioMed research international, 2013. **2013**.
26. Dai, Z., et al., *Bio-based succinic acid: an overview of strain development, substrate utilization, and downstream purification.* Biofuels, Bioproducts and Biorefining, 2020. **14**(5): p. 965-985.
27. Della-Bianca, B.E., et al., *What do we know about the yeast strains from the Brazilian fuel ethanol industry?* Applied microbiology and biotechnology, 2013. **97**(3): p. 979-991.
28. Trinh, C.T., P. Unrean, and F. Srienc, *Minimal Escherichia coli cell for the most efficient production of ethanol from hexoses and pentoses.* Applied and environmental microbiology, 2008. **74**(12): p. 3634-3643.
29. Woodruff, L.B., et al., *Towards a metabolic engineering strain "commons": an Escherichia coli platform strain for ethanol production.* Biotechnology and bioengineering, 2013. **110**(5): p. 1520-1526.