CFSA: Comparative Flux Sampling Analysis as a Guide for Strain Design

van Rosmalen, R.P. Moreno-Paz, S. Duman-Özdamar, Z.E.
Suarez-Diez, M.

Abstract

Genome-scale metabolic models of microbial metabolism have extensively been used to guide the design of microbial cell factories, still, many of the available strain design algorithms often fail to produce a reduced list of targets for improved performance that can be implemented and validated in a step-wise manner. We present Comparative Flux Sampling Analysis (CFSA), a strain design method based on the extensive comparison of complete metabolic spaces corresponding to maximal or near-maximal growth and production phenotypes. The comparison is complemented by statistical analysis to identify reactions with altered flux that are suggested as targets for genetic interventions including up-regulations, down-regulations and gene deletions. We apply CFSA to the production of lipids by *Cutaneotrichosporon oleaginosus* and naringenin by *Saccharomyces cerevisiae* identifying engineering targets in agreement with previous studies as well as new interventions. CFSA is an easy-to-use, robust method that suggests potential metabolic engineering targets for growth-uncoupled production that can be integrated in Design-Build-Test-Learn cycles for the design of microbial cell factories.

1 Introduction

Microbial cell factories (MCF) are microorganisms engineered for the production of bio-molecules that thrive on renewable carbon sources. A broad range of bio-molecules can be produced using MCF from non-edible feedstocks such as recalcitrant biomass or industrial waste streams thereby providing sustainable replacements for production systems based on fossil fuels [1]. MCF design requires the choice of an appropriate host strain (chassis), and selection of a suitable available pathway, the discovery of new pathways, or the design of synthetic pathways for new-to-nature compounds. Still, industrial feasibility requires extensive engineering to improve MCF performance [2].

GEnome-scale Metabolic models (GEM) comprehensively represent cellular metabolism and allow the simulation of metabolic fluxes and the prediction of cellular phenotypes. They have largely been used to identify metabolic engineering targets to optimize pathway performance [3]. Machado et al. [4] classify strain design algorithms relying on GEMs in two main groups: methods based on the analysis of elementary flux modes (EFM), and those based on the optimization of an objective function.

EFMs are minimal sets of reactions that can jointly operate at steady-state such that all steady-state solutions can be described as a combination of EFMs [5]. They provide an unbiased framework to explore the metabolic space but have limited scaling potential and applicability to larger models. Instead, related approaches, based on minimal cut sets (MCS) such as minimal metabolic functionality (MMF) and FluxDesign, scale better to genome-scale [4, 6].

Optimization-based approaches rely on the simulation of metabolic fluxes of *wild type* and/or mutant strains using an objective function. Many of these methods use Flux Balance Analysis (FBA) to calculate fluxes and therefore only explore one of the multiple flux distributions that can lead to the optimal objective. In this way, OptKnock [7] and derived algorithms such as RobustKnock [8], OptGene [9], and OptCouple [10], aim at identifying gene knock-outs to couple the production of the

compound of interest to the production of biomass using growth as the objective function. Other methods such as OptForce [11] and OptDesign [12] are used to predict modulation of gene expression, including up and down-regulations. They compare simulated fluxes of growth and production phenotypes and minimize the number of interventions required for the overproduction phenotype. Although flux variability analysis (FVA) might be used to explore feasible flux ranges and constrain the solution space, comparisons among phenotypes are based on non-unique FBA solutions. Alternatively, flux sampling allows the exploration of the full space of feasible flux distribution of each reaction given a set of constraints on the metabolic model [13]. Although flux sampling has been used to evaluate metabolic flux differences between different conditions [**ravi2021**], it has not yet been applied for strain design.

From the strain design perspective, growth-coupling strategies are suitable for experimental implementation and further optimization using adaptive laboratory evolution but require multiple simultaneous interventions. Alternatively, Design-Build-Test-Lern (DBTL) cycles are useful to iteratively improve the performance of strains with growth-uncoupled production [**Carbonell2018**]. These strains can be used in two-stage fermentation processes where the growth and production phases are sequential in order to alleviate metabolic stress and improve productivity [14, 15].

We present Comparative Flux Space Analysis (CFSA), a model-guided strain design approach based on extensive sampling of the feasible solution space in alternative scenarios. Growth and production phenotypes are simulated and compared, also with a growth-limited scenario, which serves as a negative control for down-and up-regulation targets. Flux distributions are statistically compared resulting on the identification of potential down-regulations, knock-outs and over-expressions targets leading to growth-uncoupled increased production. As a proof of concept, we use CFSA to identify metabolic engineering targets for lipid production by *C. oleaginosus* and naringenin production by *S. cerevisiae* and compare them with available data.

2 Methods

2.1 Comparative Flux Sampling Analysis (CFSA)

To implement CFSA a GEM of the desired organism including the production pathway of choice is required. As first step, media conditions (e.g. substrate uptake rate, aerobic/anaerobic growth...) are specified. Model reactions are grouped in seven categories to facilitate later filtering: *required* reactions including the growth and maintenance reactions; *not biological* reactions including boundary, exchange, sink and demand reactions; *blocked* reactions (unable to carry flux); reactions *without associated genes*; *essential* reactions; reactions *containing essential genes* and *transport* reactions.

Reaction fluxes are sampled from the metabolic solution space in three scenarios: growth, slow growth and production. In the growth scenario, a constraint on growth (lower bound of biomass synthesis reaction) ensures that sampled flux distributions result in at least 90% of the optimal growth predicted by FBA. This scenario is used as a reference for comparison. In the production scenario, a lower bound in the production reaction ensures that flux distributions result in at least 90% of the optimal production predicted by FBA. In the slow growth scenario, the maximum growth rate compatible to a production rate of 90% of the optimal production is calculated and used as an upper bound for the biomass synthesis reaction. To limit the solution space a parsimonious FBA approach [16] is implemented by introducing an additional constraint to limit the total sum of fluxes to the minimum value compatible with maximal growth. This extra constraint is applied in the production and slow growth scenario simulations to limit unrealistic futile cycles.

The Optimal Gaussian Process (OptGP) sampler [17] implemented in cobrapy [18] is used to model the distribution of the target space and iteratively sample from this distribution. A thinning parameter is used to reduce the correlation between samples. Invalid samples (i.e. those that do not meet constraints specific to each scenario) are discarded. Last, the Geweke diagnostic is used to calculate the chain convergence for each process, and samples corresponding to reactions whose distributions have not converged are discarded.

For each reaction, the two-sample Kolmogorov-Smirnov (KS) test is used to compare samples from different scenarios and determine if they belong to the same continuous distribution. Potential

targets are selected if, for a specific reaction, distributions differ in the scenarios based on KS-statistics and p-values corrected for multiple testing using Bonferroni. As default, reactions with KS-statistics above 0.5 for comparing the growth and production scenarios, KS-statistics above 0.25 for comparing the production and slow growth scenarios and corrected p-values below 0.05 are considered significant. Besides, reactions whose fluxes correlate with fluxes through the biomass synthesis reaction, or that are not associated to genes (through a gene-protein-reaction association) are discarded as potential targets. Lastly, only reactions whose absolute change in flux between the growth and production scenario is bigger than a threshold ($0.01 mmol \cdot g_{dw}^{-1} \cdot h^{-1}$ as default) are considered to be suitable targets.

Potential targets are then divided into over-expression or down-regulation targets depending on whether the mean fold change comparing growth and production scenarios is above or below 1. Knock-down targets that correspond to non-essential genes are classified as possible knock-out targets. Reactions are clustered based on the correlation of the absolute fluxes between samples in order to identify redundant targets (i.e. belonging to the same metabolic pathway).

2.2 Selected applications

2.2.1 Production of lipids by *C. oleaginosus*

The *i*NP636_*Coleaginosus*_ATCC20509 genome-scale metabolic model was manually curated to provide all fatty acid elongation reactions [19]. In total 7 reactions were added, 1 of them in the cytoplasm and 6 of them in the mitochondria and all reactions were associated with the corresponding genes. Details of the model curation can be found on GitLab (GEMs folder). Glycerol and urea were used as the carbon and nitrogen sources respectively and nitrogen-depleted biomass composition was assumed. CFSA was used with the default parameters using the lipid synthesis reaction (lipid_synthesis) as a target for the production scenario. Targets were further filtered by removing reactions with ATP and/or NADPH as substrates. These metabolites are required for maintenance and lipid synthesis and increased consumption might result in an energy burden on the microorganism. Reactions catalyzed by genes annotated for transport reactions or protein complexes were not considered due to potential cumbersome implementation *in vivo*. The feasibility of selected targets was evaluated based on previous studies.

2.2.2 Production of naringenin by *S. cerevisiae*

The Yeast8 genome-scale metabolic model of *S. cerevisiae* [20] was modified to include the naringenin production pathway (Table 1) and glucose was the selected carbon source. CFSA was used with default parameters with the naringenin exchange reaction (EX_NAR) as an objective for the production scenario. Two proteomic datasets representative of *S. cerevisiae* aerobic growth on glucose during shake flask and chemostat fermentations were used as additional filters for the sampling results [21, 22]. Undetected proteins were not considered relevant targets.

3 Results

3.1 Comparative Flux Sampling Analysis (CFSA)

CFSA facilitates target identification through the comparison of the metabolic spaces obtained in three simulation scenarios: maximal growth, maximal production and slow growth, a scenario used as a negative control. This comparison is based on the flux distributions through each model reaction obtained using flux sampling which ensures the complete exploration of the solution space and leads to the identification of intervention targets.

Figure 1 (top left) shows genes classified as over-expression targets as the absolute flux through the corresponding reaction is higher in the production scenario regardless of the growth rate represented by the growth and slow growth scenarios. Similarly, genes are classified as down-regulation targets when the absolute flux through the corresponding reactions is lower in the production scenario compared to the growth scenario and does not overlap with the slow growth scenario (Fig

Table 1: Naringenin pathway reactions added to Yeast8. Reaction names in the model, reaction equations and Gene-Protein-Reaction (GPR) rules are included. PAL, phenylalanine ammonia-lyase; TAL, tyrosine ammonia-lyase; C4H, cinnamate 4-hydroxylase; CL, 4-coumarate-CoA ligase; CHS, naringenin-chalcone synthase; CHI, chalcone isomerase; NARt, naringenin transport; EX_NAR, naringenin exchange; L-Phe, L-phenylalanine; CIN, cinnamate; L-Tyr, L-tyrosine; 4-CUA, 4-coumarate; CoA, Coenzyme-A; CUACoA, 4-coumaroyl-CoA; MaCoA, malonyl-CoA; NAR-C, naringenin chalcone; NAR, naringenin. _c and _e designate metabolites in the cytoplasm and the extracellular space respectively.

Reaction name	**Reaction equation**	**GPR**
PAL	L-Phe_c → CIN_c + NH_4_c	*pal*
TAL	L-Tyr_c → 4-CUA_c + NH_4_c	*tal*
C4H	CIN_c + H^+_c + NADPH_c + O_2_c → 4-CUA_c + H_2O_c + NADP_c	*c4h* and *YHR042W*
CL	4-CUA_c + ATP_c + CoA_c + 4.0 H^+_c → 4-CUACoA_c + AMP_c + PPi_c	*4cl*
CHS	4-CUACoA_c + 3.0 MaCoA_c → NAR-C_c + 3.0 CO_2_c + 4.0 CoA_c + H^+_c	*chs*
CHI	NAR-C_c → NAR_c	chi
NARt	NAR_c ↔ NAR_e	
EX_NAR	NAR_e →	

1, top right). The most extreme case of a down-regulation, a knock-out, is obtained when, for a down-regulation target, the flux in the production scenario is zero and the gene is not classified as essential (Fig 1, bottom left).

Growth and production are competing objectives and often low fluxes obtained in the production scenario are not related to an increased production but to a decreased growth rate (e.g. fluxes through reactions for biomass components). In order to avoid the identification of these genes as down-regulation targets, reactions where the production and slow growth distributions overlap are considered false positives (Fig 1, bottom right). CFSA has been implemented in python v.3.7 us-

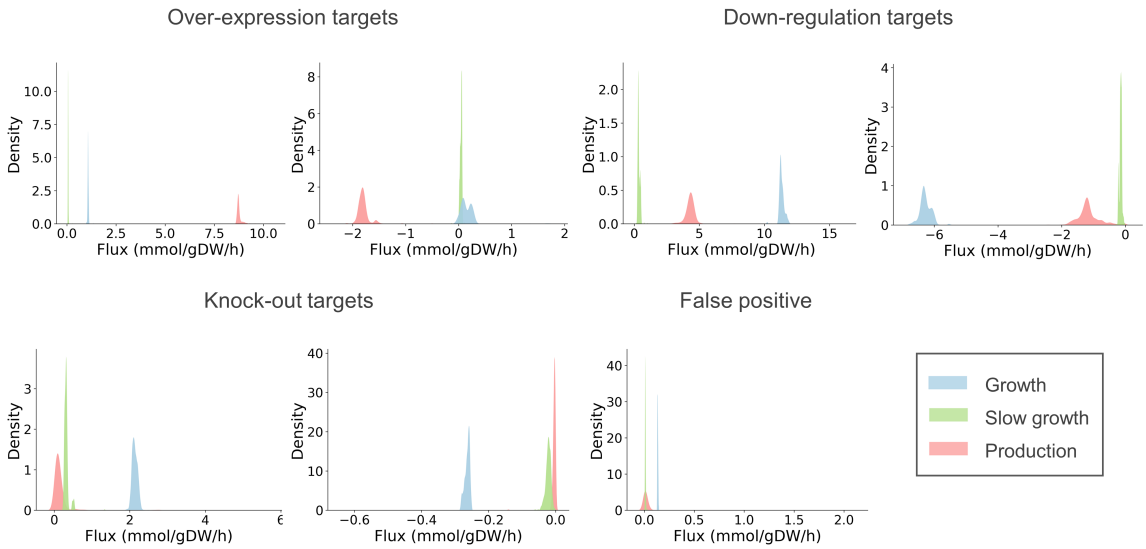
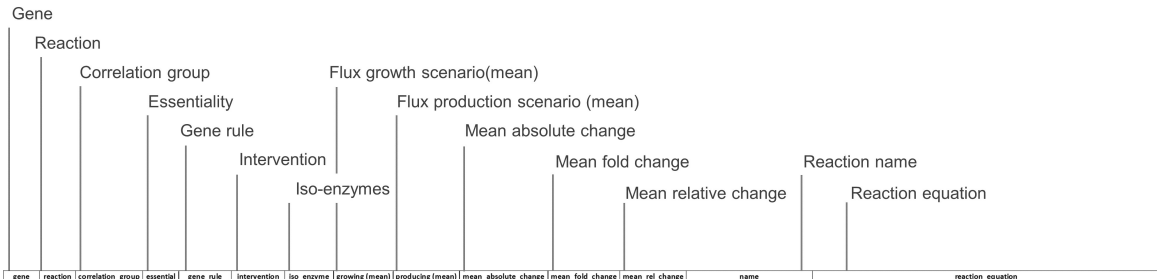


Figure 1: Example of flux distribution plots illustrating behaviours of candidate targets. In each panel, the x-axis represents possible flux values, and the y-axis represents the frequency of each flux value obtained when sampling the solution space, normalized to an area of 1.

ing the cobrapy v 0.26.2 toolbox and is available on GitLab. Statistical analysis methods based on Kolmogorov-Smirnov test have been implemented using scipy stats package. Besides distribution plots, the output of CFSA includes a summary Excel file (Figure 2). The summary includes predicted gene targets sorted by the mean fold change between growth and production scenarios, their corresponding reactions, their essentially, correlation group, and the presence or absence of isoenzymes which might complicate *in vivo* implementation of down-regulations .



gene	reaction	correlation_group	essential	gene_rule	intervention	iso_enzyme	growing (mean)	producing (mean)	mean_absolute_change	mean_fold_change	mean_rel_change	name	reaction_equation
YFL045C	r_0902	251	TRUE	YFL045C	knock-down	FALSE	-0.5826	-0.0031	0.5795	0.0053	0.9947	phosphomannomutase	D-mannose 1-phosphate <=> D-mannose 6-phosphate
YER003C	r_0723	250	TRUE	YER003C	knock-down	FALSE	-0.5851	-0.0038	0.5813	0.0065	0.9935	mannose-6-phosphate isomerase	D-mannose 6-phosphate <=> D-fructose 6-phosphate
YER026C	r_2446	1430	FALSE	YER026C		FALSE	0.0083	0.0026	0.0057	0.3152	0.6848	PS synthase (1-16:0, 2-16:1), ERM	CDP-diacylglycerol (1-16:0, 2-16:1) + L-serine --> H+ + CMP + phosphatidyl-L-serine (1-16:0, 2-16:1)
	r_2447	440	FALSE	YER026C	knock-out	FALSE	0.0131	0.0001	0.0130	0.0076	0.9924	PS synthase (1-16:1, 2-16:1), ERM	CDP-diacylglycerol (1-16:1, 2-16:1) + L-serine --> H+ + CMP + phosphatidyl-L-serine (1-16:1, 2-16:1)
	r_2448	294	FALSE	YER026C		FALSE	0.0045	0.0000	0.0044	0.0102	0.9898	PS synthase (1-18:0, 2-16:1), ERM	CDP-diacylglycerol (1-18:0, 2-16:1) + L-serine --> H+ + CMP + phosphatidyl-L-serine (1-18:0, 2-16:1)
	r_2449	1072	FALSE	YER026C		FALSE	0.0039	0.0021	0.0018	0.5459	0.4541	PS synthase (1-18:1, 2-16:1), ERM	CDP-diacylglycerol (1-18:1, 2-16:1) + L-serine --> H+ + CMP + phosphatidyl-L-serine (1-18:1, 2-16:1)
	r_2450	1441	FALSE	YER026C		FALSE	0.0059	0.0028	0.0031	0.4784	0.5216	PS synthase (1-16:0, 2-18:1), ERM	CDP-diacylglycerol (1-16:0, 2-18:1) + L-serine --> H+ + CMP + phosphatidyl-L-serine (1-16:0, 2-18:1)
	r_2451	475	FALSE	YER026C		FALSE	0.0043	0.0007	0.0035	0.1729	0.8271	PS synthase (1-16:1, 2-18:1), ERM	CDP-diacylglycerol (1-16:1, 2-18:1) + L-serine --> H+ + CMP + phosphatidyl-L-serine (1-16:1, 2-18:1)
	r_2452	514	FALSE	YER026C		FALSE	0.0038	0.0002	0.0036	0.0622	0.9378	PS synthase (1-18:0, 2-18:1), ERM	CDP-diacylglycerol (1-18:0, 2-18:1) + L-serine --> H+ + CMP + phosphatidyl-L-serine (1-18:0, 2-18:1)
	r_2453	448	FALSE	YER026C		FALSE	0.0043	0.0011	0.0032	0.2608	0.7392	PS synthase (1-18:1, 2-18:1), ERM	CDP-diacylglycerol (1-18:1, 2-18:1) + L-serine --> H+ + CMP + phosphatidyl-L-serine (1-18:0, 2-18:1)

Figure 2: Example of output Excel file.

3.2 Production of lipids by *C. oleaginosus*.

C. oleaginosus is an oleaginous yeast able to accumulate lipids above 40% (w/w) of its biomass when growing on a nitrogen-limited cultivation medium [**Duman-Ozdamar2022**]. The composition of the produced fatty acids, stored as triacylglycerols (TAGs) within lipid bodies in the cell, is comparable to commonly used plant-derived-oils and has therefore been flagged as an auspicious MCF for sustainable lipid production at industrial scale [23, 24]. The mechanism of lipid accumulation starts with the transport of citrate into the cytosol where it is cleaved into acetyl-CoA by ATP citrate lyase (ACL). Acetyl-CoA is converted into malonyl-CoA by acetyl-CoA carboxylase (ACC) leading to the activation of the lipid synthesis and elongation pathways (Fig 3 A) [25, 26].

CFSA was applied to improve lipid production, resulting in 3 genes (6 reactions) candidates for down-regulation and 151 genes (662 reactions belonging to 142 groups) for over-expression. Due to the challenges related to the experimental implementation of down-regulation targets in this organism, we focused on over-expression candidates. Visual inspection of the distribution graphs and the removal of reactions that consume NADPH and ATP, key cofactors for lipid synthesis, produced a list of 15 targets for over-expression (Supplementary Data). As expected, the key reactions ACL and ACC, as well as reactions responsible for fatty acid synthesis and elongation were predicted as up-regulation targets. Similarly, over-expression of ACL and ACC in *Yarrowia lipolytica*, also an oleaginous yeast, improved lipid accumulation between 2 and 3.5-folds [27, 28].

In addition to targets related to the lipid synthesis pathway, reactions from the mevalonate and amino acid synthesis pathways were suggested as beneficial up-regulations. A detailed inspection of *C. oleaginosus* GEM showed that the mevalonate pathway provides an alternative route that improves citrate and isocitrate supply to the cytosol which could potentially be used for lipid synthesis (Fig 3 A). This pathway has been engineered in *Y. lipolytica* to improve the production of acetyl-CoA derived products such as α-farnesane and β-carotene [29, 30]. Similarly, over-expression of genes in amino acid synthesis pathways could potentially increase the activity of the TCA cycle providing citrate for lipid production [31]. Moreover, amino acid metabolism has a significant role in the regulation of lipid synthesis of oleaginous yeasts and *in silico* simulations of amino acid addition for *Y. lipolytica* resulted in improved TAG production [32].

3.3 Production of naringenin by *S. cerevisiae*.

S. cerevisiae is a model organism with in-depth genetic and physiological characterization, ample application in industrial bioprocesses and *Generally Regarded as Safe* (GRAS) status [33, 34]. Flavonoids such as naringenin are precursors of anthocyanins and have traditionally been used for fragrance, flavour, and colour in various food types [35, 36]. Naringenin is derived from the shikimate (SHK) pathway for aromatic amino acid biosynthesis, which starts with the condensation of erythrose-4-phosphate (E4P) and phosphoenolpyruvate (PEP) as illustrated in Figure 2 B. Production of this compound requires the heterologous expression of 4CL, CHS, CHI and either PAL, C4H and CPR or TAL (Table 1), as well as an appropriate supply of malonyl-CoA [37].

We used CFSA to find metabolic engineering strategies that improve naringenin production. We obtained 216 gene candidates for down-regulation corresponding to 210 unique reactions belonging to 144 groups and 168 gene targets for up-regulation (590 unique reactions belonging to 492 groups). Two proteomic datasets covering 48.4% of the genes and 23.6% of the reactions in the model were used as additional filtering steps. Most of the reactions with significant differences in flux distributions with a non-detected protein were involved in the transport of metabolites and were therefore not deemed as appropriate metabolic engineering candidates. Finally, the reactions catalyzed by detected proteins (122 over-expressions and 139 down-regulations) were manually curated resulting in 34 and 20 candidate reactions to over-express and down-regulate respectively (Supplementary Data).

As expected, reactions from the shikimate and naringenin production pathways were predicted as over-expression targets, with a preference for the tyrosine branch. Although the model predicts down-regulation of *trp2* to decrease by-product formation, it is unable to suggest the deletion of Ehrlich pathway (EP) genes, involved in the degradation of intermediates that have shown to improve the production of other aromatic compounds [36, 38–40]. Similarly, CFSA suggested experimentally validated strategies to increase the production of acetyl-CoA and malonyl-CoA syntheses such as down-regulation of PDH and CIT2 and up-regulation of PDCS, ALD, ACS and ACC [41]. However, it fails to predict the down-regulation of fatty acid synthesis as a strategy to improve malonyl-CoA availability [34, 38, 41] (Fig 3B).

CFSA predicted the up-regulation of genes involved in the non-oxidative pentose phosphate pathway (RKI, RPE), which provides E4P and PEP (Fig 3B). Interestingly, CFSA suggested a decreased flux through the phosphofructokinase (PFK) and aldolase (FBA) reactions that result in the conversion of fructose-6-phosphate (F6P) to dihydroxyacetone-phosphate (DAP) and glyceraldehyde-3-phosphate (GAP). Instead, it proposed F6P conversion to E4P and DAP via sedoheptulose-1,7-biphosphate (S1,7bP). In the cells, PFK and FBA are responsible for both pathways and conversion of F6P to DAP, and GAP is favoured due to a higher affinity of FBA towards F1,6-bP. According to the simulations, expressing *pfk* genes with higher affinity towards S7P such as *ppi-pfk* from *Clostridium thermosuccionogenes* [42] could be a new strategy to improve naringenin production.

Besides, the model suggest down-regulation of *ser3*, the first step of serine biosynthesis, increasing the glycolytic flux towards PEP; over-expression of adenylate kinase (ADK) which has been reported to increase production of malonyl-CoA derived products [43]; and down-regulation of reactions involving tetrahydrofolic acid (THF) and glycine, likely to reduce chorismate consumption for THF formation (Fig 3B).

4 Discussion

Inspired by other strain design algorithms that suggest metabolic engineering targets based on the comparison of predicted fluxes between *wild type* and production phenotypes, we present CFSA, a tool based on flux sampling that, by analyzing the complete GEM solution space, can guide the design of MCF for growth-uncoupled production. We apply CFSA to improve the production of lipids by *C. oleaginosus*, an endogenous product in an emerging MCF, as well as the production of naringenin in *S. cerevisiae*, a heterologous product in an established MCF. In both cases, CFSA suggested experimentally validated targets including evident up-regulations belonging to the product synthesis pathway, distant targets that improve precursor availability, as well new engineering

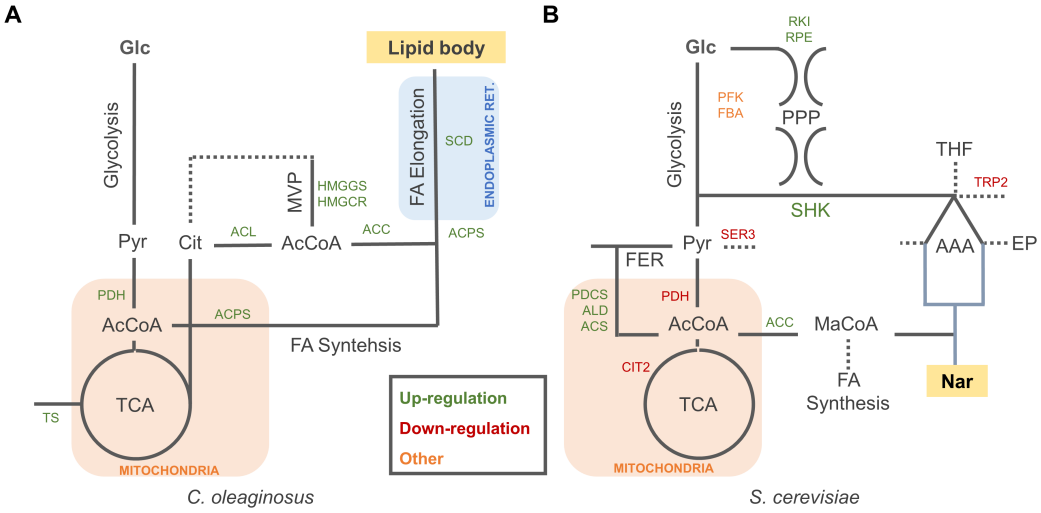


Figure 3: Summary of selected metabolic engineering targets for lipid production in *C. oleaginosus* (A) and naringenin (NAR) production in *S. cerevisiae* (B). Endogenous and heterologous metabolic pathways are simplified and presented in grey and blue respectively: TCA, tricarboxylic acid cycle; MVP, mevalonate pathway; FA, fatty acids; PPP, pentose phosphate pathway; SHK, shikimate pathway; FER, fermentation pathway; AAA, aromatic amino acid pathway; EP, Ehrlich pathway. Metabolite abbreviations: Glc, glucose; Py, pyruvate; AcCoA, acetyl-coenzyme A; Cit, Citrate; MaCoA, malonyl-coenzyme A.

strategies. However, CFSA fails to predict the down-regulation of some of the pathways involved in the degradation of production pathway intermediates (e.g. Ehrlich pathway genes deletion for naringenin production [36, 38–40]). This pitfall is shared with other GEM-based strain design approaches since, although active *in vivo*, these reactions remain inactive in GEM simulations.

In addition to the simulation of growth and production phenotypes, we include the simulation of slow growth. This scenario is used to differentiate between fluxes that change due to the simulated low-growth in the production scenario from fluxes that are potentially related to increased production. The simulation of this phenotype is used as negative control potentially reducing false positive target predictions (Figure 1). While other strategies assume maximum growth or production, we used 90% as a lower bound for the reactions to ensure that not only the optimal flux profile is sampled, but less efficient behaviour is also tolerated leading to increased robustness.

CFSA is easy to implement and will result in numerous potential targets for up-, down- regulations, and knock-outs for subsequent inspection. Default parameters can easily be modified, different filtering criteria can be applied based on user needs (e.g. use of proteomic data), and visual examination of the sampling results is a recommended approach for suitable targets selection. Other strain design methods provide minimal intervention strategies that, although useful in theory, might not return the expected results in practice. For example, when entire pathways are predicted as up-regulation targets, GEM cannot identify limiting reactions as enzyme kinetics and regulation are not considered in constraint-based models [4]. Instead, we provide a complete list of possible interventions and endow the user with additional information to make a decision based on the most feasible suggestions.

Disadvantages of CFSA include the difficulty to estimate the quantitative effect of the suggested manipulations which are identified based on statistical testing. There is no guarantee that the effect size in flux correlates to the strength of the knock-down or over-expression required. Besides, targets are suggested as individual interventions and the effect of possible combinations of targets is not provided. Still, after a first round of CFSA, the algorithm can be re-applied using GEMs including desired interventions (i.e. with modified reaction bounds) to find potential complementary targets.

As with other methods based on GEMs, errors in the models can lead to unexpected outcomes. For example, loops in the metabolic network can lead to artificially high fluxes and thus false predictions. Often parsimonious flux balance analysis [16] or loop-less flux balance analysis [44, 45] is used to limit these errors, however, implementing these in flux sampling is non-trivial. We mimicked the parsimonious approach by adding a total sum of fluxes constraint based on the growing phenotype. However, this approach, while biologically reasonable for the reference condition at a high growth rate, does not necessarily apply for the production condition, as the assumption of minimal flux and thus protein usage does not hold for this artificial scenario. Instead, loop-less flux sampling approaches such as the loopless Artificially Centered Hit-and-Run on a Box algorithm (ll-ACHRB) [46] could be used to improve this process.

CFSA is the first strain design algorithm based on flux sampling that explores the whole solution space of a GEM and suggests metabolic engineering targets for growth-uncoupled production. Its robustness, simplicity and flexibility makes it ideal to complement and systematize DBTL cycles for MCF design. CFSA predictions, including non-obvious targets, can be sequentially tested using high-throughput approaches as automated platforms and biosensor-aid screening accelerating and broadening the strain design process.

5 Acknowledgement

This project has been founded by the Netherlands Organization for Scientific Research (NWO; project number GSGT.2019.008) and the Dutch Ministry of Agriculture through the “TKI-toeslag” project LWV19221 “Tailor-made microbial oils and fatty acids”.

References

[1] S. Y. Lee et al. "A comprehensive metabolic map for production of bio-based chemicals". en. In: *Nature Catalysis* 2.1 (2019). Number: 1 Publisher: Nature Publishing Group, pp. 18–33. DOI: 10.1038/s41929-018-0212-4. URL: https://www.nature.com/articles/s41929-018-0212-4.

[2] J. S. Cho et al. "Designing Microbial Cell Factories for the Production of Chemicals". In: *JACS Au* 2.8 (2022). Publisher: American Chemical Society, pp. 1781–1799. DOI: 10.1021/jacsau.2c00344. URL: https://doi.org/10.1021/jacsau.2c00344.

[3] X. Fang, C. J. Lloyd, and B. O. Palsson. "Reconstructing organisms in silico: genome-scale models and their emerging applications". en. In: *Nature Reviews Microbiology* 18.12 (2020). Number: 12 Publisher: Nature Publishing Group, pp. 731–743. DOI: 10.1038/s41579-020-00440-4. URL: https://www.nature.com/articles/s41579-020-00440-4.

[4] D. Machado and M. J. Herrgård. "Co-Evolution of Strain Design Methods Based on Flux Balance and Elementary Mode Analysis". In: *Metabolic Engineering Communications* 2 (2015), pp. 85–92. DOI: 10.1016/j.meteno.2015.04.001.

[5] C. T. Trinh, A. Wlaschin, and F. Srienc. "Elementary Mode Analysis: A Useful Metabolic Pathway Analysis Tool for Characterizing Cellular Metabolism". In: *Applied Microbiology and Biotechnology* 81.5 (2009), pp. 813–826. DOI: 10.1007/s00253-008-1770-1.

[6] A. von Kamp and S. Klamt. "Enumeration of Smallest Intervention Strategies in Genome-Scale Metabolic Networks". In: *PLOS Computational Biology* 10.1 (2014), e1003378. DOI: 10.1371/journal.pcbi.1003378.

[7] A. P. Burgard, P. Pharkya, and C. D. Maranas. "Optknock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization". In: *Biotechnology and Bioengineering* 84.6 (2003), pp. 647–657. DOI: 10.1002/bit.10803.

[8] N. Tepper and T. Shlomi. "Predicting Metabolic Engineering Knockout Strategies for Chemical Production: Accounting for Competing Pathways". In: *Bioinformatics* 26.4 (2010), pp. 536–543. DOI: 10.1093/bioinformatics/btp704.

[9] K. R. Patil et al. "Evolutionary Programming as a Platform for in Silico Metabolic Engineering". In: *BMC Bioinformatics* 6.1 (2005), pp. 1–12. DOI: 10.1186/1471-2105-6-308.

[10] K. Jensen et al. "OptCouple: Joint Simulation of Gene Knockouts, Insertions and Medium Modifications for Prediction of Growth-Coupled Strain Designs". In: *Metabolic Engineering Communications* 8 (2019), e00087. DOI: 10.1016/j.mec.2019.e00087.

[11] S. Ranganathan, P. F. Suthers, and C. D. Maranas. "OptForce: An Optimization Procedure for Identifying All Genetic Manipulations Leading to Targeted Overproductions". In: *PLOS Computational Biology* 6.4 (2010), e1000744. DOI: 10.1371/journal.pcbi.1000744.

[12] S. Jiang et al. "OptDesign: Identifying Optimum Design Strategies in Strain Engineering for Biochemical Production". In: *ACS Synthetic Biology* (2022), acssynbio.1c00610. DOI: 10.1021/ACSSYNBIO.1C00610. URL: https://pubs.acs.org/doi/full/10.1021/acssynbio.1c00610.

[13] H. A. Herrmann et al. "Flux sampling is a powerful tool to study metabolism under changing environmental conditions". In: *npj Systems Biology and Applications* 5.1 (2019), pp. 1–8. DOI: 10.1038/s41540-019-0109-0.

[14] T. M. Lo et al. "A Two-Layer Gene Circuit for Decoupling Cell Growth from Metabolite Production". In: *Cell Systems* 3.2 (2016), pp. 133–143. DOI: 10.1016/j.cels.2016.07.012. URL: http://dx.doi.org/10.1016/j.cels.2016.07.012.

[15] K. Raj, N. Venayak, and R. Mahadevan. "Novel two-stage processes for optimal chemical production in microbes". In: *Metabolic Engineering* 62 (2020), pp. 186–197. DOI: 10.1016/j.ymben.2020.08.006. URL: https://www.sciencedirect.com/science/article/pii/S1096717620301269.

[16] N. E. Lewis et al. "Omic Data from Evolved E. Coli Are Consistent with Computed Optimal Growth from Genome-Scale Models". In: *Molecular Systems Biology* 6.1 (2010), p. 390. DOI: 10.1038/msb.2010.47.

[17] W. Megchelenbrink, M. Huynen, and E. Marchiori. "optGpSampler: An Improved Tool for Uniformly Sampling the Solution-Space of Genome-Scale Metabolic Networks". In: *PLoS ONE* 9.2 (2014). Ed. by S. Rogers, e86587. DOI: 10.1371/journal.pone.0086587.

[18] A. Ebrahim et al. "COBRApy: COnstraints-Based Reconstruction and Analysis for Python." In: *BMC systems biology* 7.1 (2013), p. 74. DOI: 10.1186/1752-0509-7-74.

[19] N. Pham et al. "Genome-Scale Metabolic Modeling Underscores the Potential of Cutaneotrichosporon Oleaginosus ATCC 20509 as a Cell Factory for Biofuel Production". In: *Biotechnology for Biofuels* 14.1 (2021), p. 2. DOI: 10.1186/s13068-020-01838-1.

[20] H. Lu et al. "A Consensus S. Cerevisiae Metabolic Model Yeast8 and Its Ecosystem for Comprehensively Probing Cellular Metabolism". In: *Nature Communications* 10.1 (2019), pp. 1–13. DOI: 10.1038/s41467-019-11581-3.

[21] R. Costenoble et al. "Comprehensive Quantitative Analysis of Central Carbon and Amino-acid Metabolism in Saccharomyces Cerevisiae under Multiple Conditions by Targeted Proteomics". In: *Molecular Systems Biology* 7.1 (2011), p. 464. DOI: 10.1038/msb.2010.122.

[22] P. .-. Lahtvee et al. "Absolute Quantification of Protein and mRNA Abundances Demonstrate Variability in Gene-Specific Translation Efficiency in Yeast". In: *Cell Systems* 4.5 (2017), 495–504.e5. DOI: 10.1016/j.cels.2017.03.003.

[23] J. Zhang et al. "Microbial Lipid Production by the Oleaginous Yeast Cryptococcus Curvatus O3 Grown in Fed-Batch Culture". In: *Biomass and Bioenergy* 35.5 (2011), pp. 1906–1911. DOI: 10.1016/j.biombioe.2011.01.024.

[24] N. Di Fidio et al. "Cutaneotrichosporon Oleaginosus: A Versatile Whole-Cell Biocatalyst for the Production of Single-Cell Oil from Agro-Industrial Wastes". In: *Catalysts* 11.11 (2021), p. 1291. DOI: 10.3390/catal11111291.

[25] F. Bracharz et al. "Opportunities and Challenges in the Development of Cutaneotrichosporon Oleaginosus ATCC 20509 as a New Cell Factory for Custom Tailored Microbial Oils". In: *Microbial Cell Factories* 16.1 (2017), p. 178. DOI: 10.1186/s12934-017-0791-9.

[26] C. Görner et al. "Genetic Engineering and Production of Modified Fatty Acids by the Non-Conventional Oleaginous Yeast Trichosporon Oleaginosus ATCC 20509". In: *Green Chemistry* 18.7 (2016), pp. 2037–2046. DOI: 10.1039/C5GC01767J.

[27] H. Zhang et al. "Enhanced Lipid Accumulation in the Yeast Yarrowia Lipolytica by Over-Expression of ATP:Citrate Lyase from Mus Musculus". In: *Journal of Biotechnology* 192 (2014), pp. 78–84. DOI: 10.1016/j.jbiotec.2014.10.004.

[28] F. X. Yan et al. "Overexpression of $\triangle$12, $\triangle$15-Desaturases for Enhanced Lipids Synthesis in Yarrowia Lipolytica". In: *Frontiers in Microbiology* 11 (2020). DOI: 10.3389/fmicb.2020.00289.

[29] I. H. Jacobsen, R. Ledesma-Amaro, and J. L. Martinez. "Recombinant β-Carotene Production by Yarrowia Lipolytica – Assessing the Potential of Micro-Scale Fermentation Analysis in Cell Factory Design and Bioreaction Optimization". In: *Frontiers in Bioengineering and Biotechnology* 8 (2020). DOI: 10.3389/fbioe.2020.00029.

[30] S. Shi and H. Zhao. "Metabolic Engineering of Oleaginous Yeasts for Production of Fuels and Chemicals". In: *Frontiers in Microbiology* 8 (2017). DOI: 10.3389/fmicb.2017.02185.

[31] M. Kim et al. "In Silico Identification of Metabolic Engineering Strategies for Improved Lipid Production in Yarrowia Lipolytica by Genome-Scale Metabolic Modeling". In: *Biotechnology for Biofuels* 12.1 (2019), pp. 1–14. DOI: 10.1186/s13068-019-1518-4.

[32] S. Wei et al. "Reconstruction of Genome-Scale Metabolic Model of Yarrowia Lipolytica and Its Application in Overproduction of Triacylglycerol". In: *Bioresources and Bioprocessing* 4.1 (2017), pp. 1–9. DOI: 10.1186/s40643-017-0180-6.

[33] M. Gottardi et al. "Pathway Engineering for the Production of Heterologous Aromatic Chemicals and Their Derivatives in Saccharomyces Cerevisiae: Bioconversion from Glucose". In: *FEMS Yeast Research* 17.4 (2017), p. 35. DOI: 10.1093/femsyr/fox035.

[34] J. Lian, S. Mishra, and H. Zhao. "Recent Advances in Metabolic Engineering of Saccharomyces Cerevisiae: New Tools and Their Applications". In: *Metabolic Engineering* 50 (2018), pp. 85–108. DOI: 10.1016/J.YMBEN.2018.04.011.

[35] P. S. Nigam and J. S. Luke. "Food Additives: Production of Microbial Pigments and Their Antioxidant Properties". In: *Current Opinion in Food Science* 7 (2016), pp. 93–100. DOI: 10.1016/J.COFS.2016.02.004.

[36] A. Rodriguez et al. "Metabolic Engineering of Yeast for Fermentative Production of Flavonoids". In: *Bioresource Technology* 245 (2017), pp. 1645–1654. DOI: 10.1016/j.biortech.2017.06.043.

[37] X. Lyu et al. "Enhancement of Naringenin Biosynthesis from Tyrosine by Metabolic Engineering of Saccharomyces Cerevisiae". In: *Journal of Agricultural and Food Chemistry* 65.31 (2017), pp. 6638–6646. DOI: 10.1021/acs.jafc.7b02507.

[38] S. Li et al. "Recent Progress in Metabolic Engineering of Saccharomyces Cerevisiae for the Production of Malonyl-CoA Derivatives". In: *Journal of Biotechnology* 325 (2021), pp. 83–90. DOI: 10.1016/j.jbiotec.2020.11.014.

[39] L. Milke et al. "Production of Plant-Derived Polyphenols in Microorganisms: Current State and Perspectives". In: *Applied Microbiology and Biotechnology* 102.4 (2018), pp. 1575–1585. DOI: 10.1007/s00253-018-8747-5.

[40] N. D. Gold et al. "Metabolic Engineering of a Tyrosine-Overproducing Yeast Platform Using Targeted Metabolomics". In: *Microbial Cell Factories* 14 (2015), p. 73. DOI: 10.1186/s12934-015-0252-2.

[41] Y. Chen et al. "Coupled Incremental Precursor and Co-Factor Supply Improves 3-Hydroxypropionic Acid Production in Saccharomyces Cerevisiae". In: *Metabolic Engineering* 22 (2014), pp. 104–109. DOI: 10.1016/j.ymben.2014.01.005.

[42] J. G. G. Koendjbiharie et al. "The Pentose Phosphate Pathway of Cellulolytic Clostridia Relies on 6-Phosphofructokinase Instead of Transaldolase". In: *Journal of Biological Chemistry* 295.7 (2020), pp. 1867–1878. DOI: 10.1074/jbc.RA119.011239.

[43] R. Ferreira et al. "Model-Assisted Fine-Tuning of Central Carbon Metabolism in Yeast through dCas9-Based Regulation". In: *ACS Synthetic Biology* 8.11 (2019), pp. 2457–2463. DOI: 10.1021/acssynbio.9b00258.

[44] J. Schellenberger, N. E. Lewis, and B. Ø. Palsson. "Elimination of Thermodynamically Infeasible Loops in Steady-State Metabolic Models". In: *Biophysical Journal* 100.3 (2011), pp. 544–553. DOI: 10.1016/j.bpj.2010.12.3707.

[45] A. A. Desouki et al. "CycleFreeFlux: Efficient Removal of Thermodynamically Infeasible Loops from Flux Distributions". In: *Bioinformatics* 31.13 (2015), pp. 2159–2165. DOI: 10.1093/bioinformatics/btv096.

[46] P. A. Saa and L. K. Nielsen. "Ll-ACHRB: A Scalable Algorithm for Sampling the Feasible Solution Space of Metabolic Networks". In: *Bioinformatics* 32.15 (2016), pp. 2330–2337. DOI: 10.1093/bioinformatics/btw132.