Variability analysis of the biomass pseudoreaction coefficients of the genome scale metabolic model of *Saccharomyces cerevisiae* using genetic algorithm

Roberto Olivares-Hernández[1]*, Ivan Domenzain[2], Verena Siewers[2], Eduard J. Kerkhoven[2], Jens Nielsen[2]

[1] Departamento de Procesos y Tecnologı´a, Universidad Auto´noma Metropolitana, Unidad Cuajimalpa, Av. Vasco de Quiroga 4871, Col. Santa Fe Cuajimalpa, Delegación Cuajimalpa, Ciudad de Mexico, 05348, Mexico. E-mail: rolivares@cua.uam.mx

[2] Department of Biology and Biological Engineering, Chalmers University of Technology, SE41296 Gothenburg, Swede

Abstract

Genome scale metabolic models (GSMM) are common tools for the exploration of metabolic engineering approaches for the overproduction of desired industrial chemicals. Their reconstruction implies an intensive data searching and manual curation. Additionally, the gaps in the metabolic network are filled with experimental designs. The mathematical approach to use GSMM for the metabolic flux calculations and evaluation of metabolic engineering strategies is the flux balance analysis. The mathematical algorithm behind this approach is linear programming where in first instance the objective function is a biomass pseudo-reaction accounting for the precursors of the main macromolecules composing the cell. This equation is determined experimentally but, in many cases, this is taking from other organisms or not for the experimental range of analysis. In this work, we present an intensive analysis of the sensitivity of the composition of this pseudo-reaction and the metabolic flux calculations. We observed that in some ranges of growth rate, the biomass mass composition impacts the flux distribution.

Introduction

Flux balance analysis is widely used for predicting metabolic fluxes using genomes scale metabolic models. It relies on linear programming where the restrictions of the linear optimization model are the mass balances for the internal metabolites and the exchange fluxes

are the rates of the external metabolites crossing the cellular barrier. The matrix formulation of the linear constrains renders what is the so called the stoichiometric matrix; a sparse matrix where non-zero elements are the stoichiometric coefficients from the reactions present in the metabolism. This constrain-based modelling has been explored whit different approaches, some of them are focused in the objective using quadratic programming, bilevel o multiobjective formulations. Other approaches worked in the restrictions to include more information and restrict the space solution, such a regulatory information or enzymatic restrictions. In FBA the mathematical formulation of the optimization model has a key aspect which is the formulation of the biomass reaction. This is not a reaction that conserves the gene-protein-reaction (GPR) annotation, as the rest of enzymatic reactions in the metabolic network, it is defined as a pseudoreaction and it is constructed *ad hoc* with the purpose of including the molar content of the biomass precursors, such as, lipids, carbohydrates, amino acids, cofactors and energetic requirements (Feist and Palsson, 2010).This molar base formulation allows to include this pseudoreaction as part of the stoichiometric matrix.

The details of the biomass composition are dependent on the experimental capabilities of the responsible in constructing the metabolic model and the growth rate conditions under study. In *Saccharomyces cerevisiae*, the content (%w/w) in RNA and protein increases linearly with the increment in growth rate, in opposition to the carbohydrates content. DNA, free amino acids and lipids have less variations (Nissen *et al.*, 1997). In contrast, Escherichia coli protein composition decreases under different growth rates (Pramanik and Keasling, 1997). For both cases, the adjustment of the biomass composition was used to evaluate the flux distribution highlighting the sensibility of the calculations to the differences in biomass composition as function of growth rate.

The exploration of this genetic algorithms (GA) is a population-based algorithm for global optimization that is inspired on natural selection. It mimics the process of the genetic combination among individuals within a population and the selection of the best genotypes that ensure the survival (Mirjalili, 2019b) . Many engineering discilines have make use the problems terms of mathematical modelling of biological process, it hase been used for the parameter stimatiation of

Methodology

Metabolic models of yeast

To evaluate flux distribution and growth rate calculations, three version of the genome scale metabolic model of yeast were used. The Yeast8 metabolic model (Lu *et al.*, 2019), and two derivates of this model when enzyme restrictions are imposed. One without protein pool constrain (ecYeast8) and the other one with protein pool constrain (ecYeast8_batch). The models were accessed through SysBioChalmers github repository and they were built with the Gecko toolbox (Sánchez *et al.*, 2017; Domenzain *et al.*, 2022). The idea to evalute the threes version was to evaluate the viability of the genetic algorithm

Experimental data

The experimental data used for the simulations with the three metabolic models are in Table 1. The experimental specific rates for eight different growth rates chemostat cultivations and one batch cultivation of the CENPK.PK strain, all of them in an aerobic condition (van Hoek *et al.*, 1998; Bakker *et al.*, 2000), see table in supplementary.

Biomass pseudoreaction coefficients

The biomass pseudoreaction is expressed as the sum of the precursors in molar proportions and it is related to the growth rate (flux) as:

$$v_{growth}: \quad \sum_{i=1}^{n} \gamma_i B_i \rightarrow biomass \tag{1}$$

Where γ_i are the molar coefficients of the biomass precursors B_i; in the biomass equation the precursors are grouped in lipids, carbohydrates, DNA, RNA, cofactors, ions and the energetic requirement as ATP. The growth rate (v_{growth}) is part of optimization problems in the FBA approach:

$$max/min \quad Z = v_{growth} \tag{2}$$

$$s.t \sum_{j=1}^{M} S_{ij}\, v_j = 0 \qquad \forall\, i \in M$$

$$v_j = v_{j\_uptake}$$

$$\alpha \leq v_j \leq \beta. \quad \alpha, \beta \in \mathbb{R}$$

In total the biomass pseudoreaction formulated for the yeast metabolic models contains 67 coefficients. ATP and water are equimolar, as well as the coefficients for nucleosides. Deoxyadenosine monophosphate (dAMP) and deoxythymidine monophosphate (dTMP) are equimolar, similarly, deoxyguanosine monophosphate (dGMP) and deoxycytidine monophosphate (dCMP) are in equimolar proportions.

Genetic algorithm

The genetic algorithm consists of the basic process initiation, selection, crossover, mutation (Mirjalili, 2019a) and in this case the rate of elitism to maintain a subset of best solutions and transferred to the next generation without modification, see flow diagram in supplementary figure. During the initialization and the mutation, biomass coefficients are modified within a defined interval like:

$$a = \gamma_i + \varepsilon * \gamma_i \tag{3}$$

$$b = \gamma_i - \varepsilon * \gamma_i \tag{4}$$

$$\gamma_{i,new} = a + (b - a) * rand()| \tag{5}$$

Where ε, is the variation of the coefficient of the biomass precursor. The variation, up or down, is relative to the values of the coefficients in the genome scale metabolic model used.

Initially, all the biomass coefficients had the same coefficient of variation, after performing the first simulations and evaluate the results, it was necessary to differentiate this for ATP and the rest of the molecules, as the values of the coefficients after running the genetic algorithm were too high for ATP.

The fitness function is evaluated with the growth rate calculated using FBA (μ_{FBA}) and the experimental value (μ_{exp}) in the follow expression:

$$fitness\ function = \frac{1}{|\mu_{exp} - \mu_{FBA}|} \quad (6)$$

The parameters of genetic algorithm are in Table 4. They were adjusted to have the final convergence, see supplementary data for the evaluations.

Table 4. Parameters of the genetic algorithm

Parameter	**value**
Number of genes	M=100
Generations	MaxGen = 500
Crossover rate	Pc = 0.9
Mutation rate	Pm = 0.05
Elitism rate	Er = 0.1
Coefficient variation	$\epsilon = \pm 30\%$
ATP variation	$\epsilon = \pm 10\%$

RESULTS AND DISCUSSION

Growth rate calculations

The growth rate calculations in the genome scale metabolic is presented in Table 1. It is possible to observer the deviation

Table 1. Genome scale models evaluations at different growth rates and conditions. Only two significant digits are taken for comparison.

model	experimental values of μ (h^{-1})								
	chemostat								batch
	0.10	0.15	0.2	0.25	0.30	0.33	0.35	0.38	0.38
Yeast8	0.09	0.10	0.11	0.16	0.28	0.34	0.35	0.39	0.43
ecYeast	0.09	0.10	0.11	0.16	0.28	0.33	0.33	0.37	0.36
ecYeast_batch	0.09	0.09	0.11	0.15	0.27	0.32	0.34	0.35	0.40

Initially, the prediction capability of the genome scale models was evaluated taking the experimental specific rates available for yeast. The predictions for growth rates in three scenarios, 0.15, 0.20 and 0.25 h^{-1}, where in the range of 27-47% higher, whereas for the rest, the deviation was less than 10%.

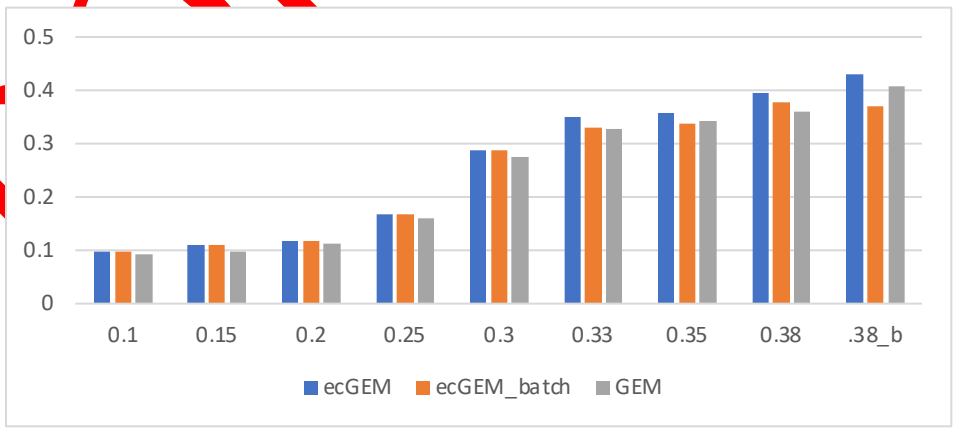


Figure 1. growth rate predictions using flux balance analysis. Experimental data from

A genetic algorithm is used to adjust the biomass pseudoreaction coefficients resembling the changes in biomass composition as function of different growth rates and cultivation conditions. Eight scenarios for aerobic chemostat cultivation and one scenario for an aerobic batch cultivation were evaluated. According to genetic algorithms concepts, the size of the chromosome was 67 genes, which corresponds to the number of precursors coefficients in the biomass pseudo-reactions and the population size was 100 chromosomes.

The first parameter of the evaluations were the crossover rate and mutation rate. Genetic algorithms are based on heuristics, which means that it is necessary the explore the impact of the parameters in the simulations. In general, there is a consensus for some of the parameters, for instance, the mutation rates might be in the range of 0.01-0.1 in order to avoid the destabilization of stable solutions; the value for the present simulations was 0.05. For the crossover rates we could identified diversity of the solution for the process. Several calculations of the algorithm were made in order to evaluate the progression of the fitness function, we could identify that for the case mutation rate and elitism rate, there was no incidence on the number of generations needed.

In the initialization and mutation process, the algorithm variates the values withing predefined range, relative to original value in the model, searching to the best fit that predicts the experimental growth rate. After the initial evaluations, all the cases to evaluate we could finally standardize the set of parameters to perform the comparations. This are in the table and include the deviation for the biomass coefficients ($\epsilon = \pm 30\%$) and ATP ($\epsilon_{ATP} = \pm 10\%$).

The conversion of the simulation for the cases 0.15, 0.20 and 0.25 h^{-1} reached the maximum number of generations and the deviations were reduced to a range 4-23%, whereas for the rest of the scenarios the number generation needed for a stable solution were in the range of 100 generations and the experimental value was reached, see Figure 2.

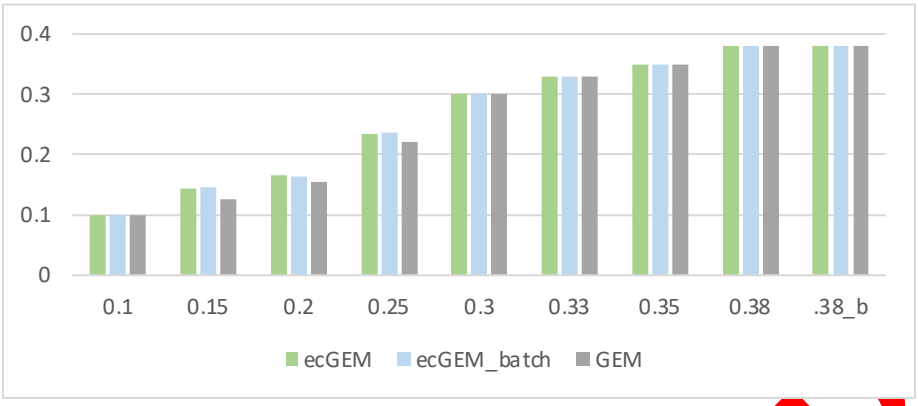


Figure 2. The performance of the initial parameters.

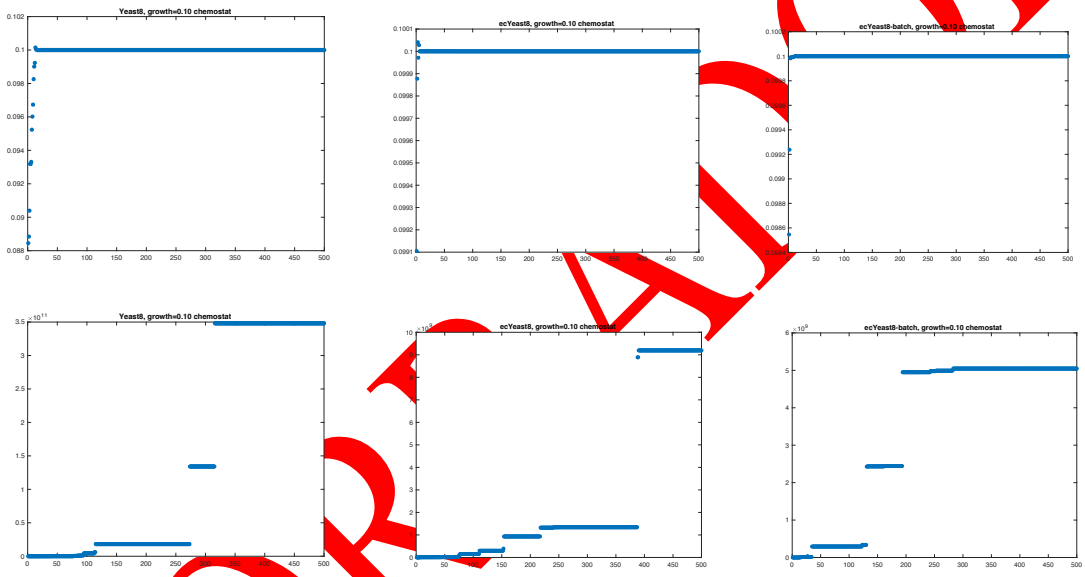
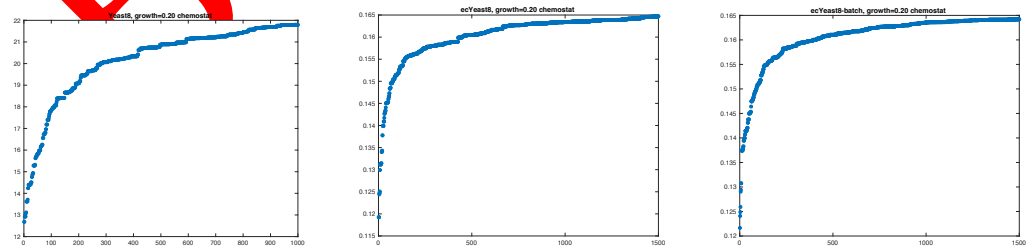


Figure 3. Fitness function for the case when mutation rate is 0.1 for the three yeast models.



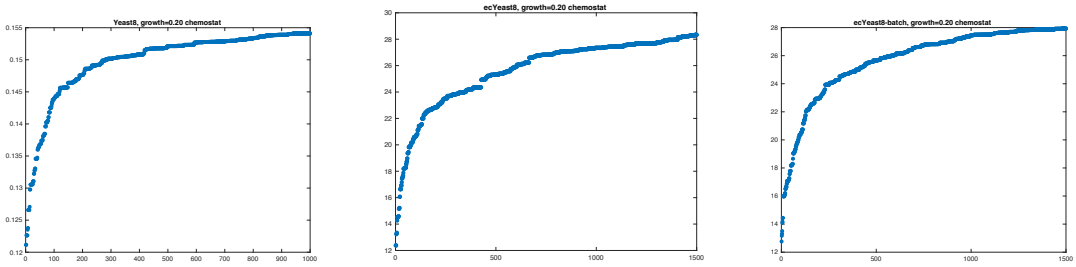


Figure 4. For the case when mutation rate was 0.2.

The progression of the solution across the generations can be seen in Figure 3. The biomass coefficients area grouped according to the type of molecules. In the plots it can be seen the variations (ϵ) in each coefficient relative to its original value. The fitness function is the invers of the difference between the experimental value and the predicted value of the growth rate. The definition also implies that a value of 10^5 of the fitness function represents is a difference of 0.00001 between predicted and experimental value. If with define a tolerance value of the order of 10^5 for the fitness function and used as criteria to stop the calculations, some scenarios rapidly converge and fulfill this criterion.

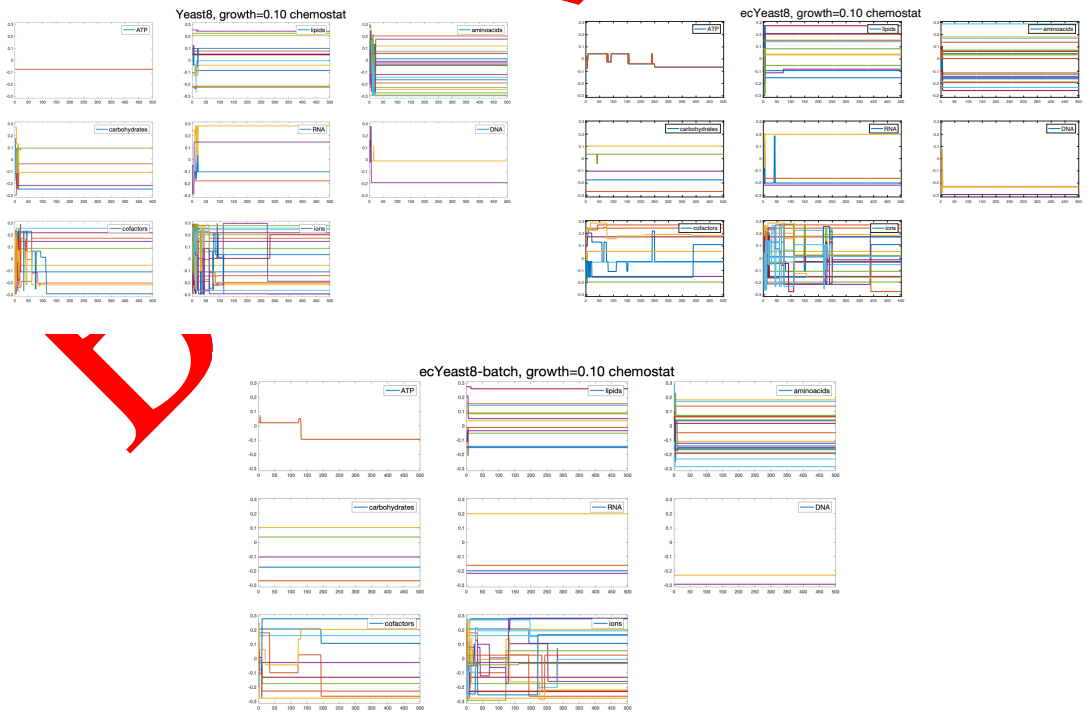


Figure 5. For the case when growth is 0.1 h^{-1} we can see that ions and cofactors presenta a wide variation across the generations.

In Figure 5 it is presented the progression of the solution in terms of the number of generations when growth rate is 0.1 h^{-1}. It is interesting to observe that ATP, lipids, amino acids carbohydrates, RNA and DNA precursor have stable values whereas cofactors and ions variates almost stochastically. We present 500 generations. The variation of the cofactors are not affecting the production of growth rate, wheatear this feature is related to ion and cofactor accounted in the biomass models or it is the result of a biological impact has to be evaluated.

In Figure 6 we represent the case for growth rate equal 0.2 h^{-1}, in this case it was necessary the 1500 generation to reach the convergence.

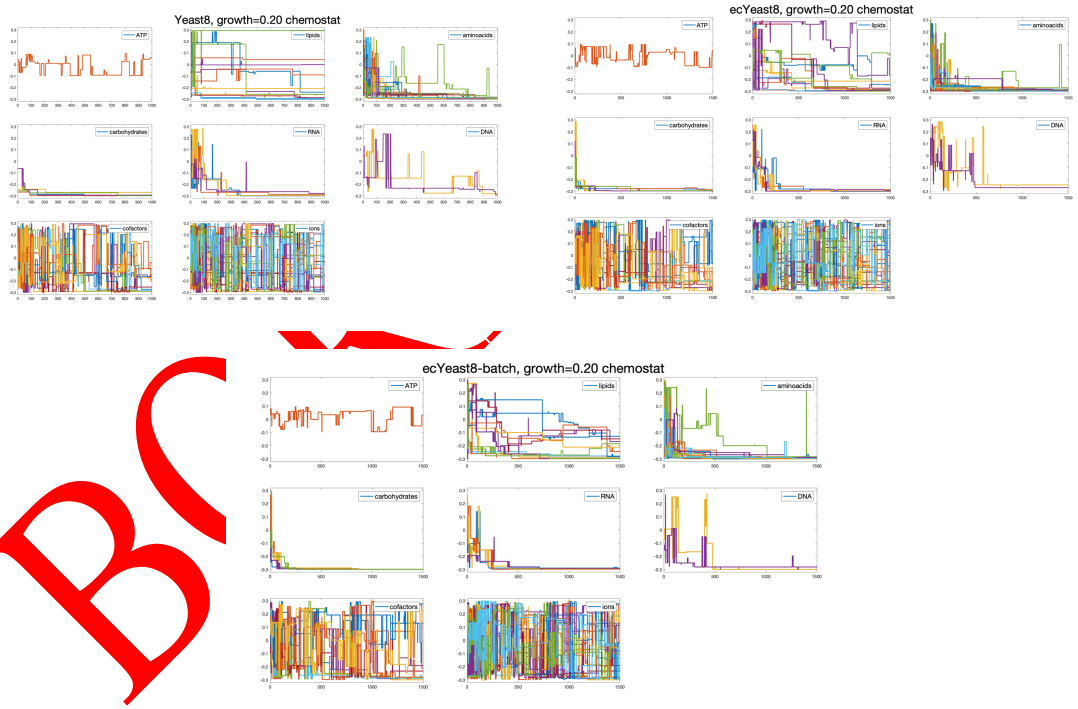
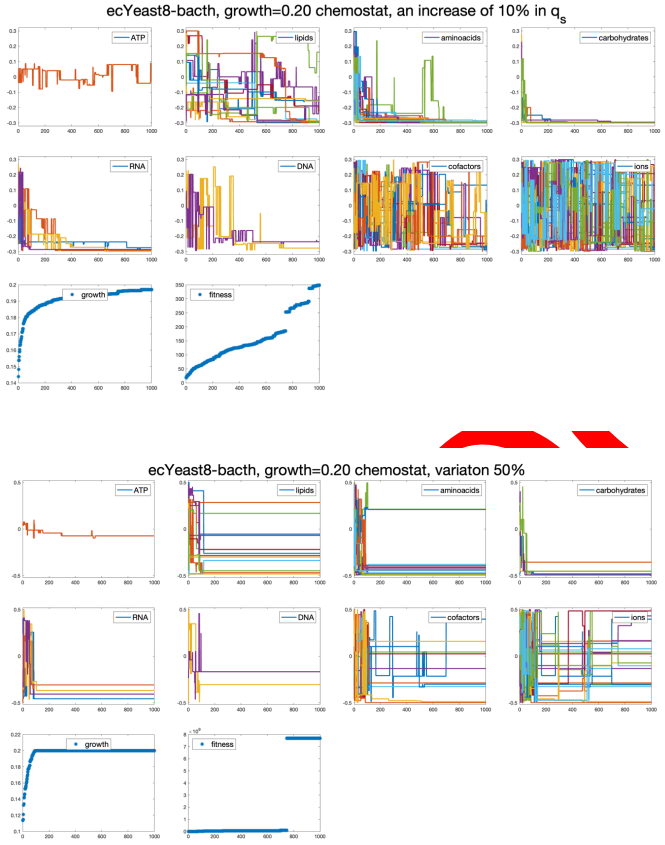


Figure 6. Coefficient variations when growth rate is 0.2 h^{-1}.

For the case of 0.2, the solution still not reached up to 500 generation and the value never reached the experimental value. To achieve the convergence of the values 0.15-0.25 h-1. In

the first case we increase the glucose consumption up to 10% considering the experimental deviation. Again, the value was not achieved it.



Figures 7. Variation of the coefficients when we increase the glucose consumption

Conclusions

SUPLEMENTARY

Table S.1 Experimental value for the analysis of the flux calculations

	μ (h^{-1})								
	chemostat								batch
mmol/(gDW h^{-1})	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**	**0.33**	**0.35**	**0.38**	**0.38**
$q_{glucose}$	1.17	1.50	1.70	2.36	3.88	6.11	7.20	13.19	15.7
q_{O2}	2.9	4.1	5.2	7	9.5	7.2	6.9	3.9	
q_{CO2}	2.8	4.2	5.3	7.3	9.8	12.2	13.1	21	
$q_{ethanol}$					0.5	4.8	6	18.3	22.3
$q_{glycerol}$									0.97
$q_{acetate}$									0.52
$q_{lactate}$									0.12
$q_{pyruvate}$									0.12
$q_{succinate}$									<0.1

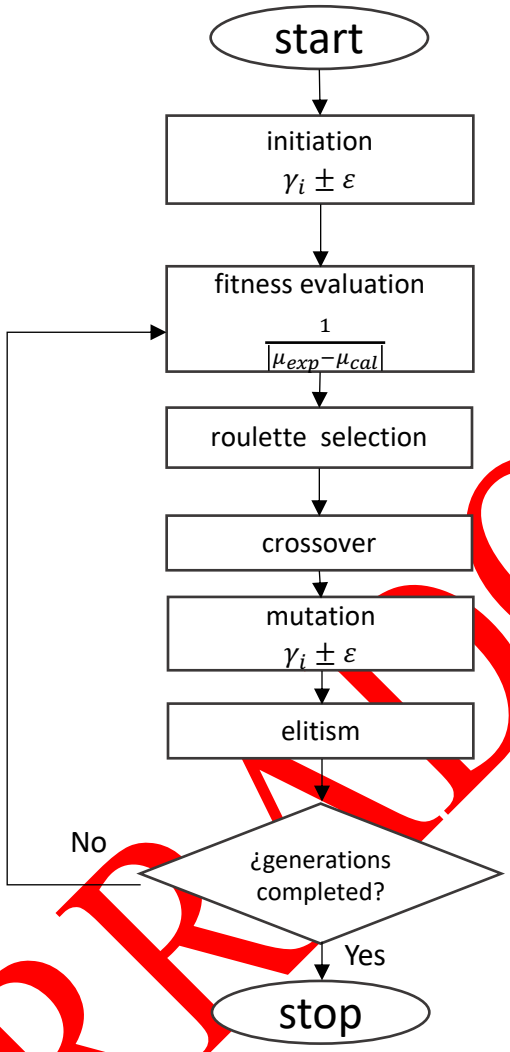


Figure S.2. Flow diagram of the genetic algorithm. This algorithm uses and elitism step.

Table S1. Coefficients of the biomass pseudoreaction in the three GSMM of yeast. The Yeast and ecYeast difference for the ecYeast_batch is part of the rescaling process when protein pool is included. Water is part of the precursors but its not listed, its coefficient is equimolar to ATP.

#	Biomass precursors	Molecules types	Yeast8	ecYeast8	ecYeast8_batch
1	ATP	ATP	55.3	55.3	56.6883

2	C16:0 chain	lipids	0.00808584	0.00808584	0.0073947
3	C16:1 chain	lipids	0.0237302	0.0237302	0.0217019
4	C18:0 chain	lipids	0.00226632	0.00226632	0.0020726
5	C18:1 chain	lipids	0.00870664	0.00870664	0.00796243
6	1-phosphatidyl-1D-myo-inositol backbone	lipids	0.0069103	0.0069103	0.00631964
7	ergosterol	lipids	0.0265829	0.0265829	0.0243107
8	ergosterol ester backbone	lipids	0.0068058	0.0068058	0.00622407
9	fatty acid backbone	lipids	0.0014801	0.0014801	0.00135359
10	phosphatidyl-L-serine backbone	lipids	0.0059508	0.0059508	0.00544215
11	phosphatidylcholine backbone	lipids	0.025783	0.025783	0.0235792
12	phosphatidylethanolamine backbone	lipids	0.0069293	0.0069293	0.00633701
13	triglyceride backbone	lipids	0.0068571	0.0068571	0.00627098
14	Ala-tRNA(Ala)	aminoacids	0.527012	0.527012	0.57284
15	Arg-tRNA(Arg)	aminoacids	0.184592	0.184592	0.200644
16	Asn-tRNA(Asn)	aminoacids	0.11682	0.11682	0.126979
17	Asp-tRNA(Asp)	aminoacids	0.341731	0.341731	0.371447
18	Cys-tRNA(Cys)	aminoacids	0.00758126	0.00758126	0.0082405
19	Gln-tRNA(Gln)	aminoacids	0.12107	0.12107	0.131598
20	Glu-tRNA(Glu)	aminoacids	0.34667	0.34667	0.376816
21	Gly-tRNA(Gly)	aminoacids	0.333575	0.333575	0.362582
22	His-tRNA(His)	aminoacids	0.0761572	0.0761572	0.0827796
23	Ile-tRNA(Ile)	aminoacids	0.22135	0.22135	0.240598
24	Leu-tRNA(Leu)	aminoacids	0.340467	0.340467	0.370073
25	Lys-tRNA(Lys)	aminoacids	0.328751	0.328751	0.357338
26	Met-tRNA(Met)	aminoacids	0.0582379	0.0582379	0.063302
27	Phe-tRNA(Phe)	aminoacids	0.153808	0.153808	0.167182
28	Pro-tRNA(Pro)	aminoacids	0.189187	0.189187	0.205638
29	Ser-tRNA(Ser)	aminoacids	0.212964	0.212964	0.231483
30	Thr-tRNA(Thr)	aminoacids	0.219857	0.219857	0.238974
31	Trp-tRNA(Trp)	aminoacids	0.0326224	0.0326224	0.0354591
32	Tyr-tRNA(Tyr)	aminoacids	0.117165	0.117165	0.127353
33	Val-tRNA(Val)	aminoacids	0.30394	0.30394	0.330369
34	(1->3)-beta-D-glucan	carbohydrates	0.748515	0.748515	0.684535
35	(1->6)-beta-D-glucan	carbohydrates	0.250092	0.250092	0.228715
36	glycogen	carbohydrates	0.361415	0.361415	0.330522
37	mannan	carbohydrates	0.71094	0.71094	0.650171
38	trehalose	carbohydrates	0.138276	0.138276	0.126456
39	AMP	RNA	0.0445348	0.0445348	0.0445348
40	CMP	RNA	0.0432762	0.0432762	0.0432762
41	GMP	RNA	0.0445348	0.0445348	0.0445348
42	UMP	RNA	0.0579921	0.0579921	0.0579921
43	dAMP	DNA	0.0036	0.0036	0.0036

44	dCMP	DNA	0.0024	0.0024	0.0024
45	dGMP	DNA	0.0024	0.0024	0.0024
46	dTMP	DNA	0.0036	0.0036	0.0036
47	coenzyme A	cofactors	0.00019	0.00019	0.00019
48	FAD	cofactors	1.00E-05	0.00001	0.00001
49	NAD	cofactors	0.00265	0.00265	0.00265
50	NADH	cofactors	0.00015	0.00015	0.00015
51	NADP(+)	cofactors	0.00057	0.00057	0.00057
52	NADPH	cofactors	0.0027	0.0027	0.0027
53	riboflavin	cofactors	0.00099	0.00099	0.00099
54	TDP	cofactors	1.20E-06	0.0000012	0.0000012
55	THF	cofactors	6.34E-05	0.0000634	0.0000634
56	heme a	cofactors	1.00E-06	0.000001	0.000001
57	iron(2+)	ions	3.04E-05	0.0000304	0.0000304
58	potassium	ions	0.00363	0.00363	0.00363
59	sodium	ions	0.00397	0.00397	0.00397
60	sulphate	ions	0.02	0.02	0.02
61	chloride	ions	0.00129	0.00129	0.00129
62	Mn(2+)	ions	0.00273	0.00273	0.00273
63	Zn(2+)	ions	0.000748	0.000748	0.000748
64	Ca(2+)	ions	0.000217	0.000217	0.000217
65	Mg(2+)	ions	0.00124254	0.00124254	0.00124254
66	Cu2(+)	ions	0.000659	0.000659	0.000659

References

Bakker, B.M. *et al.* (2000) 'The Mitochondrial Alcohol Dehydrogenase Adh3p Is Involved in a Redox Shuttle in Saccharomyces cerevisiae', *Journal of Bacteriology*, 182(17), pp. 4730–4737. Available at: http://www.rz.uni-frankfurt.de/FB/fb16/mikro/euroscarf/index.html.

Domenzain, I. *et al.* (2022) 'Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using GECKO 2.0', *Nature Communications*, 13(1), p. 3766. Available at: https://doi.org/10.1038/s41467-022-31421-1.

Feist, A.M. and Palsson, B.O. (2010) 'The biomass objective function', *Current Opinion in Microbiology*, pp. 344–349. Available at: https://doi.org/10.1016/j.mib.2010.03.003.

van Hoek, P. *et al.* (1998) 'Effects of Pyruvate Decarboxylase Overproduction on Flux Distribution at the Pyruvate Branch Point in Saccharomyces cerevisiae', *Applied and Environmental Microbiology*, 64(6), pp. 2133–2140.

Lu, H. *et al.* (2019) 'A consensus S. cerevisiae metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism', *Nature Communications*, 10(1). Available at: https://doi.org/10.1038/s41467-019-11581-3.

Mirjalili, S. (2019a) 'Genetic Algorithm', in *Evolutionary Algorithms and Neural Networks: Theory and Applications*. Cham: Springer International Publishing, pp. 43–55. Available at: https://doi.org/10.1007/978-3-319-93025-1_4.

Mirjalili, S. (2019b) *Studies in Computational Intelligence 780 Evolutionary Algorithms and Neural Networks Theory and Applications*. Berlin/Heidelberg, Germany. Available at: http://www.springer.com/series/7092.

Nissen, T.L. *et al.* (1997) *Flux distributions in anaerobic, glucose-limited continuous cultures of Saccharomyces cerevisiae*, *Microbiology*.

Pramanik, J. and Keasling, J.D. (1997) 'Stoichiometric model of Escherichia coli metabolism: Incorporation of growth-rate dependent biomass composition and mechanistic energy requirements', *Biotechnology and Bioengineering*, 56(4), pp. 398–421. Available at: https://doi.org/10.1002/(SICI)1097-0290(19971120)56:4<398::AID-BIT6>3.0.CO;2-J.

Sánchez, B.J. *et al.* (2017) 'Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints', *Molecular Systems Biology*, 13(8), p. 935. Available at: https://doi.org/10.15252/msb.20167411.