

## RESEARCH ARTICLE

# Metabolic flux configuration determination using information entropy

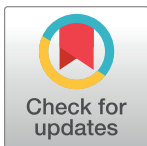
Marcelo Rivas-Astroza <sup>1,2\*</sup>, Raúl Conejeros <sup>1</sup>

Escuela de Ingeniería Bioquímica, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

<sup>1</sup> Current address: Departamento de Biotecnología, Universidad Tecnológica Metropolitana, Ñuñoa, Chile\* [marcelo.rivas@utem.cl](mailto:marcelo.rivas@utem.cl)

## Abstract

Constraint-based models use steady-state mass balances to define a solution space of flux configurations, which can be narrowed down by measuring as many fluxes as possible. Due to loops and redundant pathways, this process typically yields multiple alternative solutions. To address this ambiguity, flux sampling can estimate the probability distribution of each flux, or a flux configuration can be singled out by further minimizing the sum of fluxes according to the assumption that cellular metabolism favors states where enzyme-related costs are economized. However, flux sampling is susceptible to artifacts introduced by thermodynamically infeasible cycles and it is not clear if the economy of fluxes assumption (EFA) is universally valid. Here, we formulated a constraint-based approach, MaxEnt, based on the principle of maximum entropy, which in this context states that if more than one flux configuration is consistent with a set of experimentally measured fluxes, then the one with the minimum amount of unwarranted assumptions corresponds to the best estimation of the non-observed fluxes. We compared MaxEnt predictions to *Escherichia coli* and *Saccharomyces cerevisiae* publicly available flux data. We found that the mean square error (MSE) between experimental and predicted fluxes by MaxEnt and EFA-based methods are three orders of magnitude lower than the median of 1,350,000 MSE values obtained using flux sampling. However, only MaxEnt and flux sampling correctly predicted flux through *E. coli*'s glyoxylate cycle, whereas EFA-based methods, in general, predict no flux cycles. We also tested MaxEnt predictions at increasing levels of overflow metabolism. We found that MaxEnt accuracy is not affected by overflow metabolism levels, whereas the EFA-based methods show a decreasing performance. These results suggest that MaxEnt is less sensitive than flux sampling to artifacts introduced by thermodynamically infeasible cycles and that its predictions are less susceptible to overfitting than EFA-based methods.



## OPEN ACCESS

**Citation:** Rivas-Astroza M, Conejeros R (2020) Metabolic flux configuration determination using information entropy. PLoS ONE 15(12): e0243067. <https://doi.org/10.1371/journal.pone.0243067>

**Editor:** Antonio Calcagni, University of Padova, ITALY

**Received:** May 30, 2020

**Accepted:** November 14, 2020

**Published:** December 4, 2020

**Copyright:** © 2020 Rivas-Astroza, Conejeros. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and its [Supporting information](#) files.

**Funding:** This work was supported by Fondo Nacional de Ciencia y Tecnología (FONDECYT, Chile; <http://www.conicyt.cl/fondecyt/>), grant number 3170488.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

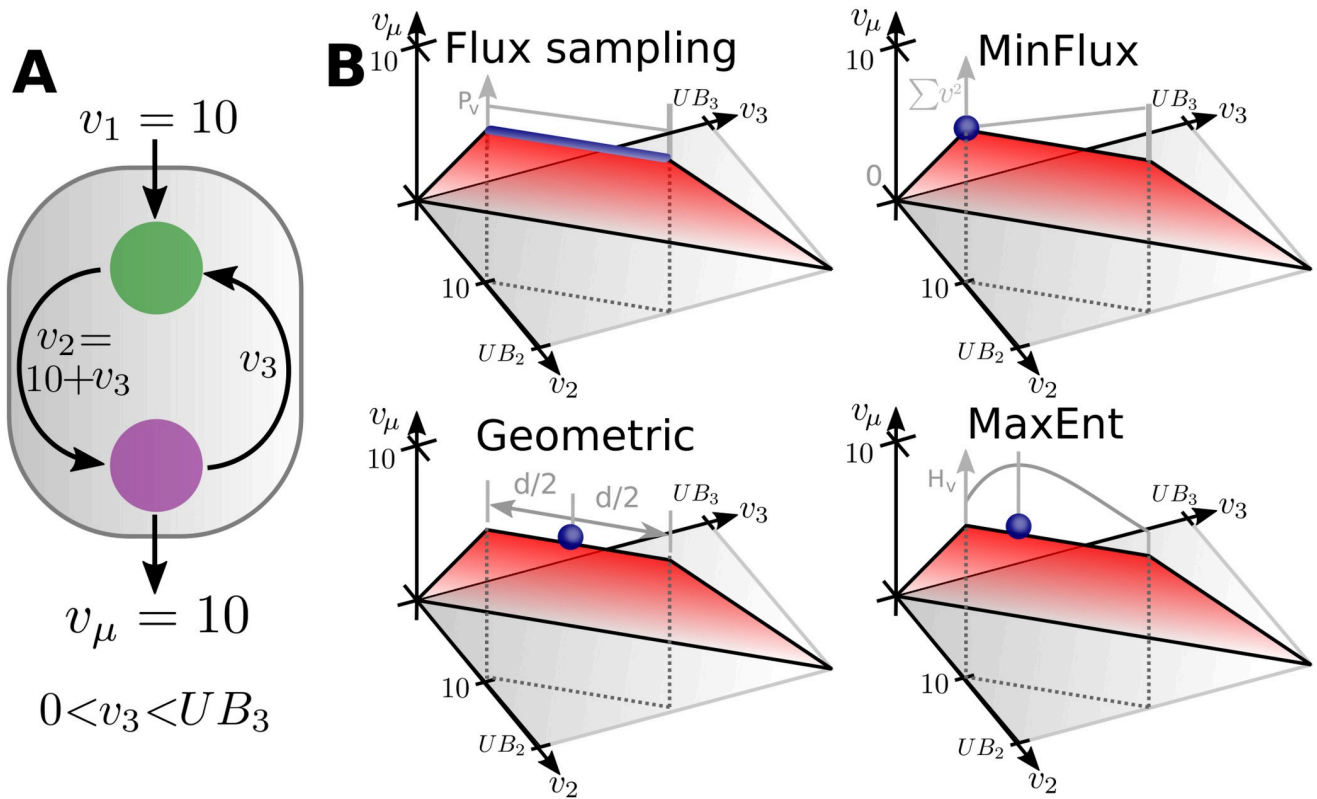
Genome-scale metabolic networks provide the basis for reconstructing the set of metabolic reactions occurring within a living organism. These reactions carry the flux of materials distributing the building blocks for macro-molecules production and, ultimately, biomass

formation. However, flux configurations cannot be determined with complete certainty as experiments can only probe a fraction of all the states allowed by the various concentration levels that enzymes, RNAs, and metabolites can reach within a cell [1].

Alternatively, the mass balance principle can be applied to obtain a mathematical model describing the variation of all concentrations for the metabolic system. This model can be further simplified by assuming a steady-state condition, resulting in a linear mathematical model which provides a solution space for all possible flux configuration that comply with the constraints set by the stoichiometry of the reaction network. In this framework, metabolic networks can be considered as providing a space of possible flux configurations where some adaptive regulatory mechanism of the reaction rates resolves into the one that maximizes cellular fitness [2, 3]. An implementation of this idea has been Flux Balance Analysis (FBA) [4], which encodes known uptake rates and metabolites mass balances as constraints of a linear optimization problem where biomass growth rate is maximized. FBA has been an influential approach as sequencing technologies have allowed inferring the topology of metabolic networks at a genome-scale [5–7]. As FBA is formulated as a linear programming problem, it does not necessarily yield a single solution [8, 9]. This is typically the case for metabolic networks as they contain loops and alternative pathways [10] that accept various flux configurations to be compatible with a given set of known uptake rates and maximized objective function values [11]. Linear programming can be used to efficiently select a single flux configuration within this space, but its pick is based on the implementation of the algorithm performing the optimization rather than on biological considerations. This is problematic as different implementations may result in different solutions, which affect the reproducibility of results, and it can produce conflicting outcomes. For instance, two algorithmic implementations can predict flux through mutually exclusive pathways. Using alternative metrics of cellular fitness, e.g. ATP production, or measuring extra uptake rates is typically insufficient to reduce the solution space to a single flux configuration.

Various methods have been proposed to deal with the ambiguity of these alternative solutions. One of them is flux sampling, where a sequence of random samples from the space of alternative solutions is generated until the entire space is analyzed [12, 13]. Unlike FBA, flux sampling does not require defining an objective function [12]. This method has been applied to small catabolic networks [14] and to genome-scale networks [12] to infer the range and probability distribution values for each flux. However, the mass balances of the metabolic network are not enough to prevent thermodynamically infeasible flux cycles. For instance, in the example network presented in Fig 1, an arbitrarily large flux value can be cycled between the metabolites of the inner loop. Only the upper bounds imposed over these fluxes prevent them from reaching ever larger values. These bounds are not meant to be biologically meaningful so that the sampling space is arbitrarily biased. Considering each reaction Gibbs free energy can prevent thermodynamically infeasible cycles [15], but this information is not always available for each reaction of a genome-scale metabolic network. To circumvent this problem extra constraints can be added in order to rule out the formation of closed cycles [16] but this renders the problem computationally intractable at genome-scale [17]. As a consequence, the presence of thermodynamically infeasible loops remains an open problem that can severely bias the inferences drawn from flux sampling.

Alternatively, infeasible cycles can be avoided by considering only the subspace of alternative solution where the sum of all fluxes magnitudes reach their minimum [17]. In this subspace, reactions forming close loops attain flux values equal to zero, effectively preventing thermodynamically infeasible cycles. If fluxes magnitudes are measured by their squared values, the space of alternative solutions is reduced to a single flux configuration [18]. We will refer to this method as MinFlux. Alternatively, if fluxes magnitudes are measured by their



**Fig 1. Polytope of possible flux configurations.** (A) Example metabolic network with known exchange fluxes  $v_1$  and  $v_\mu$ , and two unknown inner fluxes  $v_2$  and  $v_3$  forming a loop. The upper bounds for fluxes  $v_2$  and  $v_3$  are  $UB_2$  and  $UB_3$ . Any value of  $v_3 \in [0, UB_3]$  satisfy the metabolites' mass balances, resulting in an infinite set of solutions. (B) Different methods to estimate the inner fluxes. Flux sampling estimates flux configurations from random samples from the set of alternative flux configurations (represented by the points forming the blue line). MinFlux selects the flux configuration where  $v_2^2 + v_3^2$  reach their minimum value according to the assumption that cells uses the the minimum amount of fluxes to economize enzyme synthesis. Geometric selects the flux configuration located at the geometric center of the space formed by the alternative solutions. MaxEnt selects flux configuration with the maximum information entropy ( $H_v$ ).

<https://doi.org/10.1371/journal.pone.0243067.g001>

absolute values, more than one flux configuration can achieve the same minimum sum of fluxes, rendering a subspace of alternative solutions. In this case, further assumptions can be made to select a single flux configuration from this subspace. In particular, Geometric is a constraint-based model that selects the flux configuration located at the geometric center of the polytope formed by these alternative solutions [19]. Regardless of how flux magnitudes are measured, the minimization of fluxes assumption is coherent with the hypothesis that under optimal growing conditions cells save energy by producing the minimum amount of enzyme-related proteins [20]. Compared to MinFlux, Geometric predicts flux configurations where more reactions have zero flux. This is a result of Geometric measuring fluxes magnitudes by their absolute values, which is a known sparsity inducing norm [21].

MinFlux and Geometric each yield reproducible results while avoiding thermodynamically infeasible cycles, but they also have limitations. On the one hand, it is not known if the minimization of fluxes' assumption is universally valid. For instance, it has been observed in *Saccharomyces cerevisiae* and *Escherichia coli* that high glucose consumption rates are accompanied by the activation of otherwise shut off pathways, resulting in the production of overflow metabolites through non oxidative pathways [22–24]. Numerous explanations have been offered, including ATP savings for the production of non-oxidative enzymes (which by being smaller, compared to their oxidative counterparts, requires less ATP in their synthesis)

[25, 26], limited uptake rates capacity [27], and an upper limit on the dissipation of Gibbs energy [28]. On the other hand, minimizing the sum of fluxes can break thermodynamically feasible cycles. For example, the isocitrate dehydrogenase reaction of the glyoxylate cycle would be systematically predicted to be inactive, although it is known to be active in *E. coli* [29–31]. Also, when a metabolite can be converted into another by more than one pathway, the minimization of the sum of the fluxes absolute values leads to the inactivation of all but one of these alternative pathways [17]. Thus, introducing the risk of overestimating the flux through the remaining active pathway.

Thus, current methods to estimate flux configurations are either overly sensitive to artifacts introduced by thermodynamically infeasible cycles or rely on assumptions that may not be universally valid. To overcome this, we propose to use statistical inference methods, specifically the principle of maximum entropy [32], which in general terms states that the best state of knowledge of a system –expressed as a probability distribution– is the one that admits the most ignorance besides prior information. This principle has been applied in biological sciences [33], including the metabolism of bacterial populations to infer statistical models from limited data. De Martino *et al.* (2016) [34] modeled the fluctuations of growth rates in *E. coli* using a Boltzmann probability distribution as this is the one that maximizes entropy under the constraint that the population’s average growth rate equals its experimental value. A sampling of the solution space of *E. coli* metabolic network based on Boltzmann distribution is proposed, producing distributions of growth rates that closely resemble experimental data. De Martino *et al.* (2018) [1] have applied this procedure to sample the space of flux configurations of the catabolic core of *E. coli* metabolism. This method produces flux distributions whose averages are closer to experimental data than those produced using FBA or uniform sampling. Fernandez de Cossio Diaz and Mulet (2019) [35] applied this approach to address cell-to-cell metabolic variability in Chinese hamster ovary cells population as a function of the dilution rate in a chemostat. Since the sampling procedure is intractable at genome-scale sized metabolic networks, the authors reduced the network by pruning reactions that do not carry flux when computed using FBA at various dilution rates. As FBA typically produces multiple alternative solutions, the authors selected the one where enzymatic costs are minimized. Tourigny (2020) [36] has expanded the application of these ideas, proposing that the maximum entropy principle can assign the best allocation of resources among elementary flux modes for maximizing expected return on investment of metabolic resources in the face of uncertain environmental conditions. As the number of elementary flux modes explodes with the network’s size [37], this approach was applied to a simplified model of yeast metabolism, reproducing the observed behavior of a cellular population in continuous and batch cultures.

Here, we use the principle of maximum entropy to determine a single flux configuration for a genome-scale metabolic network based on information theory. In particular, we proposed that the best estimation of the cellular flux configuration is the one with the minimum amount of unwarranted assumptions. Each flux configurations within the polytope of alternative solutions can be encoded as a probability distribution. The information entropy of these probability distributions can be interpreted as the average level of information inherent to each flux configuration. It follows that out of all flux configurations that are consistent with experimentally measured fluxes (for instance, glucose uptake), we should select the one with the largest value of information entropy [38], as it requires the fewest prior assumptions, and hence corresponds to the least biased solution. This idea was implemented at a genome-scale as a constraint-based model, which we called MaxEnt. MaxEnt finds the flux configuration with the most homogeneous distribution of fluxes that is consistent with the restrictions imposed by the constraint-based model. This makes MaxEnt less sensitive than flux sampling to the artifacts introduced by thermodynamically infeasible cycles as their fluxes are prevented

from reaching their upper bounds. At the same time, MaxEnt predictions neither eliminate thermodynamically feasible cycles nor alternative pathways. The latter of which are biases introduced by MinFlux and Geometric methods.

In the methods section, we provide a formalism to apply information entropy to flux configurations. In the results section, we used this formalism to set up and test quantitative predictions for *E. coli* and *S. cerevisiae*. Here, we provide evidence that MaxEnt can improve our estimation of single flux configurations.

## Materials and methods

### General background

For a metabolic network of  $N$  reactions and  $M$  metabolites, we define the scalar vector of metabolites concentrations  $c$ , its time derivatives  $\dot{c}$  and the scalar vector of fluxes  $v$ , where  $c$  and  $\dot{c} \in \mathbb{R}^M$ , and  $v \in \mathbb{R}_+^N$ . We have considered all fluxes to be non-negative, with reversible reactions been split-up into forward and reverse reactions. The stoichiometric matrix  $S$  has  $(M \times N)$  elements, so that the product  $Sv$  results in a  $(M \times 1)$  matrix. Thus, the time dependent mass balance is written as:

$$\dot{c} = Sv \quad (1)$$

Assuming steady-state condition for the metabolism,  $\dot{c} = 0$ , and including further constraints for reversibility of reactions, uptake fluxes of nutrient, and kinetic limits in the form of lower  $LB \in \mathbb{R}_+^N$  and upper bounds  $UB \in \mathbb{R}_+^N$  on fluxes, a convex polytope  $\mathcal{P}$  of alternative flux configurations is defined:

$$\mathcal{P} = \{v \in \mathbb{R}_+^N | Sv = 0, LB \leq v \leq UB\} \quad (2)$$

Biomass growth rate is integrated into the metabolic network using a biomass reaction in the form of a linear combination of metabolic fluxes  $v_\mu = \sum_i b_i v_i$ , where  $b_i$  correspond to the mass proportion of the metabolite  $i$  in biomass.

### Genome-scale metabolic networks

Genome-scale metabolic networks reconstructions for *E. coli* and *S. cerevisiae* were used iJR904 ( $N = 1075$  reactions, and  $M = 761$  metabolites) [39] and iMM904 ( $N = 1577$ ,  $M = 1226$ ) [40], respectively. Both were obtained from the BiGG Models database [41].

### Information entropy modeling

Let us consider an experiment where reactions are randomly sampled from a metabolic network. The outcome of each sample can be encoded by the random variable  $X \in \{x_1, \dots, x_N\}$ , where  $x_i$  is the identity of reaction  $i$  (for instance the name of the enzyme catalyzing the reaction), and the probability of observing  $x_i$  is given by  $P_v(X = x_i)$ . If  $Q$  enzymes are distributed among  $N$  reactions, such that  $Q = \sum_j q_j$ , where  $q_i$  are the enzyme units catalyzing reaction  $i$ , then:

$$P_v(X = x_i) = \frac{q_i}{\sum_j q_j} \quad (3)$$

For reaction  $i$ , its flux  $v_i$  is a function of the amount of enzymes, namely  $v_i = k_i \eta_i q_i$ , where  $k_i$  is the maximum turnover of catalyst  $q_i$  and  $\eta_i$  is a function ranging from 0 to 1 describing the decrease in catalytic rate due to intracellular conditions (for example, incomplete substrate

saturation) [42]. Then,  $q_i$  in Eq 3 can be replaced by  $v_i/(k_i \eta_i)$ . Unfortunately, values of  $k_i \eta_i$  are unknown for the vast majority of reactions [42], so that we assume their values to be similar to one another in order rewrite Eq 3 as:

$$P_v(X = x_i) = \frac{v_i}{\sum_j v_j} \quad (4)$$

This is an approximation that can be improved if values of  $k_i \eta_i$  became available. Still, each  $v \in \mathcal{P}$  defines a different probability distribution for  $X$ , and the average level of information inherent to the various possible outcomes of  $X$  is given by its information entropy [38]:

$$H_v(X) = - \sum_{i=1}^N P_v(X = x_i) \log P_v(X = x_i) \quad (5)$$

$H_v(X)$  has two extreme values. Its minimum is 0 and is obtained when all but one flux in  $v$  are 0. In this case, we would be certain of the outcome of any random sample as  $P_v(x_k) = 1$  for the non-zero flux, and 0 otherwise. On the other hand, the maximum value of  $H_v(X)$  is obtained when all fluxes of  $v$  have the same value, generating a uniform probability distribution  $P_v(X) = 1/N$ . In this case,  $H_v(X) = \log(N)$ , which is the maximum uncertainty for the outcome of a random sample. These two limits are not biologically realistic but illustrate the notion that  $H_v(X)$  corresponds to the average uncertainty contained in the outcome of this random sampling.

According to the principle of maximum entropy [32, 43], the  $v \in \mathcal{P}$  that best represents our knowledge of the flux configuration of the cell is the one with the largest value of  $H_v(X)$ .  $H$  can be interpreted as the average number of yes/no questions that we would need to ask in order to determine the outcomes of  $X$  (when using two as the base of log). It follows that out of all  $v \in \mathcal{P}$ , the one with the largest value of  $H$  should be selected, as this is the one that would require the fewest prior assumptions. Alternatively, the  $v$  that maximizes  $H$  can also be interpreted as the flux configuration that can happen in the greatest number of ways when a cell assigns a given amount of catalysts among its  $N$  reactions. This is the case if we assume that any two units of  $Q$  can be exchanged between reactions, for instance, by recycling the amino acids from one enzyme to produce another. Then, the greatest number of permutations in which these units of  $Q$  can be distributed among  $N$  reactions is given by the probability distribution that maximizes Boltzmann entropy [44],  $S = -k_b \sum_i P(X = x_i) \log(P(X = x_i))$ , where  $k_b$  is the Boltzmann constant and  $P(X = x_i)$  is the same as defined in Eq 4. Since  $\text{argmax}(H) = \text{argmax}(S)$  [45], it can be concluded that the  $v \in \mathcal{P}$  maximizing  $H$  also allows the units of  $Q$  to be assigned in the greatest number of ways. Therefore, such  $v$  would be the most likely to be observed.

$H_v(X)$  is a strictly convex function [46], therefore there is only one  $v \in \mathcal{P}$  that maximizes  $H_v(X)$ . Hence, we formulate MaxEnt as the following constraint-based problem:

$$\begin{aligned} & \max_v H_v(X) \\ \text{subject to :} & \\ & S v = 0 \\ & LB \leq v \leq UB \end{aligned} \quad (6)$$



## Computational implementation

For the implementation, it is important to note that splitting reversible fluxes into non-negative forward and reverse fluxes introduces cycles, such as the one depicted in Fig 1A. These cycles admit an arbitrarily large amount of flux to be added to the forward and reverse fluxes and still produce a flux configuration compatible with the metabolites' mass balances. This problem can be avoided by setting at least one flux to its experimentally observed value. Since MaxEnt finds the most uniform flux configuration that is compatible with the restrictions defined in Eq 6, the forward and reverse values of the unknown fluxes result in magnitudes similar to the experimentally known. For MaxEnt and all other methods, all reaction were assigned  $LB = 0$  and  $UB = 1000$ , except for the biomass reaction,  $v_{\mu}$ , and exchange flux of glucose,  $v_{EX\_glu}$ , which were set to match their corresponding measured values:

$$LB_{EX\_glu} = UB_{EX\_glu} = v_{EX\_glu}^{Obs}, \text{ and } LBv_{\mu} = UBv_{\mu} = v_{\mu}^{Obs}.$$

All methods (MaxEnt, flux sampling, MinFlux, and Geometric) were implemented using the COBRAPY 0.16.0 [47] library in Python 3.7. MaxEnt non-linear maximization was done using IPOPT 3.12.3 [48] optimizer through the CasADi 3.4.5 [49] interface. Flux sampling and Geometric were performed using the optGpSampler [13] and `geometric_fba` functions implemented in COBRAPY. The minimization of MinFlux's quadratic objective function was done using CPLEX 12.9.

To perform MaxEnt optimization via IPOPT we added a small number,  $\epsilon$ , to each flux value in Eq 5 to avoid the undefined value of  $\text{Log}P_v(x_i)$  when  $v_i = 0$ . For all computations we used  $\epsilon = 10^{-8}$ . We used Flux Variability Analysis (`flux_variability_analysis` function of COBRAPY) to narrow down the lower and upper bounds of each flux within the polytope of alternative solutions. Fluxes narrowed down to a single value were considered constants in MaxEnt, thus reducing the number of variables.

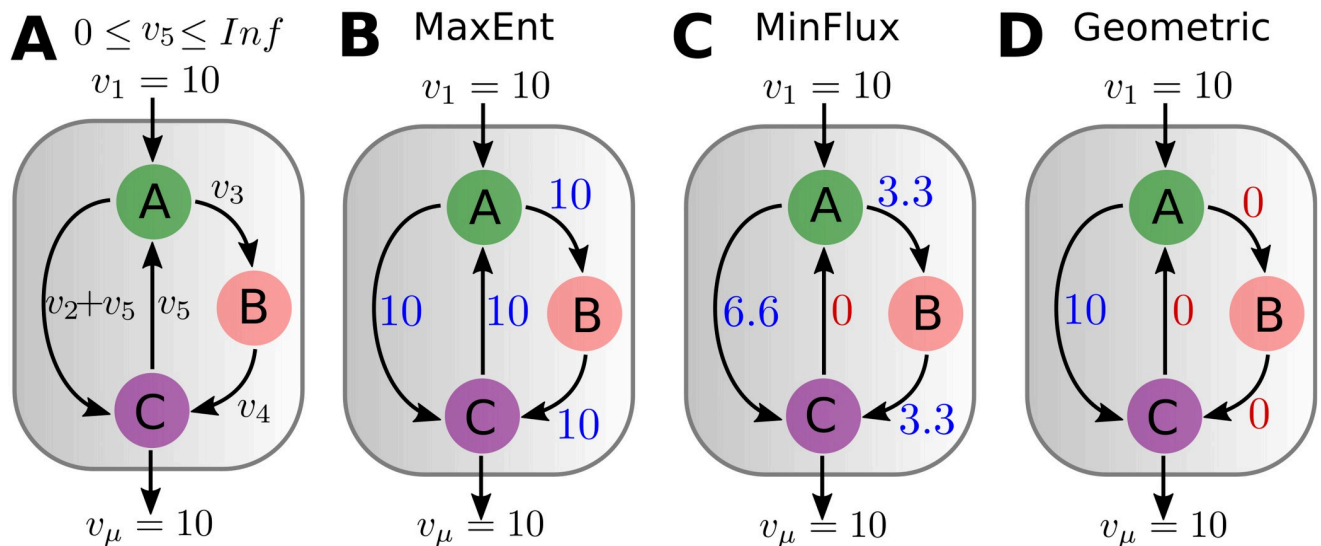
## Results

### Analysis of metabolic networks loops

Loops are an important feature of metabolic networks. They have been proposed as essentials to explain the self-amplification capacity of metabolism and necessary to re-concentrate pathways' inputs into a finite number of metabolic intermediates [50]. However, as information of the Gibbs free energy is not always available to determine the direction of reversible reactions, thermodynamically infeasible cycles can arise within the polytope of alternative solutions. This is illustrated by the example metabolic network presented in Fig 2A, which despite being constrained by its uptake ( $v_1 = 10$ ) and production ( $v_{\mu} = 10$ ) fluxes, can have an arbitrarily large flux value  $v_5$  cycling between metabolites A and C. Therefore, we started by analyzing how MaxEnt accounts for metabolic loops.

For the example network Fig 2A, MaxEnt yields a uniform distribution of fluxes (Fig 2B). By maximizing the information entropy, MaxEnt selects the most homogeneous configuration compatible with the observed fluxes, naturally tending to veer away from flux configurations where one or more fluxes have large value differences compared to the measured uptake and production fluxes. In this scenario, flux sampling would result in a distribution of values for  $v_5$  that would only be bounded by the upper limit imposed on this flux, which in itself is not biologically meaningful.

On the other hand, MinFlux and Geometric can avoid the artificially large flux values of  $v_5$  as they select flux configurations with the minimum sum of fluxes (Fig 2C and 2D). Although it is plausible to assume that cells minimize the energy costs associated with the production of the enzymes carrying out the metabolic reactions, it is not clear if these fluxes should be zero.



**Fig 2. Example network with alternatives routes and a loop.** (A) The metabolic network has a direct route from metabolites A to C and an indirect one mediated by metabolite B. A loop is formed by the reaction that goes from C back to A. The flux  $v_5$  can reach an arbitrarily large value without violating the metabolites' mass balances. (B) Constrained by the uptake and production fluxes, MaxEnt predicts a network with all fluxes being equal to 10. Any deviation from this flux configuration results in a reduction in the value of  $H(v)$ , thus ruling out arbitrarily large values of  $v_5$ . (C) MinFlux predicts a zero flux value going from C to A. (D) Geometric also predicts zero flux from C to A, but as it measures fluxes' magnitudes by their absolute values, it also predicts zero flux from the A to B and B to C reactions.

<https://doi.org/10.1371/journal.pone.0243067.g002>

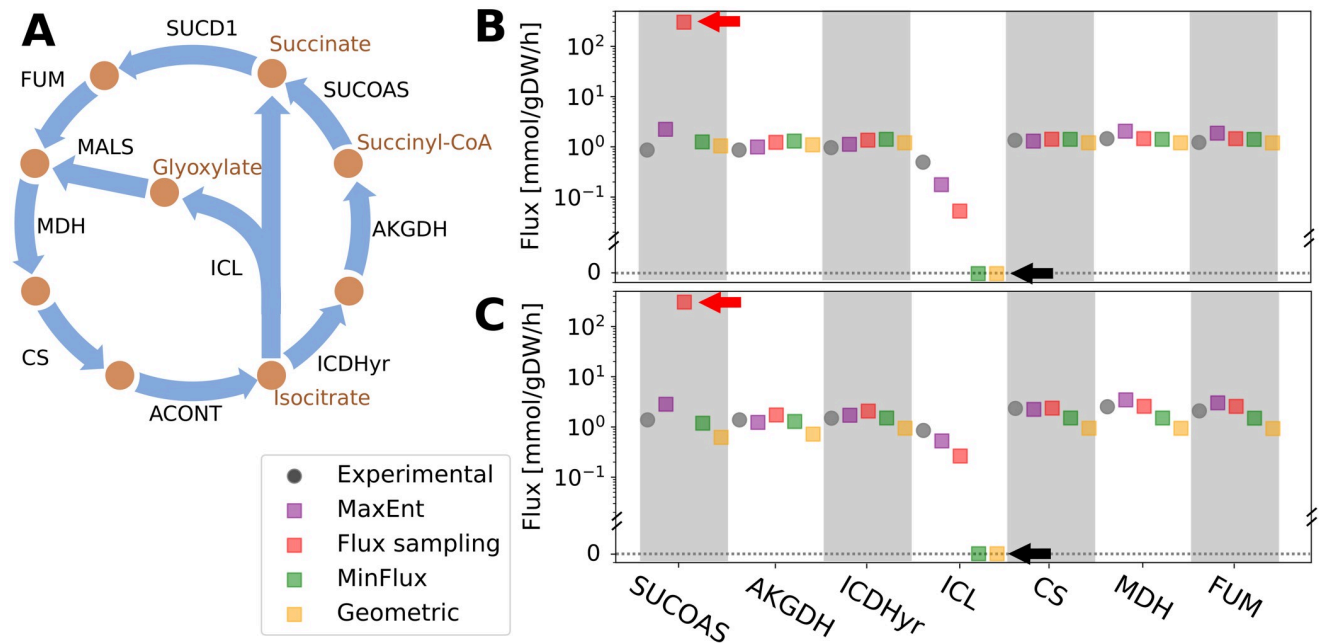
A known case is isocitrate lyase (ICL) reaction, which in *E. coli* creates a nested cycle within the Krebs cycle, the glyoxylate shunt (see Fig 3A). This reaction has been observed to have positive flux [29–31].

To test if MaxEnt would produce non zero flux value through *E. coli*'s glyoxylate shunt at a genome-scale level, we compared its predictions to 25 measured fluxes including ICL [29–31, 51–54] at two growth rates (data was retrieved from CeCaFBD [55], see also S1 File). To have a reference point, we also computed the flux configuration using flux sampling (1,350,000 samples taken within the space  $\mathcal{P}$ ), and the single predictions of MinFlux and Geometric. The results, presented in Fig 3B and 3C, show that only MaxEnt and flux sampling predict flux through the glyoxylate shunt. Comparing to experimental fluxes of the Krebs cycle, MaxEnt predicted values in the same order of magnitude (Fig 3B and 3C). On the contrary, flux sampling predicts an average SUCOAS flux that is two orders of magnitude above the experimental result, being this the result of flux sampling not able to rule out thermodynamically infeasible flux values [17]. On the other hand, MinFlux and Geometric were unable to predict flux through the glyoxylate shunt as expected by their underlying economy of fluxes assumption.

To analyze if current methods already produced flux configurations with high information entropy levels, we determined their information entropy by using Eqs 4 and 5 to their predictions (in the case of flux sampling, we used the average flux values). We found that MaxEnt predictions (Fig 4A and 4C) have a significantly larger information entropy ( $p$ -values  $< 10^{-5}$ , one-tailed test of a normal distribution) than the mean information entropy obtained by flux sampling, with MinFlux and Geometric having information entropy values in between these two methods.

We further studied MaxEnt predictions by comparing them to the rest of the 25 experimentally observed metabolic fluxes, which span the central catabolic core of *E. coli*. We quantified





**Fig 3. *E. coli* Krebs cycle and glyoxylate shunt.** (A) The glyoxylate shunt is a two-step metabolic pathway (isocitrate lyase, ICL; and malate synthase, MALS) that bypasses the Krebs cycle carbon dioxide-producing steps [56]. As it forms a loop, the economy of fluxes assumption would predict no flux through it, which contradicts experimental data. (B) and (C) show the predicted and experimental fluxes of the glyoxylate shunt and Krebs cycle at two specific biomass growth rates, 0.1 [1/h] and 0.2 [1/h], respectively. In both cases, MaxEnt and flux sampling correctly predict flux through ICL. However, flux sampling predicts a flux magnitude through succinyl-CoA synthetase (SUCOAS) which is two orders of magnitude above the observed value (red arrows). On the other hand, MinFlux and Geometric predict zero flux through ICL, see black arrows. For flux sampling, the values reported correspond to the median of 1,350,000 samples and thinning = 1000 (see S1 and S2 Figs).

<https://doi.org/10.1371/journal.pone.0243067.g003>

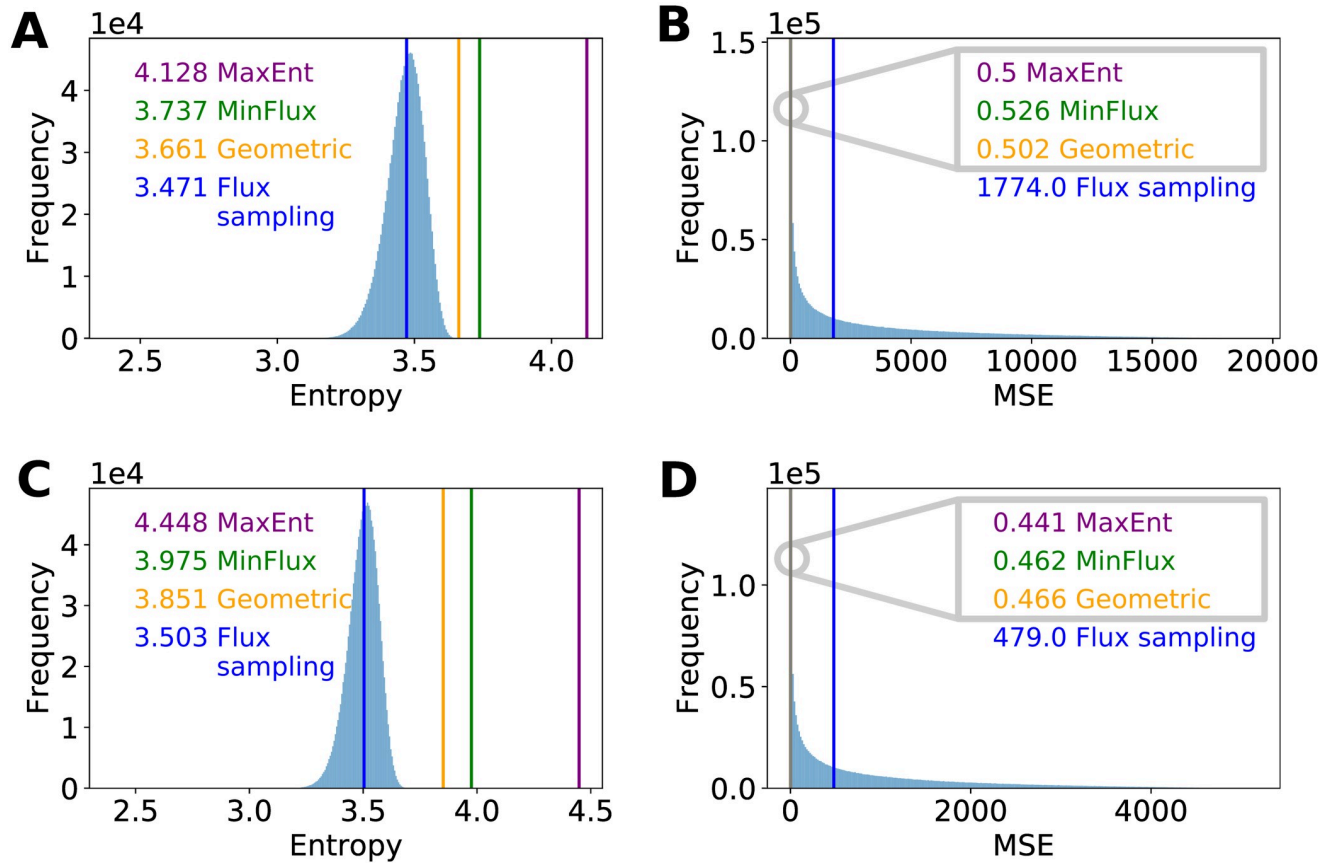
the similarity between predicted and measured fluxes using mean-squared error (MSE):

$$MSE = \frac{1}{N_f} \sum_{i=1}^{N_f} \left( \frac{v_i}{|v_{Ex-glu}^{Obs}|} - \frac{v_i^{Obs}}{|v_{Ex-glu}^{Obs}|} \right)^2, \quad (7)$$

where the measured  $v_i^{Obs}$  and predicted  $v_i$  flux are normalized by the measured  $|v_{Ex-glu}^{Obs}|$  and predicted  $|v_{Ex-glu}^{Obs}|$  magnitude of the exchange glucose rate.

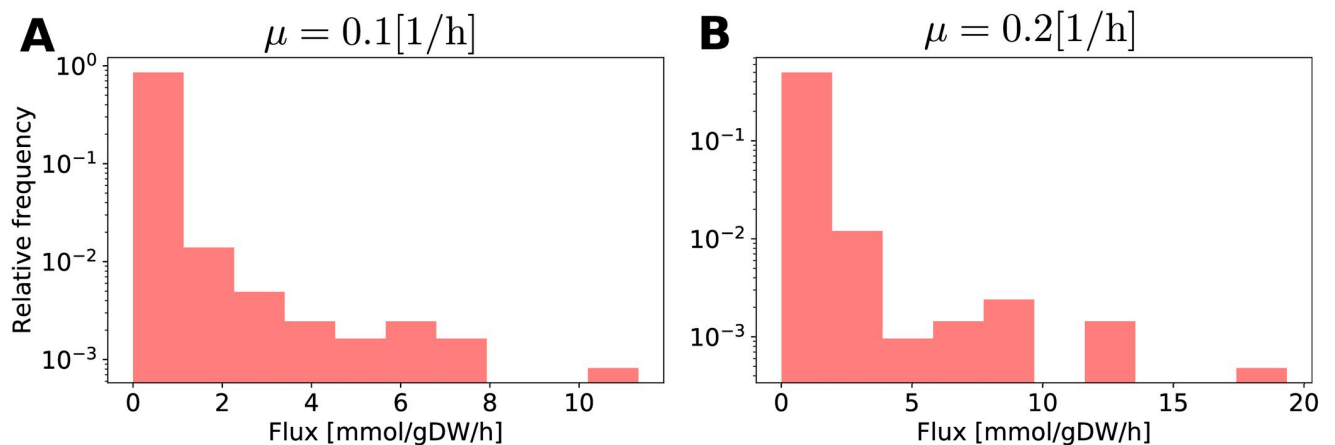
We found that MaxEnt, MinFlux, and Geometric outperforms the average solution of flux sampling (Fig 4), producing MSE values that are more than 3 orders of magnitude lower than the median MSE of flux sampling (Fig 4B and 4D). This suggests that the accuracy of flux sampling predictions is highly sensitive to artifacts introduced by thermodynamically infeasible cycles.

Previous results have shown that *E. coli* fluxes are distributed according to a power-law distribution [57], with most reactions having zero flux and only a few of them having large flux values. On the other hand, MaxEnt finds the most uniform distribution of fluxes compatible with the optimization problem's restrictions defined in Eq 6. To verify whether MaxEnt formulation produces a rich and structured distribution of fluxes or not, histograms of the flux values for *E. coli* predicted by MaxEnt at growth rates of 0.1 and 0.2 [1/h] are presented in Fig 5A and 5B. These results show that MaxEnt solution follows a power-law distribution, which was verified by plotting the same results in a log-log scale (see S3 Fig).



**Fig 4. Entropy and MSE at two *E. coli* growth rates.** (A) and (B) show the entropy and MSE values at a growth rate of 0.1 [1/h]. (C) and (D) show the same indices at a growth rate of 0.2 [1/h]. For both growth rates, MaxEnt predictions have a statistically significantly larger information entropy ( $p$ -values  $< 10^{-5}$ , one-tailed test of a normal distribution) compared to the median entropy of flux sampling, with the entropy of MinFlux and Geometric predictions falling between them. MaxEnt predictions have an MSE value in the same order of magnitude as the ones obtained by MinFlux and Geometric, but at least three orders of magnitude lower than the median MSE of flux sampling, supporting MaxEnt capacity to predict inner metabolic fluxes. For flux sampling, 1,350,000 samples from the space of alternative solutions (thinning = 1000) were taken, and for each sample, entropy and MSE were computed. The resulting distributions are shown in light blue.

<https://doi.org/10.1371/journal.pone.0243067.g004>



**Fig 5. Relative frequency of fluxes estimated by MaxEnt for *E. coli*.** (A) and (B) show histograms of fluxes of the solution of MaxEnt at growth rates 0.1 and 0.2 [1/h] in *E. coli*. For both histograms 10 bins were used.

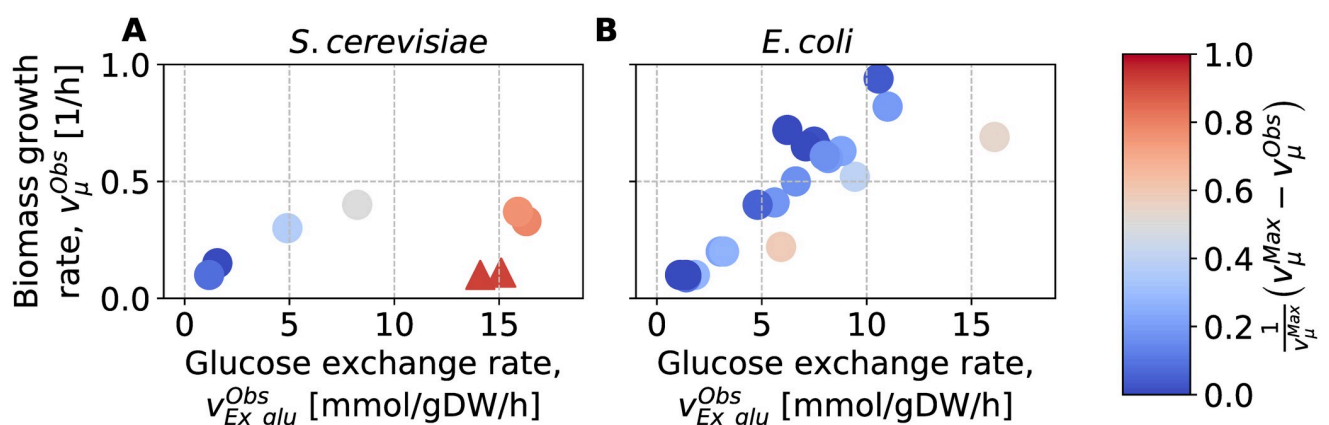
<https://doi.org/10.1371/journal.pone.0243067.g005>

## Predicting flux configuration considering overflow metabolism

MaxEnt, MinFlux, and Geometric produced flux configuration predictions with low MSE at specific growth rates of 0.1 and 0.2 [1/h], suggesting that under these growing conditions, they can predict the fluxes of the central carbon system. However, it has been observed in *S. cerevisiae* and *E. coli* that high glucose uptake rates are accompanied with partial oxidation pathways, resulting in the production of overflow metabolites: acetate, ethanol and lactate, respectively [22–24]. As a result, overflow metabolism produces a redistribution of fluxes through previously inactive pathways, which is at odds with the economy of fluxes assumption underlying MinFlux and Geometric.

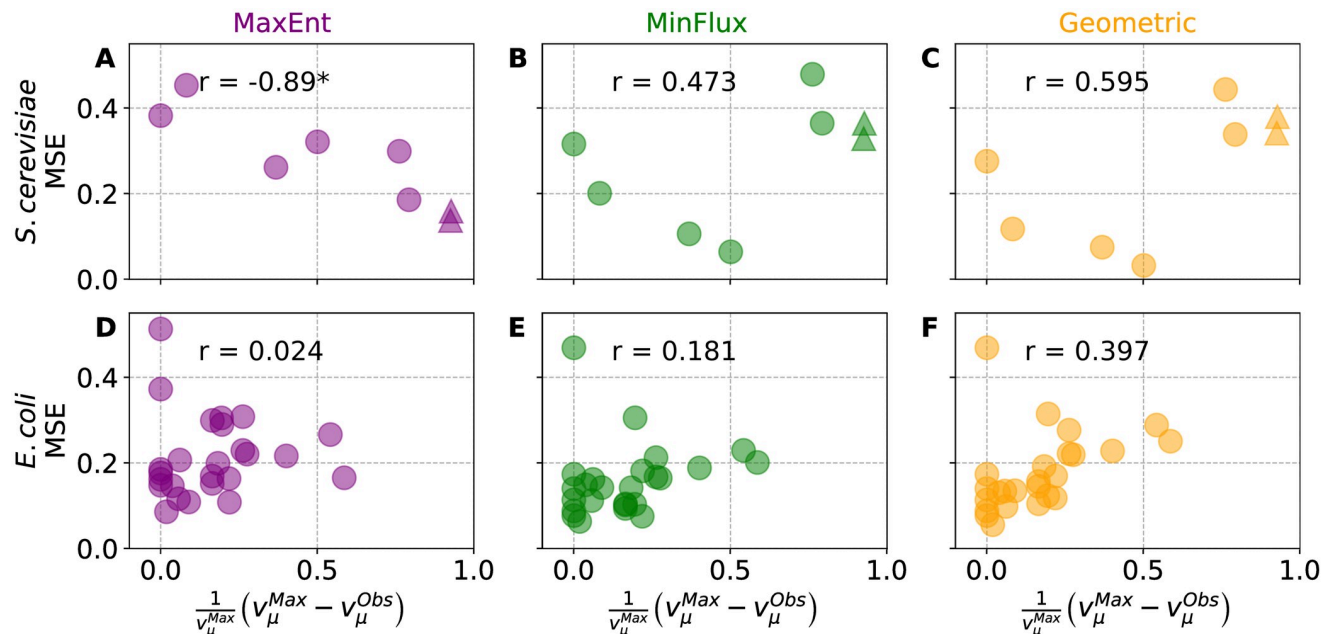
To investigate if MaxEnt is able to produce reasonable predictions at various levels of overflow metabolism, we compared its predictions against a set of 8 growth conditions in *S. cerevisiae* [58–61], and a set of 25 growth conditions in *E. coli* [29, 30, 52–54, 62–74] (data was retrieved from CeCaFBD [55], see also S2 File). In each set, the data-points vary in terms of both the uptake rate of glucose and biomass growth rate. FBA is a good predictor for the growth rate in the absence of overflow metabolism, but it overestimates the growth rates otherwise. We took advantage of this to quantify the level of overflow metabolism in the experimental data by measuring the difference between the maximum theoretical growth rate (as computed by FBA) and the actual growth rate, with the difference normalized by the maximum theoretical growth rate. The results (see Fig 6) show that the datasets of *S. cerevisiae* and *E. coli* span growth conditions with various levels of overflow metabolism.

Then, for *S. cerevisiae* and *E. coli*, we set their specific growth and glucose uptake rates to match their observed values and used MaxEnt to predict the corresponding flux configurations. The results (see Fig 7) show that MinFlux and Geometric predictions have lower MSE values compared to MaxEnt when the level of overflow metabolism is close to zero, but that the situation reverts at levels of overflow metabolism close to 1. On the contrary, MaxEnt prediction performance seems unaffected by higher levels of overflow metabolism. Although not statistically significant, these results support the use of MaxEnt over MinFlux and Geometric when metabolic pathways deviate from biomass production to generation of overflow metabolites.



**Fig 6. Metabolism of *S. cerevisiae* and *E. coli* data at various levels of overflow metabolism.** We estimated the level of overflow metabolism as the normalized difference between observed  $v_{\mu}^{Obs}$  and theoretical maximum  $v_{\mu}^{Max}$  biomass growth rates. (A) Various growth conditions of *S. cerevisiae* are represented by their observed rates of glucose exchange (uptake) and biomass growth.  $\Delta$  correspond to mutated strains overproducing acetate. (B) Various growth conditions of *E. coli*.

<https://doi.org/10.1371/journal.pone.0243067.g006>



**Fig 7. Correlation between overflow metabolism and Mean Square Error (MSE).** (A), (B), and (C) show *S. cerevisiae*'s MSE between observed and predicted inner metabolic fluxes. MinFlux and Geometric predictions outperform MaxEnt when the level of overflow metabolism is below 0.5 but thereafter the situation inverts, suggesting that the minimization of fluxes assumption of MinFlux and Geometric may not be universally suitable. (D), (E), and (F) show a similar trend in *E. coli*. MinFlux and Geometric predictions have lower MSE than MaxEnt at low levels of overflow metabolism, but their MSE shows a positive Pearson correlation,  $r$ , at higher levels of overflow metabolism. On the other hand, MaxEnt shows a close to zero correlation with the level of overflow metabolism in both species. \* indicates statistically significant Pearson correlation values ( $p$ -value  $\leq 0.005$ ). Data points are color coded as ● MaxEnt, ● Geometric, and ● MinFlux.  $\Delta$  correspond to mutated strains overproducing acetate.

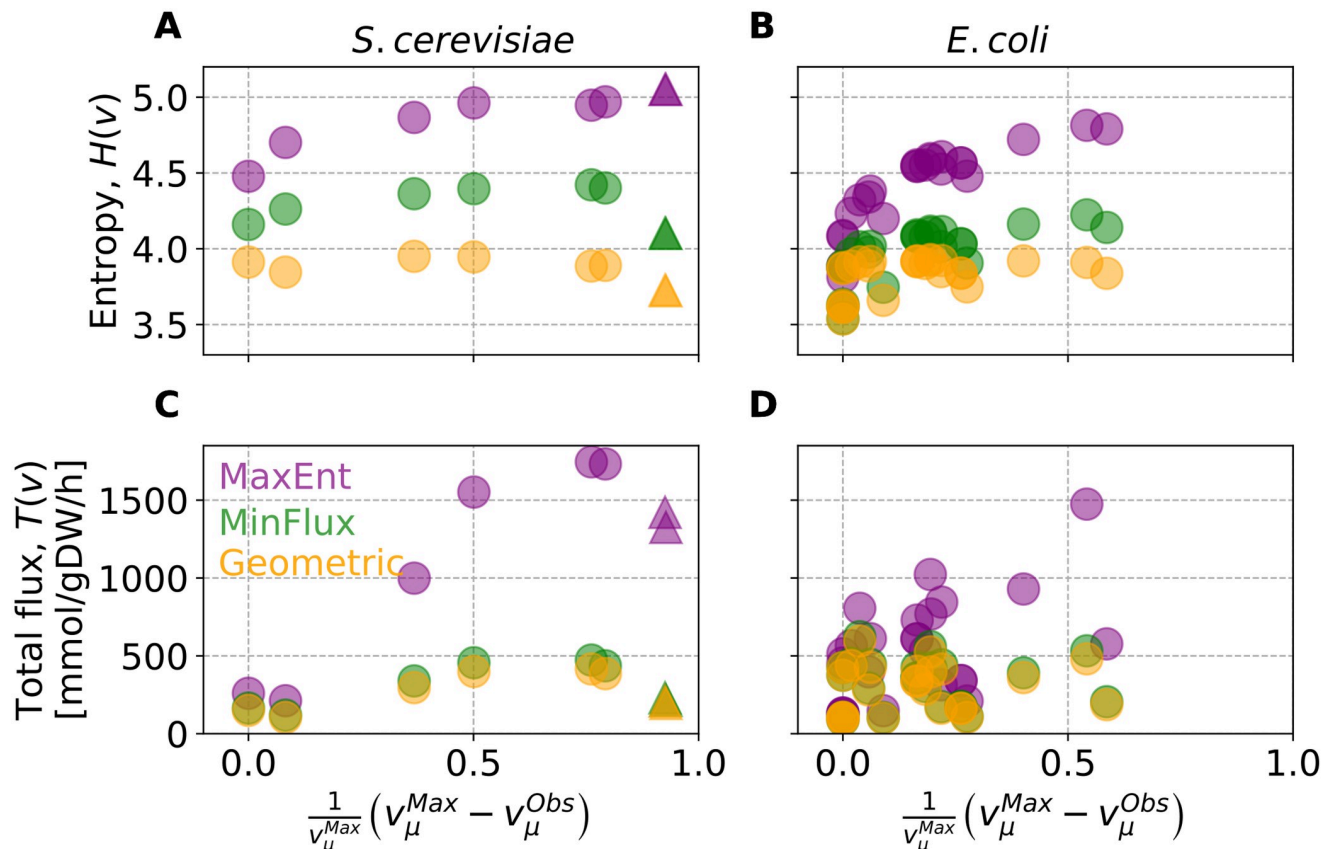
<https://doi.org/10.1371/journal.pone.0243067.g007>

There has been extensive work to explain the simultaneous use of ATP-efficient and inefficient pathways during overflow metabolism [75]. Several constraint-based models have been able to reproduce overflow metabolism behavior [76] by integrating relevant information coming from proteomics [77], gene expression [78], limitation in oxygen uptake rates [79], and free energy dissipation [28]. The extra information used by these models reduces the solution space but typically does not single out a unique flux configuration. As a constraint-based model, MaxEnt can be swiftly integrated into these models, helping the study of overflow metabolism by estimating a single flux configuration without adding extra assumptions.

To explore the differences between MaxEnt, on the one hand, and MinFlux and Geometric, on the other, that could explain their divergent behavior at higher levels of overflow metabolism, we analyzed the information entropy of their predictions. The results (see Fig 8A and 8C) show that the information entropy increases with the level of overflow metabolism. This likely stems from the additional metabolic fluxes activated to divert flux from biomass to produce overflow metabolites. To test this, at each growth condition, we measured the total flux:

$$T(v) = \sum_{i=1}^N v_i \quad (8)$$

We found that all methods increase their total flux with overflow metabolism (see Fig 8C and 8D), forming a saturation curve as the level of overflow metabolisms increases. Compared to MinFlux and Geometric, MaxEnt predicts larger increments in total flux, this being



**Fig 8. Information entropy and flux through the metabolic network predicted by various methods.** (A) and (B) show the information entropy of the flux configurations for *S. cerevisiae* and *E. coli*, respectively. For each time point, MaxEnt predictions always have greater information entropy compared to MinFlux and Geometric. (C) and (D) show the total flux of the metabolic configurations predicted for *S. cerevisiae* and *E. coli*, respectively. MaxEnt predicts larger total flux compared to the other two methods, as the latter rely on the minimization of fluxes assumption. Data points are color coded as ● MaxEnt, ● Geometric, and ● MinFlux. Δ corresponds to mutated strains overproducing acetate.

<https://doi.org/10.1371/journal.pone.0243067.g008>

coherent with its tendency to predict a more homogeneous distribution of fluxes, and it results in more reactions carrying flux.

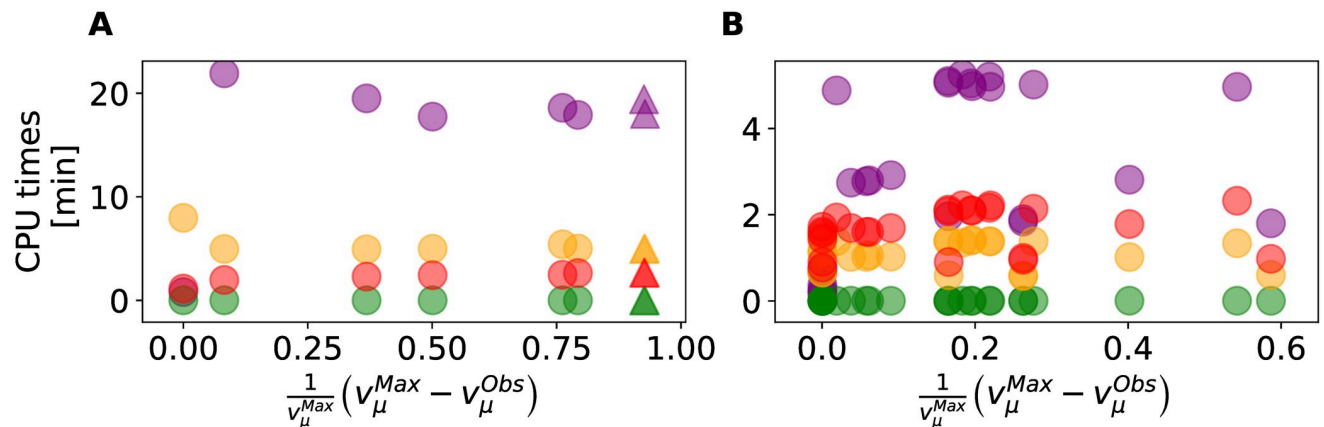
### Computing times

Finally, we compared MaxEnt and alternative methods CPU times for various levels of overflow metabolism. The results (see Fig 9) show that the CPU times of MaxEnt, Geometric, and flux sampling are all within the same order of magnitude. Only MinFlux resulted in CPU times within fractions of a second. MaxEnt CPU times do not increase linearly with the level of overflow metabolism but caps on average at 20 min for the iMM904 network of *S. cerevisiae* and 6 min for the iJR904 network of *E. coli*. MaxEnt was implemented using out of the shelf algorithms, and its CPU times may be further reduced if a tailored implementation is used.

### Conclusion

Given a set of measured fluxes, constraint-based models typically predict a consistent space of flux configurations. In this work, we present a method based on the principle of maximum entropy, which in this context states that the best estimation of fluxes is the one with the least amount of unwarranted assumptions. We searched for the least bias flux configuration by





**Fig 9. CPU times for various levels of overflow metabolism.** (A) iMM904 (*Saccharomyces cerevisiae*), and (B) iJR904 (*Escherichia coli*). Data points are color coded as ● MaxEnt, ● Geometric, ● flux sampling and ● MinFlux. Δ corresponds to mutated strains overproducing acetate. 1000 samples were used in flux sampling.

<https://doi.org/10.1371/journal.pone.0243067.g009>

computing its information entropy. Based on this, we formulated a constraint-based approach, MaxEnt, to find a single flux configuration that maximizes the information entropy within the space of alternative solutions. We found that MaxEnt predictions avoid artificially large flux values due to thermodynamically infeasible cycles in the metabolic networks. MaxEnt correctly predicted flux through the ICL reaction of the glyoxylate shunt of *E. coli*, which the alternative methods, MinFlux and Geometric, missed as they systematically avoid the formation of cycles. Unlike flux sampling, MaxEnt predicts fluxes in the same order of magnitude as the experimentally observed ones. MaxEnt also produces accurate estimations of the fluxes of the central carbon systems of *E. coli* and *S. cerevisiae* at various levels of overflow metabolism. In all these cases, MaxEnt does not require prior assumptions about the distribution of fluxes or their bounds, both of which can introduce observer bias in the results. By selecting the least bias flux configuration, MaxEnt is less prone to over-fitting, which is its main advantage over alternative methods, and may prove useful for estimating flux configurations when there is not sufficiently available *bona fide* information to constraint the solution space to a single point.

## Supporting information

**S1 Fig. Flux value distributions for reactions of the Krebs cycle and glyoxylate shunt at growth rate 0.1 [1/h].** For each reaction, a distribution of 1,350,000 flux values was obtained using flux sampling (thinning = 1000).

(TIF)

**S2 Fig. Flux value distributions for reactions of the Krebs cycle and glyoxylate shunt at growth rate 0.2 [1/h].** For each reaction, a distribution of 1,350,000 flux values was obtained using flux sampling (thinning = 1000).

(TIF)

**S3 Fig. Frequency of the flux configuration values predicted by MaxEnt for *E. coli* at growth rates 0.1 and 0.2 [1/h].** The figures are scatter plots in log-log scale.

(TIF)

**S1 File. Flux data for *E. coli* at growth rates 0.1 and 0.2 [1/h].**

(XLS)



**S2 File. Flux data for *S. cerevisiae* and *E. coli* at various growth conditions.**  
(XLSX)

## Acknowledgments

The authors would like to acknowledge Prof. Pamela Wilson (Escuela de Ingeniería Industrial, Pontificia Universidad Católica de Valparaíso, Chile) for the discussions about the theoretical aspects of this work.

## Author Contributions

**Conceptualization:** Marcelo Rivas-Astroza, Raúl Conejeros.

**Formal analysis:** Marcelo Rivas-Astroza.

**Funding acquisition:** Marcelo Rivas-Astroza.

**Investigation:** Raúl Conejeros.

**Methodology:** Raúl Conejeros.

**Project administration:** Marcelo Rivas-Astroza.

**Supervision:** Raúl Conejeros.

**Writing – original draft:** Marcelo Rivas-Astroza.

**Writing – review & editing:** Marcelo Rivas-Astroza, Raúl Conejeros.

## References

1. De Martino D, Andersson AMC, Bergmiller T, Guet CC, Tkačik G. Statistical mechanics for metabolic networks during steady state growth. *Nature Communications*. 2018; 9(1).
2. Fell DA, Small JR. Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *Biochemical Journal*. 1986; 238(3):781–786. <https://doi.org/10.1042/bj2380781>
3. Kacser H, Burns JA, Fell DA. The control of flux. In: *Biochemical Society Transactions*. vol. 23. Portland Press Ltd; 1995. p. 341–366.
4. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, et al. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Molecular Systems Biology*. 2011; 7:535. <https://doi.org/10.1038/msb.2011.65> PMID: 21988831
5. Cuevas DA, Edirisinghe J, Henry CS, Overbeek R, O'Connell TG, Edwards RA. From DNA to FBA: How to Build Your Own Genome-Scale Metabolic Model. *Frontiers in Microbiology*. 2016; 7:907.
6. Fondi M, Liò P. Genome-scale metabolic network reconstruction. *Methods in Molecular Biology*. 2015; 1231:233–256. [https://doi.org/10.1007/978-1-4939-1720-4\\_15](https://doi.org/10.1007/978-1-4939-1720-4_15)
7. Thiele I, Palsson B. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*. 2010; 5(1):93–121. <https://doi.org/10.1038/nprot.2009.203>
8. Ibarra RU, Edwards JS, Palsson BO. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature*. 2002; 420(6912):186–189. <https://doi.org/10.1038/nature01149>
9. Edwards JS, Ibarra RU, Palsson BO. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnology*. 2001; 19(2):125–130. <https://doi.org/10.1038/84379>
10. Sambamoorthy G, Raman K. Understanding the evolution of functional redundancy in metabolic networks. *Bioinformatics*. 2018; 34(17):i981–i987. <https://doi.org/10.1093/bioinformatics/bty604>
11. Ebrahim A, Almaas E, Bauer E, Bordbar A, Burgard AP, Chang RL, et al. Do genome-scale models need exact solvers or clearer standards? *Molecular Systems Biology*. 2015; 11(10):831. <https://doi.org/10.15252/msb.20156157> PMID: 26467284
12. Herrmann HA, Dyson BC, Vass L, Johnson GN, Schwartz JM. Flux sampling is a powerful tool to study metabolism under changing environmental conditions. *npj Systems Biology and Applications*. 2019; 5(1):1–8.

13. Megchelenbrink W, Huynen M, Marchiori E. optGpSampler: An improved tool for uniformly sampling the solution-space of genome-scale metabolic networks. *PLoS ONE*. 2014; 9(2):e86587. <https://doi.org/10.1371/journal.pone.0086587>
14. De Martino A, De Martino D. An introduction to the maximum entropy approach and its application to inference problems in biology. *Heliyon*. 2018; 4(4):e00596. <https://doi.org/10.1016/j.heliyon.2018.e00596>
15. Henry CS, Broadbelt LJ, Hatzimanikatis V. Thermodynamics-based metabolic flux analysis. *Biophysical Journal*. 2007; 92(5):1792–1805. <https://doi.org/10.1529/biophysj.106.093138>
16. Schellenberger J, Lewis NE, Palsson B. Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophysical Journal*. 2011; 100(3):544–553. <https://doi.org/10.1016/j.bpj.2010.12.3707>
17. Desouki AA, Jarre F, Gelius-Dietrich G, Lercher MJ. CycleFreeFlux: efficient removal of thermodynamically infeasible loops from flux distributions. *Bioinformatics*. 2015; 31(13):2159–2165. <https://doi.org/10.1093/bioinformatics/btv096>
18. Kim MK, Lane A, Kelley JJ, Lun DS. E-Flux2 and sPOT: Validated methods for inferring intracellular metabolic flux distributions from transcriptomic data. *PLoS ONE*. 2016; 11(6):e0157101. <https://doi.org/10.1371/journal.pone.0157101>
19. Smallbone K, Simeonidis E. Flux balance analysis: A geometric perspective. *Journal of Theoretical Biology*. 2009; 258(2):311–315. <https://doi.org/10.1016/j.jtbi.2009.01.027>
20. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, et al. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Molecular Systems Biology*. 2010; 6(1):390. <https://doi.org/10.1038/msb.2010.47> PMID: 20664636
21. Jenatton R, Audibert JY, Bach F. Structured variable selection with sparsity-inducing norms. *The Journal of Machine Learning Research*. 2011; 12:2777–2824.
22. Wolfe AJ. The Acetate Switch. *Microbiology and Molecular Biology Reviews*. 2005; 69(1):12–50. <https://doi.org/10.1128/MMBR.69.1.12-50.2005>
23. Postma E, Verduyn C, Scheffers WA, Van Dijken JP. Enzymic analysis of the crabtree effect in glucose-limited chemostat cultures of *Saccharomyces cerevisiae*. Applied and environmental microbiology. 1989; 55(2):468–477. <https://doi.org/10.1128/AEM.55.2.468-477.1989>
24. Warburg O. On the origin of cancer cells. *Science*. 1956; 123(3191):309–314. <https://doi.org/10.1126/science.123.3191.309>
25. Molenaar D, Van Berlo R, De Ridder D, Teusink B. Shifts in growth strategies reflect tradeoffs in cellular economics. *Molecular Systems Biology*. 2009; 5:323. <https://doi.org/10.1038/msb.2009.82>
26. Basan M, Hui S, Okano H, Zhang Z, Shen Y, Williamson JR, et al. Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. *Nature*. 2015; 528(7580):99–104. <https://doi.org/10.1038/nature15765> PMID: 26632588
27. Zhuang K, Vemuri G, Mahadevan R, Kim H, Robin K, Tung C, et al. Metabolic constraints on the evolution of antibiotic resistance. *Molecular Systems Biology*. 2017; 7(1):500–500.
28. Niebel B, Leupold S, Heinemann M. An upper limit on Gibbs energy dissipation governs cellular metabolism. *Nature Metabolism*. 2019; 1(1):125–132. <https://doi.org/10.1038/s42255-018-0006-7>
29. Nanchen A, Schicker A, Revelles O, Sauer U. Cyclic AMP-dependent catabolite repression is the dominant control mechanism of metabolic fluxes under glucose limitation in *Escherichia coli*. *Journal of Bacteriology*. 2008; 190(7):2323–2330. <https://doi.org/10.1128/JB.01353-07>
30. Al Zaid Siddiquee K, Arauzo-Bravo MJ, Shimizu K. Metabolic flux analysis of pykF gene knockout *Escherichia coli* based on <sup>13</sup>C-labeling experiments together with measurements of enzyme activities and intracellular metabolite concentrations. *Applied Microbiology and Biotechnology*. 2004; 63(4):407–417. <https://doi.org/10.1007/s00253-003-1357-9>
31. Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, et al. Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science*. 2007; 316(5824):593–597. <https://doi.org/10.1126/science.1132067> PMID: 17379776
32. Jaynes ET. Information theory and statistical mechanics. *Physical Review*. 1957; 106(4):620–630. <https://doi.org/10.1103/PhysRev.106.620>
33. McGlenn DJ, Xiao X, Kitzes J, White EP. Exploring the spatially explicit predictions of the Maximum Entropy Theory of Ecology. *Global Ecology and Biogeography*. 2015; 24(6):675–684. <https://doi.org/10.1111/geb.12295>
34. De Martino D, Capuani F, De Martino A. Growth against entropy in bacterial metabolism: the phenotypic trade-off behind empirical growth rate distributions in *E. coli*. *Physical Biology*. 2016; 13(3):036005. <https://doi.org/10.1088/1478-3975/13/3/036005>

35. Fernandez-de Cossio-Diaz J, Mulet R. Maximum entropy and population heterogeneity in continuous cell cultures. *PLoS computational biology*. 2019; 15(2):e1006823. <https://doi.org/10.1371/journal.pcbi.1006823>
36. Tourigny DS. Dynamic metabolic resource allocation based on the maximum entropy principle. *Journal of Mathematical Biology*. 2020; 80:2395–2430. <https://doi.org/10.1007/s00285-020-01499-6>
37. Gerstl MP, Ruckerbauer DE, Mattanovich D, Jungreuthmayer C, Zanghellini J. Metabolomics integrated elementary flux mode analysis in large metabolic networks. *Scientific reports*. 2015; 5(1):1–8.
38. Shannon CE. A mathematical theory of communication. *The Bell System Technical Journal*. 1948; 27(3):379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
39. Reed JL, Vo TD, Schilling CH, Palsson BO. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome biology*. 2003; 4(9):r54. <https://doi.org/10.1186/gb-2003-4-9-r54>
40. Mo ML, Palsson B, Herrgård MJ. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Systems Biology*. 2009; 3:37. <https://doi.org/10.1186/1752-0509-3-37>
41. King ZA, Lu J, Dräger A, Miller P, Federowicz S, Lerman JA, et al. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic acids research*. 2016; 44(D1):D515–D522. <https://doi.org/10.1093/nar/gkv1049> PMID: 26476456
42. Davidi D, Noor E, Liebermeister W, Bar-Even A, Flamholz A, Tummli K, et al. Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro kcat measurements. *Proceedings of the National Academy of Sciences*. 2016; 113(12):3401–3406. <https://doi.org/10.1073/pnas.1514240113> PMID: 26951675
43. Jaynes ET. Information theory and statistical mechanics. II. *Physical Review*. 1957; 108(2):171–190. <https://doi.org/10.1103/PhysRev.108.171>
44. Boltzmann L. On the relationship between the second fundamental theorem of the mechanical theory of heat and probability calculations regarding the conditions for thermal equilibrium. *Entropy*. 2015; 17(4):1971–2009.
45. Sohrab SH. Boltzmann entropy of thermodynamics versus Shannon entropy of information theory. *International Journal of Mechanics*. 2014; 8(1):73–84.
46. Rao CR. Convexity properties of entropy functions and analysis of diversity. In: *Inequalities in Statistics and Probability*. Institute of Mathematical Statistics; 1984. p. 68–77.
47. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: COstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology*. 2013; 7(1):74. <https://doi.org/10.1186/1752-0509-7-74>
48. Wächter A, Biegler LT. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*. 2006; 106(1):25–57. <https://doi.org/10.1007/s10107-004-0559-y>
49. Andersson JAE, Gillis J, Horn G, Rawlings JB, Diehl M. CasADi: a software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation*. 2019; 11(1):1–36. <https://doi.org/10.1007/s12532-018-0139-4>
50. Braakman R, Smith E. The compositional and evolutionary logic of metabolism. *Physical Biology*. 2012; 10(1):11001. <https://doi.org/10.1088/1478-3975/10/1/011001>
51. Nanchen A, Schicker A, Sauer U. Nonlinear dependency of intracellular fluxes on growth rate in miniaturized continuous cultures of *Escherichia coli*. *Applied and Environmental Microbiology*. 2006; 72(2):1164–1172. <https://doi.org/10.1128/AEM.72.2.1164-1172.2006>
52. Hua Q, Yang C, Baba T, Mori H, Shimizu K. Responses of the central metabolism in *Escherichia coli* to phosphoglucose isomerase and glucose-6-phosphate dehydrogenase Knockouts. *Journal of Bacteriology*. 2003; 185(24):7053–7067. <https://doi.org/10.1128/JB.185.24.7053-7067.2003>
53. Toya Y, Ishii N, Hirasawa T, Naba M, Hirai K, Sugawara K, et al. Direct measurement of isotopomer of intracellular metabolites using capillary electrophoresis time-of-flight mass spectrometry for efficient metabolic flux analysis. *Journal of Chromatography A*. 2007; 1159(1-2):134–141. <https://doi.org/10.1016/j.chroma.2007.04.011> PMID: 17462663
54. Li M, Ho PY, Yao S, Shimizu K. Effect of IpdA gene knockout on the metabolism in *Escherichia coli* based on enzyme activities, intracellular metabolite concentrations and metabolic flux analysis by <sup>13</sup>C-labeling experiments. *Journal of Biotechnology*. 2006; 122(2):254–266. <https://doi.org/10.1016/j.jbiotec.2005.09.016>
55. Zhang Z, Shen T, Rui B, Zhou W, Zhou X, Shang C, et al. CeCaFDB: a curated database for the documentation, visualization and comparative analysis of central carbon metabolic flux distributions explored by <sup>13</sup>C-fluxomics. *Nucleic acids research*. 2015; 43(D1):D549–D557. <https://doi.org/10.1093/nar/gku1137> PMID: 25392417
56. Ahn S, Jung J, Jang IA, Madsen EL, Park W. Role of glyoxylate shunt in oxidative stress response. *Journal of Biological Chemistry*. 2016; 291(22):11928–11938. <https://doi.org/10.1074/jbc.M115.708149>

57. Almaas E, Kovacs B, Vicsek T, Oltvai ZN, Barabási AL. Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature*. 2004; 427(6977):839–843. <https://doi.org/10.1038/nature02289>
58. Fendt SM, Sauer U. Transcriptional regulation of respiration in yeast metabolizing differently repressive carbon substrates. *BMC systems biology*. 2010; 4(1):12. <https://doi.org/10.1186/1752-0509-4-12>
59. Frick O, Wittmann C. Characterization of the metabolic shift between oxidative and fermentative growth in *Saccharomyces cerevisiae* by comparative <sup>13</sup>C flux analysis. *Microbial Cell Factories*. 2005; 4(1):30. <https://doi.org/10.1186/1475-2859-4-30>
60. Gombert AK, Dos Santos MM, Christensen B, Nielsen J. Network identification and flux quantification in the central metabolism of *Saccharomyces cerevisiae* under different conditions of glucose repression. *Journal of Bacteriology*. 2001; 183(4):1441–1451. <https://doi.org/10.1128/JB.183.4.1441-1451.2001>
61. Papini M, Nookaew I, Siewers V, Nielsen J. Physiological characterization of recombinant *Saccharomyces cerevisiae* expressing the *Aspergillus nidulans* phosphoketolase pathway: Validation of activity through <sup>13</sup>C-based metabolic flux analysis. *Applied Microbiology and Biotechnology*. 2012; 95(4):1001–1010. <https://doi.org/10.1007/s00253-012-3936-0>
62. Emmerling M, Dauner M, Ponti A, Fiaux J, Hochuli M, Szyperski T, et al. Metabolic flux responses to pyruvate kinase knockout in *Escherichia coli*. *Journal of Bacteriology*. 2002; 184(1):152–164. <https://doi.org/10.1128/JB.184.1.152-164.2002> PMID: 11741855
63. Farmer WR, Liao JC. Reduction of aerobic acetate production by *Escherichia coli*. *Applied and Environmental Microbiology*. 1997; 63(8):3205–3210. <https://doi.org/10.1128/AEM.63.8.3205-3210.1997>
64. Fischer E, Zamboni N, Sauer U. High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived <sup>13</sup>C constraints. *Analytical Biochemistry*. 2004; 325(2):308–316. <https://doi.org/10.1016/j.ab.2003.10.036>
65. Fong SS, Nanchen A, Palsson BO, Sauer U. Latent pathway activation and increased pathway capacity enable *Escherichia coli* adaptation to loss of key metabolic enzymes. *Journal of Biological Chemistry*. 2006; 281(12):8024–8033. <https://doi.org/10.1074/jbc.M510016200>
66. Haverkorn van Rijsewijk BRB, Nanchen A, Nallet S, Kleijn RJ, Sauer U. Large-scale <sup>13</sup>C-flux analysis reveals distinct transcriptional control of respiratory and fermentative metabolism in *Escherichia coli*. *Molecular Systems Biology*. 2011; 7(1):477. <https://doi.org/10.1038/msb.2011.9>
67. Holms H. Flux analysis and control of the central metabolic pathways in *Escherichia coli*. *FEMS Microbiology Reviews*. 1996; 19(2):85–116. <https://doi.org/10.1111/j.1574-6976.1996.tb00255.x>
68. Jiao Z, Baba T, Mori H, Shimizu K. Analysis of metabolic and physiological responses to *gnd* knockout in *Escherichia coli* by using C-13 tracer experiment and enzyme activity measurement. *FEMS Microbiology Letters*. 2003; 220(2):295–301. [https://doi.org/10.1016/S0378-1097\(03\)00133-2](https://doi.org/10.1016/S0378-1097(03)00133-2)
69. Meza E, Becker J, Bolivar F, Gosset G, Wittmann C. Consequences of phosphoenolpyruvate:sugar phosphotransferase system and pyruvate kinase isozymes inactivation in central carbon metabolism flux distribution in *Escherichia coli*. *Microbial cell factories*. 2012; 11(1):127. <https://doi.org/10.1186/1475-2859-11-127>
70. Peng L, Arauzo-Bravo MJ, Shimizu K. Metabolic flux analysis for a *ppc* mutant *Escherichia coli* based on <sup>13</sup>C-labelling experiments together with enzyme activity assays and intracellular metabolite measurements. *FEMS Microbiology Letters*. 2004; 235(1):17–23. <https://doi.org/10.1111/j.1574-6968.2004.tb09562.x>
71. Perrenoud A, Sauer U. Impact of global transcriptional regulation by *ArcA*, *ArcB*, *Cra*, *Crp*, *Cya*, *Fnr*, and *Mlc* on glucose catabolism in *Escherichia coli*. *Journal of Bacteriology*. 2005; 187(9):3171–3179. <https://doi.org/10.1128/JB.187.9.3171-3179.2005>
72. Sauer U, Canonaco F, Heri S, Perrenoud A, Fischer E. The soluble and membrane-bound transhydrogenases *UdhA* and *PntAB* have divergent functions in NADPH metabolism of *Escherichia coli*. *Journal of Biological Chemistry*. 2004; 279(8):6613–6619. <https://doi.org/10.1074/jbc.M311657200>
73. Zhao J, Shimizu K. Metabolic flux analysis of *Escherichia coli* K12 grown on <sup>13</sup>C-labeled acetate and glucose using GC-MS and powerful flux calculation method. *Journal of Biotechnology*. 2003; 101(2):101–117. [https://doi.org/10.1016/S0168-1656\(02\)00316-4](https://doi.org/10.1016/S0168-1656(02)00316-4)
74. Zhao J, Baba T, Mori H, Shimizu K. Global metabolic response of *Escherichia coli* to *gnd* or *zwf* gene-knockout, based on <sup>13</sup>C-labeling experiments and the measurement of enzyme activities. *Applied Microbiology and Biotechnology*. 2004; 64(1):91–98. <https://doi.org/10.1007/s00253-003-1458-5>
75. Mori M, Marinari E, De Martino A. A yield-cost tradeoff governs *Escherichia coli*'s decision between fermentation and respiration in carbon-limited growth. *NPJ systems biology and applications*. 2019; 5(1):1–9.
76. De Groot DH, Lischke J, Muolo R, Planqué R, Bruggeman FJ, Teusink B. The common message of constraint-based optimization approaches: overflow metabolism is caused by two growth-limiting constraints. *Cellular and Molecular Life Sciences*. 2020; 77(3):441–453.

77. Mori M, Hwa T, Martin OC, De Martino A, Marinari E. Constrained Allocation Flux Balance Analysis. *PLoS Computational Biology*. 2016; 12(6). <https://doi.org/10.1371/journal.pcbi.1004913> PMID: [27355325](https://pubmed.ncbi.nlm.nih.gov/27355325/)
78. O'Brien EJ, Lerman JA, Chang RL, Hyduke DR, Palsson BØ. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology*. 2013; 9(1):693. <https://doi.org/10.1038/msb.2013.52>
79. Carlson R, Sreenc F. Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: creation of overall flux states. *Biotechnology and Bioengineering*. 2004; 86(2):149–162. <https://doi.org/10.1002/bit.20044>