

Constrained Allocation Flux Balance Analysis

*Original*

Constrained Allocation Flux Balance Analysis / Mori, M.; Hwa, T.; Martin, O. C.; De Martino, A.; Marinari, E.. - In: PLOS COMPUTATIONAL BIOLOGY. - ISSN 1553-734X. - 12:6(2016), p. e1004913. [[10.1371/journal.pcbi.1004913](https://doi.org/10.1371/journal.pcbi.1004913)]

*Availability:*

This version is available at: 11583/2976760 since: 2023-03-10T09:27:53Z

*Publisher:*

Public Library of Science

*Published*

DOI:[10.1371/journal.pcbi.1004913](https://doi.org/10.1371/journal.pcbi.1004913)

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)

RESEARCH ARTICLE

# Constrained Allocation Flux Balance Analysis

Matteo Mori<sup>1,2,3</sup>, Terence Hwa<sup>3,4</sup>, Olivier C. Martin<sup>5</sup>, Andrea De Martino<sup>1,6,7,8</sup>\*, Enzo Marinari<sup>1,6,9</sup>

**1** Dipartimento di Fisica, Sapienza Università di Roma, Rome, Italy, **2** Departamento de Bioquímica y Biología Molecular I, Universidad Complutense de Madrid, Madrid, Spain, **3** Department of Physics, University of California at San Diego, La Jolla, California, United States of America, **4** Institute for Theoretical Studies, ETH Zurich, Switzerland, **5** GQE - Le Moulon, INRA, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Gif-sur-Yvette, France, **6** Soft and Living Matter Lab, Istituto di Nanotecnologia (CNR-NANOTEC), Consiglio Nazionale delle Ricerche, Rome, Italy, **7** Center for Life Nano Science@Sapienza, Istituto Italiano di Tecnologia, Rome, Italy, **8** Human Genetics Foundation, Turin, Italy, **9** INFN, Sezione di Roma 1, Rome, Italy

\* These authors contributed equally to this work.

\* [andrea.demartino@roma1.infn.it](mailto:andrea.demartino@roma1.infn.it)



**OPEN ACCESS**

**Citation:** Mori M, Hwa T, Martin OC, De Martino A, Marinari E (2016) Constrained Allocation Flux Balance Analysis. *PLoS Comput Biol* 12(6): e1004913. doi:10.1371/journal.pcbi.1004913

**Editor:** Kiran Raosaheb Patil, EMBL-Heidelberg, GERMANY

**Received:** October 5, 2015

**Accepted:** April 11, 2016

**Published:** June 29, 2016

**Copyright:** © 2016 Mori et al. 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 paper and its Supporting Information files.

**Funding:** ADM and EM are supported by the DREAM Seed Project of the Italian Institute of Technology (IIT <https://www.iit.it/>), by the joint IIT/Sapienza Lab “Nanomedicine” (<http://lns.iit.it/>), by the Marie Curie Action ITN NETADIS (FP7/Grant 290038 <http://netadis.eu/>), and by the PRIN project “Statistical mechanics of disordered complex systems” of the Italian Ministry of University and Research (<http://www.istruzione.it/>). TH is supported by grant #330378 from the Simons Foundation (<http://www.simonsfoundation.org/>). OM is supported by the Investissement d’Avenir Bio-informatique program

## Abstract

New experimental results on bacterial growth inspire a novel top-down approach to study cell metabolism, combining mass balance and proteomic constraints to extend and complement Flux Balance Analysis. We introduce here Constrained Allocation Flux Balance Analysis, CAFBA, in which the biosynthetic costs associated to growth are accounted for in an effective way through a single additional genome-wide constraint. Its roots lie in the experimentally observed pattern of proteome allocation for metabolic functions, allowing to bridge regulation and metabolism in a transparent way under the principle of growth-rate maximization. We provide a simple method to solve CAFBA efficiently and propose an “ensemble averaging” procedure to account for unknown protein costs. Applying this approach to modeling *E. coli* metabolism, we find that, as the growth rate increases, CAFBA solutions cross over from respiratory, growth-yield maximizing states (preferred at slow growth) to fermentative states with carbon overflow (preferred at fast growth). In addition, CAFBA allows for quantitatively accurate predictions on the rate of acetate excretion and growth yield based on only 3 parameters determined by empirical growth laws.

## Author Summary

The intracellular protein levels of exponentially growing bacteria are known to vary strongly with growth conditions, as described by quantitative “growth laws”. This work introduces a computational genome-scale framework (Constrained Allocation Flux Balance Analysis, CAFBA) which incorporates growth laws into canonical Flux Balance Analysis. Upon introducing 3 parameters based on established growth laws for *E. coli*, CAFBA accurately reproduces empirical results on the growth-rate dependent rate of carbon overflow and growth yield, and generates testable predictions about cellular energetic strategies and protein expression levels. CAFBA therefore provides a simple, quantitative approach

under project RESET (ANR-11-BINF-0005, <http://www.agence-nationale-recherche.fr>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

to balancing the trade-off between growth and its associated biosynthetic costs at genome-scale, without the burden of tuning many inaccessible parameters.

## Introduction

The coupling between the physiology of cell growth and cellular composition has been actively investigated since the 1940s. In exponentially growing bacteria, whose growth state is conveniently associated to a single parameter, namely their growth rate, such interdependence is best expressed in a quantitative way by the bacterial ‘growth laws’ that directly relate the protein, DNA and RNA content of a cell to the growth rate. Many such laws have been experimentally characterized [1–4] and many more are currently being probed at increasingly high resolution [5, 6]. The emerging scenario suggests that proteome organization in bacteria is actively regulated in response to the growth conditions. Recent experiments have in particular provided validation to the picture according to which, as the growth rate changes, bacteria adjust the relative amounts of ribosome-affiliated, nutrient scavenging and metabolic proteins (enzymes), so as to optimize their growth performance and energy production strategy [6–8]. At present, several phenomenological models explain the origin of different growth laws at a coarse-grained level [5, 7]. In contrast, genome-scale approaches probing such relationships at molecular levels are less developed.

Constraint-based models (CBMs) are powerful *in silico* tools that can be used to examine metabolic networks at genome scale. Starting from a non-equilibrium steady state assumption for metabolic fluxes, CBMs define the space of feasible reaction profiles through simple physico-chemical constraints like mass-balance. Once physiologically or thermodynamically motivated bounds of variability are assigned to fluxes, the solution space is essentially determined by the stoichiometry of the network alone. On the other hand, in genome-scale models stoichiometric constraints usually generate high-dimensional solution spaces in which physiologically relevant flux patterns may be hard to isolate. In many cases, optimal flux patterns can be defined through the maximization of specific objective functions. Flux Balance Analysis (FBA) [9–15] allows for instance to compute optimal flux configurations by means of linear programming (LP), employing biomass production as a standard objective function [16]. This approach is widely used to describe microbial growth in lab conditions.

It is clear that in order to capture the phenomenology of growth laws one needs to go beyond the basic elements of CBMs, and incorporate the costs associated with gene expression and protein synthesis into models of cellular metabolism. Resource Balance Analysis (RBA) [17, 18] and ME-models [19, 20] have taken important steps in this direction. These approaches propose a data-based optimization scheme to predict the growth-maximizing metabolic flux configurations under a variety of constraints, including stoichiometric mass-balance, ‘demand functions’ characterizing how the amounts of cellular components change with the growth rate, and specific prescriptions that relate fluxes to enzyme levels. The resulting schemes are more involved than FBA (resulting in nonlinear optimization problems) and require a large number of parameters. It is therefore important to devise a theoretical framework with the conceptual appeal and computational simplicity of FBA, in particular one that is more resilient to the choice of parameters and in which the interplay between metabolism and regulation is expressed through a more intuitive and transparent framework.

In this work we present a generalized FBA scheme, called Constrained Allocation FBA or CAFBA, in which (optimal) regulation is accounted for effectively through a single additional global constraint on fluxes that encodes for the relative adjustment of proteome sectors at

different growth rates. In a nutshell, the CAFBA-specific constraint describes the tug-of-war in the allocation of cellular resources across ribosomal, transport and biosynthetic proteins that has been observed in experiments. By imposing that the ribosomal share of the proteome behaves in accordance with empirically established growth laws [5, 21, 22], CAFBA is able to reproduce observed behaviors without requiring parameter tuning. In addition, CAFBA generates a variety of testable predictions, including about the usage of metabolic pathways, despite lacking the level of biochemical detail that characterizes ME-models or RBA.

Cellular strategies for energy production are the central focus of CAFBA. It is well known that fast-growing microorganisms tend to avoid using high-yield respiratory pathways to generate ATP even in the presence of oxygen, relying instead on aerobic fermentation [23–28]. The preference for low-yield pathways is manifested in the secretion of fermentation products like acetate for *E. coli* or ethanol for *S. cerevisiae* [23, 25, 26, 29]. This phenomenon, known as ‘overflow metabolism’, is captured by standard FBA schemes at a qualitative level when additional capacity constraints on respiratory pathways [30] or density constraints for soluble [31, 32] or membrane-bound [33] enzymes are included. However, certain quantitative aspects of potential interest for industrial applications, like the rate of metabolic overflow and the growth rate at which it occurs, have so far eluded comprehensive mechanistic models. By effectively modeling the trade-off between growth and its biosynthetic costs, CAFBA naturally produces cellular states with suboptimal growth yields, where carbon overflow is obtained with quantitative accuracy.

This paper focuses on the scenario obtained by CAFBA for carbon-limited growth of *E. coli*. We find in particular that acetate secretion appears in *E. coli* at fast growth rates, whereas yield-maximizing FBA-like solutions dominate at slow growth rates. In spite of the nominal need for a large number of uncharacterized parameters in genome-wide models, CAFBA solutions remarkably depend only on a few global parameters. In particular, overflow metabolism is obtained consistently with quantitative accuracy, while all results are robust against 10-fold changes in the values of the enzymatic efficiency parameters. From a technical viewpoint, CAFBA effectively turns out to be an LP problem even when one accounts for growth-rate dependent biomass composition. This, together with its simple conceptual framework, makes CAFBA a very convenient scheme to analyze the interplay of metabolism and gene expression at genome scale.

## Model

### Proteome sectors

Phenomenological studies of bacterial growth physiology suggest that the bacterial proteome is organized into “sectors” whose mass fractions adjust linearly with the growth rate in response to specific environmental and intracellular changes, including carbon limitation, anabolic limitation and translational inhibition [5, 6, 8]. Proteome organization and optimal growth constitute in essence an intertwined allocation problem, with the cell trying to optimally partition its proteome so as to maximize its growth performance. Based on empirical evidence on *E. coli* growth in carbon-limited media, CAFBA posits a 4-sector partitioning of the proteome in

- ribosome-affiliated proteins (R-sector);
- biosynthetic enzymes (E-sector);
- proteins devoted to carbon intake and transport (C-sector);
- core housekeeping proteins whose expression level is independent of the growth rate (Q-sector).

The corresponding proteome fractions (denoted by  $\phi_X$  for the X-sector) should sum up to 1, i.e.

$$\phi_C + \phi_E + \phi_R + \phi_Q = 1. \quad (1)$$

We shall now provide an explicit characterization of the different terms in the above sum.

**The ribosomal sector.**  $\phi_R$  is experimentally found to be linearly dependent on the growth rate  $\lambda$  when growth is nutrient-limited [3–5], namely

$$\phi_R = \phi_{R,0} + w_R \lambda, \quad (2)$$

where  $\phi_{R,0}$  is a strain-dependent constant representing the extrapolated ribosomal proteome fraction at zero growth rate, and  $w_R$  is a strain-independent constant related to the ribosome's translational efficiency [5, 6]. Phenomenologically,  $w_R$  describes the proteome fraction allocated to ribosomal proteins per unit of growth rate. At the molecular level, the linear relation (2) is enforced by a regulatory mechanism involving the alarmone ppGpp [34–37]. When focusing on carbon-limited growth, one can set  $w_R$  equal to the empirical value  $w_{R,0} \simeq 0.169$  h [5]. The effects of translational inhibition can instead be studied by increasing  $w_R$  from the value  $w_{R,0}$ , so as to model the increasingly slowed-down translation induced by antibiotics [22]. As will soon become clear, the value of the offset  $\phi_{R,0}$  is immaterial for the formulation of CAFBA.

**The carbon catabolic sector.** We focus on balanced growth in a minimal medium containing a single carbon source (e.g. glucose). Based on experimental findings [6, 8], we assume that  $\phi_C$  depends linearly on the carbon intake flux  $v_C$ , i.e.

$$\phi_C = \phi_{C,0} + w_C v_C, \quad (3)$$

where, by analogy with Eq (2),  $\phi_{C,0}$  is a  $\lambda$ -independent offset and  $w_C$  characterizes the proteome fraction allocated to the C-sector per unit of carbon influx. Recent proteomic studies [8, 38] suggest that the C-sector should include not only the specific transport system taking up the sugar, but also other proteins that are co-expressed in response to carbon limitation through mediation by the pleiotropic regulator cAMP-Crp [6], like intake proteins for other nutrients, motility proteins, etc. Therefore, Eq (3) should be seen as an effective prescription accounting for the fact that several types of proteins intended for nutrient scavenging and intake are co-expressed in carbon limitation. All of these should be expected to contribute to  $\phi_C$ , even if certain proteins, like motility proteins, may not be required for growth in laboratory conditions. The offset  $\phi_{C,0}$  thus represents a basal level of proteins not due to carbon intake only.

In order to better characterize  $w_C$  (i.e. the carbon-intake dependent part of  $\phi_C$ ), we assume that the carbon influx  $v_C$  at a given extracellular sugar level  $[g]$  is described by a Michaelis-Menten kinetics of the form

$$v_C = \frac{V}{M_{DW}} k_{cat,g} [E_g] \frac{[g]}{[g] + K_{M,g}}, \quad (4)$$

where  $[E_g]$  stands for the level of the intake protein(s) specific to  $g$  that are not in  $\phi_{C,0}$ ,  $k_{cat,g}$  and  $K_{M,g}$  are kinetic constants, and  $V$  and  $M_{DW}$  represent the cell volume and dry weight, respectively. (The ratio  $V/M_{DW}$  is introduced so that the flux units are mmol/ $g_{DW}h$ .) Denoting the total protein mass by  $M_{TP}$  and the enzyme's molecular weight by  $\mu_g$ , and letting  $\alpha_g$  be the mass fraction of enzyme  $E_g$  in the C-sector, we can express  $[E_g]$  in terms of the C-sector's proteome fraction  $\phi_C$  as  $[E_g] = (\phi_C - \phi_{C,0}) \alpha_g M_{TP} / (\mu_g V)$ . In turn, Eq (4) can be rewritten as

$$v_C = \kappa_{cat,g} (\phi_C - \phi_{C,0}) \frac{[g]}{[g] + K_{M,g}}, \quad (5)$$

where  $\kappa_{\text{cat},g} \equiv \alpha_g(k_{\text{cat},g}/\mu_g)(M_{TP}/M_{DW})$ . The factor  $M_{TP}/M_{DW} \simeq 60\%$  is roughly constant for a wide range of growth rates [39, 40],  $k_{\text{cat},g}/\mu_g$  is instead an enzyme-specific property, while the proportion  $\alpha_g$  is determined genetically by the expression level of the enzyme  $E_g$  relative to those of the other co-expressed C-sector proteins. Comparing Eq (5) to Eq (3), one sees that  $w_C$  can be represented as

$$w_C = w_{C,0} \left( 1 + \frac{K_{M,g}}{[g]} \right), \quad (6)$$

with  $w_{C,0} \equiv 1/\kappa_{\text{cat},g}$ .

The above analysis suggests that  $w_C$  can be conveniently used to control the carbon influx: it takes on a sugar-specific value  $w_{C,0}$  at saturating sugar concentrations (i.e. for  $[g] \gg K_{M,g}$ ) and the effect of reducing extracellular sugar levels can be modeled by simply increasing its value. Hence, as a proxy of varying the abundance of the carbon source, we will simply dial  $w_C$ . The importance of using  $w_C$  as control parameter, as opposed to varying the maximum nutrient intake capacity, is discussed in Note B in S1 Text. Note that the maximal growth rate achievable in the medium we consider (referred to as  $\lambda_{\text{max}}$  below, and obtained for  $w_C = 0$  or, equivalently,  $w_{C,0} = 0$ ) is experimentally determined by the extrapolated growth rate at which C-sector protein expression vanishes [6, 41].

**The biosynthesis sector.** The flux through enzymatic reactions involved in biosynthesis (E-sector) can be generally written in the form

$$v_i = \frac{V}{M_{DW}} k_{\text{cat},i} [E_i] f_i([s_i], [p_i]), \quad (7)$$

where  $[E_i]$  denotes the concentration of enzyme  $i$  and we considered explicitly an additional dependence on the concentrations of the substrates ( $[s_i]$ ) and products ( $[p_i]$ ) through the function  $f_i$ . For an elementary irreversible reaction with a single substrate and a single product,  $f_i$  is a Michaelis-Menten function of  $[s_i]$ , while for reactions close to thermodynamic equilibrium  $f_i \simeq [s_i] - [p_i]/K_{\text{eq}}$  [42]. In full analogy with the previous case, we can express  $[E_i]$  in terms of the proteome fraction  $\phi_i$  of enzyme  $E_i$  as  $[E_i] = \phi_i M_{TP}/(\mu_i V)$ . Defining  $\kappa_{\text{cat}} \equiv (k_{\text{cat},i}/\mu_i)(M_{TP}/M_{DW})$ , we then have

$$v_i = \kappa_{\text{cat}} \phi_i f_i([s_i], [p_i]). \quad (8)$$

Motivated by the observed linear dependence between enzyme abundance and growth rate in carbon-limited growth [6, 8] and assuming the generic linear dependence between biosynthetic flux and growth rate, we set

$$\phi_i = \phi_{i,0} + w_i |v_i|, \quad (9)$$

with a fixed offset  $\phi_{i,0}$ . The “weight”  $w_i$  represents the proteome fraction to be invested in enzyme  $E_i$  per unit flux of reaction  $i$ . The absolute value instead reflects the fact that, for a reversible process, a protein cost has to be faced independently of the net direction. Note that, in principle, the values of the weights  $w_i$  can be determined experimentally by fitting, for each reaction, proteomic and flux measurements at different growth rates to Eq (9).

The linear relation (9) can be directly obtained from Eq (8) assuming that reaction  $i$  is irreversible and that the enzyme  $E_i$  is operating in the saturated regime. In this case,  $\phi_{i,0} = 0$  and  $w_i = 1/\kappa_{\text{cat},i}$ . However, such reactions would be incapable of balancing flux in the event of transient changes, leading to the accumulation of intermediate metabolites. Therefore, most intracellular reactions in physiological conditions should not be expected to operate in the saturated regime. Reactions carrying a flux proportional to the substrate level (as in flux sensors [5, 43]

and charged tRNAs [35]) can again be described by Eq (9), albeit with an offset  $\phi_{i,0}$ ; see Note A in S1 Text. In this view, the offset  $\phi_{i,0}$  provides a mathematically simple way to capture the fact that, at slow growth, the flux approaches zero due to adjustments in metabolite pools while enzyme levels remain finite. As for the other sectors, the values of these offsets play no role in CAFBA (see below). Summing up the contributions of each reaction, the proteome fraction of the E-sector  $\phi_E \equiv \sum_i \phi_i$  can be written as

$$\phi_E = \phi_{E,0} + \sum_i w_i |v_i|, \tag{10}$$

where the sum runs over all enzyme-catalyzed reactions and  $\phi_{E,0} \equiv \sum_i \phi_{i,0}$  contributes to a core,  $\lambda$ -independent proteome fraction for baseline expression levels [6, 8].

### Proteome-wide constraint

Putting the different terms together, the sum rule Eq (1) for proteome fractions can be recast as

$$w_C v_C + \sum_i w_i |v_i| + w_R \lambda = \phi_{\max}, \tag{11}$$

where  $\phi_{\max} = 1 - \phi_Q - \phi_{C,0} - \phi_{E,0} - \phi_{R,0}$  denotes the proteome fraction accessible to growth-rate dependent components of the protein sectors, which was estimated to be of the order of 50% for *E. coli* [5]. The linear constraint (11) encodes for the tug-of-war that ultimately determines optimal growth and proteome allocation, as depicted in Fig 1A: as  $\lambda$  increases, so does the proteome fraction of the R-sector, and the E- and C- sectors will concomitantly have to adjust their shares so as to satisfy Eq (11), forcing in turn a remodeling of the underlying flux and nutrient intake patterns.

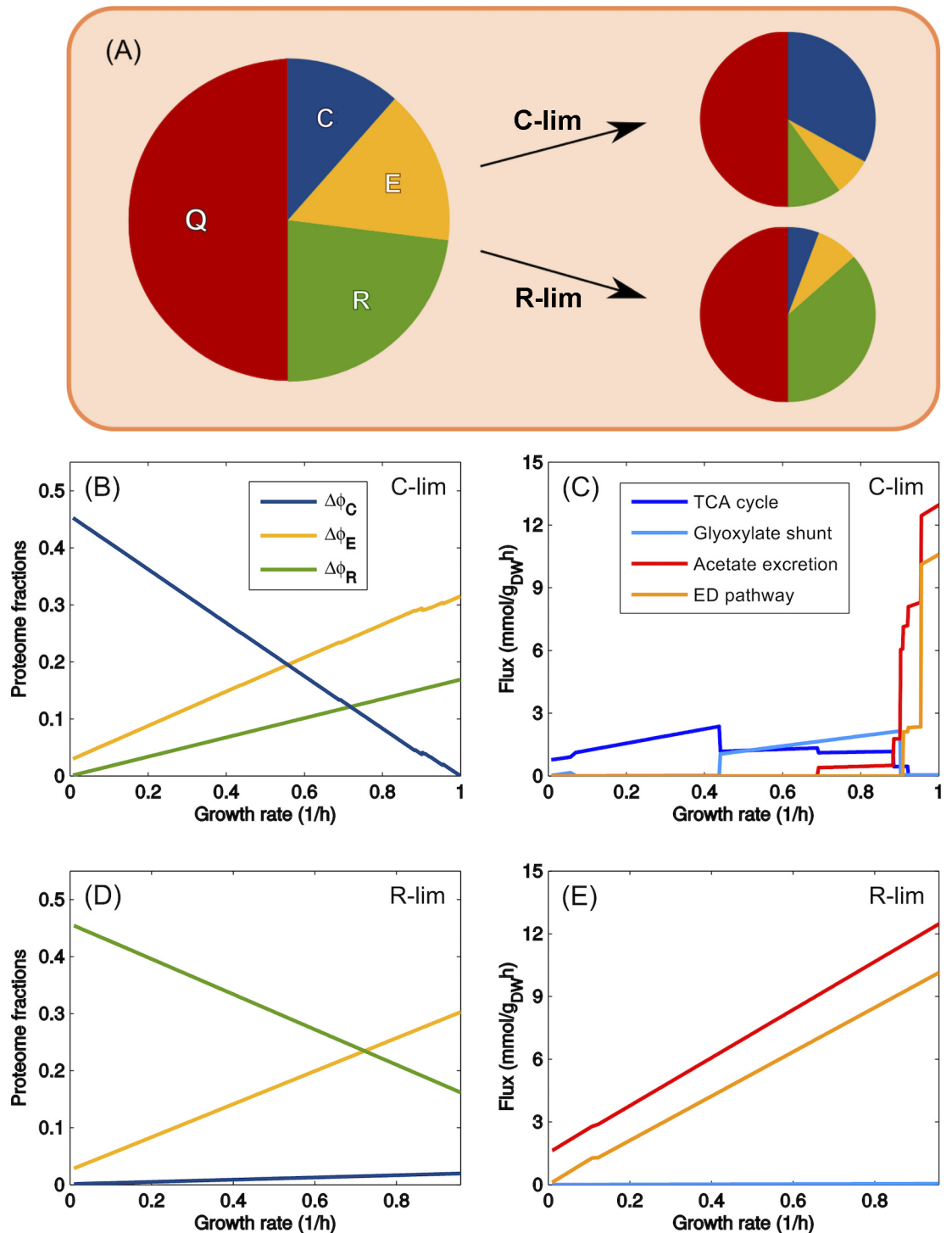
Formally, the proteome allocation constraint (11) resembles the molecular crowding constraint defined in [31, 32], which essentially enforces a global upper bound on fluxes due to finite solvent capacity and was also adopted in RBA [17, 18]. However, the intracellular density is empirically known to be (roughly) constant across different growth conditions [44], suggesting that cells can adapt their volume to accommodate additional metabolites and macromolecules when necessary. In this respect, a hard constraint on solvent capacity is not fully justified. The CAFBA constraint (1) is instead derived from the normalization of protein fractions, due ultimately to the limited translational capacity of the ribosomes [5]. Note that the growth rate  $\lambda$  is explicitly involved in Eq (11).

### Constrained allocation FBA

Summing up, CAFBA is defined by the following optimization problem:

$$\begin{aligned} \max_{\mathbf{v}} \lambda \quad \text{subject to} \quad & \text{(i)} \quad \sum_{\mu} S_{\mu i} v_{\mu} = 0 \quad \forall \mu \\ & \text{(ii)} \quad \ell_i \leq v_i \leq u_i \quad \forall i \\ & \text{(iii)} \quad w_C v_C + \sum_i w_i |v_i| + w_R \lambda = \phi_{\max} \end{aligned} \tag{12}$$

where  $S_{\mu i}$  stands for elements of the metabolic network's stoichiometric matrix (with  $\mu$  indexing metabolites and  $i$  indexing reactions),  $\ell_i$  and  $u_i$  denote lower and upper bounds for the flux  $v_i$ , while the value of  $\lambda$  is defined by the flux of the biomass reaction [16].



**Fig 1. CAFBA solutions for *E. coli* in the homogeneous case for carbon limitation and translational limitation.** (A) Proteome organization in CAFBA: R-sector of ribosome-affiliated proteins (growth rate dependent), E-sector of “enzymes” (flux-dependent), C-sector of catabolic proteins (dependent on the carbon influx), and a fixed Q-sector of “housekeeping” proteins. The fractions in these four sectors sum up to one. C-, E- and R- sectors adjust their size depending on the environmental conditions, while the Q-sector accounts for roughly 50% of the proteome. We model the three growth-dependent sectors as a constant plus a variable part, i.e.  $\phi_X = \phi_{X,0} + \Delta\phi_X$  with  $X \in \{C, E, R\}$ .  $\Delta\phi_C = w_C v_C$  is proportional to the carbon intake flux;  $\Delta\phi_E = \sum_j w_j |v_j|$  is a weighted sum of non-catabolic fluxes;  $\Delta\phi_R = w_R \lambda$  is proportional to the growth rate  $\lambda$ . (B) Growth rate-dependent parts of proteome sectors plotted versus  $\lambda$  in carbon limitation (C-lim). As

the external glucose concentration is reduced, more catabolic proteins are needed per unit of carbon influx. The cell allocates a larger share to C-proteins, while reducing the E- and R-sector shares. (C) CAFBA fluxes as a function of  $\lambda$ , obtained by varying the degree of carbon (glucose) limitation (C-lim). A transition from fermentation to respiration appears when growth rate is in the range 0.7–0.9/h. The Embden-Doudoroff pathway and the glyoxylate shunt are both operated at high growth rates. (D) The  $\lambda$ -dependent parts of the proteome sectors plotted against growth rate in translational limitation (R-lim). This is obtained by keeping  $w_C$  constant while increasing  $w_R$ , thereby simulating increasing levels of translation-inhibiting antibiotics. The cell allocates more proteins to the ribosomal sector while reducing the proteome share devoted to carbon metabolism and biosynthesis. (E) CAFBA fluxes as a function of  $\lambda$  obtained in R-limitation for increasing values of  $w_R$ , at constant  $w_C$ . Acetate is secreted at low growth rates if the extracellular carbon level is large enough. Fluxes through the TCA cycle, the glyoxylate shunt and the Entner-Doudoroff pathway are represented by  $\alpha$ KG dehydrogenase, malate synthase and 6-phosphogluconate dehydratase fluxes, respectively. In panels B and C, corresponding to C-limitation,  $w_R$  was set to  $w_{R,0} = 0.169$  h, while R-limitation (panels D and E) was obtained using  $w_C = 1.4 \times 10^{-3}$  gh/mmol, corresponding to a carbon source with high nutritional capacity. In both cases we set all weights in the E-sector to  $w_E = 8.3 \times 10^{-4}$  gh/mmol and  $\phi_{\max} = 48.4\%$ .

doi:10.1371/journal.pcbi.1004913.g001

## Results

We have studied CAFBA solutions for the *E. coli* iJR904 GSM/GPR reconstruction [45] assuming growth limited by a single carbon source (glucose). (See Note C in [S1 Text](#) for details about CAFBA in different growth-limiting conditions.) We started from the case of  $\lambda$ -independent biomass composition, in which CAFBA can be solved exactly by LP (see [Materials and Methods](#)), and then considered the more general case of growth-rate dependent biomass. Throughout this study, we set the parameters  $w_R$  and  $\phi_{\max}$  to the values  $w_{R,0} = 0.169$  h and  $\phi_{\max} = 48.4\%$  found empirically for *E. coli* K-12 MG1655 [5]. A nearly identical  $w_{R,0}$  and a slightly smaller  $\phi_{\max}$  have been reported in [6, 29] for *E. coli* K-12 strain (NCM3722). (The values of  $w_{R,0}$  and  $\phi_{\max}$  need not be fine-tuned. In fact, variations in one of these parameters can be compensated by rescaling the weights of reactions in the E-sector. The dependence of the fluxes on  $w_{R,0}$  and  $\phi_{\max}$  is described in detail in Note D in [S1 Text](#).) Carbon limitation is enforced by increasing the value of  $w_C$  from its minimum  $w_{C,0}$ , corresponding to saturating glucose concentrations. (Likewise, translational limitation can be studied by increasing the value of  $w_R$  from its minimum,  $w_{R,0}$  [22].) For each choice of  $w_C$  and  $w_R$  (and of the set of  $w_i$ 's), we solve [Eq \(12\)](#) for the fluxes  $v_i$  that maximize the growth rate  $\lambda$ .

As said above, the weights  $w_i$  could in principle be determined by combining proteomic studies [8, 37, 46] with direct flux measurements taken in the appropriate growth conditions. However, the coverage by mass spectroscopy is still limited, and the accuracy of protein abundance is often no better than 2-fold. Much better estimates of protein abundances have been obtained recently using ribosome profiling [47], which also provides a near complete coverage of *E. coli* proteins. This method is however much less versatile compared to proteomics and only a few conditions have been probed so far. Given the lack of reliable empirical estimates, we have focused on two limiting situations allowing us to characterize intrinsic properties of CAFBA for *E. coli* in the most transparent manner.

- The homogeneous case: here,  $w_i$ 's are uniformly set to the same value, denoted by  $w_E$ , for each reaction  $i$ .  $w_E$  is chosen so that the fastest growth rate achievable  $\lambda_{\max}$ , corresponding to  $w_C = 0$ , matches the corresponding empirical value.
- The heterogeneous case: here,  $w_i$ 's are taken at random, to reflect one's lack of knowledge of their specific value. More precisely,  $w_i$  is drawn from a prescribed probability distribution independently for each reaction  $i$ . The mean value  $\langle w \rangle$  of the distribution essentially plays the role played by  $w_E$  in the homogeneous case, in that it sets the average growth rate obtained for  $w_C = 0$  to  $\lambda_{\max}$ . Clearly, the quantitative details of CAFBA solutions will depend on the specific values taken by the  $w_i$ 's. However, as there is no reason to concentrate on a

single set of weights, we will focus our analysis on two aspects, namely (i) the “average” behaviour obtained by averaging solutions over different realizations of the  $w_i$ 's, and (ii) the fluctuations of solutions around this average.

## Homogeneous weights and patterns of flux

We set  $w_E$  so that the extrapolated maximal growth rate in unlimited carbon supply, corresponding to  $w_C = 0$ , is close to the value 1.1–1.2/h found in [6, 41]. In the case of glucose as the sole carbon source, as well as for a number of other glycolytic carbon sources (see Table A in [S1 Text](#)), the value  $w_E = 8.3 \times 10^{-4}$  gh/mmol turned out to yield  $\lambda_{\max} = 1/h$ . (Slightly larger growth rates are obtained with phosphorylated carbon sources due to the fact that the extra energy carried by these carbon sources allow for a reduced flux in the E-sector.) To capture the effects of changing the glucose level, we simply increased the value of  $w_C$  from zero. For the sakes of completeness, the values of  $w_C$  leading to empirically observed growth rates for *E. coli* growth on different carbon sources are reported in Table B in [S1 Text](#).

[Fig 1](#) reports results obtained for growth on glucose with this choice of parameters, while results for growth on other carbon sources are shown in Fig A in [S1 Text](#). One sees in [Fig 1B](#) that the growth-dependent fraction of C-proteins ( $\Delta\phi_C$ , blue line) increases almost linearly with decreasing  $\lambda$  as the carbon concentration is limited, in line with the experimentally observed expressions of catabolic proteins [6, 8] and PTS activity [20]. Both the proteome fractions of the E- and R-sectors ( $\Delta\phi_E$  and  $\Delta\phi_R$ , yellow and green lines, respectively) instead decrease linearly with growth rate. CAFBA therefore confirms the findings from a coarse grained model of proteome allocation [7]: in the optimal state, the cell invests more and more of its proteomic resources in intake systems as nutrient becomes limiting, while translational machinery and biosynthetic pathways are favored at high growth rates.

[Fig 1C](#) displays the main fluxes of central carbon metabolism. The rate of acetate secretion and the flux through the Entner-Doudoroff pathway (red and orange colors, respectively) both drop fast as the growth rate decreases. Respiration, represented by the flux through the TCA cycle (blue color) is the predominant energy-producing pathway at small growth rates, while at high growth rates fermentation is preferred and acetate is secreted. Note that the acetate onset point is within 10% of the one observed experimentally for NCM3722 [29] roughly independently of the specific carbon source (see Fig A in [S1 Text](#))—a remarkable result given the simplicity of the homogeneity assumption for the  $w_i$ 's.

Translational limitation [5] is modeled by increasing  $w_R$  from the value of  $w_{R,0}$  while keeping all other parameters fixed, including  $w_C$ . In this case (see [Fig 1D](#)), the ribosomal proteome fraction ( $\Delta\phi_R$ ) increases as translation is increasingly inhibited, while the other growth-dependent sectors ( $\Delta\phi_C$  and  $\Delta\phi_E$ ) shrink almost linearly. Acetate secretion extends to the slowest growth rates in accordance with experimental findings [29], while the respiratory flux (see [Fig 1E](#)) is negligible.

It is interesting to compare CAFBA results with the phenomenological proteome allocation model introduced in [5], which describes how proteome is allocated in different environments. There, the growth rate was predicted to be a Michaelis-Menten function of the “nutritional capacity”  $\kappa_n$  and “translation capacity”  $\kappa_t$ , independent phenomenological parameters that can be estimated from empirical growth laws. CAFBA recovers this result within a genome-scale model, with  $1/w_C$  playing the role of  $\kappa_n$  and  $1/w_R$  acting as  $\kappa_t$  (see Fig B in [S1 Text](#) and Note D in [S1 Text](#) for a detailed discussion).

It transpires from [Fig 1C](#) and Fig A in [S1 Text](#) that the optimal flux configurations in carbon limitation vary discontinuously with the growth rate. This is due to the fact that the control

parameter is not a flux (as in standard FBA), but, rather, the weight of the C-sector  $w_C$ , which, as discussed above, is a proxy for either the external carbon concentration or the amount of glucose intake proteins [6]. Even though  $w_C$  is varied continuously, growth-rate maximization can induce large rearrangements of the active pathways in response to small changes of the control parameter. This behavior is ultimately a mathematical feature due to the way in which the optimal solution in constraint-based models like CAFBA changes as one modifies  $w_C$ .

## Heterogeneous weights and patterns of average flux

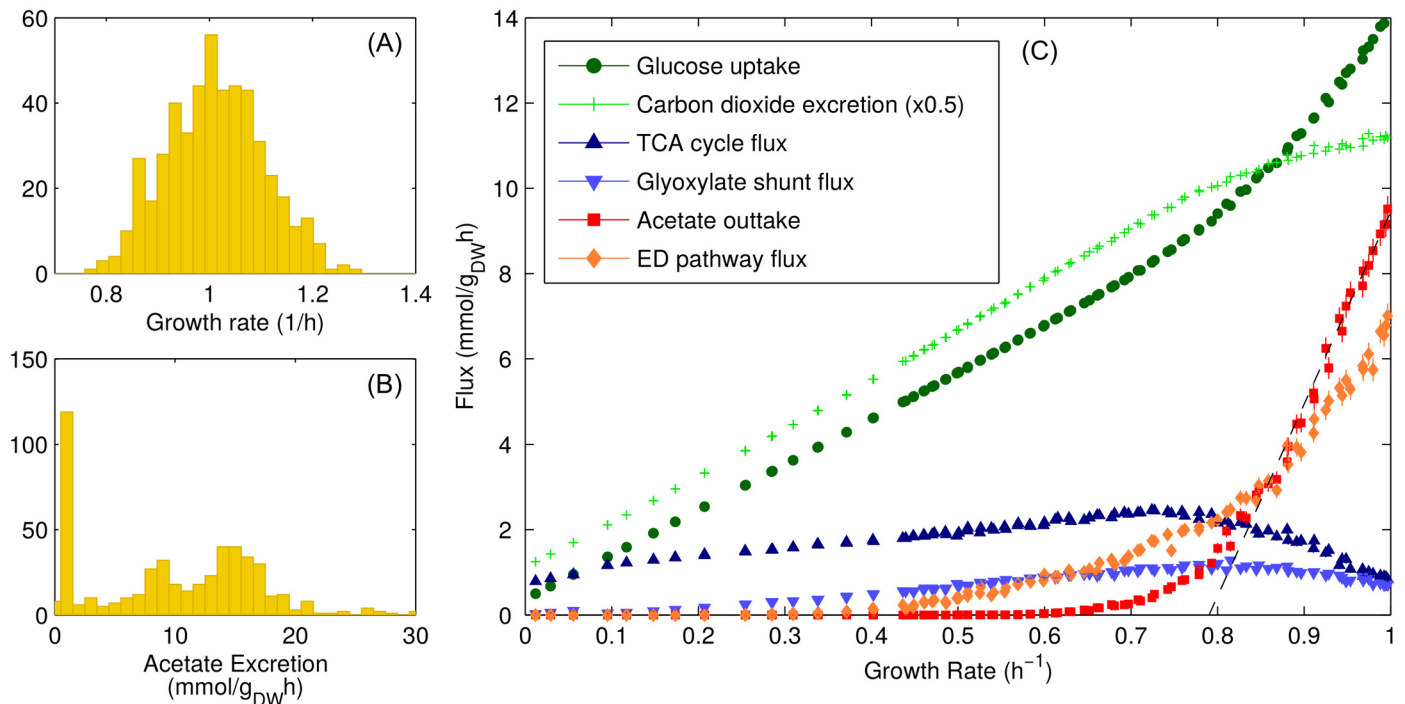
For the heterogeneous case, for each value of  $w_C$  we generated 1000 models, each with a random set of weights  $w_i$  independently drawn from the same probability density

$$p(w) \propto 1/w \quad , \quad w_{\min} \leq w \leq w_{\max} \quad , \quad (13)$$

which corresponds to a uniform density for the logarithm of  $w$ .  $p(w)$  is fully determined by its average  $\langle w \rangle$  and width  $\delta \equiv \log_{10}(w_{\max}/w_{\min})$ . We set  $\langle w \rangle$  so that the average value of the maximum achievable growth rate  $\lambda_{\max}$  (obtained for  $w_{C,0} = 0$ ) equals 1/h. This fixes  $\langle w \rangle = 8.8 \times 10^{-4}$  gh/mmol, a value that is remarkably close to  $w_E = 8.3 \times 10^{-4}$  gh/mmol as determined in the homogeneous case.  $\delta$  was instead fixed to 1, implying that the weights are assumed to span one order of magnitude (results obtained for different values of  $\delta$  are discussed in Note E in [S1 Text](#)).

Each set of weights  $\{w_i\}$  leads to a corresponding optimal flux pattern, growth rate, acetate secretion rate, etc. The distribution of growth rates obtained from many realizations of the weights is shown in [Fig 2A](#). Note that in spite of the 10-fold variability of the weights, the growth rate remains within a modest range of  $\pm 20\%$ . The distribution for acetate secretion rates is instead conveyed in [Fig 2B](#): it is rather heterogeneous, with a marked peak for phenotypes with very low acetate secretion. While individual fluxes can fluctuate significantly across solutions, average fluxes are strikingly well-behaved. This phenomenon is illustrated in [Fig 2C](#) where we show a set of average fluxes plotted against the average growth rate. The average acetate secretion rate (red symbols) has an approximate linear dependence on the growth rate starting from  $\lambda_{ac} \simeq 0.79$ /h. Average fluxes through TCA and the glyoxylate shunt (blue up- and down- triangles) reach their respective maxima close to  $\lambda_{ac}$ . Notice that a smooth transition from a predominantly fermentative to a predominantly respiratory mode of energy production clearly emerges, in full agreement with empirical evidence. It is especially remarkable that this scenario does not seem to depend on the specific choice of  $p(w)$ . For instance, a log-normal distribution gives qualitatively similar results (see Fig C in [S1 Text](#)).

Despite the crude approximations, CAFBA solutions appear to reproduce experimental findings with surprising accuracy. [Fig 3A](#) shows how the average acetate excretion rate compares with data from different experiments [29, 48–51]. Secretion rates from experiments using the MG1655 strain are consistent among each other (open triangles), as are results obtained with the NCM3722 and ML308 strains (open circles). CAFBA predictions are shown as solid circles for the two classes of strains. Data obtained with NCM3722 and ML308 were compared with CAFBA solutions obtained by setting  $\lambda_{\max} = 1$ /h and hence  $\langle w \rangle = 8.8 \times 10^{-4}$  gh/mmol. Instead, based on experimental evidence suggesting that MG1655 cells grow about 1/3 slower than the other two strains (see Fig D in [S1 Text](#)), for MG1655 we set  $\lambda_{\max} = 0.67$ /h, leading to  $\langle w \rangle = 1.55 \times 10^{-3}$  gh/mmol. With this choice, CAFBA quantitatively reproduces the growth-rate dependence of acetate secretion. Growth yields, instead, are less consistent across different experiments and/or strains, see [Fig 3B](#). Without any further parameter tuning, CAFBA solutions capture the growth yields for NCM3722 and MG1655 at a quantitative level,

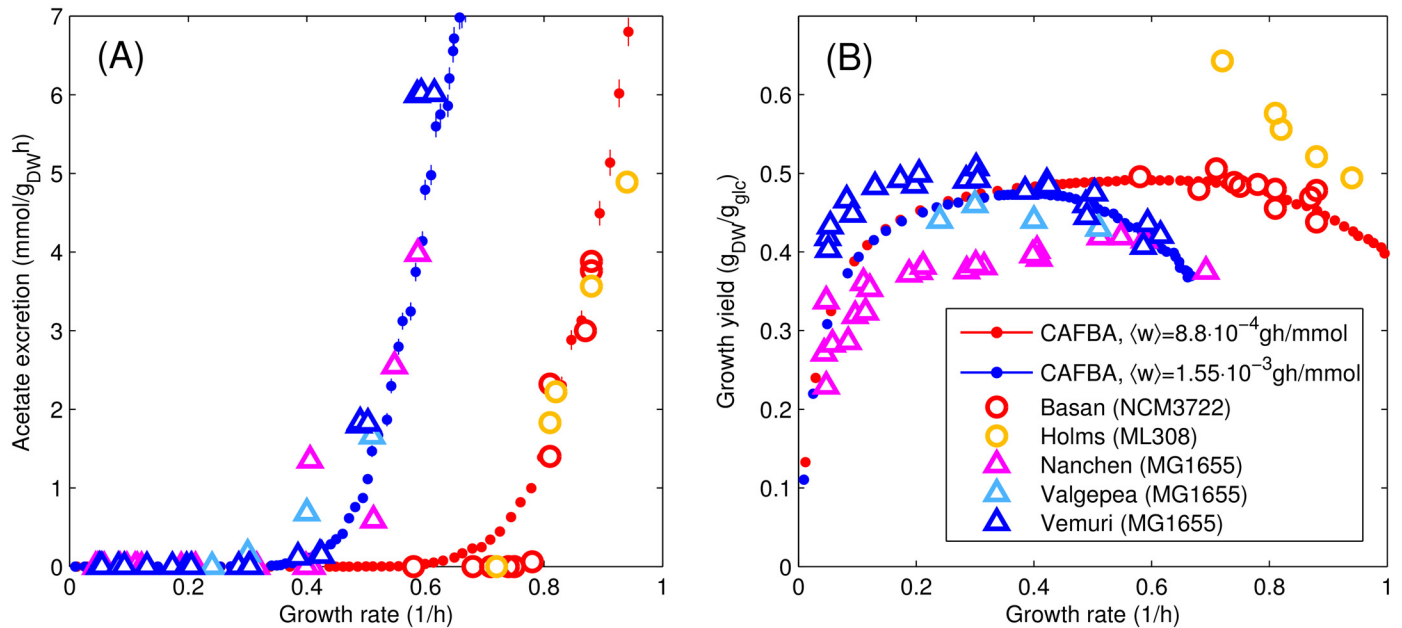


**Fig 2. CAFBA solutions for *E. coli* in the heterogeneous case in carbon limitation.** Fluxes in glucose minimal medium computed at fixed  $w_C \geq 0$  for different realizations of the E-sector weights, using  $\langle w \rangle = 8.8 \times 10^{-4}$  gh/mmol and  $w_{max}/w_{min} = 10$  ( $\delta = 1$ ). (A) Histogram of the growth rates obtained from 1000 CAFBA solutions obtained using different randomly drawn weights for reactions in the E-sector and  $w_C = 0$ .  $\lambda$  peaks around  $\lambda_{max} = 1/h$ . (B) Histogram of the acetate secretion rates in the same conditions. Two classes of solutions are clearly visible, with roughly 25% of states excreting less than 0.5 mmol/g<sub>DW</sub>h of acetate. The average secretion flux is close to 10 mmol/g<sub>DW</sub>h. (C) Average fluxes (glucose intake, carbon excretion, TCA, glyoxylate shunt, acetate excretion and ED pathway) versus the average growth rate. Each point represents the average of 1000 CAFBA solutions obtained with the same  $w_C$  and different realizations of the weights of reactions in the E-sector. Both x and y error bars are shown. Different points are obtained by using different  $w_C$  values. Acetate secretion is approximately linear at large values of  $\lambda$ . A line  $v_{ac} = s \times (\lambda - \lambda_{ac})$  with  $s = 45$  mmol/g<sub>DW</sub> and  $\lambda_{ac} = 0.79/h$  is shown for comparison.

doi:10.1371/journal.pcbi.1004913.g002

although they fail for ML308. It should be noted that the differences in yield among experiments done on the same strain (MG1655) suggest that other factors beyond the scope of this simple model might be at play, such as differences in growth conditions and/or maintenance requirements.

We have also analyzed how the flux patterns of various intracellular pathways are modulated by the growth rate. Results for the central carbon pathways are summarized in Fig E in [S1 Text](#), with the fluxes through the TCA cycle and glyoxylate shunt consistently increasing in proportion as glucose is limited. A similar behavior has been observed in measured expression levels of the corresponding enzymes [8, 51]. Glycolytic fluxes are heterogeneously regulated, due to the interplay between the EMP pathway, the ED pathway and the switch between glyoxylate shunt and the phosphoenolpyruvate carboxylase (PPC) reaction. The redox balance of the cell appears to be affected, as described in Fig F in [S1 Text](#). Indeed we find that NADP transhydrogenase switches on at high growth rates, oxidizing NADH and reducing NADP<sup>+</sup>, in agreement with the different roles of the two transhydrogenases, UdhA and PntA, as quantified by transcription data [52]. Moreover we observe a switch between two separate ubiquinol oxidase reactions, characterized by different abilities to generate proton-motive force, in agreement with studies focused on the crowding of the cell's membrane [33].



**Fig 3. Comparison between CAFBA predictions and experimental data.** (A) Acetate secretion rates for *E. coli* cells grown in minimal glucose media, with data obtained from different datasets [29, 48–51]. Full dots represent average CAFBA solutions (heterogeneous case) obtained with different degrees of carbon limitation (different  $w_C$ , averages over 500 solutions). Results were obtained with two different values for the average E-sector weight, namely  $\langle w \rangle = 8.8 \times 10^{-4}$  gh/mmol (red) and  $\langle w \rangle = 1.55 \times 10^{-3}$  gh/mmol (blue). These choices reproduce the acetate secretion rates of NCM3722 and ML308 (open circles) and MG1655 (open triangles) strains, respectively. (B) Same as panel (A), but for the growth yield. CAFBA predictions (red and blue filled circles) are obtained by averaging the ratio of the growth rate to the glucose intake flux, divided by the molecular weight of glucose  $\mu_{glc} = 0.18$  g/mmol. Data points from [29] have been converted using  $1 \text{ mM}/\text{OD}_{600}/\text{h} = 2 \text{ mmol}/\text{g}_{\text{DW}}/\text{h}$ . x- and y-error bars for the average CAFBA solutions are too small to be visible.

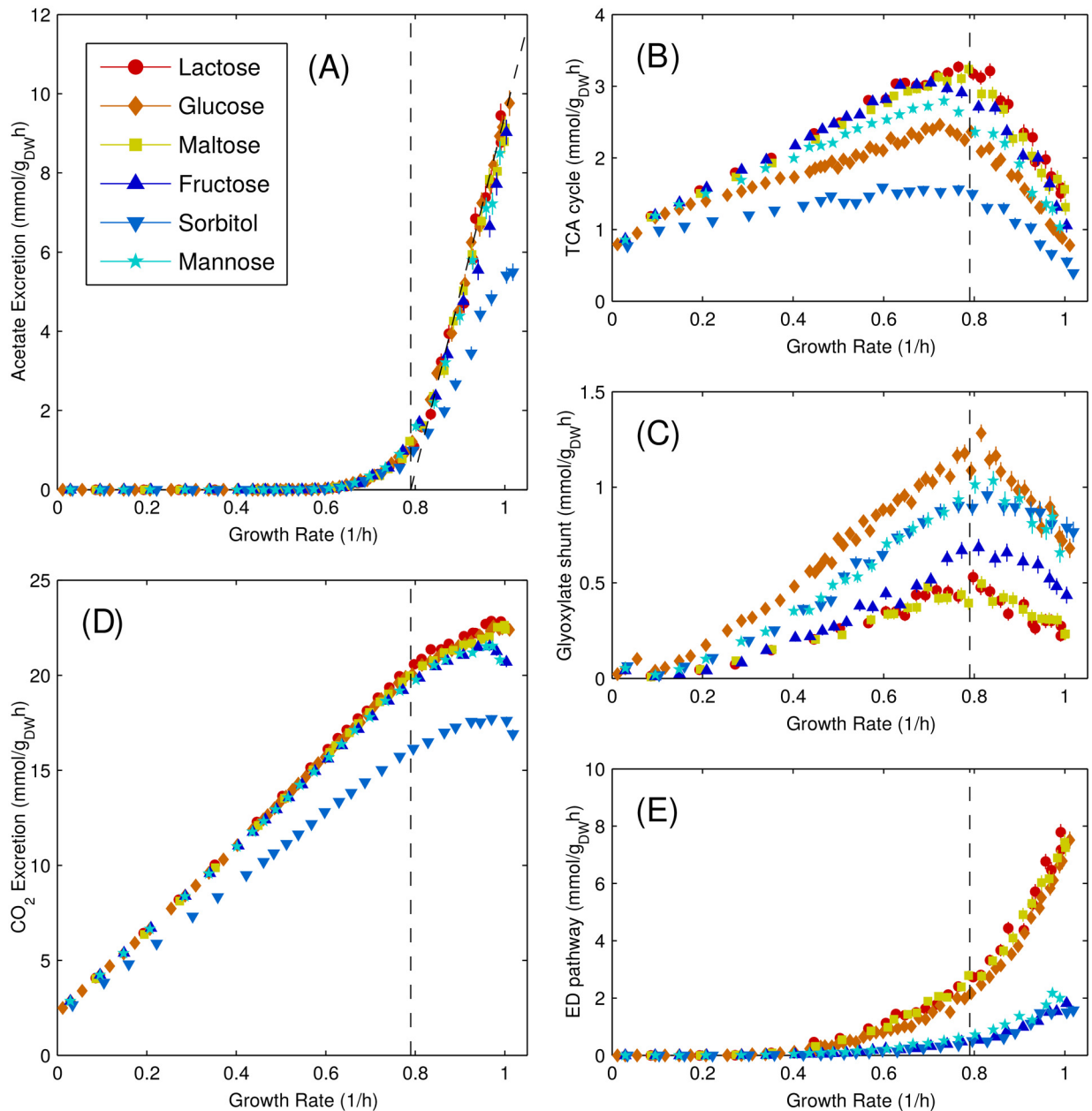
doi:10.1371/journal.pcbi.1004913.g003

### Patterns of average flux for different carbon sources

We further tested CAFBA’s ability to describe *E. coli* growth on carbon sources other than glucose. For illustration purposes, for each carbon source studied we have varied  $w_C$  from zero to high values, so as to produce result in the entire range of growth rates  $0 \leq \lambda \leq \lambda_{\max}$ , even though growth rates measured on individual carbon sources are always smaller than  $\lambda_{\max}$  due to non-zero values of  $w_{C,0}$  (see Table B in S1 Text).

The typical behaviour of CAFBA solutions with different glycolytic carbon sources is remarkably consistent (see Fig 4). For each of the nutrients we tested, as the carbon supply becomes limiting, acetate excretion (Fig 4A) decreases almost linearly with growth rate, extrapolating to zero roughly at  $\lambda_{ac} \simeq 0.79/\text{h}$  (continuous black line). By contrast, fluxes through TCA and glyoxylate shunt (Fig 4B and 4C) rise linearly with decreasing growth rate at fast growth, reaching a maximum close to  $\lambda_{ac}$  before decreasing at slower growth. The secretion rate of  $\text{CO}_2$  (Fig 4D) almost always diminishes as  $\lambda$  is reduced. For  $\lambda < \lambda_{ac}$  the decrease is linear, while it is non-linear for  $\lambda > \lambda_{ac}$ . Altogether, for all carbon sources, results point to two distinct types of behaviors arising, respectively, below and above  $\lambda_{ac}$ .

The Entner-Doudoroff (ED) pathway, an alternative to the Embden-Meyerhoff-Parnass (EMP) pathway, is used in *E. coli* for glucose catabolism at high growth rates [53, 54]. CAFBA solutions reproduce this feature, relying on the ED pathway from medium ( $\lambda \simeq 0.3/\text{h}$ ) to high growth rates as shown in Fig 4E. Interestingly, average fluxes are consistent for lactose, glucose and maltose on the one hand, and for fructose, sorbitol and mannose on the other. The reason is that, in the former group of substrates, the carbon source enters glycolysis as glucose-6P, which can be processed either by upper glycolysis or by the ED and pentose phosphate



**Fig 4. Average CAFBA solutions (heterogeneous case) for six different glycolytic carbon sources.** Each plot shows a different average flux, specifically: (A) Acetate secretion, (B) TCA cycle flux (represented by  $\alpha$ KG dehydrogenase), (C) Glyoxylate shunt flux (malate synthase), (D) CO<sub>2</sub> secretion, (E) ED pathway flux, (6-phosphogluconate dehydratase). Each point represents the average over 500 solutions obtained with the same  $w_C \geq 0$  and  $\langle w \rangle = 8.8 \times 10^{-4}$  gh/mmol. Vertical lines at  $\lambda_{ac} = 0.79$ /h are shown for clarity. In panel (A), acetate secretion can be approximated, for  $\lambda \geq \lambda_{ac}$ , with a straight line  $v_{ac} = s \times (\lambda - \lambda_{ac})$  with  $s = 45$  mmol/g<sub>DW</sub>·h.

doi:10.1371/journal.pcbi.1004913.g004

pathways. In the latter group, instead, carbon is transformed into fructose-6P, which is more conveniently processed into fructose biphosphate. A similar behavior is observed for phosphorylated carbon sources or other substrates of the glycolytic or pentose pathways, see Fig G in [S1 Text](#). The ED pathway, despite having a smaller ATP yield, requires a much smaller number of enzymes than the EMP pathway. Therefore, the use of ED over EMP may be the result of a proteome-saving strategy. Our findings thus agree with the conclusions of [42, 54, 55]. The switch

between the EMP and ED pathways sets in at a growth rate close to 0.3/h, well below  $\lambda_{ac}$ , suggesting that it is independent of acetate secretion. Nonetheless, both features appear in CAFBA in order to cope with increasingly expensive proteins, in agreement with quantitative proteomics data [46].

On the other hand, CAFBA shows that a variety of strategies exist for cells growing on carbon substrates belonging to the lower part of glycolysis or to the TCA cycle, see Fig H in S1 Text. What these strategies share is an increased production of CO<sub>2</sub> at faster growth, and a vanishing activity of the ED pathway. The latter is of course due to the intrinsic glycolytic, as opposed to gluconeogenic, nature of the ED pathway.

### Comparison between CAFBA and FBA solutions

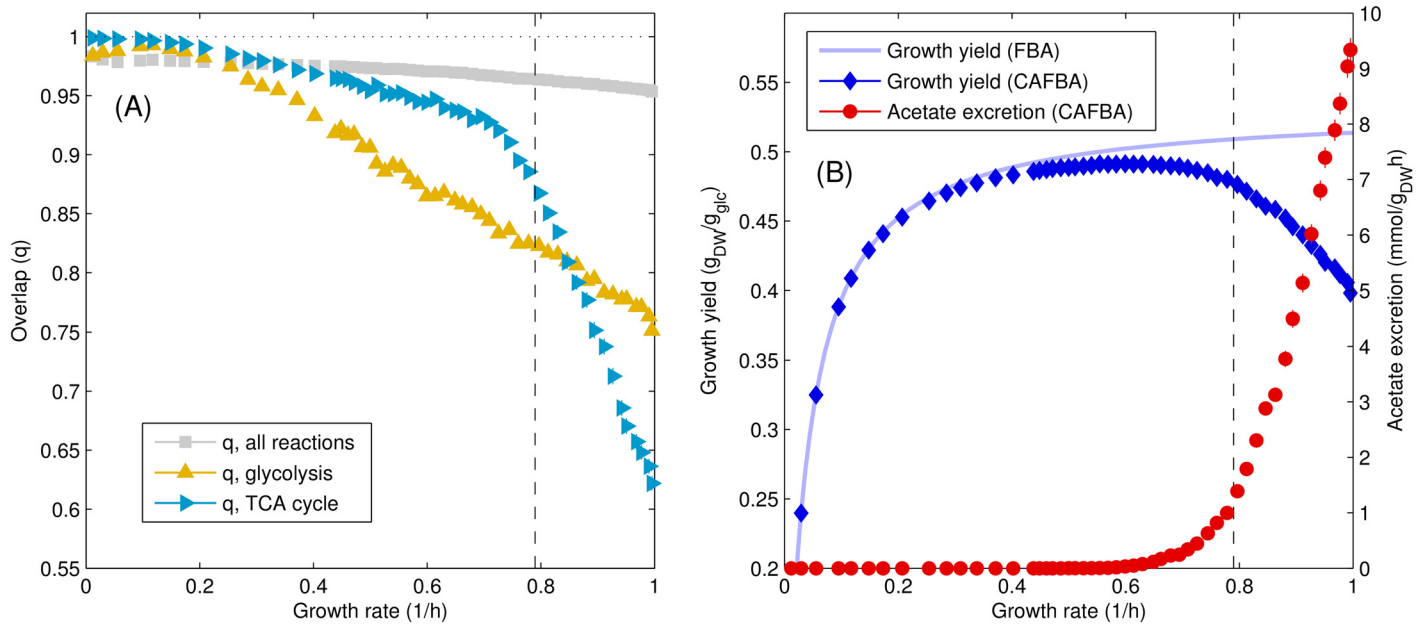
Standard FBA optimizes the growth yield subject to constraining the carbon intake flux. It is useful to compare CAFBA solutions with solutions obtained by FBA at the same growth rate and with glucose as the sole carbon source for both models. To do so, we have first solved Parsimonious Enzyme Usage FBA (pFBA, see [56]) varying the bounds on glucose intake so as to obtain FBA solutions as a function of the growth rate. We shall denote them as  $\mathbf{z}(\lambda) = \{z_i(\lambda)\}$ . CAFBA solutions found upon varying  $w_C$  lead instead to  $w_C$ -dependent mean growth rates  $\bar{\lambda}(w_C)$ . We shall denote CAFBA solutions obtained for a value of  $w_C$  such that  $\bar{\lambda}(w_C) = \lambda$  by  $\mathbf{v}(\lambda) = \{v_i(\lambda)\}$ . We have then computed, for a given set  $\mathcal{R}$  of reactions of interest, the similarity index  $q_{\mathcal{R}}$  called “overlap” and defined as [57]

$$q_{\mathcal{R}}(\lambda) = \frac{1}{N_{\mathcal{R}}} \sum_{i \in \mathcal{R}} \left\langle \frac{2v_i(\lambda)z_i(\lambda)}{v_i(\lambda)^2 + z_i(\lambda)^2} \right\rangle, \quad (14)$$

where the sum is restricted to reactions in  $\mathcal{R}$  and the brackets  $\langle \dots \rangle$  denote an average over 1000 different CAFBA solutions  $\mathbf{v}(\lambda)$ . If in each solution  $v_i = z_i$  for each  $i \in \mathcal{R}$ , then  $q_{\mathcal{R}} = 1$ . Conversely, the more the two flux vectors differ, the smaller  $q$  gets. In particular, if in each solution  $v_i = -z_i$  for each  $i \in \mathcal{R}$ , one finds  $q_{\mathcal{R}} = -1$ . Fig 5A shows the behavior of  $q_{\mathcal{R}}$  versus  $\lambda$  for different choices of  $\mathcal{R}$ . When all reactions are accounted for,  $q$  is generally very large at low growth rates and decreases slowly as  $\lambda$  increases. When focusing on individual pathways, one sees that the overlap for TCA fluxes (cyan) drops abruptly above the acetate onset point  $\lambda_{ac} \simeq 0.79/h$ , where the growth yield of CAFBA solutions starts to reduce significantly compared to that of FBA solutions (Fig 5B, shown in red and blue symbols respectively). The overlap of fluxes in the glycolytic pathway instead diminishes with  $\lambda$  in a more gradual way, corresponding to the smooth increase in the activity of the ED pathway, see Fig 4E. Thus, as the growth rate increases, CAFBA solutions cross over from flux distributions that maximize the growth yield (slow growth) to a regime in which low-yield fermentation, accompanied by carbon overflow and energy spilling, is favored (fast growth).

### Growth rate-dependent biomass composition

In CBMs, the energetic cost of anabolic pathways is accounted for by the stoichiometry of the network. By contrast, the energetic requirements of growth (e.g. protein synthesis) and homeostasis must be included separately as an additional ATP hydrolysis flux  $v_{ATP}$ . In metabolic models, the latter is assumed to be linearly related to the growth rate, i.e.  $v_{ATP} = v_{ATPM} + \beta_{ATP} \lambda$  [15]. The first term is a growth-rate independent maintenance flux that represents the energy required to sustain basal cellular activities. The second term, instead, accounts for  $\lambda$ -dependence through a coefficient  $\beta_{ATP}$  that fixes the moles of ATP to be hydrolyzed per gram of dry weight. The values of  $v_{ATPM}$  and  $\beta_{ATP}$  are usually fitted from growth yield curves [16], and



**Fig 5. Overlap between FBA and CAFBA solutions.** (A) Overlap  $q$  between pFBA and CAFBA solutions as a function of growth rate, computed for three different reaction sets: all reactions included in the reconstruction, reactions in the glycolysis/gluconeogenesis pathway, and reactions in the TCA cycle. FBA fluxes were computed for the same glucose influx as the CAFBA solution and then interpolated at the growth rates of CAFBA solutions in order to plot the overlap as a function of  $\lambda$ . (B) Growth yield and acetate secretion from CAFBA, together with the FBA-predicted growth yield. In both panels the value  $\lambda_{ac} = 0.79/h$  is marked by a vertical dashed line.

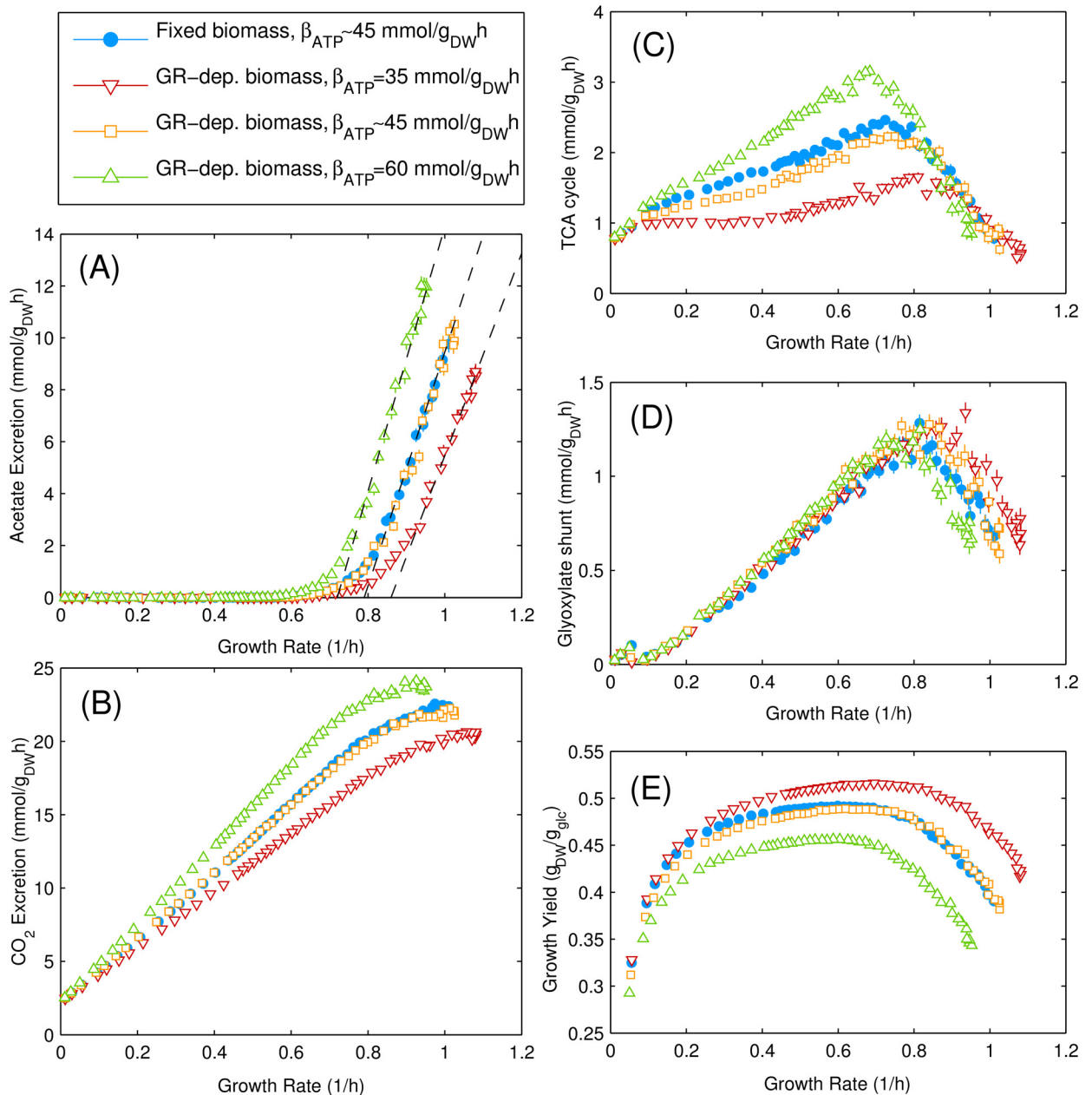
doi:10.1371/journal.pcbi.1004913.g005

different metabolic reconstructions of *E. coli* use different numerical values for both of them, see [45, 58, 59] and Table C in S1 Text. However, as the cell's composition (and specifically the amounts of RNA, DNA, proteins, fatty acids, etc.) adjusts with the growth rate, biomass coefficients, including the demand of growth-related ATP, are in general  $\lambda$ -dependent [39]. A natural question to ask at this point is how cellular ATP requirements impact the shift between respiration and fermentation.

Results obtained by solving CAFBA with  $\lambda$ -dependent biomass composition are shown in Fig 6 (open symbols), together with the solution obtained for constant biomass composition at the same  $\langle w \rangle = 8.8 \times 10^{-4}$  gh/mmol (filled blue circles). We tested CAFBA predictions with three different values of  $\beta_{ATP}$  while keeping  $v_{ATPM}$  fixed: (i) the default value for iJR904 model, (ii) the default value for the iAF1260 model, which is 30% larger than (i) [58], and (iii) a value 30% smaller than (i). One sees that, for the same ATP hydrolysis parameter (open and filled blue symbols), solutions for the two versions of CAFBA nearly overlap. On the other hand, both the slope and the onset growth rate  $\lambda_{ac}$  for acetate secretion appear to depend on the value of  $\beta_{ATP}$ . Likewise, the flux through TCA increases with  $\beta_{ATP}$  so as to satisfy energetic requirements. The growth yield and maximum growth rate  $\lambda_{max}$  obtained at  $w_C = 0$  decrease accordingly. However, if we tune  $\langle w \rangle$  to fix  $\lambda_{max} = 1/h$  for each value of  $\beta_{ATP}$ , acetate secretion starts consistently at  $\lambda_{ac} \simeq 0.8/h$  (Fig J in S1 Text), implying that energetic costs do not affect the ratio  $\lambda_{ac}/\lambda_{max}$ .

## Discussion

In this work we have introduced CAFBA, an extension of FBA inspired by the proteome allocation scenario underpinned by bacterial growth laws [3, 5, 21, 60]. By integrating a single additional global constraint in FBA, CAFBA formulates the interplay of growth and expression in



**Fig 6. CAFBA solutions with growth rate-dependent biomass composition.** Representative fluxes obtained by CAFBA for *E. coli* growth in glucose minimal medium with fixed (blue points) and variable biomass composition (in open red, yellow and green markers for three different values of the  $\lambda$ -dependent ATP hydrolysis rate  $\beta_{ATP}$ ). (A) Acetate secretion rate, (B)  $CO_2$  secretion rate, (C) flux through TCA cycle ( $\alpha$ KG dehydrogenase), (D) flux through glyoxylate shunt (Malate synthase), (E) growth yield. No significant differences are observed between the constant and  $\lambda$ -dependent cases for  $\beta_{ATP} = 45.5608$  mmol $_{ATP}$ /g $_{DW}$ , corresponding to the default value for the iJR904 model. We also show, for comparison, results obtained for larger and smaller values of  $\beta_{ATP}$ . The acetate secretion rate can always be fitted by a linear function of  $\lambda$ , i.e.  $v_{ac} = s \times (\lambda - \lambda_{ac})$ , albeit with different slopes and intercepts. The three dashed lines correspond to  $s = 39, 45, 51$  mmol/g $_{DW}$ , respectively, while  $\lambda_{ac} = 0.86, 0.79, 0.72$ /h, respectively. We also indicate  $\lambda_{ac} = 0.79$ /h with a vertical dashed line in all panels. In all cases we set  $\langle w \rangle = 8.8 \times 10^{-4}$  gh/mmol,  $w_C \geq 0$  and  $w_{max}/w_{min} = 10$ .

doi:10.1371/journal.pcbi.1004913.g006

metabolism as a simple and elegant growth-rate optimization problem, with the same computational complexity as standard FBA. States of optimal growth found by CAFBA therefore encode for optimality from both an energetic and a proteome allocation perspective. A most distinctive feature of CAFBA lies in the extremely simple empirical inputs required to make

quantitatively accurate predictions. The 3 parameters on which the proteome allocation constraint relies, namely  $w_{R,0}$ ,  $\phi_{\max}$  and  $\lambda_{\max}$ , can be easily obtained in experiments [5, 6]. All other parameters can be set based on these 3 numbers.

CAFBA predictions obtained for *E. coli* by averaging solutions over protein costs are found to be close to growth-yield maximizing solutions at slow growth. As growth gets faster, a continuous switch to a regime characterized by carbon overflow occurs. The onset of carbon overflow (at a growth rate denoted as  $\lambda_{ac}$ ) turns out to be largely independent of the nature of the glycolytic substrate. The ratio  $\lambda_{ac}/\lambda_{\max}$ , with  $\lambda_{\max}$  the fastest achievable growth rate, is indeed a remarkably robust quantity, that is roughly independent of the empirical parameters that characterize the proteome allocation constraint. Rather, it is mainly influenced by the weights of the biosynthetic reactions,  $\{w_i\}$ . These results strongly support the picture according to which acetate secretion is part of an optimal strategy to cope with increasing protein costs at high growth rates [7, 8, 29, 46].

CAFBA easily allows to model cellular metabolic activity in a variety of conditions, including translational limitation [8, 22] and protein overexpression [5]. As more growth laws are being characterized in different organisms [61–63], CAFBA's application range is likely to expand significantly. Note C in [S1 Text](#) details how to port CAFBA to growth conditions and/or bacterial species different from those considered here. Other growth maximizing organisms may also be studied by CAFBA if the required ingredients (network structure, biomass composition, empirical inputs) are available. Going beyond cell-autonomous models, CAFBA may prove highly effective for characterizing trophic interactions (e.g. cross-feeding) in microbial communities by treating excreted metabolites as potential nutrients [64]. CAFBA therefore provides a conceptually simple and computationally efficient platform that can be easily adapted and calibrated to describe the metabolism and growth of different organisms, making it a versatile tool for the computational modelling of interacting species in complex environments.

## Comparison of CAFBA to other models

Under carbon limitation, solutions of CAFBA are obtained by varying the parameter  $w_C$ , a proxy for the extracellular carbon level representing the proteome cost of the C-sector (carbon intake). In essence, for any given substrate level, CAFBA allocates the C-sector proteins per unit flux by simultaneously optimizing the allocation of the proteome fractions required to sustain biosynthesis and translation in order to maximize growth. The use of  $w_C$  as a control parameter as opposed to directly dialing the nutrient intake flux is one of the elements that distinguish CAFBA from closely related CBMs like FBA [15], FBAwMC [31, 32] and ME-models [19, 20]. In fact, the CAFBA constraint effectively reduces to a finite capacity constraint similar to the one that characterizes FBAwMC when an upper bound on the glucose intake flux is used to modulate growth at fixed  $w_C = 0$  (see Note B in [S1 Text](#)). (Note however that tuning nutrient levels as opposed to nutrient influx was employed in RBA to model the substitution between low affinity and high affinity cysteine transporters in *B. subtilis* [18].)

Secondly, CAFBA does not provide the detailed mechanistic description of gene expression and protein synthesis conveyed by ME-models and RBA, whose definition includes, for instance, explicit variables for macromolecular concentrations (ribosomes, DNA, RNA, etc.). Rather, it relies on an effective formulation based on empirical growth laws and (when desired) on a growth-rate dependent biomass composition. In this light, while less comprehensive than its closely related CBMs, the CAFBA scheme highlights the key biological ingredients constraining proteome allocation.

On a more technical level, both RBA and ME-models are intrinsically non-linear and handle non-linearity by approximating their underlying optimization problems through sequences of

linear feasibility problems. In CAFBA, even the worst case is solved through a fast iterative algorithm involving a small number of LP problems.

Finally, the optimal proteome allocation problem posed by CAFBA can be seen as an assumption of “optimal enzymatic efficiency”, close to that underlying FBA approaches based on flux minimization [65, 66].

## Choice of parameters

One of the strong points of standard FBA consists in its reliance on the stoichiometric matrix and on thermodynamic reversibility constraints alone, making kinetic parameters unnecessary. CAFBA’s proteome allocation constraint in principle introduces a large number of additional parameters related to reaction and/or transport kinetics that, for the most part, are either uncharacterized or inferred from *in vitro* studies performed in different biochemical conditions [67]. This raises the issue of parameter selection.

Two of the constants entering the proteome constraints (namely  $w_R$  and  $\phi_{\max}$ ) are obtained directly from empirical growth laws. With  $w_C$  acting as the control variable, the only free parameters left are the weights  $w_i$  characterizing intracellular reactions. While quantitative CAFBA predictions appear to be dependent on their specific values, the qualitative behaviour of the solutions is not. Furthermore, the scenario obtained by averaging CAFBA solutions over different choices of the  $w_i$  quantitatively reproduces experimental findings for acetate secretion and growth yield. These results point to a considerable degree of robustness of the CAFBA framework against fluctuations in parameter values. Notice however that the CAFBA picture can be further improved upon tuning the weights of individual reactions. For example, by increasing the average weight of reactions involved in respiration one sees a shift in the onset of acetate secretion and the value of  $\lambda_{ac}/\lambda_{\max}$  changes, see Fig K in [S1 Text](#). On the other hand, parameters can also be tuned according to empirical evidence so as to allow for a more thorough comparison with experiments performed on different strains and/or growth conditions, e.g. concerning intracellular fluxes (see Fig L in [S1 Text](#)).

In perspective, detailed flux measurements may allow to estimate typical weights for each pathway, and possibly even for individual enzymes, opening for the possibility to better calibrate the model and obtain completely quantitative predictions. Our work here has aimed at keeping the number of parameters as small as possible. In this light, many emerging features of the interplay between metabolism and gene expression appear to be mostly determined by the topology of the metabolic network. Elucidating the origin of this simplification is a foremost theoretical challenge for future studies of metabolic systems.

## On using average fluxes

The need to resort to an averaging procedure in order to reproduce bulk measurements for the growth yield and the acetate excretion forces us to ask whether the CAFBA averaging may have some further meaning. We offer here two possible scenarios.

The first one is based on the fact that, even in well controlled growth conditions, cells in a population are normally heterogeneous, as transcription levels, protein abundances, reaction fluxes and instantaneous growth rates may change significantly from one to the other [68–70]. This would in turn reflect in fluctuations in the values of each  $w_i$  across cells. Averaging over different choices for the weights could then simply be interpreted as averaging over a population of heterogeneous cells (as would seem appropriate in modelling batch culture or chemostats).

The alternative scenario presupposes that, even in absence of any cell-to-cell variability, cells may not be able to perfectly adjust fluxes to the distribution maximizing the instantaneous growth rate. This may occur for different reasons. First, the regulatory machinery needed to

perform protein allocation requires by itself an investment of metabolic and proteomic resources [71–73]. This burden becomes more severe as the regulatory system gets more sensitive and fine-tuned, and, clearly, CAFBA does not account for it. Secondly, environments where cells grow are always fluctuating. Any regulatory machinery implementing fast adjustments in response to small environmental changes will necessarily come at a cost that will negatively affect the growth rate. Under such constraints, regulatory programs selected over evolutionary time scales may prefer to maximize an *average* growth rate, the average being taken over life process history. The actual regulatory programs implemented would then balance the trade-off between the costs of not being exactly in the instantaneously optimal growth state and the costs of adjusting regulation too frequently in “natural” conditions (not those provided in the laboratory). An interpretation of the CAFBA averaging prescription would then be that it is a way to implement an “average” strategy that smooths the output upon variations of the environmental conditions.

In both of the above scenarios, CAFBA points to the emergence of acetate excretion as triggered by regulatory system(s) sensing the abundance of the carbon source and the balance of biomass synthesis and energy generation [29]. We note that carbon overflow in *E. coli* has been proposed to be modulated by catabolite repression mediated ACS down-regulation [49]. Discriminating between the two scenarios we have just presented could be achieved already in bulk experiments, by changing the weight of specific enzymes over time (e.g., by expressing useless proteins specific to certain pathways) and monitoring whether the associated fluxes adjust dynamically in real time. Naturally, tests of cell-to-cell heterogeneities would also allow to favor one scenario over the other.

Finally, we address the magnitude of fluctuations of the weights  $w_i$ . The spread in the enzyme catalytic rates  $k_{cat,i}$  as tabulated in databases, is notoriously broad, exceeding 3 orders of magnitude [32, 67]. In the absence of more refined information, it is reasonable to expect that the weights  $w_i$  should fluctuate by about the same amount. However, we have seen that unrealistic results are generated by CAFBA if weights are allowed to fluctuate more than 10-fold. Therefore, either the true width of the distribution of the weights  $w_i$  is much smaller than what is suggested by the values of  $k_{cat,i}$  estimated in vitro, or weights are subtly distributed across pathways in such a way that strong compensatory effects occur that reduce fluctuations. With steady improvements in proteomic methods, it may soon be possible to quantitatively determine these parameters empirically and elucidate this puzzle.

## Materials and Methods

### Optimization problem

Given a metabolic network encoded by a stoichiometric matrix  $\mathbf{S} = \{S_{\mu i}\}$ , CAFBA is stated in the case of carbon limitation as

$$\max_{\mathbf{v}} \lambda \quad \text{subject to} \quad \text{(i)} \quad \sum_i S_{\mu i} v_i = 0 \quad \forall \mu \quad (15)$$

$$\text{(ii)} \quad \ell_i \leq v_i \leq u_i \quad \forall i \quad (16)$$

$$\text{(iii)} \quad w_C v_C + \sum_i w_i |v_i| + w_R \lambda = \phi \max \quad , \quad (17)$$

where  $\lambda$  denotes the growth rate,  $\mathbf{v} = \{v_i\}$  is a flux vector, and  $(\ell_i, u_i)$  represent lower and upper bounds for each flux  $v_i$ , respectively. [Condition \(17\)](#) corresponds to the proteome allocation constraint  $\phi_C + \phi_E + \phi_R + \phi_Q = 1$ , with  $v_C \geq 0$  being the active glucose intake flux and with the

sum in  $\sum_i w_i |v_i|$  running over all enzyme-catalyzed reactions except for transports, exchanges and carbon intake pathways. The biomass flux  $\lambda$  and ATP maintenance reaction are also excluded from Eq (17).

In principle, CAFBA is a MILP (Mixed Integer-Linear Programming) problem due to the presence of absolute values in Eq (17). However, in CAFBA they can be disposed of by splitting each flux  $v_i$  into a forward  $v_i^+$  and a backward  $v_i^-$  component, both non-negative. Note that if either  $v_i^+$  or  $v_i^-$  can be set to zero for each  $i$ , net fluxes  $v_i = v_i^+ - v_i^-$  are univocally determined, one has  $|v_i| = v_i^+ + v_i^-$  for absolute values, and CAFBA becomes equivalent to

$$\max_{v^+, v^-} \lambda \quad \text{subject to} \quad \text{(i)} \quad \sum_i S_{\mu i} (v_i^+ - v_i^-) = 0 \quad \forall \mu \quad (18)$$

$$\text{(ii)} \quad 0 \leq v_i^- \leq -\ell_i \quad 0 \leq v_i^+ \leq u_i \quad \forall i \quad (19)$$

$$\text{(iii)} \quad w_C v_C + \sum_i w_i (v_i^+ + v_i^-) + w_R \lambda = \phi_{\max} \quad , \quad (20)$$

which is a simple LP problem rather than a MILP. The key observation is that, as long as  $\lambda$  is maximized, CAFBA actually adjusts fluxes so that either the forward or the backward component vanish for each  $i$ . Indeed, a necessary condition for maximizing  $\lambda$  is that the quantity  $v_i^+ + v_i^-$  is minimized for each  $i$ , which, at fixed  $v_i = v_i^+ - v_i^-$ , is achieved by setting either  $v_i^+$  or  $v_i^-$  to zero. Therefore CAFBA reduces from a MILP to a LP problem.

Note that, because of the tight link between CAFBA and flux minimization, degeneracies in CAFBA solutions can only arise from the presence of (a) futile loops or (b) pathways that perform the same overall chemical conversion with the same flux at the same proteome cost, and which therefore can be used alternatively. In CAFBA with heterogeneous weights, however, the chance that two equivalent flux configurations have exactly the same total weight is negligible, since weights are i.i.d. random variables. On the other hand, futile loops only concern transports that do not involve the main carbon source and therefore are not included explicitly in the CAFBA constraint. These loops however do not affect other fluxes and are easily spotted and removed, either by manually shutting off redundant processes or by including them into the proteome allocation constraint with an arbitrarily small but non-zero weight. Therefore, each instance of the inhomogeneous CAFBA scheme has a unique solution almost surely. Because our main results are obtained in this framework, alternate optima are in practice not an issue in CAFBA.

## Implementation

We implemented CAFBA on the E. coli iJR904 genome-scale model [45], comprehensive of 761 metabolites and 1075 reactions, as a Matlab function, using the COBRA Toolbox [74] to load the network reconstruction with a minor modification. Specifically, we shut off the glucose dehydrogenase reaction, since it is only functional if the cofactor pyrroloquinoline quinone is supplied in the environment (see the Ecocyc [75] entry on the enzyme). Both GLPK- and Gurobi-compatible CAFBA solvers for Matlab are provided as S1 Code, along with a small set of utility functions. Running times for a single CAFBA optimization of the iJR904 network with a common laptop (single thread of an Intel Core i7-2630QM CPU @ 2.00GHz) are around 0.12 s for the GLPK (version 4.47) LP solver and 0.05 s the Gurobi Optimizer (version 5.6) solver. For comparison, the time required to compute the standard FBA solution for the same network with the COBRA toolbox using GLPK is around 0.06 s.

## Case of growth-rate dependent biomass

The fact that cells adapt their composition with the growth rate [2, 4, 5] implies that biomass composition is itself  $\lambda$ -dependent. Growth-rate dependent biomass coefficients (see e.g. [39] for *E. coli*) indeed reflect empirical knowledge of how the amounts of RNA, DNA, proteins, fatty acids, etc. are modulated by  $\lambda$ . While constraint-based models such as FBA and CAFBA with growth-dependent biomass are non-linear, approximate solutions can be obtained efficiently by simple iterated LP protocols as follows: (a) starting from a given biomass vector, solve the model by optimizing the growth rate; (b) update the biomass composition to the computed optimal growth rate using the prescribed set of  $\lambda$ -dependent biomass coefficients; (c) iterate until a solution is reached, such that further iterations do not change the optimal growth rate within a desired precision. For CAFBA, this procedure typically converges in a very small number of iterations (see Fig I in [S1 Text](#)). Further details about the case of growth rate-dependent biomass composition and the iterative algorithm for computing the optimal FBA or CAFBA solutions are given in Note F in [S1 Text](#).

## Extension to different growth media and/or organisms

Besides the full study of the *E. coli* iJR904 model, we have tested CAFBA on the more recent reconstructions iAF1260 [58] and iJO1366 [59], obtaining very similar results. COBRA-compatible Matlab functions [74] to run CAFBA on these models are provided as [S1 Code](#). Note C in [S1 Text](#) describes in detail how to port CAFBA to different growth media, nutrient limitations and/or bacterial species. However, provided the input coming from empirical growth laws is available together with the network structure and the biomass composition, the CAFBA framework can in principle be extended to growth-maximizing organisms other than bacteria.

## Supporting Information

**S1 Text. Supporting text.** Supplementary notes, tables, figures and references. (PDF)

**S1 Code. Supplementary MatLab code.** COBRA-compatible Matlab functions implementing CAFBA. (ZIP)

## Acknowledgments

MM gratefully acknowledges the hospitality for his visit to Hwa labs at UCSD. TH gratefully acknowledges the hospitality of U. Paris-Sud and Sapienza, U. Roma for his visits. The authors acknowledge members of Hwa lab for contributing growth rate data for comparing between MG1655 and NCM3722 strains. Useful discussions with S. Hui, R. Mulet, A. Pagnani, A. Samal, A. Tramontano and Z. Yang are gratefully acknowledged.

## Author Contributions

Conceived and designed the experiments: TH OCM ADM EM. Performed the experiments: MM. Analyzed the data: MM. Contributed reagents/materials/analysis tools: MM TH OCM ADM EM. Wrote the paper: MM TH OCM ADM EM.

## References

1. Monod J (1949) The growth of bacterial cultures. *Ann Rev Microbiol* 3: 371–394 doi: [10.1146/annurev.mi.03.100149.002103](https://doi.org/10.1146/annurev.mi.03.100149.002103)

2. Schaechter M, Maaløe O, Kjeldgaard N (1958) Dependency on medium and temperature of cell size and chemical composition during balanced growth of *Salmonella typhimurium*. *J Gen Microbiol* 19: 592–606 doi: [10.1099/00221287-19-3-592](https://doi.org/10.1099/00221287-19-3-592) PMID: [13611202](https://pubmed.ncbi.nlm.nih.gov/13611202/)
3. Kjeldgaard N, Kurland C (1963) The distribution of soluble and ribosomal RNA as a function of growth rate. *J Mol Biol* 6: 341–348 doi: [10.1016/S0022-2836\(63\)80093-5](https://doi.org/10.1016/S0022-2836(63)80093-5)
4. Bremer H, Dennis PP (1996) Modulation of chemical composition and other parameters of the cell by growth rate. *Escherichia coli and Salmonella: Cellular and Molecular Biology* 2: 1553–1569
5. Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T (2010) Interdependence of cell growth and gene expression: origins and consequences. *Science* 330: 1099–1102 doi: [10.1126/science.1192588](https://doi.org/10.1126/science.1192588) PMID: [21097934](https://pubmed.ncbi.nlm.nih.gov/21097934/)
6. You C, et al. (2013) Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature* 500: 301–306 doi: [10.1038/nature12446](https://doi.org/10.1038/nature12446) PMID: [23925119](https://pubmed.ncbi.nlm.nih.gov/23925119/)
7. Molenaar D, van Berlo R, de Ridder D, Teusink B (2009) Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol Sys Biol* 5: 323
8. Hui S, et al. (2015) Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Mol Sys Biol* 11: 784 doi: [10.15252/msb.20145697](https://doi.org/10.15252/msb.20145697)
9. Papoutsakis ET (1984) Equations and calculations for fermentations of butyric acid bacteria. *Biotechnol Bioeng* 26: 174–187 doi: [10.1002/bit.260260210](https://doi.org/10.1002/bit.260260210) PMID: [18551704](https://pubmed.ncbi.nlm.nih.gov/18551704/)
10. Varma A, Boesch BW, Palsson BØ (1993) Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. *Appl Environ Microbiol* 59: 2465–2473. PMID: [8368835](https://pubmed.ncbi.nlm.nih.gov/8368835/)
11. Varma A, Palsson BØ (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl Environ Microbiol* 60: 3724–3731. PMID: [7986045](https://pubmed.ncbi.nlm.nih.gov/7986045/)
12. Edwards JS, Ibarra RU, Palsson BØ (2001) In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnol* 19: 125–130 doi: [10.1038/84379](https://doi.org/10.1038/84379)
13. Ibarra RU, Edwards JS, Palsson BØ (2002) *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* 420: 186–189 doi: [10.1038/nature01149](https://doi.org/10.1038/nature01149) PMID: [12432395](https://pubmed.ncbi.nlm.nih.gov/12432395/)
14. Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? *Nature Biotechnol* 28: 245–248 doi: [10.1038/nbt.1614](https://doi.org/10.1038/nbt.1614)
15. Palsson BØ (2015) *Systems biology*. Cambridge University Press.
16. Feist AM, Palsson BØ (2010) The biomass objective function. *Curr Opin Microbiol* 13: 344–349 doi: [10.1016/j.mib.2010.03.003](https://doi.org/10.1016/j.mib.2010.03.003) PMID: [20430689](https://pubmed.ncbi.nlm.nih.gov/20430689/)
17. Goelzer A, Fromion V (2011) Bacterial growth rate reflects a bottleneck in resource allocation. *Biochimica et Biophysica Acta (BBA)-General Subjects* 1810: 978–988 doi: [10.1016/j.bbagen.2011.05.014](https://doi.org/10.1016/j.bbagen.2011.05.014)
18. Goelzer A, Fromion V, Scorletti G (2011) Cell design in bacteria as a convex optimization problem. *Automatica* 47: 1210–1218 doi: [10.1016/j.automatica.2011.02.038](https://doi.org/10.1016/j.automatica.2011.02.038)
19. Lerman JA et al. (2012) In silico method for modelling metabolism and gene product expression at genome scale. *Nature Comm* 3: 929 doi: [10.1038/ncomms1928](https://doi.org/10.1038/ncomms1928)
20. O'Brien EJ, Lerman JA, Chang RL, Hyduke DR, Palsson BØ (2013) Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol Sys Biol* 9: 693 doi: [10.1038/msb.2013.52](https://doi.org/10.1038/msb.2013.52)
21. Maaløe O (1979) Regulation of the protein-synthesizing machinery—ribosomes, tRNA, factors, and so on. In *Biological regulation and development*. Springer, pp. 487–542
22. Deris JB, et al. (2013) The innate growth bistability and fitness landscapes of antibiotic-resistant bacteria. *Science* 342: 1237435 doi: [10.1126/science.1237435](https://doi.org/10.1126/science.1237435) PMID: [24288338](https://pubmed.ncbi.nlm.nih.gov/24288338/)
23. Gottschalk G (1986) *Bacterial Metabolism*, chap. 8. Springer Science & Business Media, pp. 208–282
24. Clark DP (1989) The fermentation pathways of *Escherichia coli*. *FEMS Microbiol Rev* 5: 223–234 doi: [10.1111/j.1574-6968.1989.tb03398.x](https://doi.org/10.1111/j.1574-6968.1989.tb03398.x) PMID: [2698228](https://pubmed.ncbi.nlm.nih.gov/2698228/)
25. Wills C (1990) Regulation of sugar and ethanol metabolism in *Saccharomyces cerevisiae*. *Critical Reviews Biochem Mol Biol* 25: 245–280 doi: [10.3109/10409239009090611](https://doi.org/10.3109/10409239009090611)
26. Wolfe AJ (2005) The acetate switch. *Microbiol Mol Biol Rev* 69: 12–50 doi: [10.1128/MMBR.69.1.12-50.2005](https://doi.org/10.1128/MMBR.69.1.12-50.2005) PMID: [15755952](https://pubmed.ncbi.nlm.nih.gov/15755952/)
27. Diaz-Ruiz R, Rigoulet M, Devin A (2011) The Warburg and Crabtree effects: On the origin of cancer cell energy metabolism and of yeast glucose repression. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1807: 568–576 doi: [10.1016/j.bbabi.2010.08.010](https://doi.org/10.1016/j.bbabi.2010.08.010)

28. Dashko S, Zhou N, Compagno C, Piškur J (2014) Why, when, and how did yeast evolve alcoholic fermentation? *FEMS Yeast Res* 14: 826–832 doi: [10.1111/1567-1364.12161](https://doi.org/10.1111/1567-1364.12161) PMID: [24824836](https://pubmed.ncbi.nlm.nih.gov/24824836/)
29. Basan M, et al. (2015) Efficient allocation of proteomic resources for energy metabolism results in acetate overflow. *Nature* 528: 99–104
30. Majewski R, Domach M (1990) Simple constrained-optimization view of acetate overflow in *E. coli*. *Biotechnol Bioeng* 35: 732–738 doi: [10.1002/bit.260350711](https://doi.org/10.1002/bit.260350711) PMID: [18592570](https://pubmed.ncbi.nlm.nih.gov/18592570/)
31. Beg Q, et al. (2007) Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity. *Proc Natl Acad Sci USA* 104: 12663–12668 doi: [10.1073/pnas.0609845104](https://doi.org/10.1073/pnas.0609845104) PMID: [17652176](https://pubmed.ncbi.nlm.nih.gov/17652176/)
32. Vazquez A, et al. (2008) Impact of the solvent capacity constraint on *E. coli* metabolism. *BMC Syst Biol* 2: 7 doi: [10.1186/1752-0509-2-7](https://doi.org/10.1186/1752-0509-2-7) PMID: [18215292](https://pubmed.ncbi.nlm.nih.gov/18215292/)
33. Zhuang K, Vemuri GN, Mahadevan R (2011) Economics of membrane occupancy and respiro-fermentation. *Mol Syst Biol* 7 doi: [10.1038/msb.2011.34](https://doi.org/10.1038/msb.2011.34) PMID: [21694717](https://pubmed.ncbi.nlm.nih.gov/21694717/)
34. Paul BJ, Ross W, Gaal T, Gourse RL (2004) rRNA transcription in *Escherichia coli*. *Annu Rev Genet* 38: 749–770 doi: [10.1146/annurev.genet.38.072902.091347](https://doi.org/10.1146/annurev.genet.38.072902.091347) PMID: [15568992](https://pubmed.ncbi.nlm.nih.gov/15568992/)
35. Klumpp S, Scott M, Pedersen S, Hwa T (2013) Molecular crowding limits translation and cell growth. *Proc Natl Acad Sci USA* 110: 16754–16759 doi: [10.1073/pnas.1310377110](https://doi.org/10.1073/pnas.1310377110) PMID: [24082144](https://pubmed.ncbi.nlm.nih.gov/24082144/)
36. Potrykus K, Cashel M (2008) (p)ppGpp: Still Magical? *Annu Rev Microbiol* 62: 35–51 doi: [10.1146/annurev.micro.62.081307.162903](https://doi.org/10.1146/annurev.micro.62.081307.162903) PMID: [18454629](https://pubmed.ncbi.nlm.nih.gov/18454629/)
37. Scott M, Klumpp S, Mateescu EM, Hwa T (2014) Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Mol Syst Biol* 10(8): 747 doi: [10.15252/msb.20145379](https://doi.org/10.15252/msb.20145379)
38. Li Z, Nimtz M, Rinas U (2014) The metabolic potential of *Escherichia coli* BL21 in defined and rich medium. *Microb Cell Fact* 13: 45 doi: [10.1186/1475-2859-13-45](https://doi.org/10.1186/1475-2859-13-45) PMID: [24656150](https://pubmed.ncbi.nlm.nih.gov/24656150/)
39. Pramanik J, Keasling J (1998) Effect of *Escherichia coli* biomass composition on central metabolic fluxes predicted by a stoichiometric model. *Biotechnol Bioeng* 60: 230–238 doi: [10.1002/\(SICI\)1097-0290\(19981020\)60:2%3C230::AID-BIT10%3E3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1097-0290(19981020)60:2%3C230::AID-BIT10%3E3.0.CO;2-Q) PMID: [10099424](https://pubmed.ncbi.nlm.nih.gov/10099424/)
40. Taymaz-Nikerel H, Borujeni AE, Verheijen PJ, Heijnen JJ, van Gulik WM (2010) Genome-derived minimal metabolic models for *Escherichia coli* MG1655 with estimated in vivo respiratory ATP stoichiometry. *Biotechnol Bioeng* 107: 369–381 doi: [10.1002/bit.22802](https://doi.org/10.1002/bit.22802) PMID: [20506321](https://pubmed.ncbi.nlm.nih.gov/20506321/)
41. Hermesen R, Okano H, You C, Werner N, Hwa T (2015) A growth-rate composition formula for the growth of *E. coli* on co-utilized carbon substrates. *Mol Syst Biol* 11: 801 doi: [10.15252/msb.20145537](https://doi.org/10.15252/msb.20145537)
42. Noor E, et al. (2014) Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLoS Comput Biol* 10: e1003483 doi: [10.1371/journal.pcbi.1003483](https://doi.org/10.1371/journal.pcbi.1003483) PMID: [24586134](https://pubmed.ncbi.nlm.nih.gov/24586134/)
43. Kochanowski K, et al. (2013) Functioning of a metabolic flux sensor in *Escherichia coli*. *Proc Natl Acad Sci USA* 110: 1130–1135 doi: [10.1073/pnas.1202582110](https://doi.org/10.1073/pnas.1202582110) PMID: [23277571](https://pubmed.ncbi.nlm.nih.gov/23277571/)
44. Woldringh C, Binnerts J, Mans A (1981) Variation in *Escherichia coli* buoyant density measured in Percoll gradients. *J Bacteriol* 148: 58–63 PMID: [6270065](https://pubmed.ncbi.nlm.nih.gov/6270065/)
45. Reed JL, Vo TD, Schilling CH, Palsson BØ (2003) An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol* 4: R54 doi: [10.1186/gb-2003-4-9-r54](https://doi.org/10.1186/gb-2003-4-9-r54) PMID: [12952533](https://pubmed.ncbi.nlm.nih.gov/12952533/)
46. Peebo K, et al. (2015) Proteome reallocation in *Escherichia coli* with increasing specific growth rate. *Mol Biosyst* 11: 1184–1193 doi: [10.1039/C4MB00721B](https://doi.org/10.1039/C4MB00721B) PMID: [25712329](https://pubmed.ncbi.nlm.nih.gov/25712329/)
47. Li GW, Burkhardt D, Gross C, Weissman JS (2014) Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* 157: 624–635 doi: [10.1016/j.cell.2014.02.033](https://doi.org/10.1016/j.cell.2014.02.033) PMID: [24766808](https://pubmed.ncbi.nlm.nih.gov/24766808/)
48. Holms H (1996) Flux analysis and control of the central metabolic pathways in *Escherichia coli*. *FEMS Microbiol Rev* 19: 85–116 doi: [10.1111/j.1574-6976.1996.tb00255.x](https://doi.org/10.1111/j.1574-6976.1996.tb00255.x) PMID: [8988566](https://pubmed.ncbi.nlm.nih.gov/8988566/)
49. Nanchen A, Schicker A, Sauer U (2006) Nonlinear dependency of intracellular fluxes on growth rate in miniaturized continuous cultures of *Escherichia coli*. *Appl Environ Microbiol* 72: 1164–1172 doi: [10.1128/AEM.72.2.1164-1172.2006](https://doi.org/10.1128/AEM.72.2.1164-1172.2006) PMID: [16461663](https://pubmed.ncbi.nlm.nih.gov/16461663/)
50. Valgepea K, et al. (2010) Systems biology approach reveals that overflow metabolism of acetate in *Escherichia coli* is triggered by carbon catabolite repression of acetyl-CoA synthetase. *BMC Syst Biol* 4: 166 doi: [10.1186/1752-0509-4-166](https://doi.org/10.1186/1752-0509-4-166) PMID: [21122111](https://pubmed.ncbi.nlm.nih.gov/21122111/)
51. Vemuri G, Altman E, Sangurdekar D, Khodursky A, Eiteman M (2006) Overflow metabolism in *Escherichia coli* during steady-state growth: transcriptional regulation and effect of the redox ratio. *Appl Environ Microbiol* 72: 3653–3661 doi: [10.1128/AEM.72.5.3653-3661.2006](https://doi.org/10.1128/AEM.72.5.3653-3661.2006) PMID: [16672514](https://pubmed.ncbi.nlm.nih.gov/16672514/)
52. Sauer U, Canonaco F, Heri S, Perrenoud A, Fischer E (2004) The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of *Escherichia coli*. *J Biol Chem* 279: 6613–6619 doi: [10.1074/jbc.M311657200](https://doi.org/10.1074/jbc.M311657200) PMID: [14660605](https://pubmed.ncbi.nlm.nih.gov/14660605/)

53. Fuhrer T, Fischer E, Sauer U (2005) Experimental identification and quantification of glucose metabolism in seven bacterial species. *J Bacteriol* 187: 1581–1590 doi: [10.1128/JB.187.5.1581-1590.2005](https://doi.org/10.1128/JB.187.5.1581-1590.2005) PMID: [15716428](https://pubmed.ncbi.nlm.nih.gov/15716428/)
54. Flamholz A, Noor E, Bar-Even A, Liebermeister W, Milo R (2013) Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc Natl Acad Sci USA* 110: 10039–10044 doi: [10.1073/pnas.1215283110](https://doi.org/10.1073/pnas.1215283110) PMID: [23630264](https://pubmed.ncbi.nlm.nih.gov/23630264/)
55. Stettner AI, Segrè D (2013) The cost of efficiency in energy metabolism. *Proc Natl Acad Sci USA* 110: 9629–9630 doi: [10.1073/pnas.1307485110](https://doi.org/10.1073/pnas.1307485110) PMID: [23729810](https://pubmed.ncbi.nlm.nih.gov/23729810/)
56. Lewis NE, et al. (2010) Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol Sys Bio*, 6: 390
57. Martelli C, De Martino A, Marinari E, Marsili M, Perez Castillo I (2009) Identifying essential genes in *Escherichia coli* from a metabolic optimization principle. *Proc Natl Acad Sci USA* 106: 2607–2611 doi: [10.1073/pnas.0813229106](https://doi.org/10.1073/pnas.0813229106) PMID: [19196991](https://pubmed.ncbi.nlm.nih.gov/19196991/)
58. Feist AM, et al. (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* 3: 121 doi: [10.1038/msb4100155](https://doi.org/10.1038/msb4100155) PMID: [17593909](https://pubmed.ncbi.nlm.nih.gov/17593909/)
59. Orth JD, et al. (2011) A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism. *Mol Sys Biol* 7: 535 doi: [10.1038/msb.2011.65](https://doi.org/10.1038/msb.2011.65)
60. Scott M, Hwa T (2011) Bacterial growth laws and their applications. *Current Opin Biotechnol* 22: 559–565 doi: [10.1016/j.copbio.2011.04.014](https://doi.org/10.1016/j.copbio.2011.04.014)
61. Fisher SH (1999) Regulation of nitrogen metabolism in *Bacillus subtilis*: vive la difference! *Mol Microbiol* 32: 223–232 doi: [10.1046/j.1365-2958.1999.01333.x](https://doi.org/10.1046/j.1365-2958.1999.01333.x) PMID: [10231480](https://pubmed.ncbi.nlm.nih.gov/10231480/)
62. Levy S, Barkai N (2009) Coordination of gene expression with growth rate: a feedback or a feed-forward strategy? *FEBS Letters* 583: 3974–3978 doi: [10.1016/j.febslet.2009.10.071](https://doi.org/10.1016/j.febslet.2009.10.071) PMID: [19878679](https://pubmed.ncbi.nlm.nih.gov/19878679/)
63. Wegener KM, et al. (2010) Global proteomics reveal an atypical strategy for carbon/nitrogen assimilation by a cyanobacterium under diverse environmental perturbations. *Mol Cell Proteomics* 9: 2678–2689 doi: [10.1074/mcp.M110.000109](https://doi.org/10.1074/mcp.M110.000109) PMID: [20858728](https://pubmed.ncbi.nlm.nih.gov/20858728/)
64. Harcombe WR, et al. (2014). Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. *Cell Rep* 7: 1104–1115 doi: [10.1016/j.celrep.2014.03.070](https://doi.org/10.1016/j.celrep.2014.03.070) PMID: [24794435](https://pubmed.ncbi.nlm.nih.gov/24794435/)
65. Holzhütter HG (2004) The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *Eur J Biochem* 271: 2905–2922 doi: [10.1111/j.1432-1033.2004.04213.x](https://doi.org/10.1111/j.1432-1033.2004.04213.x) PMID: [15233787](https://pubmed.ncbi.nlm.nih.gov/15233787/)
66. Tarlak F, Sadıkoğlu H, Çakır T (2014) The role of flexibility and optimality in the prediction of intracellular fluxes of microbial central carbon metabolism. *Mol BioSyst* 10: 2459 doi: [10.1039/C4MB00117F](https://doi.org/10.1039/C4MB00117F) PMID: [24993806](https://pubmed.ncbi.nlm.nih.gov/24993806/)
67. Schomburg I, et al. (2012) BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Research* 41: D764–D772 doi: [10.1093/nar/gks1049](https://doi.org/10.1093/nar/gks1049) PMID: [23203881](https://pubmed.ncbi.nlm.nih.gov/23203881/)
68. Labhsetwar P, Cole JA, Roberts E, Price ND, Luthy-Schulten ZA (2013) Heterogeneity in protein expression induces metabolic variability in a modeled *Escherichia coli* population. *Proc Natl Acad Sci USA* 110: 14006–14011 doi: [10.1073/pnas.1222569110](https://doi.org/10.1073/pnas.1222569110) PMID: [23908403](https://pubmed.ncbi.nlm.nih.gov/23908403/)
69. Kiviet DJ, et al. (2014) Stochasticity of metabolism and growth at the single-cell level. *Nature* 514: 376–379 doi: [10.1038/nature13582](https://doi.org/10.1038/nature13582) PMID: [25186725](https://pubmed.ncbi.nlm.nih.gov/25186725/)
70. Solopova A, et al. (2014) Bet-hedging during bacterial diauxic shift. *Proc Natl Acad Sci USA* 111: 7427–7432 doi: [10.1073/pnas.1320063111](https://doi.org/10.1073/pnas.1320063111) PMID: [24799698](https://pubmed.ncbi.nlm.nih.gov/24799698/)
71. Lan G, Sartori P, Neumann S, Sourjik V, Tu Y (2012). The energy-speed-accuracy trade-off in sensory adaptation. *Nat Phys* 8: 422–428 doi: [10.1038/nphys2276](https://doi.org/10.1038/nphys2276) PMID: [22737175](https://pubmed.ncbi.nlm.nih.gov/22737175/)
72. Mehta P, Schwab DJ (2012) Energetic costs of cellular computation. *Proc Natl Acad Sci USA*, 109: 17978–17982 doi: [10.1073/pnas.1207814109](https://doi.org/10.1073/pnas.1207814109) PMID: [23045633](https://pubmed.ncbi.nlm.nih.gov/23045633/)
73. Govern CC, ten Wolde PR (2014). Optimal resource allocation in cellular sensing systems. *Proc Natl Acad Sci USA*, 111: 17486–17491 doi: [10.1073/pnas.1411524111](https://doi.org/10.1073/pnas.1411524111) PMID: [25422473](https://pubmed.ncbi.nlm.nih.gov/25422473/)
74. Schellenberger J, et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature Protoc* 6: 1290–1307 doi: [10.1038/nprot.2011.308](https://doi.org/10.1038/nprot.2011.308)
75. Keseler IM, et al. (2011) EcoCyc: a comprehensive database of *Escherichia coli* biology. *Nucleic Acids Res* 39: D583–D590 doi: [10.1093/nar/gkq1143](https://doi.org/10.1093/nar/gkq1143) PMID: [21097882](https://pubmed.ncbi.nlm.nih.gov/21097882/)