

Research



Cite this article: Kleijn IT, Marguerat S, Shahrezaei V. 2023 A coarse-grained resource allocation model of carbon and nitrogen metabolism in unicellular microbes. *J. R. Soc. Interface* **20**: 20230206.
<https://doi.org/10.1098/rsif.2023.0206>

Received: 10 April 2023
Accepted: 8 September 2023

Subject Category:

Life Sciences—Mathematics interface

Subject Areas:biomathematics, systems biology,
computational biology**Keywords:**microbial growth, mathematical modelling,
proteome allocation, metabolism, cellular
growth laws, gene expression**Author for correspondence:**Vahid Shahrezaei
e-mail: v.shahrezaei@imperial.ac.uk

[†]Present address: UCL Cancer Institute,
University College London, London, UK.

Electronic supplementary material is available
online at <https://doi.org/10.6084/m9.figshare.c.6837579>.

A coarse-grained resource allocation
model of carbon and nitrogen
metabolism in unicellular microbesIstvan T. Kleijn^{1,2,3,4,5}, Samuel Marguerat^{2,3,†} and Vahid Shahrezaei¹

¹Department of Mathematics, Faculty of Natural Sciences, and ²Institute of Clinical Sciences, Faculty of Medicine, Imperial College London, London, UK

³MRC London Institute of Medical Sciences, London, UK

⁴Division of Breast Cancer Research, The Institute of Cancer Research, London, UK

⁵Ralph Lauren Centre for Breast Cancer Research, The Royal Marsden NHS Foundation Trust, London, UK

ITK, 0000-0001-9466-9167; SM, 0000-0002-2402-3165; VS, 0000-0002-4013-5458

Coarse-grained resource allocation models (C-GRAMs) are simple mathematical models of cell physiology, where large components of the macromolecular composition are abstracted into single entities. The dynamics and steady-state behaviour of such models provides insights on optimal allocation of cellular resources and have explained experimentally observed cellular growth laws, but current models do not account for the uptake of compound sources of carbon and nitrogen. Here, we formulate a C-GRAM with nitrogen and carbon pathways converging on biomass production, with parametrizations accounting for respirofermentative and purely respiratory growth. The model describes the effects of the uptake of sugars, ammonium and/or compound nutrients such as amino acids on the translational resource allocation towards proteome sectors that maximized the growth rate. It robustly recovers cellular growth laws including the Monod law and the ribosomal growth law. Furthermore, we show how the growth-maximizing balance between carbon uptake, recycling, and excretion depends on the nutrient environment. Lastly, we find a robust linear correlation between the ribosome fraction and the abundance of amino acid equivalents in the optimal cell, which supports the view that simple regulation of translational gene expression can enable cells to achieve an approximately optimal growth state.

1. Introduction

Unicellular organisms are remarkably efficient self-replicators as they are under selective pressure to grow fast or risk being outcompeted by rival colonies or species [1,2]. On the other hand, microbial cells are faced with internal constraints limiting their growth, because each metabolite, macromolecule, or unit of membrane area can only be used for one reaction at any given time. Fast-growing cells must therefore possess an ability to allocate these limited resources in varied environments [3–6]. The interplay between gene expression and growth can be studied reproducibly in states of balanced growth, where cells are maintained in the same environment for many generations [7].

Observed patterns of gene expression and the resulting rate of growth depend heavily on the growth environment, in particular on the presence of external stresses and, importantly, the nutrient make-up. A well-established feature of gene expression in multiple model organisms is the presence of linear correlations between broad classes of macromolecules and the cellular growth rate [8–12]. The most pervasive positive correlation is between the abundance of translational resources (chiefly ribosomes) and the growth rate [13–17]. Classes of proteins that are negatively correlated with the growth rate are

associated with stress, or induced by the specific cause of growth inhibition—stressors or reductions in the nutrient quality or quantity [15,17–19].

These observed correlations between the total abundance of proteome sectors and the cellular growth rate have been explored in several coarse-grained phenomenological and mechanistic models [20–28]. These coarse-grained resource allocation models (C-GRAMs) make explicit the intuition that the conversion of nutrient into biomass can be done in more or less efficient ways. Abundant proteins are costly to produce and therefore have a large effect on the growth rate, which is a measure of fitness [6].

Extended summaries of these models are provided in electronic supplementary material, text S1. All were used to study trade-offs between resource allocation towards metabolic versus ribosomal gene expression and they all account for the experimental relation between ribosomal gene expression and the growth rate. Another common feature of these earlier models is the hyperbolic dependence of the growth rate on the concentration of external nutrients, observed first experimentally in [29]. C-GRAMs have been used to model the burden of synthetic circuits in the host cell [24,26]. Another powerful application of the coarse-grained approach came from [27], which explained how differences in cell size across different growth modulations could be explained by the underlying proteome composition.

The coarse-graining approach of C-GRAMs entails that large sectors of the proteome are abstracted into a single protein, whose kinetics are explicitly described. The coarse-graining approach boils down assumptions about metabolism, growth, gene expression and cellular physiology into only a handful of parameters. Fitting nonlinear models is a challenging problem in general and, in the case of whole-genome models, further complicated by the large number of parameters to be estimated. By contrast, a coarse-grained approach is much more computationally tractable than explicitly accounting for the complexity of metabolism. This is particularly true for minimal approaches such as proposed in [20], which constructed a C-GRAM that incorporated nitrogen metabolism and described proteome allocation under optimal growth, and [27], which additionally included metabolites into the system size.

Because of their tractability, C-GRAMs are also well suited to exploring hypotheses about the interplay between metabolism, growth and gene expression [30], and finding insights on the general principles behind the physiology of unicellular organisms [6]. They can furthermore be designed in an organism-agnostic manner and thereby provide an opportunity to compare microbes based on those model parametrizations that best explain global observations about growth in each organism. In summary, the coarse-grained approach allows one to directly interpret and explore model parameters, with minimal need for explicit or large-scale parameter inference.

Earlier C-GRAMs chiefly considered carbon modulations representative for the effect of the nutrient quality in general. In such models, metabolism was typically considered as a linear pathway from nutrient to protein production. This included the models proposed in [24] and in [27]. Contrasting with this earlier work, a strategy commonly employed in *Schizosaccharomyces pombe* to modulate the growth rate uses ammonium chloride and a variety of amino acids as nitrogen sources [17,31]. Using this strategy, we have recently reported

differential expression for many enzymes involved in carbon metabolism across nitrogen sources, even though abundant glucose was provided in all conditions [17]. Different amino acids have also been used as sole nitrogen sources in order to modulate growth while studying the proteome of *Saccharomyces cerevisiae* [32] and of three bacterial strains found in the *Arabidopsis* rhizosphere [33].

We aimed to better understand the effect of nitrogen-source modulation on resource allocation in a coarse-grained modelling context. We took a minimal approach, opting to construct a model with fewer parameters and therefore omitting transcription, the distinction between energy and amino acid metabolism. We extended the metabolic model proposed in [27] to include the uptake and metabolism of carbon-containing nitrogen sources. We studied how a growth-rate-maximizing allocation towards proteome sectors varied between steady states that were imposed by choices of parameters representing different nutrient conditions.

2. Results

In this study, we aim to investigate optimal resource allocation behaviour under growth in defined media with varying nitrogen sources. This, for example, will be relevant for growth media where different amino acids act as the sole source of nitrogen. Amino acids generally consist of an amino group and a ketoacid backbone, and we therefore focused on modelling nitrogen and carbon metabolism. We constructed a C-GRAM with pathways representing nitrogen and carbon uptake, protein biosynthesis, and the excretion and recycling of carbon from carbon-containing nitrogen sources. The model was formulated dynamically using ordinary differential equations (ODEs), their steady state representing balanced growth. The steady-state growth rate was calculated from the solution to the ODEs as an emergent property.

Pathways in the model were represented by enzymes with simple kinetics, and the parameters in this representation were chosen to reflect nutrient conditions. For example, carbon limitation was modelled by reducing the catalytic rate of the carbon uptake enzyme. Furthermore, nitrogen sources were distinguished by the catalytic rates of the enzymes metabolizing them as well as their stoichiometries, related to their elemental carbon-to-nitrogen ratio. For each nutrient condition (parametrization), we determined the resource allocation that maximized the growth rate. The details regarding the model implementations, namely the full formulation of the ODEs, considerations with regards to parameter choices, and the approach used to optimize resource allocation are described in the Methods.

As described in the remainder of this manuscript, we explored the effect of different growth-rate modulations on the optimal allocation. We first considered the behaviour of two submodels that lacked the recycling and excretion of the ketoacid backbone. This enabled us to explore trade-offs in resource allocation in media with simple sources of carbon (such as glucose) and nitrogen (such as ammonium chloride). In the first submodel, no distinction was yet made between respirofermentative and purely respiratory metabolism; in the second submodel, two parallel pathways representing the two were introduced. Finally, the full model was used to describe growth on amino-acid nitrogen sources. As expected, various nutrient limitations induced complex trade-offs between allocations to different enzyme fractions.

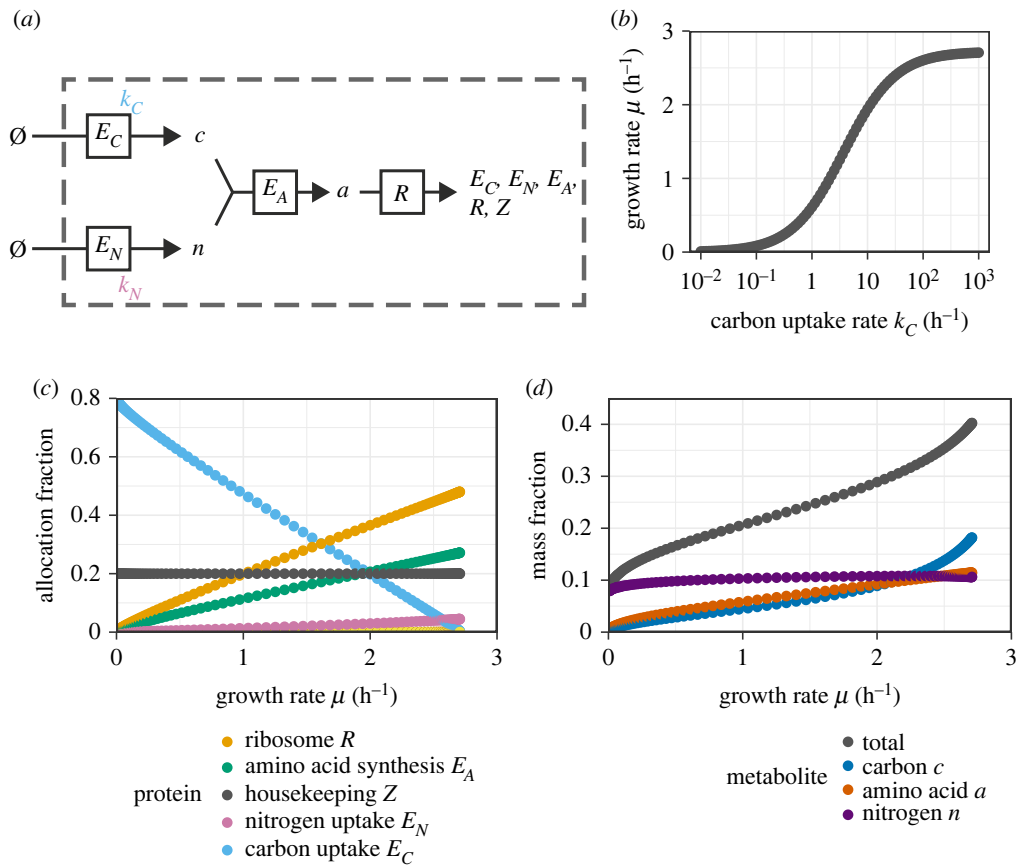


Figure 1. Analysis of the core metabolic model under modulation of the carbon uptake rate k_C . The uptake rate parameter k_C of the transporter was modulated to yield different growth rates. The allocation fractions were chosen to maximize the growth rate for each chosen k_C . (a) Illustration of the model. (b) Growth rate μ as a function of the carbon uptake rate k_C . (c) Optimal protein allocation fractions as a function of the growth rate μ . (d) Steady-state mass fractions of metabolites as a function of the growth rate μ .

Table 1. Variables (metabolites and proteins) implemented in the full model.

type	description	concentration	allocation fraction
metabolite	ketoacid	k	n.a.
	carbon	c	
	amino acid	a	
	free nitrogen	n	
protein	ketoacid recycler	$e_{K_{re}}$	$f_{K_{re}}$
	ketoacid excreter	$e_{K_{ex}}$	$f_{K_{ex}}$
	carbon uptake	e_C	f_C
	respirofermentative enzyme	e_{A_r}	f_{A_r}
	purely respiratory enzyme	e_{A_r}	f_{A_r}
	nitrogen-compound uptake	e_N	f_N
	ribosome	r	f_R
	housekeeping (non-metabolic)	z	f_Z

2.1. Modulation of the carbon uptake rate with one metabolic pathway

The first submodel we explored was one with parallel uptake of carbon and nitrogen, which were combined by a single amino-acid-producing enzyme. An illustration of the model is provided as figure 1a and the interpretation of all model constituents (proteins and metabolites) is provided in

table 1. This model was explored by varying the carbon uptake rate parameter k_C . All model parameters are listed in table 2.

As expected, the dependency of the growth rate on the modulation parameter approximated a Monod curve (figure 1b), and the optimal protein allocation fractions were approximately linearly correlated with the growth rate (figure 1c). The total relative metabolite abundance in steady state amounted to

Table 2. Parameters (Michaelis constants, enzyme efficiencies and stoichiometries) implemented in the full model.

type	description	symbol	default value
Michaelis constant	ketoacid	k_{sat}	0.0167
	carbon	c_{sat}	0.0167
	amino acid	a_{sat}	0.0167
	free nitrogen	n_{sat}	0.0167
enzyme efficiency	ketoacid recycler	$k_{K_{\text{re}}}$	10.0 h^{-1}
	ketoacid excreter	$k_{K_{\text{ex}}}$	20.0 h^{-1}
	carbon uptake	k_C	10.0 h^{-1}
	respirofermentative enzyme	k_{A_f}	15.0 h^{-1}
	purely respiratory enzyme	k_{A_r}	7.5 h^{-1}
	nitrogen-compound uptake	k_N	20.0 h^{-1}
	ribosome	k_R	6.46 h^{-1}
mass stoichiometry	carbon consumed per amino acid produced in respirofermentative metabolism	α_C	48/55
	carbon consumed per amino acid produced in purely respiratory metabolism	α_C	24/31
	ketoacid produced per nitrogen compound consumed	γ_K	0

approximately 10%–30% of biomass (figure 1d). In experimental cultures, metabolites comprise only a small fraction of the biomass (BNID 111490; [34,35]), so in this respect the growth-optimized model was a poor approximation for conditions where this occurred. This may reflect the current parametrization of the model to be inaccurate.

Importantly, under modulation of the carbon uptake enzyme, the allocation fractions of the nitrogen assimilation enzyme E_N and the amino-acid biosynthesis enzyme E_A were positively correlated with the growth rate (figure 1c). This behaviour was also implemented in [20], but it is notably different from the model described in [24], which assumed that the transporter and metabolic enzyme were both regulated identically.

Proteins associated with translation are estimated to constitute approximately 45% of the total protein mass in the fastest-growing *Escherichia coli* cultures [13,15,19], which reach maximal growth rates of around 2.0 h^{-1} . A minimal partitioning of the proteome based on growth rate correlations suggested that around half of the proteome mass does not change with the growth rate [13,21]. The earlier model in [24] reflected this by implementing protein expression regulation with a metabolic sector, a translational sector and a constitutive sector. Its proteome comprised two sequential enzymes (corresponding to our E_C and E_A), ribosomes (our R), and housekeeping proteins (our Z). In their parametrization, the latter took up around 70% of the total protein mass, and were negatively correlated with the growth rate.

By contrast, in the parametrization used to generate figure 1, the allocation towards housekeeping proteins was only 20% of the proteome. However, this parametrization was chosen such that the maximal ribosomal allocation agreed reasonably well with *E. coli* data from [13]. The discrepancy between the two models in their estimated housekeeping allocations is explained by the considerable cost of the non-modulated enzymes. At maximal growth, the amino-acid synthesis and nitrogen-uptake enzymes amount to around 25% and 5% of the proteome, respectively.

2.2. Modulation of carbon and nitrogen uptake rates with two parallel metabolic pathways

Next, we considered that a key determinant of unicellular growth is the metabolic state, with a principal difference whether fermentation occurs or whether biosynthesis is maintained in a purely respiratory fashion. Fast-growing cells express both fermentative and respiratory pathways, with the former the primary generator of free energy (in the form of ATP). When growing on poorer or less abundant carbon sources, cells rely on respiration for energy generation, but respiratory enzymes and the citric acid cycle are involved in amino acid biosynthesis even when free energy is primarily generated by fermentation [36]. Comparing the two pathways at equivalent energy generation, the fully respiratory pathway poses a greater burden on gene expression. On the other hand, the respirofermentative pathway requires more carbon uptake from the environment, because it is converted into overflow metabolites such as acetate or ethanol. We aimed to construct a minimal model allowing for these two different metabolic states, and wondered how the optimal state would be affected by nitrogen and carbon limitation.

Deciding against introducing further metabolites and enzymes to the model, we instead implemented two parallel amino-acid synthesis pathways. The parametrization of these parallel pathways reflected the above considerations about carbon uptake and gene expression burden. First, we implemented the purely respiratory enzyme E_{A_r} with a lower enzyme efficiency than the respirofermentative E_{A_f} . Second, we adjusted the carbon stoichiometry, such that E_{A_f} required additional carbon per nitrogen atom consumed (eight atoms instead of four; for details, see Methods §4.3), representing overflow metabolism without implementing it as an additional pathway. The allocation fractions to both enzymes were included in the growth rate optimization. Our parametrization implied that E_{A_r} required less carbon substrate, but a higher expression to sustain the same synthesis rate, than E_{A_f} .

We modulated the carbon and nitrogen transporter rate parameters in the model containing two parallel metabolic

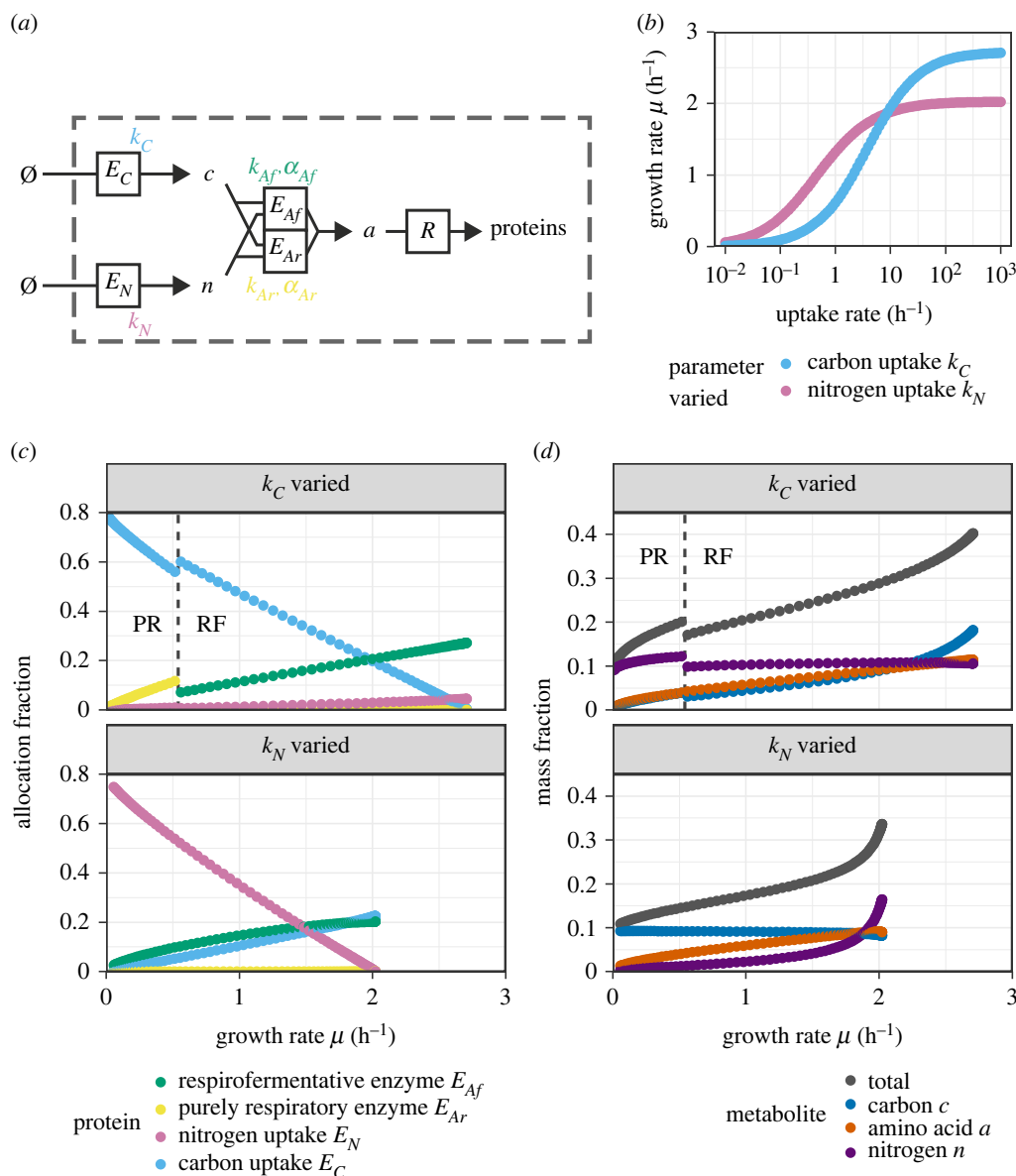


Figure 2. Analysis of the model with two parallel metabolic enzymes (respirofermentative enzyme E_{Af} and purely respiratory enzyme E_{Ar}) under the separate modulation of the carbon and nitrogen transporter uptake rates k_C and k_N , respectively. Allocation fractions were chosen to maximize the growth rate for each chosen k_C (c and d , top panels) and k_N (bottom panels). The uptake rates were varied separately between 1.0×10^{-2} and 1.0×10^3 h⁻¹. (a) Illustration of the model. (b) Growth rate μ as a function of the respective uptake rate. (c) Optimal protein allocation fractions as a function of growth rate μ for the two parameter explorations. Ribosomes R and housekeeping proteins Z were omitted from the figure to improve clarity. (d) Steady-state mass fractions of metabolites as a function of growth rate μ . The transition between purely respiratory (PR) and respirofermentative (RF) growth is indicated by a dashed line.

pathways with different parameterizations (figure 2a). As before, the growth rate depended on the modulation parameters approximately following a Monod curve (figure 2b). Because the two pathways are parallel, the usage of one of the two was preferred over the other in each parametrization, such that one of E_{Ar} and E_{Af} was expressed while the other was not. As shown in figure 2c, the purely respiratory enzyme was induced by low values of the carbon uptake rate k_C . We note a sharp and discontinuous transition from respiration to fermentation at a specific level of carbon uptake rate (figure 2c). This agrees with results of a model similar in scope to ours, where the optimal choice between a metabolically efficient or catabolically efficient pathway was shown to depend on the growth rate [20]. In contrast to this behaviour under modulations of the carbon uptake, the respirofermentative enzyme was present in all conditions where the nitrogen uptake efficiency k_N was varied from its default.

Both the internal carbon abundance c and the nitrogen abundance n varied discontinuously with the growth rate near the transition point ($\mu \approx 0.5$ h⁻¹). This indicates how purely respiratory metabolism required a higher abundance of both of its substrates (c and n) than respirofermentative metabolism to sustain the same turnover. By contrast, the abundance of the product, the amino acid equivalent a , varied almost continuously with the growth rate even at the transition point. The relation between a and the ribosomal allocation will be explored in detail in a further section of this paper.

Under perturbations of the nitrogen transporter uptake rate k_N (figure 2c, bottom panel), the allocation towards the metabolic enzyme varied nonlinearly with the growth rate. In addition to this, the amount of nitrogen metabolite n built up much more strongly with increasing growth rate via k_N than the equivalent carbon build-up under perturbations of k_C (figure 2d). Note that modulating each transporter's

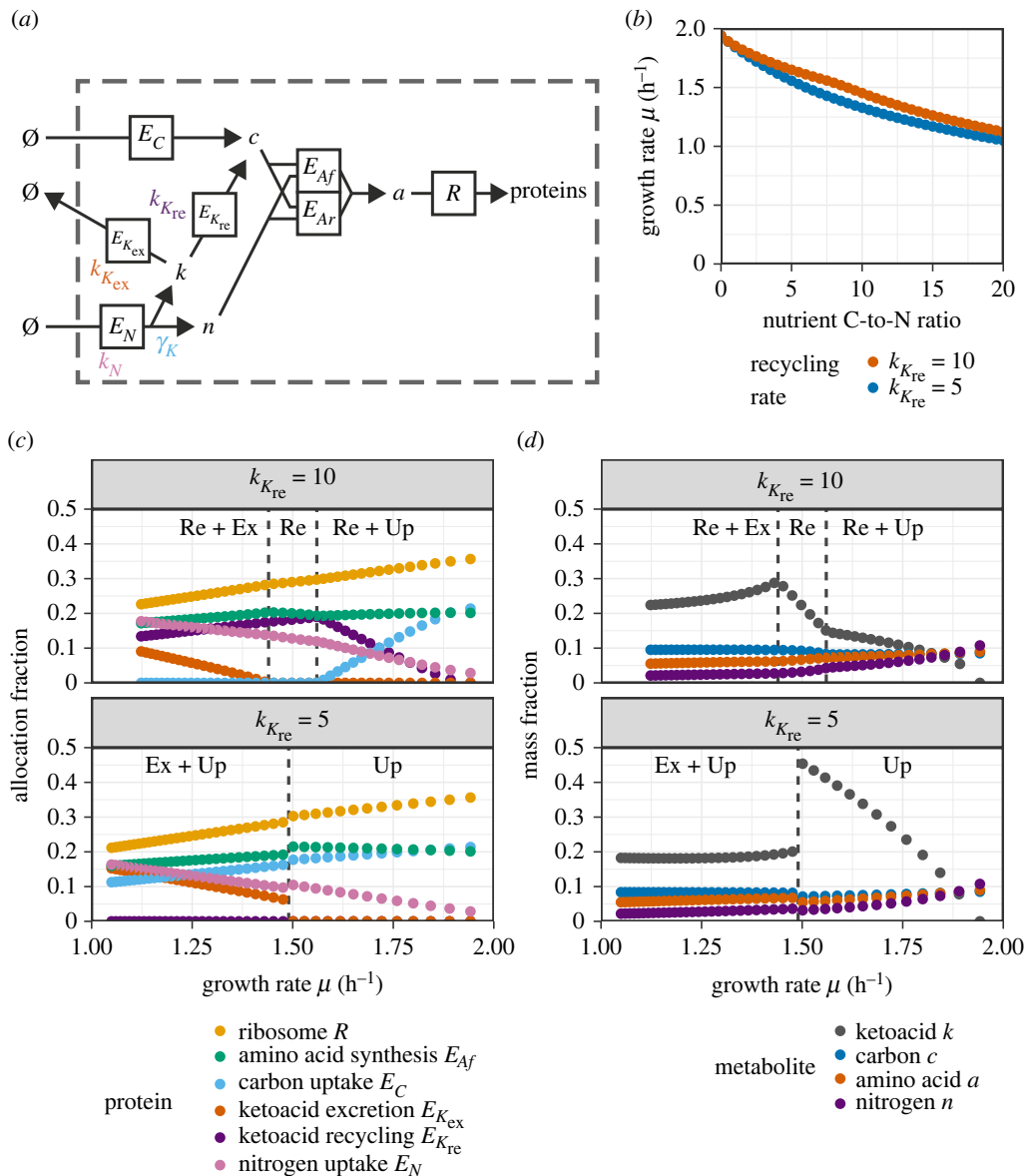


Figure 3. Analysis of the full model, including recycling and excretion of ketoacid under modulation of the carbon-to-nitrogen ratio in the nitrogen source for two choices of the ketoacid recycling rate $k_{K_{re}}$. The carbon-to-nitrogen ratio was varied between 0 and 20 in steps of 0.5 and the ketoacid stoichiometry γ_K was calculated from this as described in the text. Allocation fractions were chosen to maximize the growth rate for each γ_K and $k_{K_{re}}$. (a) Illustration of the model. (b) Growth rate μ as a function of the nitrogen source's carbon-to-nitrogen (C-to-N) ratio for the respective ketoacid recycling rates. (c) Optimal protein allocation fractions as a function of growth rate μ for the two recycling rates. Purely respiratory enzymes E_{A_r} and housekeeping proteins Z were omitted from the figure to improve clarity; E_{A_r} was not expressed for any of the parametrizations shown. (d) Steady-state mass fractions of metabolites as a function of growth rate μ . Transitions between metabolic states are indicated by dashed lines (Re, ketoacid recycling; Ex, ketoacid excretion; Up, carbon uptake).

efficiency (k_N and k_C) repressed the abundance of its own substrate (n and c , respectively). On the other hand, the modulation of one transporter rate barely affected homeostasis of the metabolite in the other pathway (c and n , respectively).

2.3. Recycling and excretion of ketoacids disturbed carbon metabolism

In the above, we restricted ourselves to the metabolism of freely usable nitrogen, which typically comes in the form of ammonium ions (NH_4^+). However, the cellular growth rate can also be perturbed considerably by using different nitrogen sources, particularly amino acids, as the sole nitrogen source instead of NH_4^+ [17,31]. These pathways typically deaminate or transaminate the amino acid, either by a single enzyme or as the net result of a longer pathway [37]. Nitrogen

is assimilated as free NH_4^+ (after deamination) or glutamate (after transamination). The leftover carbohydrate, usually a ketoacid, may be recycled into biomass or excreted.

To account for this process in the model, we implemented the nitrogen uptake pathway as an enzyme that produced both free nitrogen and a carbon-containing ketoacid metabolite. To represent different complex nitrogen sources such as amino acids, we perturbed the parameter γ_K , which represents the relative mass of recycled ketoacid with respect to the total mass taken up by the nitrogen uptake enzyme. Additionally, we introduced two enzymes, $E_{K_{re}}$ and $E_{K_{ex}}$ that, respectively, recycled the ketoacid into usable carbon precursor or excreted it from the cell. The model is illustrated in figure 3a.

The behaviour of the model that included ketoacids depended qualitatively and quantitatively on the values of key parameters, particularly on the efficiencies (catalytic

rates) of the enzymes involved and on the ketoacid stoichiometry γ_K . Recycling enzymes were only expressed if their efficiency (catalytic rate) $k_{K_{re}}$ was large. We first explored the model behaviour for two choices of this parameter, modulating γ_K but leaving fixed all other metabolic parameters. The results are shown in figure 3*b–d*.

Based on the types of proteins that were expressed, there were three qualitatively different growth regimes for $k_{K_{re}} = 10$ and two for $k_{K_{re}} = 5$. One important trade-off here is between the cellular cost of expressing ketoacid processing enzymes and the cost of carrying the excess metabolite for cell growth. When recycling enzymes were inefficient and costly ($k_{K_{re}} = 5$, bottom panels), but the carbon-to-nitrogen ratio was still below a certain threshold, the ketoacid metabolites built up considerably (up to almost half the total biomass). In this regime, this still led to faster growth than expressing either recycling or excretion enzymes would have. However, with the amino acid nitrogen source containing relatively more carbon, the growth-optimized cell excreted ketoacids and the internal concentration were approximately stable across the range of growth rates investigated. When recycling enzymes were efficient ($k_{K_{re}} = 10$, top panels), recycling replaced the uptake of carbon through the canonical carbon pathway, i.e. no carbon uptake enzyme E_C was expressed below a critical growth rate. For both choices of $k_{K_{re}}$, in the fastest-growing regime, no excretion took place, i.e. the excretion enzyme $E_{K_{ex}}$ was not expressed; but in the slowest-growing regime, ketoacid excretion was required to optimize growth. Curiously, we found a recycling-only regime with neither canonical carbon uptake nor ketoacid excretion. In this latter regime, all carbon in the biomass has its origin in the amino acid from the nutrient, and all the carbon from the nutrient made its way into the cell.

We note that across all simulations, no respiratory enzyme E_{Ar} was expressed. Furthermore, the optimal allocation to ribosomal proteins deviated discontinuously from the ribosomal growth law for both choices of $k_{K_{re}}$ when the ketoacid stoichiometry γ_K was varied.

2.4. Different nutrient environments (parametrizations) induced complex trade-offs between carbon uptake, ketoacid recycling and excretion

We further explored the five different growth regimes across a more extensive parameter sweep. The ketoacid stoichiometry γ_K is related to the ratio of carbon and nitrogen atoms in a nutrient molecule

$$\gamma_K = \frac{n_C \times 12}{n_C \times 12 + n_N \times 14} = \frac{12(n_C/n_N)}{12(n_C/n_N) + 14}, \quad (2.1)$$

where n_C and n_N are the numbers of carbon and nitrogen atoms in a nutrient molecule, and the factors 12 and 14 account for the approximate molar masses of these two elements. For example, glycine molecules contain two carbon atoms for each nitrogen atom, which would be represented by $\gamma_K = 2 \times 12 / (2 \times 12 + 14) = 12/19 \approx 0.63$, whereas each molecule of isoleucine contains nine carbon atoms, for $\gamma_K = 9 \times 12 / (9 \times 12 + 14) = 54/61 \approx 0.89$.

Alongside the ketoacid stoichiometry γ_K , we varied the maximal turnover rates (efficiencies) k_N , $k_{K_{ex}}$ and $k_{K_{re}}$ of the enzymes involved in ketoacid metabolism, as plotted in figure 4. This showed that the five regimes highlighted in

the previous section were universal. The optimally allocated cell expressed only one or two out of the ketoacid recycling, ketoacid excretion, and carbon uptake enzymes depending on the parametrization (figure 3). Neither recycling nor excretion was expressed in nitrogen sources not also containing carbon, and additionally this was optimal even for carbon-containing nitrogen sources when the ketoacid recycling and excretion enzymes were inefficient relative to the nitrogen uptake enzyme (figure 4*a,d,e*). The value of the ketoacid recycling rate $k_{K_{re}}$ below which ketoacid recycling was suboptimal depended weakly on the ketoacid stoichiometry (figure 4*a*). When recycling enzymes were expressed, the trade-off between the excretion and ketoacid uptake was heavily influenced by the nitrogen source's carbon content and all three enzyme rates (figure 4*a–c*). Low-carbon nutrient sources required additional carbon uptake whereas high-carbon nutrient sources generally required ketoacid excretion, with pure recycling being favoured in regimes with intermediate carbon content or inefficient excretion.

2.5. Approximately optimal allocation towards ribosomes could result from amino-acid regulation

Until now, we have explored the model using defined modulations of one or more rate parameters and the ketoacid stoichiometry. Across all of these, we found that the ribosomal growth law (ribosomal proteome allocation fraction $f_R \propto \mu$) was robustly satisfied. Moreover, the amino acid abundance a appeared to be similarly correlated with the growth rate. We wondered if this was an artefact of our parametrization approach or a deeper property of the model and used a random sampling parametrization strategy to further study this behaviour. We sampled 100 triplets of the rate parameters (k_C, k_N, k_{Ar}) from independent uniform distributions with support $[0, 20]$, set $k_{Ar} = 0.5k_{Ar'}$, and matched these samples to four choices of the ketoacid stoichiometry γ_K , representing carbon-to-nitrogen ratios of 0, 3, 6 and 12, and two choices of the ketoacid recycling efficiency $k_{K_{re}}$.

In the ketoacid-free model ($\gamma_K = 0$) with random rate parameters, the ribosomal growth law was nearly exactly satisfied (figure 5*a*), even though expression of the two transporters and the metabolic enzyme was highly variable between parameter choices. In a recent experimental study, we also observed this contrast between the linear ribosomal growth law and variable expression of metabolic classes [17]. In figure 5*b*, we show the expression of protein classes mapped to their three C-GRAM equivalent: ribosomes (translation and ribosome biogenesis from [17]), carbon uptake (glycolysis), and amino acid synthesis (precursors and energy generation, and amino acid metabolism).

Furthermore, the concentration of the amino acid a was closely related to growth rate as well (figure 5*c*). It follows from the growth rate correlations for both f_R and a that the two were correlated themselves as well (figure 5*d*). A linear fit $f_R \propto (a - a_0)$ closely approximated the observed relation.

We further explored the relation between f_R and a in the presence of complex nitrogen sources such as amino acids, when additionally the recycling and/or excretion of ketoacid is accounted for (figure 5*d*). For large carbon-to-nitrogen ratios, the relation between f_R and a may be better described by a nonlinear relation, although the fit for $f_R \propto (a - a_0)$ obtained for nitrogen sources without carbon was still a

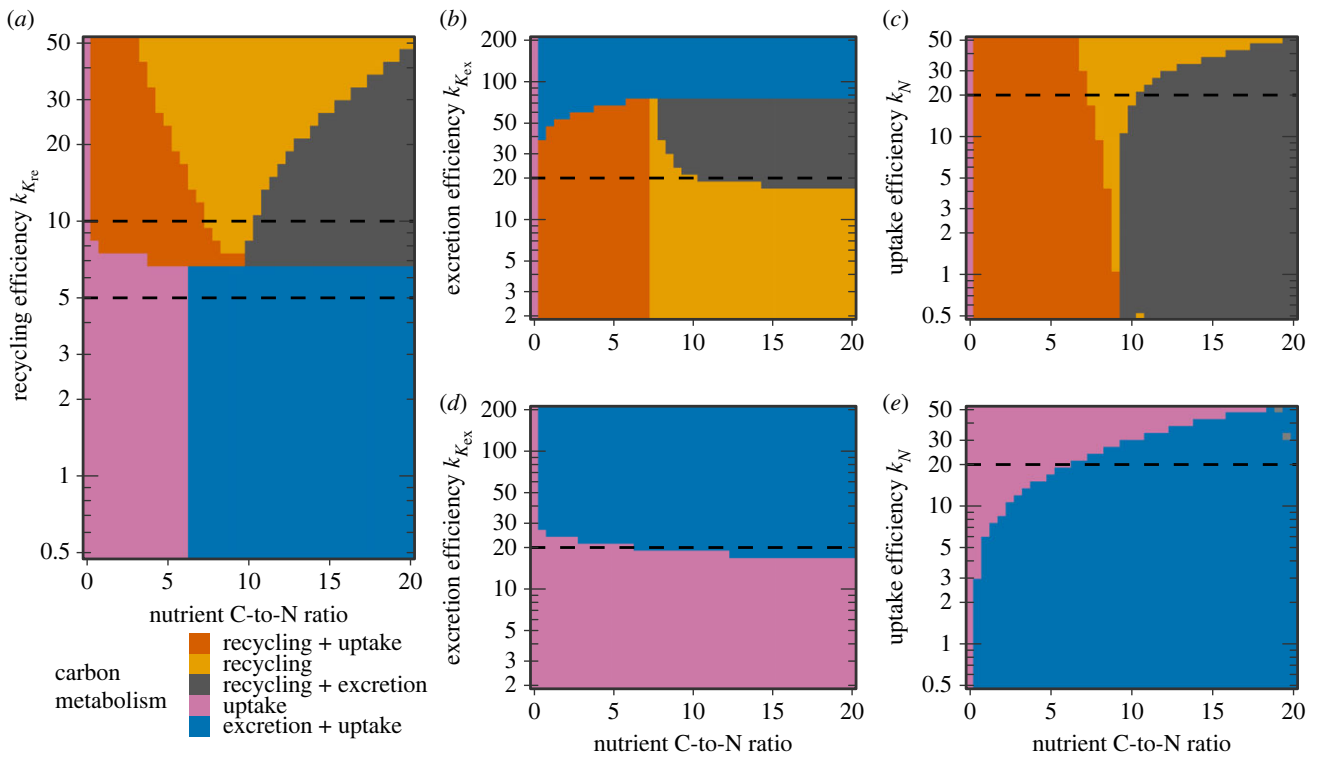


Figure 4. Phase diagrams of ketoacid recycling, excretion and carbon uptake under perturbations of enzyme efficiencies and nutrient carbon-to-nitrogen ratio. For each figure, the optimal allocation was determined for 41^2 combinations of the ketoacid stoichiometry γ_K and one of the enzyme efficiencies $k_{K_{re}}$, $k_{K_{ex}}$ and k_N . As before, the carbon-to-nitrogen ratio was varied between 0 and 20 in steps of 0.5 and the ketoacid stoichiometry γ_K was subsequently calculated. Furthermore, the enzyme efficiencies were chosen to be equidistant after log-transformation. In the figure, colours indicate whether ketoacid recycling, ketoacid excretion and carbon uptake enzymes were expressed in the optimal allocation, and dashed lines indicate parameter values that were fixed in the other panels and in figure 3. (a) Ketoacid recycling rate $k_{K_{re}}$ varied, ketoacid excretion rate $k_{K_{ex}} = 20.0 \text{ h}^{-1}$ and nutrient uptake rate $k_N = 20.0 \text{ h}^{-1}$ held fixed. (b) $k_{K_{ex}}$ varied, $k_{K_{re}} = 10.0 \text{ h}^{-1}$ and $k_N = 20.0 \text{ h}^{-1}$ fixed. (c) k_N varied, $k_{K_{re}} = 10.0 \text{ h}^{-1}$ and $k_{K_{ex}} = 20.0 \text{ h}^{-1}$ fixed. (d) and (e) as b and c, but with $k_{K_{re}} = 5.0 \text{ h}^{-1}$.

close approximation. Only when the optimal solution involved ketoacids building up in the cell without recycling or excretion did the linear fit break down entirely. Since our model does not account for the toxicity of the intermediate ketoacids beyond their passive drag on growth, this situation is not likely to occur in real cells.

All together with the previous sections, these results suggest that a single linear regulation of ribosomes by amino acids can achieve near optimal allocation of ribosomes. For the regimes explored here, a simple proportionality $f_R = \delta a$ such as proposed in [27] would work reasonably well, though our introduction of an offset a_0 (i.e. $f_R = \delta(a - a_0)$) improved the fit. Interestingly, if $f_R \propto a$ is chosen and the ribosomes are implemented with Michaelis–Menten kinetics, the ribosomal growth law follows analytically. A derivation of this statement is presented in electronic supplementary material, text S4; it holds whether or not there is an offset a_0 in the f_R – a relationship. We note that we did not observe a simple relationship between the allocation of other non-ribosomal proteins and the metabolite concentrations.

3. Discussion

A holistic understanding of growth, gene expression and resource allocation can be codified in and achieved by C-GRAMs. Here, we constructed a minimal C-GRAM of microbial metabolism that accounted for (i) the metabolism of both carbon and nitrogen, (ii) the different proteomic and kinetic efficiencies of respirofermentative and respiratory

energy metabolism and (iii) the stoichiometry of complex nitrogen sources that contain a carbon backbone, such as amino acids. We optimized the allocation towards different protein classes so as to maximize the growth rate.

Different parameter sweeps of the model allowed us to explore different nutrient environments. The optimal allocation and resulting growth rate co-varied according to strikingly regular behaviour. In particular, the Monod law and ribosomal growth law were very robust. Furthermore, we described growth on carbon-containing nitrogen sources by introducing the possibility of recycling and/or excreting residual carbon, which induced complex trade-offs in metabolism. Notably, both the optimal allocation towards ribosomes and their substrate (internal amino acids) varied approximately linearly with the growth rate, leading to an also approximately linear relationship between the two.

3.1. Protein reserves and simple feedback of free amino acids setting ribosome allocation

Our assumption that expression of all proteins was optimized for growth in any given condition considerably simplified the model parametrization. Precise optimality conditions have been formulated for the *E. coli* carbon uptake system and gene expression indeed maximized the growth rate for several carbon sources [38]. Furthermore, a recent theoretical advance pointed out a general method of adapting gene expression control towards the optimum [39]. However, recent evidence has challenged the view that all allocation is growth-optimal. It is thought that significant fractions of

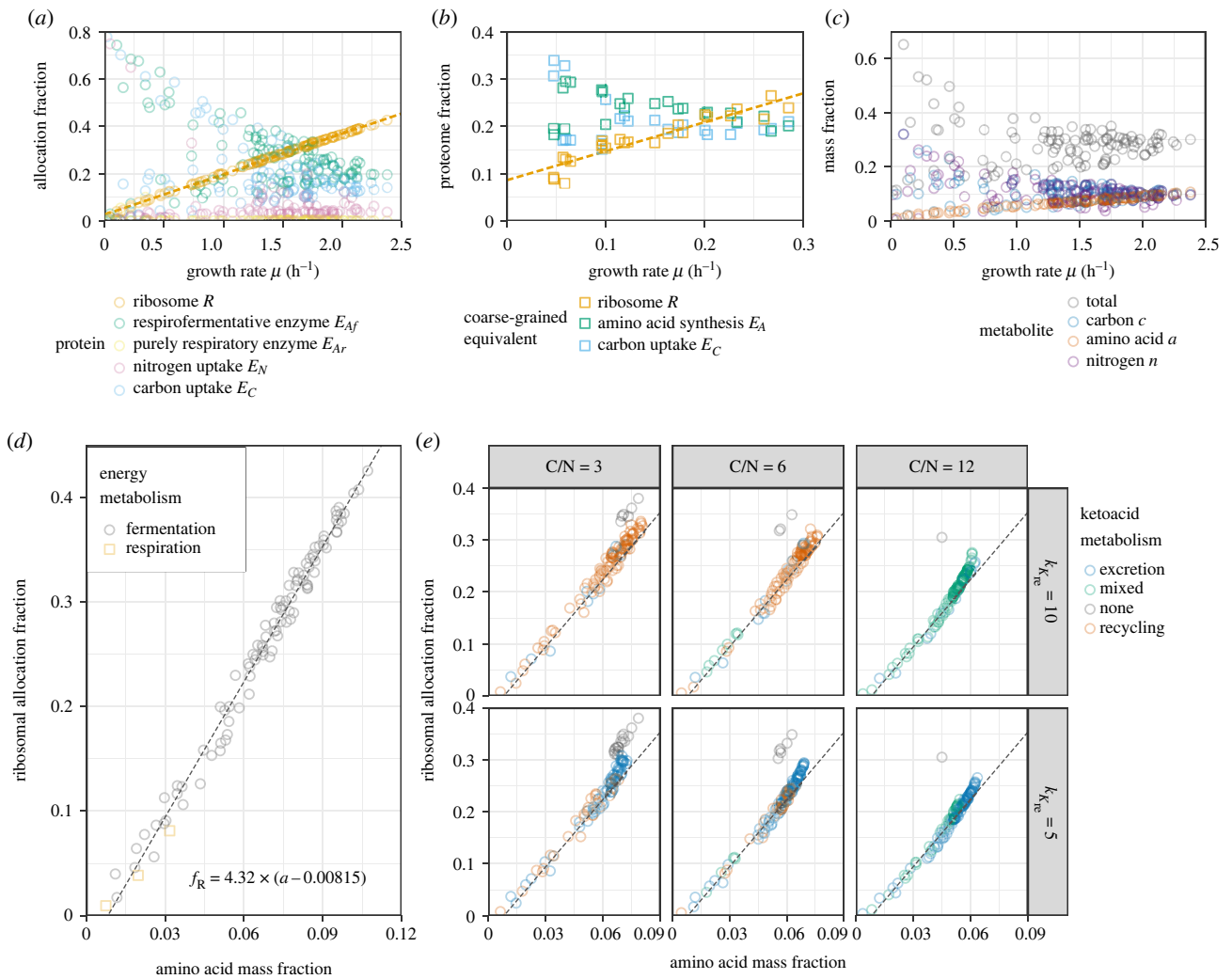


Figure 5. Analysis of the full model with randomly chosen rate parameters, exploring the relation between amino acid precursor abundance and ribosomal allocation. The enzyme efficiencies k_{A_f} , k_C and k_N were randomly drawn from separate uniform distributions ranging between 0 and 20.0 m^{-1} , and the respiratory enzyme efficiency k_{A_r} was set to $0.5k_{A_f}$. The ketoacid recycling efficiency k_{Re} and carbon-to-nitrogen ratio were initially fixed to 10.0 h^{-1} and 0, respectively (for *a*, *c* and *d*), and then varied jointly (for *e*). The allocation fractions were chosen to maximize the growth rate for each set of parameters. (*a*) Optimal protein allocation fractions as a function of the growth rate μ . Housekeeping proteins and ketoacid-processing enzymes were omitted to improve clarity. The best-fit of a linear regression model to the ribosomal allocation fraction is drawn as a dashed line to guide the eye. (*b*) Experimentally observed proteome mass fractions from a *Schizosaccharomyces pombe* study, mapped to their coarse-grained equivalents. (*c*) Steady-state biomass fractions of amino acid *a*, carbon *c*, and free nitrogen *n* plotted against the growth rate μ . (*d*) Scatter plot of the amino acid concentrations *a* and ribosomal allocation fractions f_R , indicating the type of energy metabolism that optimized growth (fermentation: only E_{A_f} expressed; respiration: only E_{A_r} expressed). The dashed line represents an ordinary least-squares linear model fit to the data. (*e*) As *d* for three non-zero choices of the carbon-to-nitrogen ratio (C/N) and two choices of the ketoacid recycling efficiency k_{Re} . Points are coloured according to whether none, one, or both of the ketoacid excretion enzyme $E_{K_{ex}}$ and ketoacid recycling enzyme $E_{K_{re}}$ was expressed. The dashed line indicates the fit from *d* (no further fitting was performed here).

the *E. coli* [40–42] and budding yeast [16,43] proteome are not necessary for sustaining the growth rate. In particular, central carbon metabolism has been suggested to have a large reserve capacity, suggesting that many enzymes are not used solely to maximize metabolic fluxes [32,44,45]. Pools of proteins held in reserve may be beneficial instead by their ability to support adaptation to environmental changes.

An intermediate step between a fully growth-optimized and a dynamically regulated allocation model may be the proportional regulation $f_R \propto a$. We showed this to be a good approximation to the growth-rate-maximizing allocation; additionally, it was robust to many variations of the parameters describing the nutrient quality. The simplicity of this relation is remarkable: in principle, optimal allocation could depend on all internal metabolite concentrations and be highly nonlinear, instead of this linear

dependence on only a single metabolite. Explicit regulation of bacterial ribosome synthesis mediated by a single metabolite, guanosine tetraphosphate (ppGpp), was thoroughly explored in a coarse-grained model by [46]. Recently, ppGpp was shown to regulate the growth rate and ribosome content in *E. coli* by sensing the instantaneous translation elongation rate [47].

Notably, neither our growth-optimized model nor one implementing ribosomal allocation proportional to internal amino acids explicitly account for a reserve pool of ribosomes not actively involved in translation. Such a pool has been held responsible for the observed offset ϕ_{R0} in the ribosomal growth law $\phi_R = \phi_{R0} + \sigma^{-1}\mu$ [16,42,48]. Although our model did not implement any inactive ribosomes, we still observed an offset greater than zero. We explain this effect by the fact that the ribosomes in our model are not fully saturated with

substrate, even when growth is optimized. Specifically, linearly correlated ribosome allocation and amino-acid abundance (either through optimization or explicit regulation), combined with nonlinear Michaelis–Menten ribosome kinetics, resulted in an offset

$$\phi_{R0} = \delta a_{\text{sat}}, \quad (3.1)$$

when $f_R = \delta(a - a_0)$ and $f_R \propto a / (a + a_{\text{sat}})$. Naturally, ribosome saturation and inactivation are not mutually exclusive, and a quantitative understanding of their relative importance will have profound implications on our understanding of the interplay between ribosome synthesis, excess translational capacity and cell growth [49].

3.2. The fate of the carbon backbone for complex nitrogen sources such as amino acids

We next discuss the fate of the carbon backbone from complex nitrogen sources, as modelled first in the model presented here. In our model, modulations of the ketoacid stoichiometry γ_K gave rise to a wide range of growth rates in a monotonically decreasing manner. However, a nitrogen source's quality is not solely determined by its carbon-to-nitrogen ratio. For example, glycine and tryptophan media gave rise to very similar growth rates in *S. pombe* ([17], but carbon is present in a 2 : 1 ratio in glycine and in a 5.5 : 1 ratio in tryptophan (see electronic supplementary material, text, table S2). This strongly suggests that each growth medium is not only associated with the ketoacid stoichiometry γ_K , but that at least one out of the enzymatic rates k_N (nitrogen uptake), $k_{K_{\text{ex}}}$ (ketoacid excretion) and $k_{K_{\text{re}}}$ (ketoacid recycling) must also be modulated by the choice of nitrogen source.

Unlike translation and central carbon metabolism, the topology of amino acid uptake pathways is rather poorly conserved between organisms. In practice, then, whether recycling or excretion is preferred for a given amino acid nitrogen source depends on which reactions are available to the organism, and how efficient they are. If ketoacid recycling effectively feeds into other synthesis pathways, this would correspond to a large $k_{K_{\text{re}}}$ in the model; on the other hand, the effectiveness (or absence) of suitable excretion pathways would influence the value of $k_{K_{\text{ex}}}$.

While our model provides a framework for understanding the optimality of recycling and excretion of carbon, we refrained from a full parametrization of growth on specific nitrogen sources. We note that despite this limitation, our model agrees with metabolic gene expression being medium-specific rather than correlated with growth rate, as has indeed been recently reported in *S. pombe* growing on different amino acids (figure 5*a,b*).

In terms of possible biotechnological applications, we note that carbon uptake from complex nitrogen sources may be advantageous from a yield perspective. However, its effect on the growth rate is generally deleterious as reflected by the model (see e.g. figure 3*b*). As our model shows, choosing a complex nitrogen source will affect carbon metabolism, though it probably will not repress fermentation. To study the interplay between metabolism and the expression of synthetic constructs, our C-GRAM could be extended by implementing an additional coarse-grained protein in the manner of [24].

Understanding the balance between carbon uptake, recycling and excretion is also important for applications where the product consists mostly of excretions, such as brewing. The metabolism of indigestible ketoacids has been well studied in *S. cerevisiae*, whose excretions of such 'fusel oil' can spoil the product [37]. While the carbon-to-nitrogen ratio is specific to the nutrient, our results shown in figure 4 suggest that the balance between excretion and recycling is not only affected by the respective efficiencies of these two processes itself, but also the uptake efficiency. In other words, if a complex nitrogen source contains relatively much carbon, it may still be effectively recycled if it is otherwise efficiently assimilated.

To get a better idea of the exact molecules and pathways involved, these general considerations should be supplemented with whole-genome models of metabolism. It is challenging to condense those into C-GRAMs due to the complexity of metabolism; the presence of many parallel pathways, moonlighting enzymes and metabolic cycles are just three examples of this complexity. However, some progress may be made in adapting our generic C-GRAM to specific organisms. Recently, high-quality metabolic maps that are aware of limited resource allocation in multiple cellular compartments have been developed for *S. cerevisiae* and *S. pombe* [50,51]. Such maps may enable comparisons between the coarse-grained proteome sectors and proteome data in the future.

3.3. Explicit overflow metabolism and energy generation

We accounted for the distinction between purely respiratory and respirofermentative growth by adjusting parameters of the enzyme representing the pathway, similar to one approach in [20]. Specifically, we adjusted α_C , which represents the stoichiometry of carbon required for biomass production relative to nitrogen, and k_E , the efficiency (maximal specific flux) of the pathway. This induced differing behaviours under nitrogen and carbon limitation: carbon limitation induced a switch to fully respiratory growth upon decreasing growth rates, but nitrogen limitation did not. However, an important feature of fermentation, namely the excretion of overflow metabolites, was not explicitly modelled. Furthermore, we did not include the generation of cellular energy in the form of ATP. These two omissions are related, as ATP is generated in different amounts by fermentative and respiratory pathways. A natural extension of the model would therefore be the addition of an explicit fermentative pathway, and parametrizing metabolic pathways by relative amounts of ATP generation and consumption.

In this light, we note that our model predicts sharp and discontinuous changes in gene expression and metabolite abundances where the switch between optimal metabolic strategies occurs (figure 2*c,d*). Because the substrates and products of the model were identical, our fully respiratory and respirofermentative enzymes could function as drop-in replacements of each other. Therefore, expression of either one or the other was optimal in our system. Such behaviour has been observed earlier in a similar C-GRAM framework [20] with linear carbon metabolism. There, adding explicit ATP metabolism-induced mixed expression of metabolically and catabolically efficient pathways.

Related to this, Basan *et al.* [52] studied fermentative fluxes under growth rate modulations in *E. coli*. They experimentally observed no fermentation below a critical growth rate, whereas fluxes increased rapidly with growth rate above the critical point. All results were quantitatively explained by a model of efficient proteome allocation balancing the need for biomass synthesis and energy generation. The finite resolution of the data does not enable distinguishing true discontinuity from a continuous transition over a narrow range of growth rates. What is clear is that fermentation does not completely replace respiration but rather supplements it at fast growth. This discussion highlights how C-GRAM proteins such as our fermentorespiratory enzyme must be interpreted cautiously as they represent multiple metabolic responsibilities.

3.4. Non-protein biomass

A second caveat to our choice modelling biosynthesis in one simple pathway is the following. While the parameter α_C represented the biomass carbon-to-nitrogen ratio in the model, it cannot be directly equated to observed dry mass compositions in real cells (some examples are collated in electronic supplementary material, text, table S3). This is partially due to the inclusion of extra carbon that real cells excrete during fermentation (see previous paragraph and electronic supplementary material, text S4.3), but also because the biomass of real cells consists of additional macromolecules besides proteins. For example, nucleotides are important cellular constituents with a higher nitrogen content than proteins, and an efficient cell must synthesize them in proportion to ribosomal proteins, as most RNA is ribosomal. Upstream, nucleotide synthesis depends on the pentose phosphate pathway (PPP), which shares multiple intermediate metabolites with glycolysis and amino acid metabolism. The trade-off between glycolytic and PPP flux has been implied to influence the relationships between growth rate, yield and oxidative stress [45,53,54]. A coarse-grained model that includes both protein and nucleotide synthesis must account for the coordination between the two carbon metabolic pathways [55], considerably increasing its complexity versus the model presented in this study.

Another large contribution to the biomass of microbes comes from their cell surface, which mostly consists of carbon. The interplay between cell surface biosynthesis, size homeostasis and growth has been explored in coarse-grained models of bacteria [56–58]. The relative importance of the cell surface changes with the size and shape of the cell, both of which depend on growth conditions and fluctuate during the cell cycle as recently reported in fission yeast [59]. Therefore, the dry mass density varies as cell mass and cell volume evolve differently. While balanced growth may be defined in terms of repeated cell cycles and the dynamic equations may be studied outside of steady state, the fluctuating dry mass density invalidates a focus on relative abundances. Accounting for the contribution of the cell surface to biomass in our model would therefore require reformulating the model into absolute abundances and including the effect of the cell cycle, which was outside the scope of this study. However, we expect that the improved understanding of concepts such as the optimization of resource allocation, the kinetic modelling of coarse-grained pathways, and the parametrization of stoichiometries, as presented in this work,

should be useful in the development of surface-aware C-GRAMs.

3.5. Conclusion

In summary, we have presented a modelling framework that describes uptake and metabolism of carbon and nitrogen in unicellular microbes in a coarse-grained manner. While the parametrization of the model presented here was chosen to facilitate comparisons with earlier models of *E. coli*, the structure of the model was deliberately kept general. The framework may therefore be applied to study the optimality of gene expression and growth across the tree of life. We hope that extensions of our model will be constructed to describe overflow metabolism, ATP, nucleotide metabolism, a cell surface and/or cell density fluctuations.

4. Methods

4.1. Dynamic equations describing the metabolic model

The model from [27] served as our starting point. We disregarded non-metabolic proteins and the inhibition of ribosomes, which were both present in the original model, because here we aimed to describe the steady-state behaviour of unperturbed wild-type cultures. We then introduced additional metabolic pathways representing (i) the uptake of carbon and nitrogen, (ii) fermentative and respiratory energy generation, represented by different parametrizations of a similar enzyme, and (iii) the recycling and excretion of carbon from complex nitrogen sources containing both nitrogen and carbon. Simplifications of the model were obtained by setting some model parameters to zero.

The reactions included in the full model are pictured in figure 3a; the interpretations of the variables and parameters in the model are given in tables 1 and 2. These reactions were modelled by the following formalism of ODEs, which is explored in more detail in electronic supplementary material, text S2. The time evolution of the concentration vector x was decomposed and given by

$$\dot{x} = \begin{pmatrix} \dot{a} \\ \dot{m} \\ \dot{p} \end{pmatrix} = S_j(x) - \mu x \begin{pmatrix} S^A \cdot \hat{j}(x) - j_R(x) - \mu a \\ \hat{S}j(x) - \mu m \\ j_R(x)f - \mu p \end{pmatrix}. \quad (4.1)$$

The concentrations described by this equation are those representing the amino acids a , the other metabolites $m = (k, c, n)$, and the proteins $p = (e_{K_{re}}, e_{K_{ex}}, e_C, e_{A_f}, e_{A_r}, e_N, r, z)$. The ribosomal allocation is given by the vector $f = (f_{K_{re}}, f_{K_{ex}}, f_C, f_{A_f}, f_{A_r}, f_N, f_R, f_Z)$ (table 1). Fluxes catalysed by the proteins are given by the vector

$$j = \begin{pmatrix} j_{K_{re}} \\ j_{K_{ex}} \\ j_C \\ j_{A_f} \\ j_{A_r} \\ j_N \\ j_R \end{pmatrix} = \begin{pmatrix} k_{K_{re}} e_{K_{re}} \frac{k}{k+k_{sat}} \\ k_{K_{ex}} e_{K_{ex}} \frac{k}{k+k_{sat}} \\ k_C e_C \\ k_{A_f} e_{A_f} \frac{c}{(c+c_{sat})(n+n_{sat})} \\ k_{A_r} e_{A_r} \frac{c}{(c+c_{sat})(n+n_{sat})} \\ k_N e_N \\ k_R r \frac{a}{a+a_{sat}} \end{pmatrix}. \quad (4.2)$$

The enzymes were chosen to follow Michaelis–Menten kinetics (for single-substrate enzymes) and products thereof, such that the rate laws were linear in the enzyme concentration, linear in the substrate concentrations at low concentrations, and saturated at high substrate concentrations. Details on the enzyme kinetics are provided in electronic supplementary material, text S3. Note that the housekeeping proteins (with

concentration z) do not catalyse any enzymatic reaction and therefore did not feature in the fluxes (table 2). Finally, the stoichiometry matrix was given by

$$S = \begin{pmatrix} (S^A)^T \\ \hat{S} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & +1 & +1 & 0 \\ -1 & -1 & 0 & 0 & 0 & \gamma_K \\ 1 & 0 & 1 & -\alpha_{C_f} & -\alpha_{C_r} & 0 \\ 0 & 0 & 0 & -\alpha_{N_f} & -\alpha_{N_r} & \gamma_N \end{pmatrix}. \quad (4.3)$$

Mass balance was maintained in all internal reactions, meaning columns in the stoichiometry matrix representing these totalled zero. However, the carbon and nitrogen transporters imported nutrients from the environment (so columns totalled +1), and the ketoacids could be excreted to the environment by the respective enzyme (so this column totalled -1). These considerations imposed the following constraints:

$$\alpha_{C_f} + \alpha_{N_f} = 1, \quad (4.4)$$

$$\alpha_{C_r} + \alpha_{N_r} = 1 \quad (4.5)$$

and

$$\gamma_K + \gamma_N = 1. \quad (4.6)$$

Using the above definitions, the ODEs describing the system can be written as

$$\dot{a} = j_{A_f} + j_{A_r} - j_R - \mu a, \quad (4.7)$$

$$\dot{k} = \gamma_K j_N - j_{K_{re}} - j_{K_{ex}} - \mu k, \quad (4.8)$$

$$\dot{c} = j_C + j_K - \alpha_{C_f} j_{A_f} - \alpha_{C_r} j_{A_r} - \mu c, \quad (4.9)$$

$$\dot{n} = \gamma_N j_N - \alpha_{N_f} j_{A_f} - \alpha_{N_r} j_{A_r} - \mu n, \quad (4.10)$$

$$\dot{e}_K = f_K j_R - \mu e_K, \quad (4.11)$$

$$\dot{e}_C = f_C j_R - \mu e_C, \quad (4.12)$$

$$\dot{e}_{A_f} = f_{A_f} j_R - \mu e_{A_f}, \quad (4.13)$$

$$\dot{e}_{A_r} = f_{A_r} j_R - \mu e_{A_r}, \quad (4.14)$$

$$\dot{e}_N = f_N j_R - \mu e_N, \quad (4.15)$$

$$\dot{r} = f_R j_R - \mu r \quad (4.16)$$

$$\text{and} \quad \dot{q} = f_Q j_R - \mu q. \quad (4.17)$$

The growth rate μ is found by taking the sum of these equations. Furthermore, using the allocation constraint $\sum_i f_i = 1$ and the concentration constraint $\sum_i x_i = 1$, such that $\sum_i \dot{x}_i = 0$, gives

$$\mu = j_C + j_N - j_{K_{ex}} = k_C e_C + k_N e_N - k_{K_{ex}} e_{K_{ex}} \frac{k}{k + k_{sat}}. \quad (4.18)$$

This is equal to the net import of nutrient from the environment (uptake minus excretion).

4.2. Balanced growth

As mentioned in the main text, we aimed to describe gene expression allocation under balanced growth in defined environments. The state of balanced growth was represented in the model by the steady state of equation (4.1). To compute the steady state numerically, we evolved these equations until a steady state was reached.

Given a model parametrization (see §4.3), an initial condition was partially guessed from the allocation fractions. Because the protein content in steady state is proportional to the proteome allocation, the initial relative abundance of each protein was chosen as a fixed multiple of its allocation fraction parameter. Furthermore, each initial relative metabolite abundance was established manually as 0.05. All together, the initial concentration vector x_0 was set to

$$x_0 = \begin{pmatrix} a_0 \\ m_0 \\ p_0 \end{pmatrix} = \begin{pmatrix} a_0 \\ (k_0, c_0, n_0) \\ p_0 \end{pmatrix} = \begin{pmatrix} 0.05 \\ (0.05, 0.05, 0.05) \\ 0.8f \end{pmatrix}, \quad (4.19)$$

with f the given allocation vector.

The simulations were implemented in the Julia programming language, using the DifferentialEquations.jl ecosystem [60]. For solving the concentration ODEs towards steady state, we used the Rodas5 solver.

4.3. Parametrization

Unicellular organisms grow in one of two principal states: respirofermentative growth and entirely respiratory growth. From the point of view of coarse-grained modelling, there are two primary differences between the two. On the one hand, the fermentation pathway consists of few different enzymes. Although the individual enzymes are highly abundant, they are highly efficient. The aggregate effect on the total expression burden is that the fermentation enzymes comprise a markedly smaller proteome fraction than the respiratory pathway would when providing equal biomass and energy production. On the other hand, fermentation requires more nutrients from the environment: carbon is consumed rapidly and converted into ethanol or acetate. We encoded the efficiency in the rate parameter k_{A_r} and the carbon usage in the parameter α_C . Specifically, with respirofermentative growth represented by the subscript f and purely respiratory growth by the subscript r , the above considerations require $k_{A_f} > k_{A_r}$ and $\alpha_{C_f} > \alpha_{C_r}$.

To obtain a rough estimate of the stoichiometry parameter in respiratory growth α_{C_r} , we used that proteinogenic amino acids contain approximately four carbon atoms per nitrogen atom. Using molar masses of 12 and 14 g mol⁻¹ for carbon and nitrogen, respectively, this ratio translated to (mass-action)

$$\alpha_{C_r} = \frac{4 \times 12}{4 \times 12 + 14} = \frac{24}{31}. \quad (4.20)$$

A further rough estimate, namely that the excess carbon excreted in fermentation is approximately equal in amount to the carbon converted to biomass, gave the stoichiometry in respirofermentative growth as

$$\alpha_{C_f} = \frac{8 \times 12}{8 \times 12 + 14} = \frac{48}{55}. \quad (4.21)$$

The full metabolic model, described in §4.1, was explored in figures 3, 4 and 5. The simplified models described earlier were obtained by setting key parameters to zero. For the model with parallel carbon and nitrogen assimilation pathways, the lack of ketoacid recycling or excretion (figure 2) was represented by the ketoacid stoichiometry parameter $\gamma_K = 0$. Additionally, the initial ketoacid fraction k_0 was set to zero, and no allocation was made to the ketoacid enzymes, i.e. $f_{K_{ex}} = f_{K_{re}} = 0$. For the initial core model, which did not distinguish respirofermentative and respiratory growth, $f_{A_r} = 0$ was additionally forced.

4.4. Optimizing ribosomal allocation to maximize steady-state growth rate

We explored the behaviour of the model under parameter modulations that represented different growth environments, while assuming that growth was optimized to suit this environment. The allocation vector f from equation (4.1) was therefore chosen not as a free parameter, but rather as the result of an optimization routine that maximized the growth rate μ . The fraction of housekeeping proteins f_Z was excluded from the optimization. The optimization problem can be defined as finding the allocation fraction f that maximizes the growth rate μ , while satisfying the allocation constraint

$$\sum_i f_i = 1. \quad (4.22)$$

The computation of the optimal allocation fraction is complicated, because the function $\mu(f)$ is only defined implicitly: a steady state must be found for each choice of f before the steady-state growth rate can be extracted. The nonlinearity of the Michaelis–Menten kinetics prohibits an explicit solution for μ in terms of f , which would be required by gradient-based optimization routines. We were therefore restricted to using a gradient-free optimization routine, for which we chose the Nelder–Mead algorithm as defined in the `Optim.jl` package [61–63]. We used tolerances of 1.0×10^{-10} and set the initial simplex to `Optim.AffineSimplex` ($a = 0.0$, $b = -0.1$). The optimization objective was set to minimize the doubling time $t_d = \log 2 / \mu$.

The allocation constraint (4.22) further complicated the optimization procedure. We eliminated one element of the allocation vector by imposing the allocation constraint. However, the Nelder–Mead sampling strategy still allowed for situations where this constrained element became negative. As a practical solution for such situations, and for allocation fractions whose sum exceeded the constraint, we evaluated the doubling time as the inverse of the machine precision, which served as a predefined very large value.

The Nelder–Mead algorithm further proved to be sensitive to numerical inaccuracies when the optimal allocation fraction had zero elements, i.e. the optimal cell entirely lacked expression of some proteins. Therefore, we ran the optimization procedure several times, each with different allocation fractions set to zero and excluded from the optimization. Each iteration resulted in one optimal allocation vector; the one with the

largest growth rate (smallest doubling time) was selected as the global optimum.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. The code used to generate the results and figures in this paper has been made available on GitHub (<https://www.github.com/istvankleijn/C-GRAM-carbon-nitrogen>) and archived on Zenodo [64].

Additional discussion about earlier C-GRAMs, and further details and derivations for the model presented in this paper are provided in electronic supplementary material [65].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. I.T.K.: conceptualization, data curation, formal analysis, investigation, methodology, software, visualization, writing—original draft, writing—review and editing; S.M.: conceptualization, funding acquisition, project administration, supervision; V.S.: conceptualization, funding acquisition, project administration, supervision, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. I.T.K. was supported by the Wellcome Trust (108908/Z/15/Z, 203968/Z/16/Z). S.M. was supported by the UK Medical Research Council. V.S. is supported by the EPSRC Centre for Mathematics of Precision Healthcare (EP/N014529/1).

Acknowledgements. We would like to thank Philipp Thomas, Nick Jones and Peter Swain for discussions related to some of the material presented here.

References

- Dekel E, Alon U. 2005 Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436**, nature03842. (doi:10.1038/nature03842)
- Lynch M, Field MC, Goodson HV, Malik HS, Pereira-Leal JB, Roos DS, Turkewitz AP, Sazer S. 2014 Evolutionary cell biology: two origins, one objective. *Proc. Natl Acad. Sci. USA* **111**, 16 990–16 994. (doi:10.1073/pnas.1415861111)
- Koch AL. 1988 Why can't a cell grow infinitely fast? *Can. J. Microbiol.* **34**, 421–426. (doi:10.1139/m88-074)
- López-Maury L, Marguerat S, Bähler J. 2008 Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat. Rev. Genet.* **9**, 583–593. (doi:10.1038/nrg2398)
- Berkhout J, Bosdriesz E, Nikerel E, Molenaar D, de Ridder D, Teusink B, Bruggeman FJ. 2013 How biochemical constraints of cellular growth shape evolutionary adaptations in metabolism. *Genetics* **194**, 505–512. (doi:10.1534/genetics.113.150631)
- Bruggeman FJ, Planqué R, Molenaar D, Teusink B. 2020 Searching for principles of microbial physiology. *FEMS Microbiol. Rev.* **44**, 821–844. (doi:10.1093/femsre/fuaa034)
- Butcher EC, Berg EL, Kunkel EJ. 2006 Systems biology in drug discovery. *Nat. Biotechnol.* **9**, 157–161.
- Schaechter M, Maaløe O, Kjeldgaard NO. 1958 Dependency on medium and temperature of cell size and chemical composition during balanced growth of *Salmonella typhimurium*. *Microbiology* **19**, 592–606. (doi:10.1099/00221287-19-3-592)
- Mitchison JM, Lark KG. 1962 Incorporation of ^3H -adenine into RNA during the cell cycle of *Schizosaccharomyces pombe*. *Exp. Cell Res.* **28**, 452–455. (doi:10.1016/0014-4827(62)90304-X)
- Waldron C, Lacroute F. 1975 Effect of growth rate on the amounts of ribosomal and transfer ribonucleic acids in yeast. *J. Bacteriol.* **122**, 855–865. (doi:10.1128/jb.122.3.855-865.1975)
- Bremer H, Dennis PP. 2008 Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *Ecosal Plus* **3**. (doi:10.1128/ecosal.5.2.3)
- Zavrel T, Faizi M, Loureiro C, Poschmann G, Stühler K, Sinetova M, Zorina A, Steurer R, Cervený J. 2019 Quantitative insights into the cyanobacterial cell economy. *Elife* **8**, e42508. (doi:10.7554/elife.42508)
- 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)
- You C *et al.* 2013 Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature* **500**, 301–306. (doi:10.1038/nature12446)
- Schmidt A *et al.* 2016 The quantitative and condition-dependent *Escherichia coli* proteome. *Nat. Biotechnol.* **34**, 104–110. (doi:10.1038/nbt.3418)
- Metzl-Raz E, Kafri M, Yaakov G, Soifer I, Gurvich Y, Barkai N. 2017 Principles of cellular resource allocation revealed by condition-dependent proteome profiling. *Elife* **6**, e28034. (doi:10.7554/elife.28034)
- Kleijn IT, Martínez-Segura A, Bertaux F, Saint M, Kramer H, Shahrezaei V, Marguerat S. 2022 Growth-rate-dependent and nutrient-specific gene expression resource allocation in fission yeast. *Life Sci. Alliance* **5**, e202101223. (doi:10.26508/lsa.202101223)
- Brauer MJ, Huttenhower C, Airoidi EM, Rosenstein R, Matese JC, Gresham D, Boer VM, Troyanskaya OG, Botstein D. 2008 Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. *Mol. Biol. Cell* **19**, 352–367. (doi:10.1091/mbc.E07-08-0779)
- Hui S, Silverman JM, Chen SS, Erickson DW, Basan M, Wang J, Hwa T, Williamson JR. 2015 Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Mol. Syst. Biol.* **11**, 784. (doi:10.15252/msb.20145697)
- Molenaar D, Van Berlo R, De Ridder D, Teusink B. 2009 Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol. Syst. Biol.* **5**, 323. (doi:10.1038/msb.2009.82)
- Scott M, Klumpp S, Mateescu EM, Hwa T. 2014 Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Mol. Syst. Biol.* **10**, 747. (doi:10.15252/msb.20145379)
- Maitra A, Dill KA. 2015 Bacterial growth laws reflect the evolutionary importance of energy efficiency. *Proc. Natl Acad. Sci. USA* **112**, 406–411. (doi:10.1073/pnas.1421138111)
- Maitra A, Dill KA. 2016 Modeling the overproduction of ribosomes when antibacterial drugs act on cells.

- Biophys. J.* **110**, 743–748. (doi:10.1016/j.bpj.2015.12.016)
24. Weiß E, Oyarzún DA, Danos V, Swain PS. 2015 Mechanistic links between cellular trade-offs, gene expression, and growth. *Proc. Natl Acad. Sci. USA* **112**, E1038–E1047. (doi:10.1073/pnas.1416533112)
 25. Pandey PP, Jain S. 2016 Analytic derivation of bacterial growth laws from a simple model of intracellular chemical dynamics. *Theory Biosci.* **135**, 121–130. (doi:10.1007/s12064-016-0227-9)
 26. Liao C, Blanchard AE, Lu T. 2017 An integrative circuit–host modelling framework for predicting synthetic gene network behaviours. *Nat. Microbiol.* **2**, 1658–1666. (doi:10.1038/s41564-017-0022-5)
 27. Bertaux F, Von Kügelgen J, Marguerat S, Shahrezaei V. 2020 A bacterial size law revealed by a coarse-grained model of cell physiology. *PLoS Comput. Biol.* **16**, e1008245. (doi:10.1371/journal.pcbi.1008245)
 28. Hu XP, Dourado H, Schubert P, Lercher MJ. 2020 The protein translation machinery is expressed for maximal efficiency in *Escherichia coli*. *Nat. Commun.* **11**, 5260. (doi:10.1038/s41467-020-18948-x)
 29. Monod J. 1949 The growth of bacterial cultures. *Annu. Rev. Microbiol.* **3**, 371–394. (doi:10.1146/annurev.mi.03.100149.002103)
 30. De Jong H, Casagrande S, Giordano N, Cinquemani E, Ropers D, Geiselmann J, Gouzé JL. 2017 Mathematical modelling of microbes: metabolism, gene expression and growth. *J. R. Soc. Interface* **14**, 20170502. (doi:10.1098/rsif.2017.0502)
 31. Carlson CR, Gallert B, Stokke T, Boye E. 1999 Regulation of the start of DNA replication in *Schizosaccharomyces pombe*. *J. Cell Sci.* **112**, 939–946. (doi:10.1242/jcs.112.6.939)
 32. Yu R, Vorontsov E, Sihlbom C, Nielsen J. 2021 Quantifying absolute gene expression profiles reveals distinct regulation of central carbon metabolism genes in yeast. *Elife* **10**, e65722. (doi:10.7554/eLife.65722)
 33. Jacoby RP, Succuro A, Kopriva S. 2020 Nitrogen substrate utilization in three rhizosphere bacterial strains investigated using proteomics. *Front. Microbiol.* **11**, 784. (doi:10.3389/fmicb.2020.00784)
 34. Milo R, Jorgensen P, Moran U, Weber G, Springer M. 2010 BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Res.* **38**, D750–D753. (doi:10.1093/nar/gkp889)
 35. Milo R, Phillips R. 2016 Concentrations and absolute numbers. In *Cell biology by the numbers* (collab. with N Orme), pp. 65–151. Abingdon, UK: Garland Science.
 36. Malecki M, Kamrad S, Ralser M, Bähler J. 2020 Mitochondrial respiration is required to provide amino acids during fermentative proliferation of fission yeast. *EMBO Rep.* **21**, e50845. (doi:10.15252/embr.202050845)
 37. Godard P, Urrestarazu A, Vissers S, Kontos K, Bontempi G, Van Helden J, André B. 2007 Effect of 21 different nitrogen sources on global gene expression in the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **27**, 3065–3086. (doi:10.1128/MCB.01084-06)
 38. Towbin BD, Korem Y, Bren A, Doron S, Sorek R, Alon U. 2017 Optimality and sub-optimality in a bacterial growth law. *Nat. Commun.* **8**, 14123. (doi:10.1038/ncomms14123)
 39. Planqué R, Hulshof J, Teusink B, Hendriks JC, Bruggeman FJ. 2018 Maintaining maximal metabolic flux by gene expression control. *PLoS Comput. Biol.* **14**, e1006412. (doi:10.1371/journal.pcbi.1006412)
 40. Valgepea K, Adamberg K, Seiman A, Vilu R. 2013 *Escherichia coli* achieves faster growth by increasing catalytic and translation rates of proteins. *Mol. Biosyst.* **9**, 2344–2358. (doi:10.1039/C3MB70119K)
 41. Peebo K, Valgepea K, Maser A, Nahku R, Adamberg K, Vilu R. 2015 Proteome reallocation in *Escherichia coli* with increasing specific growth rate. *Mol. Biosyst.* **11**, 1184–1193. (doi:10.1039/C4MB00721B)
 42. Mori M, Schink S, Erickson DW, Gerland U, Hwa T. 2017 Quantifying the benefit of a proteome reserve in fluctuating environments. *Nat. Commun.* **8**, 1225. (doi:10.1038/s41467-017-01242-8)
 43. Yu R, Campbell K, Pereira R, Björkeröth J, Qi Q, Vorontsov E, Sihlbom C, Nielsen J. 2020 Nitrogen limitation reveals large reserves in metabolic and translational capacities of yeast. *Nat. Commun.* **11**, 1–12. (doi:10.1038/s41467-020-15749-0)
 44. O'Brien EJ, Utrilla J, Palsson BO. 2016 Quantification and classification of *E. coli* proteome utilization and unused protein costs across environments. *PLoS Comput. Biol.* **12**, e1004998. (doi:10.1371/journal.pcbi.1004998)
 45. Christodoulou D, Link H, Fuhrer T, Kochanowski K, Gerosa L, Sauer U. 2018 Reserve flux capacity in the pentose phosphate pathway enables *Escherichia coli*'s rapid response to oxidative stress. *Cell Syst.* **6**, 569–578.e7. (doi:10.1016/j.cels.2018.04.009)
 46. Bosdriesz E, Molenaar D, Teusink B, Bruggeman FJ. 2015 How fast-growing bacteria robustly tune their ribosome concentration to approximate growth-rate maximization. *FEBS J.* **282**, 2029–2044. (doi:10.1111/febs.13258)
 47. Wu C, Balakrishnan R, Braniff N, Mori M, Manzanarez G, Zhang Z, Hwa T. 2022 Cellular perception of growth rate and the mechanistic origin of bacterial growth law. *Proc. Natl Acad. Sci. USA* **119**, e2201585119. (doi:10.1073/pnas.2201585119)
 48. Dai X *et al.* 2016 Reduction of translating ribosomes enables *Escherichia coli* to maintain elongation rates during slow growth. *Nat. Microbiol.* **2**, 2016231. (doi:10.1038/nmicrobiol.2016.231)
 49. Dai X, Zhu M. 2016 Coupling of ribosome synthesis and translational capacity with cell growth. *Trends Biochem. Sci.* **45**, 681–692. (doi:10.1016/j.tibs.2020.04.010)
 50. Elsemman IE *et al.* 2022 Whole-cell modeling in yeast predicts compartment-specific proteome constraints that drive metabolic strategies. *Nat. Commun.* **13**, 801. (doi:10.1038/s41467-022-28467-6)
 51. Grigaitis P, Grundel DA, van Pelt-Kleinjan E, Isaku M, Xie G, Mendoza Farias S, Teusink B, Van Heerden JH. 2022 A computational toolbox to investigate the metabolic potential and resource allocation in fission yeast. *Msystems* **7**, e00423-22. (doi:10.1128/mystems.00423-22)
 52. Basan M, Hui S, Okano H, Zhang Z, Shen Y, Williamson JR, Hwa T. 2015 Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. *Nature* **528**, 99–104. (doi:10.1038/nature15765)
 53. Cheng C *et al.* 2019 Laboratory evolution reveals a two-dimensional rate-yield tradeoff in microbial metabolism. *PLoS Comput. Biol.* **15**, e1007066. (doi:10.1371/journal.pcbi.1007066)
 54. Hurbain J, Thommen Q, Anquez F, Pfeuty B. 2022 Quantitative modeling of pentose phosphate pathway response to oxidative stress reveals a cooperative regulatory strategy. *IScience* **25**, 104681. (doi:10.1016/j.isci.2022.104681)
 55. Chubukov V, Gerosa L, Kochanowski K, Sauer U. 2014 Coordination of microbial metabolism. *Nat. Rev. Microbiol.* **12**, 327–340. (doi:10.1038/nrmicro3238)
 56. Harris LK, Theriot JA. 2016 Relative rates of surface and volume synthesis set bacterial cell size. *Cell* **165**, 1479–1492. (doi:10.1016/j.cell.2016.05.045)
 57. Ojick N, Serbanescu D, Banerjee S. 2019 Surface-to-volume scaling and aspect ratio preservation in rod-shaped bacteria. *Elife* **8**, e47033. (doi:10.7554/eLife.47033)
 58. Oldewurtel ER, Kitahara Y, Cordier B, Özbaykal G, van Teeffelen S. 2019 Bacteria control cell volume by coupling cell-surface expansion to dry-mass growth. *BioRxiv* 769786. (doi:10.1101/769786)
 59. Bertaux F, Kleijn IT, Marguerat S, Shahrezaei V. 2023 Fission yeast obeys a linear size law under nutrient titration. *microPublication Biol.* **2023**. (doi:10.17912/micropub.biology.000833)
 60. Rackauckas C, Nie Q. 2017 DifferentialEquations.jl – a performant and feature-rich ecosystem for solving differential equations in Julia. *J. Open Res. Softw.* **5**, 15. (doi:10.5334/jors.151)
 61. Nelder JA, Mead R. 1965 A simplex method for function minimization. *Comput. J.* **7**, 308–313. (doi:10.1093/comjnl/7.4.308)
 62. Gao F, Han L. 2012 Implementing the Nelder-Mead simplex algorithm with adaptive parameters. *Comput. Optim. Appl.* **51**, 259–277. (doi:10.1007/s10589-010-9329-3)
 63. Mogensen PK, Riseth AN. 2018 Optim: a mathematical optimization package for Julia. *J. Open Sour. Softw.* **3**, 615. (doi:10.21105/joss.00615)
 64. Kleijn IT, Marguerat S, Shahrezaei V. 2023 A coarse-grained resource allocation model of carbon and nitrogen metabolism in unicellular microbes. *Zenodo*. (doi:10.5281/zenodo.8308497)
 65. Kleijn IT, Marguerat S, Shahrezaei V. 2023 A coarse-grained resource allocation model of carbon and nitrogen metabolism in unicellular microbes. Figshare. (doi:10.6084/m9.figshare.c.6837579)