OPEN

Bridging substrate intake kinetics and bacterial growth phenotypes with flux balance analysis incorporating proteome allocation

Hong Zeng & Aidong Yang*

Empirical kinetic models such as the Monod equation have been widely applied to relate the cell growth with substrate availability. The Monod equation shares a similar form with the mechanistically-based Michaelis-Menten kinetics for enzymatic processes, which has provoked long-standing and un-concluded conjectures on their relationship. In this work, we integrated proteome allocation principles into a Flux Balance Analysis (FBA) model of *Escherichia coli*, which quantitatively revealed potential mechanisms that underpin the phenomenological Monod parameters: the maximum specific growth rate could be dictated by the abundance of growth-controlling proteome and growth-pertinent proteome cost; more importantly, the Monod constant (K_s) was shown to relate to the Michaelis constant for substrate transport ($K_{m,g}$), with the link being dependent on the cell's metabolic strategy. Besides, the proposed model was able to predict glucose uptake rate at given external glucose concentration through the size of available proteome resource for substrate transport and its enzymatic cost, while growth rate and acetate overflow were accurately simulated for two *E. coli* strains. Bridging the enzymatic kinetics of substrate intake and overall growth phenotypes, this work offers a mechanistic interpretation to the empirical Monod law, and demonstrates the potential of coupling local and global cellular constrains in predictive modelling.

Understanding the growth of microbial cultures occupies a central place in the study of microorganisms. How growth rate varies with internal traits or external conditions greatly affects the choice of process-level parameters such as temperature, medium composition in cell cultures for a specific purpose and the strategy for the modulation of delicate cellular attributes such as enzyme activity, pathway diversity and regulatory systems.

Ever since Jacques Monod correlated the specific growth rate (λ) with extracellular substrate concentration ($[g]$) via two kinetic parameters, the maximum specific growth rate λ_{max} and the Monod constant K_s, the Monod equation $\lambda = \lambda_{max}\frac{[g]}{[g] + K_s}$[1] becomes the best-known microbial growth kinetics for describing various bioprocesses. Other growth models (reviewed in[2]) are less prevalent due to more sophisticated formulation and insufficient experimental validation. Despite its unstructured and empirical nature[3], the Monod equation generally renders satisfactory results in terms of fitting substrate consumption and growth rate profile for a wide range of microorganisms and culture conditions. More recent studies have shown that growth kinetic constants (in Monod terms, λ_{max} and K_s) are actually a function of the culture history (e.g. sludge ages) and community composition[4,5]. In fact, twenty-years ago Kovarova-Kovar and Egli explicitly pointed out that cells could change their growth kinetics via adaptation, therefore a single set of kinetic parameters is not able to represent such variable properties[2]. The notion of variable growth kinetics is further supported by experimental evidence showing the evolution of K_s[6,7], 'intrinsic' and 'extant' λ_{max}[8] and the $\lambda_{max} - K_s$ relationship[9].

In addition to the Monod-type empirical models, the development of metabolic network reconstructions and the associated analytical approach, namely Flux Balance Analysis (FBA) facilitates the study of growth phenotypes at the metabolic flux level[10]. A FBA model is established based on the known metabolic reaction stoichiometry and gene-enzyme relationship. It is inherently mechanistic and provides more insight of cellular physiological properties compared to an empirical growth model. More importantly, the general objective of FBA is to predict the maximum growth rate at given culture conditions and possibly genetic modifications[11]. If the prediction of

Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ, UK. *email: aidong.yang@eng.ox.ac.uk

growth rate and intracellular metabolic fluxes were sufficiently accurate, FBA would become a very useful tool for engineering efficient microbial systems. Essentially, the accuracy of the flux prediction in FBA is governed by constraints. One typical example is setting uptake limits for oxygen and glucose exchange fluxes to predict aerobic acetate formation (namely the overflow metabolism) in *Escherichia coli*[12], where acetate was predicted to excrete when the system was limited by oxygen uptake rate. However, it has been proposed and validated that the aerobic acetogensis in *E. coli* is the result of an efficient proteome allocation strategy for rapid growth[13,14]. This example suggests that setting hard limitations on uptake fluxes could be a rather arbitrary and problematic treatment when there is a lack of theoretical basis, not to mention that the boundaries of myriad intracellular fluxes are often difficult to determine reliably. In principle, using experimentally determined flux values as lower and/or upper bounds could help FBA to compute more realistic flux distributions. This leads a dilemma between the overall accuracy of the simulation and the predictive power of the model (e.g. fluxes with hard-set boundaries are in a sense "weakly" predicted). Progress in incorporating macromolecular expressions[15] and intracellular resource allocation[16–19] with the classic FBA effectively alleviates the need for setting subtle uptake boundaries and improves the accuracy and scope of the predicted growth phenotypes. For instance, Resource Balance Analysis (RBA)[18] eliminates the need of defining uptake bounds of substrates by relating the enzyme efficiency with the concentration of extracellular nutrients. Multi-scale models of Metabolism and macromolecular Expression (ME-models)[20–22] enable the prediction of substrate uptake rates by incorporating growth rate-dependent demand functions and limited macromolecular synthesis machineries, e.g. limited ribosomal translation rate, limited mRNA catalytic rate and limited RNA polymerase transcription rate. Despite these successes, the prediction of substrate uptake rate in general highly relies on the accuracy of the enzyme kinetic parameters adopted in the model, most of which are not readily available at the genome scale[23].

In this work, we explore the potential links between the empirical Monod growth kinetics and the FBA-based metabolic modelling, with the latter to reveal the underlying biological mechanisms governing the phenomenological parameters of the former. Using two *E. coli* strains (NCM3722 and ML308) as case studies, we investigate how the integration of the proteome allocation principles into FBA would help depict the intriguing multi-scale mechanisms that govern the various growth phenotypes under a wide range of growth conditions. We firstly divided the overall proteome into coarse-grained functional sectors, i.e. carbon scavenging (C), energy generation (E), biomass synthesis (BM) and growth-independent offset (Q) sectors[19,24]. The size of each sector is dictated by the magnitude of the metabolic flux the sector processes and the corresponding proteome cost per unit flux. All individual proteome sectors were assembled together into a global proteome allocation constraint[24–26]. Furthermore, we parameterized a correlation between the enzyme cost for carbon transport and the abundance of carbon source. Applying these modelling concepts and parameterisation in FBA simulations, we were able to predict the glucose uptake rate, growth rate and the acetate overflow upon varying extracellular glucose level. Our theoretical model also predicted a step change in K_s after the onset of the acetate excretion. Further analysis of the model suggests that this variation originates from the change in cell's metabolic strategy, which coincides with the previous notion that cells possess variable growth kinetics due to adaptation to the environment[2]. Besides, λ_{max} is shown to be controlled by the abundance of growth-controlling proteome and the proteome cost per unit increase of growth rate. Finally, we propose that a proper connection of local and global physiological constraints may be essential for improving the predictive power of FBA models.

Results

Proteomic fraction occupied by the carbon-scavenging sector. Following the treatment in several recent studies on proteome allocation[14,19,25–27], we divide the overall cell proteome into four coarse-grained functional sectors: carbon-scavenging sector C that comprises enzymes for importing extracellular carbohydrate, energy biogenesis sector E that consists of enzymes used for respiration and (aerobic) fermentation, biomass synthesis sector BM that accounts for ribosomal proteins and enzymes carrying anabolic fluxes, and an offset sector Q that contains other proteins whose abundance do not vary with the fluxes in the above three sectors (Fig. 1a). The summation of the fractions of all sectors (i.e. ϕ_C, ϕ_E, ϕ_{BM}, ϕ_Q) equals to one

$$\phi_C + \phi_E + \phi_{BM} + \phi_Q = 1 \tag{1}$$

Let the maximum amount of proteome attainable to C, E and BM sectors be ϕ_{max}^g, which corresponds to the minimum size of the Q sector, ϕ_{Qmin}. For cells striving for maximum growth,

$$\phi_C + \phi_E + \phi_{BM} = 1 - \phi_{Qmin} = \phi_{max}^g \tag{2}$$

This global constraint is coupled to another proteome constraint that was established previously as a key mechanism governing the overflow metabolism in *E. coli*[14,28]

$$\phi_E + \phi_{BM} \leq \phi_{max}^o \tag{3}$$

ϕ_{max}^o denotes the maximum amount of proteome attainable to energy production and biomass synthesis. Equations (2 and 3) were used to derive two concise proteome allocation constraints (Eqs. (16 and 17), Methods), which were integrated into the *E. coli* metabolic network as additional rows. The resulted proteome allocation-constrained FBA model forms a mixed integer liner programing (MILP) problem (Supplementary Text). Determination of model parameters is detailed in Supplementary Text. FBA simulations were performed using the integrated model with extracellular glucose concentration as input and maximising cell growth as the objective function. Simulated optimal growth rate (μ), carbon uptake rate (v_c), acetate production rate (v_f) and tricarboxylic acid cycle flux (v_r) at μ, along with proteome cost parameters (defined in Eqs. (14–16)) were used to calculate the distribution of proteome sectors (defined in Eq. (2)) at different growth rates (Fig. 2a,b). Simulation
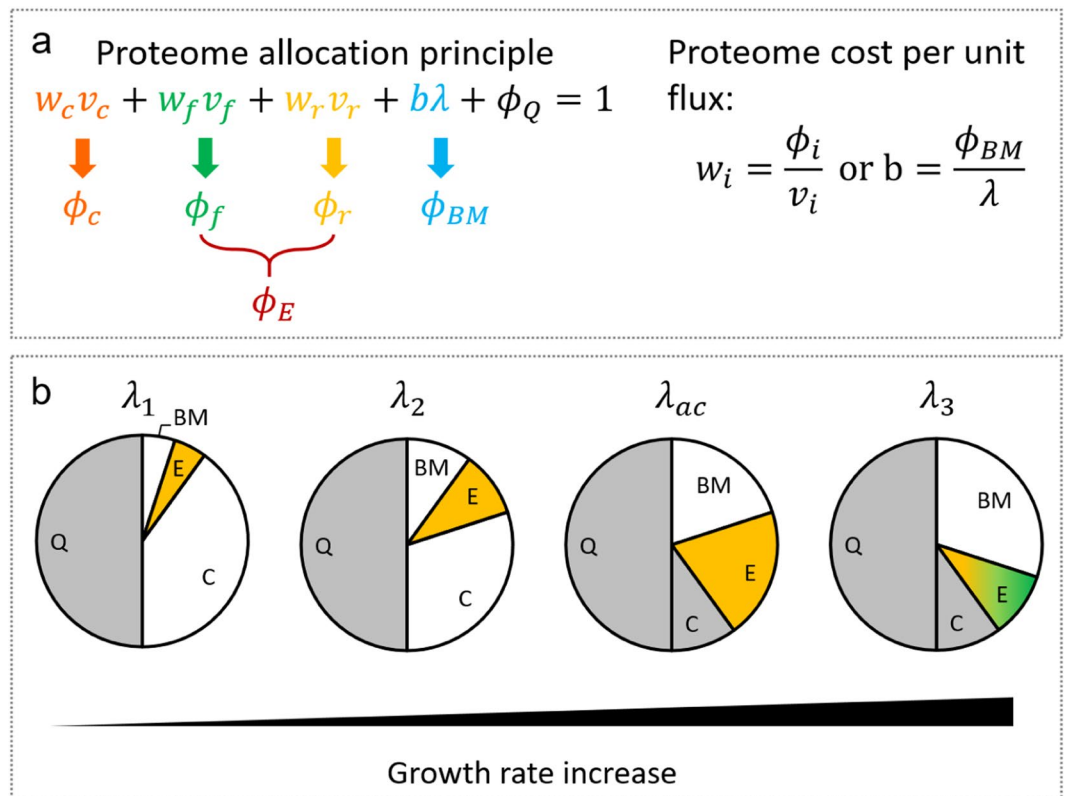
Figure 1. Schematic diagram of the proteome allocation model. (**a**) The overall proteome comprises four sectors, including carbon-scavenging ϕ_C, energy biogenesis ϕ_E, biomass synthesis ϕ_{BM} and growth-independent offset ϕ_Q. The enzymatic cost per unit flux of each sector is calculated by the ratio between proteome abundance and the corresponding flux. (**b**) The evolution of proteome component at increased growth rate. Before acetate overflow, BM and E sectors increase with growth rate while the C sector continuously declines; The E sector comprises only respiration (marked in yellow). At $\lambda = \lambda_{ac}$, $\phi_E = \phi_r$ reaches the highest value. At higher growth rates, the E sector starts to decline to save space for the increase of the BM sector and becomes a mixture of fermentation and respiration (marked with a mixture of yellow and green); the C sector reaches its lower bound and stays constant.

results show that ϕ_C gradually decreases as growth rate increases until it reaches a minimum level (ϕ_{Cmin}), which is equal to the gap between ϕ_{max}^g and ϕ_{max}^o:

$$\phi_C > \phi_{Cmin}, \quad when \ \lambda < \lambda_{ac} \tag{4a}$$

$$\phi_C = \phi_{Cmin} = \phi_{max}^g - \phi_{max}^o, \quad when \ \lambda \geq \lambda_{ac} \tag{4b}$$

The growth rate at which ϕ_C reaches its lower bound is corresponding to the threshold growth rate λ_{ac}, above which acetate overflow occurs (Fig. 2a,b). When growth rate exceeds λ_{ac} the composition of ϕ_E changes from the enzymes for pure respiration to those for a combination of fermentation and respiration (Fig. 1a). The predicted decline of ϕ_C is supported by the observed upregulation of catabolic genes under carbon limitation[25]. We did not find experimental evidence for a non-zero ϕ_{Cmin}. However combining Eqs. (2 and 3) gives $\phi_C \geq \phi_{max}^g - \phi_{max}^o$, which in theory implies a positive value of ϕ_{Cmin} given $\phi_{max}^g > \phi_{max}^o$.

Predictions of glucose uptake rate and the acetate overflow. It has been shown that the incorporation of Eq. (3) into a metabolic model of *E. coli* enables the prediction of the acetate overflow at $\lambda \geq \lambda_{ac}$[28]. Assuming a linear relationship between a proteome fraction and the metabolic flux it processes[25], the C sector could be modelled as

$$\phi_C = w_c v_c \tag{5}$$

where v_c is the carbon uptake flux. The linear coefficient w_c denotes the enzyme cost per unit carbon influx, which has been proposed to be a function of the environmental substrate levels, derived from the Michaelis-Menten equation[19]
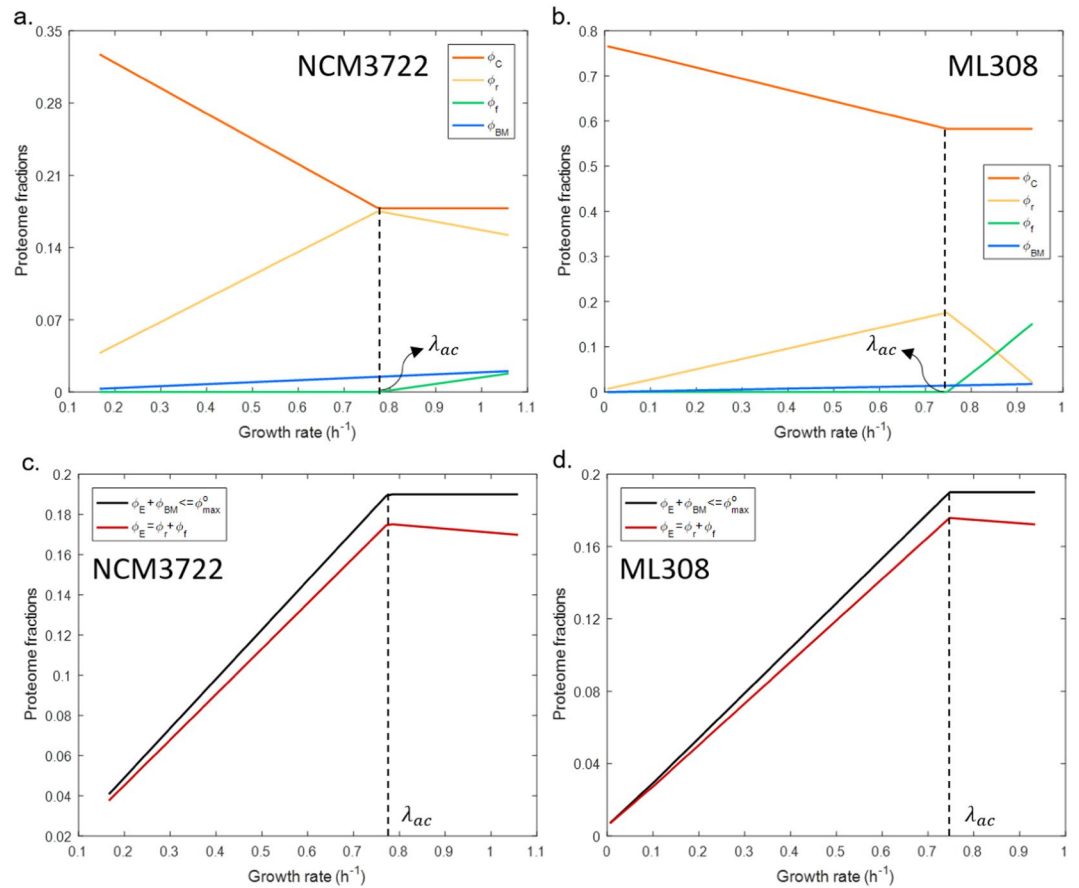
Figure 2. Simulated distribution of proteome sectors ϕ_C, ϕ_r, ϕ_f and ϕ_{BM} against different growth rates for *E. coli* NCM3722 **(a)** and ML308 **(b)**; and simulated change of the E sector (ϕ_E) and summation of E and BM sectors ($\phi_E + \phi_{BM}$) for NCM3722 **(c)** and ML308 **(d)**. ϕ_r and ϕ_f are E sector components. The dashed line indicates the onset of acetate overflow. λ_{ac} is the threshold growth rate, above which acetate overflow occurs.

$$w_c = w_{c,0}\left(1 + \frac{K_{m,g}}{[g]}\right) \tag{6}$$

The constant $w_{c,0}$ represents the minimum enzyme cost per unit carbon influx; $[g]$ is the glucose concentration, where glucose is taken as a representative substrate; $K_{m,g}$ is the Michaelis constant. We determined the value of $w_{c,0}$ using strain-specific cell culture data (Supplementary Text); with reported values of $K_{m,g}$ we were subsequently able to compute w_c as a function of $[g]$ (Supplementary Eqs. (S36 and S37)). Coupling Eq. (6) with the proteome allocation constraints (Eqs. (16 and 17)), the glucose uptake rate, acetate production rate and optimal growth rate of two *E. coli* strains were predicted by specifying $[g]$ (as opposed to setting glucose influx as pre-determined input[12]) and maximizing the growth rate. Simulation results agree well with the experimental data (Fig. 3a–d).

λ_{max} is a function of proteome resource and proteome cost. Combining the proteome allocation constraints and the established $w_c - [g]$ correlation (Eq. (6)) with additional linear assumptions for the $v_c - \lambda$ pair (Supplementary Text), we have derived

$$\lambda_{max} = \phi_{growth}/p_{growth} \tag{7}$$

ϕ_{growth} represents the maximal fraction of the growth-controlling proteome, and p_{growth} is the proteome cost per unit increase of growth rate. Equation (7) is in line with the previous proposal that λ_{max} is closely related to the protein synthesis mechanism[4]. During non-overflow growth ($\lambda < \lambda_{ac}$), we show that (Supplementary Eqs. (S16–S21))

$$\phi_{growth} = \phi_{max}^g - \sum_i \phi_{i,atpm} \tag{8}$$

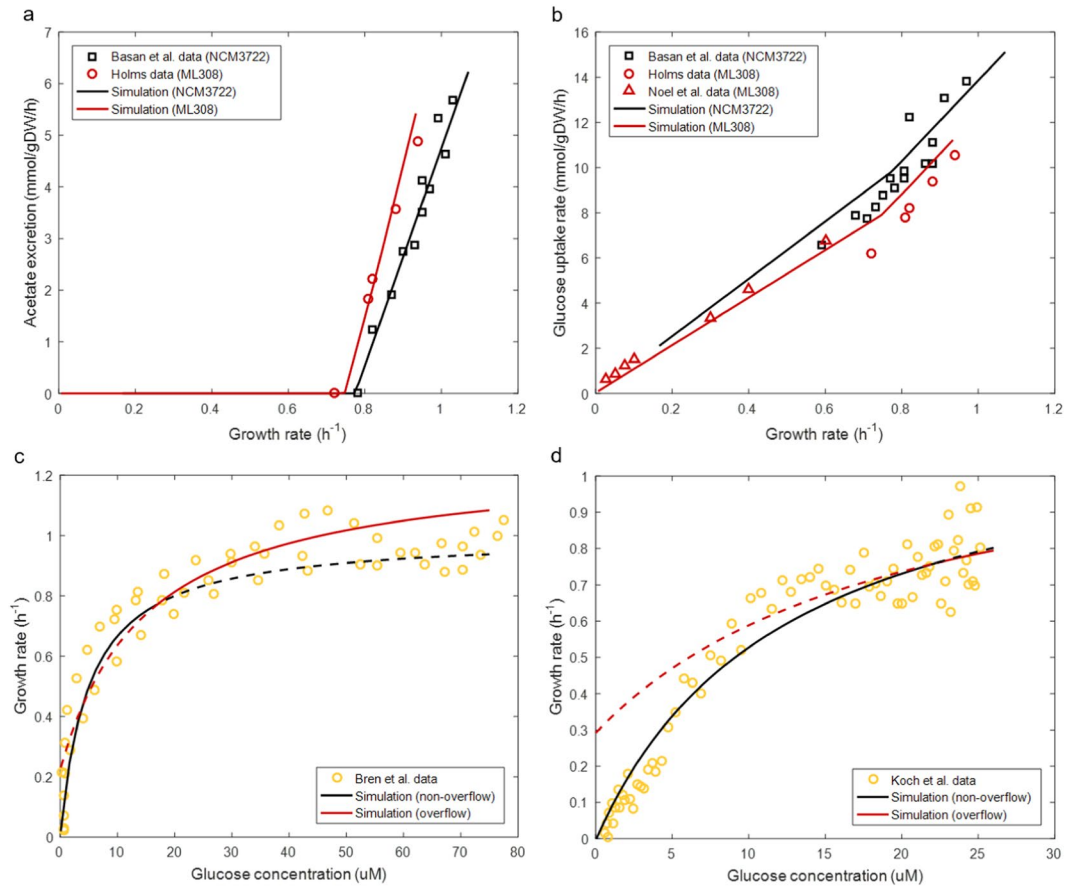$$p_{growth} = \sum_i p_i \tag{9}$$

Figure 3. Comparison between the model simulation and experimental data of the growth of *E. coli* in glucose-limited cultures. **(a)** Comparison between predicted and measured acetate excretion rate. Experimental data for NCM3722 were obtained from Fig. 1 of ref. [14]; experimental data for ML308 were obtained from Table 7 of ref. [52]. **(b)** Comparison between predicted and measured glucose uptake rate. Experimental data for NCM3722 were obtained from Fig. 3B of ref. [19]; experimental data for ML308 were obtained from Table 7 of ref. [52] and Fig. 3b of ref. [53]. **(c)** Comparison between predicted and measured specific growth rate against glucose concentration profile for NCM3722. Experimental data were obtained from SI Fig. 1 of ref. [51]. **(d)** Comparison between predicted and measured specific growth rate against glucose concentration profile for ML308. Experimental data were obtained from Fig. 1 of ref. [34]. Solid lines in **(c,d)** are actual model predictions; dashed lines are presented to show the model prediction in the region where the model is not applicable. Simulation results of optimal growth rate, glucose uptake rate and acetate production rate were obtained in FBA simulations of the proteome allocation-integrated metabolic model. In all simulations, the extracellular glucose concentration was specified as input and the growth rate was maximised.

$\phi_{i,atpm}$ is the proteome fraction of sector i occupied by non-growth-associated maintenance; p_i is the proteome cost of sector i per unit increase of growth rate; i represents C, E, or BM sector. During overflow growth ($\lambda \geq \lambda_{ac}$), we show that (Supplementary Eqs. (S22–S24))

$$\phi_{growth} = \phi'_c \tag{10}$$

$$p_{growth} = p_c \tag{11}$$

ϕ'_c denotes the fraction of the proteome allocable to the carbon-scavenging sector (adjusted with an off-set to ϕ_{Cmin}, see Supplementary Eq. (S22)). Equations (10 and 11) imply that in the overflow region λ_{max} can be dictated solely by the characteristics of the C sector.

The above results suggest two ways to improve cell growth: increase ϕ_{growth} or decrease p_i. ϕ_{growth} could be elevated by reducing maintenance or eliminating the expression of useless proteins. However, it might be difficult to directly modify ϕ_{growth} as it is more or less an intrinsic feature of a cell. Reducing p_i could be more realistic in synthetic biology. In this work p_c could gradually reduce upon the decline of w_c at increased growth rate (Supplementary Eqs. (S19 and S23) and Supplementary Fig. S1). p_E was shown to change from positive to negative due to varied ϕ_E against growth rate correlation before and after λ_{ac} (Fig. 2c,d). The variability of p_i could result from two mechanisms: the cell upregulates catabolic genes under carbon limitation (leads to variable p_c) and

changes its metabolic strategies under different culture conditions (leads to variable p_E). The latter is governed by the presence of alternative pathways. We could infer that if a cell carries multiple pathways of a similar biological function (e.g. energy biogenesis) but with distinct enzyme (or protein) costs, the cell might be able to improve its growth via directing more metabolic flux to the enzymatically low-cost pathways. This could help design synthetic microorganisms with potentially higher robustness and fitness to the growth environment. A recent study also predicted that improved enzyme activity of energy pathways leads to a higher growth rate[29], which supports the above suggestions.

K_s depends on the interplay between fractions of proteome sectors and affinity of transport enzymes.

While λ_{max} unambiguously refers to the highest specific growth rate a cell can achieve, the biological meaning of K_s is less clear. Often taken as a mathematical analogy to the Michaelis-Menten enzyme kinetics, the Monod constant K_s seems to share some link with the Michaelis constant for substrate transport $K_{m,g}$. This might explain why the Monod growth kinetics is sometimes mistermed as "Michaelis-Menten kinetics" for describing growth-associated bioprocesses[2]. While the derivation of the Monod equation is strictly empirical[30], Monod himself raised interesting comments between K_s and $K_{m,g}$: (i) 'the value of K_s should be expected to bear some more or less distant relation to the apparent dissociation constant of the enzyme involved in the first step of breakdown of a given compound'; (ii) 'since a change of conditions affecting primarily the velocity of only one rate-determining step will, in general (but not necessarily), be only partially reflected in the overall rate, one might expect K_s values to be lower than the corresponding values of the Michaelis constant of the enzyme catalyzing the reaction'[1]. Essentially Monod suggested that K_s could be a function of $K_{m,g}$ and that K_s should be generally smaller than $K_{m,g}$. In the past twenty years several attempts have been made to explore the physical meaning of K_s. One interpretation is that $1/K_s$ reflects the overall affinity of a cell to a substrate[2]. A later study suggests that K_s is a function of the overall change of free energy of the microbial growth process[31]. More recently, K_s was related to $K_{m,g}$ via investigating the control of the transport step on the specific growth rate[30]. In this work, we explicitly derived that K_s relates to $K_{m,g}$ via a proportional factor δ (Supplementary Eqs. (S15) and (S26–S29))

$$K_s = \delta K_{m,g} \tag{12}$$

$$\delta = \begin{cases} \dfrac{p_c}{\sum_i p_i} < 1, & when\ \lambda < \lambda_{ac} \\ 1, & when\ \lambda \geq \lambda_{ac} \end{cases} \tag{13}$$

The discrete feature of δ leads to a (upward) step change of K_s at the acetate switch ($K_{s,NCM3722} = 5\ \mu M$ and $K_{s,ML308} = 12\ \mu M$ for $\lambda < \lambda_{ac}$; $K_{s,NCM3722} = 15\ \mu M$ and $K_{s,ML308} = 20\ \mu M$ for $\lambda \geq \lambda_{ac}$), which perfectly validates Monod's proposition that $K_s = f(K_{m,g})$ and $K_s \leq K_{m,g}$. In fact, it has been noted that K_s values for *E. coli* grown with glucose vary significantly in different studies or different growth conditions, and there has been a lack of a satisfactory explanation of why the presumably constant K_s spans across such a wide range[32]. The discrete variable δ provides a simple explanation to this long-lasting problem: under strong carbon limitation, the cell benefits from the metabolic strategy where all carbon consumed for energy production is metabolized through respiration; $\delta = p_c/(p_c + p_E + p_{BM}) < 1$ (Supplementary Eq. (S28)) suggests the relative importance of C, E and BM sectors in dictating the cell's overall affinity to the growth-controlling substrate. When substrate becomes sufficient which potentially allows the cell to grow at a rate beyond a threshold, it becomes more advantageous for a cell to switch to a metabolic strategy where a significant portion of carbon consumed for energy production is processed through the proteome-efficient acetate pathway. In this case, the C sector touches its lower bound and decouples from E and BM sectors, $\delta = p_c/p_c = 1$ (Supplementary Eq. (S29)). Essentially, δ reflects the metabolic state of a cell. In principle, if a cell possesses a number of metabolic strategies, δ would hold the same number of discrete values, each leading to a distinct value of the apparent affinity constant K_s. This mechanism could be the underlying reason of the observed large variation in K_s under different culture conditions. Overall, our theoretical model indicates that K_s reflects the combined effect of the local characteristics of the carbon transport system and the overarching interplay between carbon-scavenging and other growth-related functions.

Discussion

In this work, we modified the classic FBA model (constrained primarily by the mass and energy balance) via embedding the coarse-grained proteome allocation constraints with an intention to reveal the potential mechanistic nature of the empirical growth kinetics proposed by Monod seventy years ago. By doing so, we showed theoretically a non-zero lower bound of the carbon-scavenging sector, which differs from the previous notion that ϕ_C could continuously decrease towards zero as growth rate increases[19]. Under carbon limitation, it seems to be agreeable that $\phi_E + \phi_{BM}$ increases with growth rate[26,27] and reaches a plateau (defined by ϕ_{max}^o) where the acetate overflow occurs so that more proteome could be allocated to biomass synthesis as required for rapid growth[14]. Consequently, the rest of the proteome ($\phi_C + \phi_Q$) has to decrease and then levels[33]. If ϕ_Q remains constant, taking for example its minimum value (ϕ_{Qmin}) as considered in this work for cells striving for maximum growth, the size of the level end of the ($\phi_C + \phi_Q$) profile will be ($\phi_{Cmin} + \phi_{Qmin}$). We argue that in this level sum, the contribution of ϕ_{Cmin} is unlikely to be zero, given the significant carbon influx to sustain the high rate of biomass synthesis and energy production under overflow conditions. On the maximum value of ϕ_C (ϕ_{Cmax}), our parameter estimation based on data assembled from literature (Supplementary Table S1) determined $\phi_{Cmax} = 0.37$ for NCM3722 and $\phi_{Cmax} = 0.77$ for ML308 ($\phi_{Cmax} \approx \phi_{max}^g$, determined by extrapolating to zero growth rate, also see Eq. (2) and

Fig. 2a,b). These specific values are affected by a number of factors, including particularly (i) the assumed value for ϕ^o_{max} (0.19, previously suggested for *E. coli* NCM3722[14]) and (ii) the value of K_s estimated from cell culture data (Supplementary Text and Supplementary Eqs. (S30) and (S34 and S35)). In reality, the former could be strain-dependent, while the latter is known to be sensitive to growth conditions and culture history[4,32,34]. Therefore, the specific results of parameterization from this work were to support the illustration of the overall modelling approach; their biological significance needs to be interpreted with caution.

In dynamic FBA (DFBA), v_c (uptake rate of growth-limiting carbon source) is generally determined by applying the Michaelis-Menten kinetics for substrate transport[35–39]. An inherent issue with this approach is that the kinetic parameters are normally valid only for a specific substrate uptake pattern, which however may vary depending on the variety of fermentative by-products (e.g. acetate, formate, ethanol or a mixture of them)[10] or more generally on the physiological state. Therefore, using Michaelis-Menten kinetics alone (with a single set of kinetic parameters) to predict the carbon intake suffers from a very limited scope of applicability – it cannot depict a physiological state that deviates from the state that it was originally fitted to. In this work, this drawback is rectified by introducing the proteome allocation principle, which couples carbon transport with other growth-associated metabolism. Through a fully parameterized $w_c − [g]$ correlation, we compute the variable w_c given the environmental glucose concentration (as opposed to directly manipulating w_c[19]). The computable w_c is involved in the proteome allocation constraints, which allows us to simultaneously predict the glucose uptake rate, specific growth rate and acetate overflow upon varying the environmental glucose level. The coupling of the local metabolic capacity (for carbon transport) and the global regulatory mechanism (on proteomic resource allocation) was shown to be essential in improving the predictive power of FBA models. It should be noted that, while in this work we have adopted a single $K_{m,g}$ for the carbon transport system, it is known that in *E. coli*, $K_{m,g}$ differs from different transporters which could be active either exclusively or in parallel[40,41]. When using a single (aggregated) $K_{m,g}$ becomes inadequate, a more accurate representation of the carbon transport could be achieved by explicitly modelling individual transporters (see below for further discussion). Besides, the proteome allocation principle could also be revised to investigate proteomic burdens imposed by the expression of heterologous genes[42].

In this work, we used the core metabolic model of *E. coli* to demonstrate the effectiveness of proteome allocation constraints (Eqs. (16 and 17)) in dictating bacterial growth strategies. However, the proposed modelling framework can be implemented with more detailed genome-scale metabolic models (GEMs)[43–45] to explore biological insights for *E. coli* and other microorganisms with similar putative mechanisms. The key advance of the GEM over the core model is the extensive information of metabolic reactions originating from genome annotation[46]. Pathways such as alternate carbon metabolism, amino acid metabolism, nucleotide metabolism, cofactor biosynthesis and fatty acid biosynthesis are exclusively modelled in GEMs. To fuse the proteome allocation constraints into a GEM, particular attention is needed for the choice of fluxes underlying v_c, v_f and v_r, as well as the determination of proteome cost parameters w_c^*, w_f^* and w_r^*.

We first focus on w_c^* and v_c. In the core model, glucose is the sole carbon source to support growth and is imported via phosphoenolpyruvate(PEP):pyruvate (PYR) phosphotransferase system (PTS). In this case v_c is related to a one-step transport flux (GLCpts). However in *E. coli* GEM *i*AF1260, glucose can be transported not only via glucose-specific PTS components, but also via alternative routes such as glucose-specific ABC system, glucose:proton symporter or simply via diffusion[43]. In *i*AF1260, glucose import (from extracellular glucose to cytosolic glucose-6-phosphate) is carried out by multiple enzymatic steps, e.g. first via the glucose:proton symporter then through hexokinase, instead of a simplified one-step process. Furthermore, 174 carbon sources were predicted to potentially support growth in *i*AF1260, meaning that growth simulation is not limited to glucose minimal media. The improved details of the metabolic network imply that (a) when modelling cells grown on glucose, v_c may need to be coupled to multiple glucose transporters and (b) when modelling cells grown on alternative or mixed carbon sources, v_c needs to correspond to various carbohydrate transport pathways. Accompanying a more detailed account of substrate transport fluxes, the (normalised) proteome cost parameter for carbon transport w_c^* has to be revised. We have shown in this work that w_c^* is a function of ϕ^o_{max}, $w_{c,0}$ and $K_{m,g}$ (Eqs. 6 and 16) and that $w_{c,0}$ is a function of ϕ^o_{max} and $K_{m,g}$ (Supplementary Eq. (S34)); combining the two makes w_c^* a function of $K_{m,g}$ (the Michaelis constant for glucose transport). When multiple transporters exist, as in GEMs, a set of $K_{m,g}$ (for glucose) or multiple sets of $K_{m,s}$ (for non-glucose carbon sources) would be required for active transporters in the specific growth conditions. Correspondingly, w_c^* will be calculated from the combination of $K_{m,g}$ for cells grown on a glucose minimal medium. For cells grown on a complex carbon medium, multiple w_c^* will be needed (to match distinct transport fluxes) and should be determined by multiple sets of $K_{m,s}$. In short, integrating the proteome cost constraints into GEM asks for more information of kinetic parameters of carbon transport.

Furthermore, the increased flux variability and potential reactions running in parallel in GEMs may alter the choice of representative fluxes of fermentation and respiration pathways. For example, v_f and v_r in *i*AF1260 can become enolase (ENO) and citrate synthase (CS)[42]. Changed v_f and v_r will subsequently affect the estimation of proteome cost parameters w_f^* and w_r^* (Supplementary Text).

In conclusion, our model offers a latest mechanistic interpretation of the seventy-year old Monod growth kinetics: the maximum specific growth rate of a microorganism could be governed by the abundance of growth-controlling proteome and the associated proteome cost per unit increase of growth rate. The Monod constant K_s was shown to be quantitatively related to not only the enzymatic affinity for substrate transport $K_{m,g}$ but also the metabolic state of a cell, which might explain the large variations in reported K_s values. Our analysis also suggests that a microorganism with a lower maintenance cost, higher fraction of growth-controlling proteome and alternative pathways with different enzyme costs likely competes successfully in a changing environment. Finally, the proposed modelling concept eliminates the need for treating substrate intake as an input in FBA simulation, which demonstrates the potential of coupling local and global physiological constrains in predictive modelling for systems biology and synthetic biology.

Methods

Modelling proteome allocation in *E. coli*.   Following the assumption that the fraction of a proteome sector is proportional to the metabolic flux it processes[26], Eq. (2) can be rewritten as

$$w_c v_c + w_f v_f + w_r v_r + b\lambda = \phi_{max}^g \tag{14}$$

w_i is the enzyme cost per unit flux processed by sector i ($i = C, f\ or\ r$). b is the proteome cost per unit biomass synthesis flux. v_c is the carbon transport flux. For *E. coli* grown with glucose as the single growth-controlling substrate, v_c is the glucose uptake flux which is normally processed by the phosphoenolpyruvate:sugar phosphotransferase system (PTS)[40,47]. v_f (v_r) is the fermentation (respiration) flux; λ is the biomass synthesis flux, which in this work is equal to the specific growth rate. From ref. [28], the fraction of proteome attainable to energy biogenesis and biomass synthesis is constrained by an upper bound ϕ_{max}^o

$$w_f^* v_f + w_r^* v_r + b^* \lambda \leq 1 \tag{15}$$

where $w_i^* \equiv \frac{w_i}{\phi_{max}^o}$ ($i = f\ or\ r$) and $b^* \equiv \frac{b}{\phi_{max}^o}$ are the (normalised) proteome cost per unit flux. Dividing both sides of Eq. (14) by ϕ_{max}^b

$$w_c^* v_c + w_f^* v_f + w_r^* v_r + b^* \lambda = \frac{\phi_{max}^g}{\phi_{max}^o} \tag{16}$$

Substituting Eq. (15) into Eq. (16), under acetate overflow (hence equal sign applied to Eq. (15)

$$w_c^* v_c = \frac{\phi_{max}^g}{\phi_{max}^o} - 1 \tag{17}$$

It follows that

$$\phi_{Cmin} = \phi_{max}^g - \phi_{max}^o \tag{18}$$

ϕ_{Cmin} is the minimum proteome fraction allocated to the carbon-scavenging sector (occurring during acetate overflow). $\phi_{max}^g > \phi_{max}^o$ (due to the inclusion of the C sector). Equations (16 and 17) are the proteome allocation constraints adopted in model simulations (see below).

Deriving the hyperbolic $\lambda - [g]$ correlation.   Starting from the Michaelis-Menten kinetics together with the proteome allocation embedded FBA model presented in this work, we have derived a hyperbolic $\lambda - [g]$ correlation that is comparable to the Monod equation and reveals the potential biological mechanisms that underpin the phenomenological Monod parameters. Detailed mathematical derivation is provided in the Supplementary Text.

Determination of GAM and ATPM.   In FBA models, the energy consumption is dictated primarily via growth-associated maintenance (GAM) and non-growth associated maintenance (ATPM), both of which are difficult to quantify accurately[43]. GAM and ATPM could also vary between different species and strains (BiGG database, http://bigg.ucsd.edu/). Therefore we modified the default values of GAM and ATPM according to strain-specific energetic data (Supplementary Table S2). To do so we associated GAM with the molar growth yield (namely the true growth yield Y_G[48]). ATPM is linked to the maintenance coefficient (m), which corresponds to the rate of substrate uptake extrapolated to zero growth rate.

Model simulation.   The following linear programing problem was solved in FBA simulations of the growth of *E. coli* under different extracellular glucose concentrations. FBA was performed based on the core *E. coli* metabolic model[49].

$$
\begin{aligned}
&maximise\ \lambda \\
&subject\ to\quad Sv = 0 \\
&\qquad lb \leq v_i \leq ub \\
&\qquad w_f^* v_f + w_r^* v_r + b^* \lambda \leq 1 \\
&\qquad w_c^* v_c + \left(w_f^* v_f + w_r^* v_r + b^* \lambda\right) = \phi_{max}^g / \phi_{max}^o \\
&\qquad w_c^* v_c = \phi_{max}^g / \phi_{max}^o - 1\ (if\ \lambda \geq \lambda_{ac})
\end{aligned}
\tag{19}
$$

v_c is represented by the first enzymatic step of glycolysis, GLCpts, v_f is the acetate synthesis reaction ACKr, v_r adopts the first reaction after biomass withdrawal in TCA cycle, AKGDH. λ is growth rate and refers to the biomass reaction. λ_{ac} is the threshold growth rate, above which acetate overflow occurs. The determination of model parameters ($w_c^*, w_f^*, w_r^*, b^*, \phi_{max}^o$ and ϕ_{max}^g) is detailed in the Supplementary Text. The lower bound of glucose and oxygen exchange fluxes were set to -1000 to avoid artificial control and to allow prediction of their uptake flux. Enzymatic reactions ICL, MALS, FRD7 were switched off under aerobic-glucose conditions according to[43]. Irrelevant exchange fluxes around the pyruvate node (EX_pyr(e), EX_lac(e), EX_acald(e), EX_etoh(e)) and the α-ketoglutarate node (EX_glu_L(e) and EX_akg(e)) were closed. Maximum carbon flow to the PP pathway was limited to 67% and 42% of the overall carbon intake for ML308 and NCM3722, respectively according to ref. [28].

Extracellular glucose concentration was adopted as sole model input (varies from 0.1–75 μM for NCM3722 and 0.1–26 μM for ML308). The optimisation model was converted to an MILP problem and solved using the Gurobi optimizer (see Supplementary Text for details). FBA was solved via COBRA Toolbox[50]. All simulations were performed in MATLAB R2016a. Experimental data used to compare with the simulation results were obtained from Fig. 1 of ref. [14], Fig. 3B of ref. [19] and SI Fig. 1 of ref. [51] for NCM3722 and from Table 7 of ref. [52], Fig. 3b of ref. [53] and Fig. 1 of ref. [34] for ML308.

Data availability

All model equations and parameter values are included in the main text or in the Supplementary Information. The MATLAB code for running simulations and the generated datasets in this study are available upon request from the corresponding author.

References

1. Monod, J. The growth of bacterial cultures. *Annu. Rev. Microbiol.* **3**, 371–394 (1949).
2. Kovárová-Kovar, K. & Egli, T. Growth kinetics of suspended microbial cells: from single-substrate-controlled growth to mixed-substrate kinetics. *Microbiol. Mol. Biol. Rev.* **62**, 646–666 (1998).
3. Tsipa, A., Koutinas, M., Usaku, C. & Mantalaris, A. Optimal bioprocess design through a gene regulatory network – Growth kinetic hybrid model: Towards replacing Monod kinetics. *Metab. Eng.* **48**, 129–137 (2018).
4. Orhon, D., Cokgor, E. U., Insel, G., Karahan, O. & Katipoglu, T. Validity of Monod kinetics at different sludge ages – Peptone biodegradation under aerobic conditions. *Bioresour. Technol.* **100**, 5678–5686 (2009).
5. Pala-Ozkok, I. *et al*. Characteristics of mixed microbial culture at different sludge ages: Effect on variable kinetics for substrate utilization. *Bioresour. Technol.* **126**, 274–282 (2012).
6. Wick, L. M., Weilenmann, H. & Egli, T. The apparent clock-like evolution of Escherichia coli in glucose-limited chemostats is reproducible at large but not at small population sizes and can be explained with Monod kinetics. *Microbiology* **148**, 2889–2902 (2002).
7. Franchini, A. G. & Egli, T. Global gene expression in Escherichia coli K-12 during short-term and long-term adaptation to glucose-limited continuous culture conditions. *Microbiology* **152**, 2111–2127 (2006).
8. Füchslin, H. P., Schneider, C. & Egli, T. In glucose-limited continuous culture the minimum substrate concentration for growth, s min, is crucial in the competition between the enterobacterium Escherichia coli and Chelatobacter heintzii, an environmentally abundant bacterium. *ISME J.* **6**, 777 (2012).
9. Insel, G. *et al*. Respirometric evaluation and modeling of glucose utilization by Escherichia coli under aerobic and mesophilic cultivation conditions. *Biotechnol. Bioeng.* **96**, 94–105 (2007).
10. Varma, A., Boesch, B. W. & Palsson, B. Ø. Stoichiometric interpretation of Escherichia coli glucose catabolism under various oxygenation rates. *Appl. Environ. Microbiol.* **59**, 2465–2473 (1993).
11. Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).
12. Varma, A. & Palsson, B. Ø. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. *Appl. Environ. Microbiol.* **60**, 3724–3731 (1994).
13. Molenaar, D., van Berlo, R., de Ridder, D. & Teusink, B. Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol Syst Biol* **5** (2009).
14. Basan, M. *et al*. Overflow metabolism in Escherichia coli results from efficient proteome allocation. *Nature* **528**, 99–104 (2015).
15. Thiele, I., Jamshidi, N., Fleming, R. M. T. & Palsson, B. O. Genome-scale reconstruction of escherichia coli's transcriptional and translational machinery: A knowledge base, its mathematical formulation, and its functional characterization. *PLoS Comput. Biol.* **5** (2009).
16. Beg, Q. K. *et al*. Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity. *P. Natl. Acad. Sci. USA.* **104** (2007).
17. Zhuang, K., Vemuri, G. N. & Mahadevan, R. Economics of membrane occupancy and respiro-fermentation. *Mol. Syst. Biol.* **7** (2011).
18. Goelzer, A. & Fromion, V. Bacterial growth rate reflects a bottleneck in resource allocation. *Biochim. Biophys. Acta (BBA)-General Subj.* **1810**, 978–988 (2011).
19. Mori, M., Hwa, T., Martin, O. C., De Martino, A. & Marinari, E. Constrained Allocation Flux Balance Analysis. *PLoS Comput. Biol.* **12** (2016).
20. Thiele, I. *et al*. Multiscale Modeling of Metabolism and Macromolecular Synthesis in E. coli and Its Application to the Evolution of Codon Usage. *PLoS One* **7**, e45635 (2012).
21. Lerman, J. A. *et al*. In silico method for modelling metabolism and gene product expression at genome scale. *Nat. Commun.* **3**, 929 (2012).
22. O'Brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R. & Palsson, B. O. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol. Syst. Biol.* **9**, 693–693 (2013).
23. Nilsson, A., Nielsen, J. & Palsson, B. O. Metabolic models of protein allocation call for the kinetome. *Cell Syst.* **5**, 538–541 (2017).
24. Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z. & Hwa, T. Interdependence of Cell Growth and Gene Expression: Origins and Consequences. *Science* **330**, 1099–1102 (2010).
25. You, C. *et al*. Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature* **500**, 301–306 (2013).
26. Hui, S. *et al*. Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Mol. Syst. Biol.* **11**, (2015).
27. Peebo, K. *et al*. Proteome reallocation in Escherichia coli with increasing specific growth rate. *Mol. BioSyst.* **11**, 1184–1193 (2015).
28. Zeng, H. & Yang, A. Modelling overflow metabolism in Escherichia coli with flux balance analysis incorporating differential proteomic efficiencies of energy pathways. *BMC Syst. Biol.* **13**, 3 (2019).
29. Chen, Y. & Nielsen, J. Energy metabolism controls phenotypes by protein efficiency and allocation. *Proc. Natl. Acad. Sci.* **116**, 17592–17597 (2019).
30. Snoep, J. L., Mrwebi, M., Schuurmans, J. M., Rohwer, J. M. & de Mattos, M. J. Control of specific growth rate in Saccharomyces cerevisiae. *Microbiology* **155**, 1699–1707 (2009).
31. Liu, Y. A simple thermodynamic approach for derivation of a general Monod equation for microbial growth. *Biochem. Eng. J.* **31**, 102–105 (2006).
32. Liu, Y. Overview of some theoretical approaches for derivation of the Monod equation. *Appl. Microbiol. Biotechnol.* **73**, 1241–1250 (2007).
33. Vazquez, A. & Oltvai, Z. N. Macromolecular crowding explains overflow metabolism in cells. *Sci. Rep.* **6**, 31007 (2016).

34. Koch, A. L. & Houston Wang, C. How close to the theoretical diffusion limit do bacterial uptake systems function? *Arch. Microbiol.* **131**, 36–42 (1982).
35. Hanly, T. J. & Henson, M. A. Dynamic metabolic modeling of a microaerobic yeast co-culture: predicting and optimizing ethanol production from glucose/xylose mixtures. *Biotechnol. Biofuels* **6**, 44 (2013).
36. Henson, M. A. & Hanly, T. J. Dynamic flux balance analysis for synthetic microbial communities. *IET Syst. Biol.* **8**, 214–229 (2014).
37. Hjersted, J. L. & Henson, M. A. Optimization of fed-batch Saccharomyces cerevisiae fermentation using dynamic flux balance models. *Biotechnol. Prog.* **22**, 1239–1248 (2006).
38. Hjersted, J. L., Henson, M. A. & Mahadevan, R. Genome-scale analysis of Saccharomyces cerevisiae metabolism and ethanol production in fed-batch culture. *Biotechnol. Bioeng.* **97**, 1190–1204 (2007).
39. Hanly, T. J. & Henson, M. A. Dynamic model-based analysis of furfural and HMF detoxification by pure and mixed batch cultures of S. cerevisiae and S. stipitis. *Biotechnol. Bioeng.* **111**, 272–284 (2014).
40. Gosset, G. Improvement of Escherichia coli production strains by modification of the phosphoenolpyruvate:sugar phosphotransferase system. *Microb. Cell Fact.* **4**, 14 (2005).
41. Stock, J. B., Waygood, E. B., Meadow, N. D., Postma, P. W. & Roseman, S. Sugar transport by the bacterial phosphotransferase system. The glucose receptors of the Salmonella typhimurium phosphotransferase system. *J. Biol. Chem.* **257**, 14543–14552 (1982).
42. Zeng, H. & Yang, A. Quantification of proteomic and metabolic burdens predicts growth retardation and overflow metabolism in recombinant Escherichia coli. *Biotechnol. Bioeng.* (2019).
43. Feist, A. M. *et al.* A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* **3** (2007).
44. Orth, J. D. *et al.* A comprehensive genome-scale reconstruction of Escherichia coli metabolism—2011. *Mol. Syst. Biol.* **7**, 535 (2011).
45. Mo, M. L., Palsson, B. Ø. & Herrgård, M. J. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst. Biol.* **3**, 37 (2009).
46. Thiele, I. & Palsson, B. Ø. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* **5**, 93–121 (2010).
47. Steinsiek, S. & Bettenbrock, K. Glucose Transport in Escherichia coli Mutant Strains with Defects in Sugar Transport Systems. *J. Bacteriol.* **194**, 5897–5908 (2012).
48. Pirt, S. J. The maintenance energy of bacteria in growing cultures. *Proc. R. Soc. Lond. B* **163**, 224–231 (1965).
49. Orth, J. D., Palsson, B. Ø. & Fleming, R. M. T. Reconstruction and Use of Microbial Metabolic Networks: the Core Escherichia coli Metabolic Model as an Educational Guide. *EcoSal Plus* **4** (2010).
50. Schellenberger, J. *et al.* Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2. 0. *Nat. Protoc.* **6**, 1290 (2011).
51. Bren, A., Hart, Y., Dekel, E., Koster, D. & Alon, U. The last generation of bacterial growth in limiting nutrient. *BMC Syst. Biol.* **7**, 27 (2013).
52. Holms, H. Flux analysis and control of the central metabolic pathways in Escherichia coli. *FEMS Microbiol. Rev.* **19**, 85–116 (1996).
53. Noel, J. T., Cox, B. & Narang, A. Identification of the growth-limiting step in continuous cultures from initial rates measured in response to substrate-excess conditions. *arXiv Prepr. q-bio/0509013* (2005).

Acknowledgements

Author contributions

H.Z. and A.Y. designed the research; H.Z. performed simulation and data analysis; H.Z. and A.Y. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-61174-0.

Correspondence and requests for materials should be addressed to A.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.