Computation of single-cell metabolite distributions using mixture models

Mona K. Tonn,[1] Philipp Thomas,[1] Mauricio Barahona,[1] and Diego A. Oyarzún[2,3,a)]
[1)] *Department of Mathematics, Imperial College London, SW7 2AZ, London, United Kingdom*
[2)] *School of Biological Sciences, University of Edinburgh, United Kingdom*
[3)] *School of Informatics, University of Edinburgh, United Kingdom*

Metabolic heterogeneity is widely recognised as the next challenge in our understanding of non-genetic variation. A growing body of evidence suggests that metabolic heterogeneity may result from the inherent stochasticity of intracellular events. However, metabolism has been traditionally viewed as a purely deterministic process, on the basis that highly abundant metabolites tend to filter out stochastic phenomena. Here we bridge this gap with a general method for prediction of metabolite distributions across single cells. By exploiting the separation of time scales between enzyme expression and enzyme kinetics, our method produces estimates for metabolite distributions without the lengthy stochastic simulations that would be typically required for large metabolic models. The metabolite distributions take the form of Gaussian mixture models that are directly computable from single-cell expression data and standard deterministic models for metabolic pathways. The proposed mixture models provide a systematic method to predict the impact of biochemical parameters on metabolite distributions. Our method lays the groundwork for identifying the molecular processes that shape metabolic heterogeneity and its functional implications in disease.

I. INTRODUCTION

Non-genetic heterogeneity is a hallmark of cell physiology. Isogenic cells can display markedly different phenotypes as a result of the stochasticity of intracellular processes and fluctuations in environmental conditions. Gene expression variability, in particular, has received substantial attention thanks to robust experimental techniques for measuring transcripts and proteins at a single-cell resolution[1,2]. This progress has gone hand-in-hand with a large body of theoretical work on stochastic models to identify the molecular processes that affect expression heterogeneity[3–7].

In contrast to gene expression, our understanding of stochastic phenomena in metabolism is still in its infancy. Traditionally, cellular metabolism has been regarded as a deterministic process on the basis that metabolites appear in large numbers that filter out stochastic phenomena[8]. But this view is changing rapidly thanks to a growing number of single-cell measurements of metabolites and co-factors[9–17] that suggest that cell-to-cell metabolite variation is much more pervasive than previously thought. The functional implications of this heterogeneity are largely unknown but likely to be substantial given the roles of metabolism in many cellular processes, including growth[18], gene regulation[19], epigenetic control[20] and immunity[21]. For example, metabolic heterogeneity has been linked to bacterial persistence[22,23], a dormant phenotype characterised by a low metabolic activity, as well as antibiotic resistance[24] and other functional effects[25]. In biotechnology applications, metabolic heterogeneity is widely recognised as a limiting factor on metabolite production with genetically engineered microbes[26–28].

A key challenge for quantifying metabolic variability is the difficulty in measuring cellular metabolites at a single-cell resolution[29–31]. As a result, most studies use other phenotypes as a proxy for metabolic variation, e.g. enzyme expression levels[32,33], metabolic fluxes[34] or growth rate[35,36]. From a computational viewpoint, the key challenge is that metabolic processes operate on two timescales: a slow timescale for expression of metabolic enzymes, and a fast timescale for enzyme catalysis. Such multiscale structure results in stiff models that are infeasible to solve with standard algorithms for stochastic simulation[37]. Other strategies to accelerate stochastic simulations, such as τ-leaping, also fail to produce accurate simulation results due to the disparity in molecule numbers between enzymes and metabolites[38]. These challenges have motivated a number of methods to optimise stochastic simulations of metabolism[39–43]. Most of these methods exploit the timescale separation to accelerate simulations at the expense of some approximation error. This progress has been accompanied by a number of theoretical results on the links between molecular processes and the shape of metabolite distributions[6,44–46]. Yet to date there are no general methods for computing metabolite distributions that can handle inherent features of metabolic pathways such as feedback regulation, complex stoichiometries, and the high number of molecular species involved.

In this paper we present a widely applicable method for approximating single-cell metabolite distributions. Our method is founded on the timescale separation between enzyme expression and enzyme catalysis, which we employ to approximate the stationary solution of the chemical master equation. The approximate solution takes the form of mixture distributions with: (i) mixture weights that can be computed from models for gene expression or single-cell expression data, and (ii) mixture components that are directly computable from deterministic pathway models. The resulting mixture model can be employed to explore the impact of biochemical parameters on metabolite variability. We illustrate the power of the method in two exemplar systems that are core building blocks of large metabolic networks. Our theory provides a quantitative basis to draw testable hypotheses on the sources of metabolite heterogeneity, which together with the ongoing efforts in single-cell metabolite measurements, will help to re-evaluate the role of metabolism as an active source of pheno-

[a)] Corresponding author: d.oyarzun@ed.ac.uk; This work was partly funded by the Human Frontier Science Program through a Young Investigator Grant (RGY0076-2015) awarded to D.O., a UKRI Future Leaders Fellowship (MR/T018429/1) awarded to P.T., and the EPSRC Centre for Mathematics of Precision Healthcare (EP/N014529/1) awarded to M.B.

typic variation.

II. GENERAL METHOD FOR COMPUTING METABOLITE DISTRIBUTIONS

We consider metabolic pathways composed of enzymatic reactions interconnected by sharing of metabolites as substrates or products. In general, we consider models with M metabolites P_i with $i \in \{1, 2, ..., M\}$ and N catalytic enzymes E_j with $j \in \{1, 2, ..., N\}$. A typical enzymatic reaction has the form

$$P_i + E_j \underset{k_{\mathrm{b},j}}{\overset{k_{\mathrm{f},j}}{\rightleftharpoons}} C_j \underset{k_{\mathrm{rev},j}}{\overset{k_{\mathrm{cat},j}}{\rightleftharpoons}} P_k + E_j, \tag{1}$$

where P_i and P_k are metabolites, and E_j and C_j are the free and substrate-bound forms of the enzyme. The parameters $(k_{\mathrm{f},j}, k_{\mathrm{b},j})$ and $(k_{\mathrm{cat},j}, k_{\mathrm{rev},j})$ are positive rate constants specific to the enzyme. In contrast to traditional metabolic models, where the number of enzyme molecules is assumed constant, here we explicitly model enzyme expression and enzyme catalysis as stochastic processes. Our models also account for dilution of molecular species by cell growth and consumption of the metabolite products by downstream processes.

Though in principle one can readily write a Chemical Master Equation (CME) for the marginal distribution $\mathbf{P}(P_1, P_2, ...P_M)$ given the pathway stoichiometry, analytical solutions of the CME are tractable only in few special cases. To overcome this challenge, we propose a method for approximating metabolite distributions that can be applied in a wide range of metabolic models. We first note that using the Law of Total Probability, the marginal distribution $\mathbf{P}(P_1, P_2, ...P_M)$ can be generally written as:

$$\mathbf{P}(P) = \sum_E \mathbf{P}(E) \times \mathbf{P}(P|E), \tag{2}$$

where $P = (P_1, P_2, ...P_M)$ and $E = (E_1, E_2, ..., E_N)$ are the vectors of metabolite and enzyme abundances, respectively. The equation in (2) describes the metabolite distribution in terms of fluctuations in gene expression, comprised in the distribution $\mathbf{P}(E)$, and fluctuations in reaction catalysis, described by conditional distribution $\mathbf{P}(P|E)$.

A key observation is that Eq. (2) corresponds to a mixture model with weights $\mathbf{P}(E)$ and mixture components $\mathbf{P}(P|E)$. To compute the mixture weights and components, we make use of the timescale separation between gene expression and metabolism. Gene expression operates on a much slower timescale than catalysis[40,44,47], with protein half-lives typically comparable to cell doubling times and catalysis operating in the millisecond to second range. Therefore, in the fast timescale of catalysis we can write a conservation law for the total amount of each enzyme (free and bound):

$$E_{\mathrm{t},j} = E_j + C_j, \tag{3}$$

where $E_{\mathrm{t},j}$ is the total number of enzymes E_j. Note that since our models integrate enzyme kinetics with enzyme expression, the variables $E_{\mathrm{t},j}$ follow their own, independent stochastic dynamics. It is important to note that in our approach, the conservation relation in (3) holds only in the fast timescale of catalysis. This contrasts with classic deterministic models for metabolic reactions, which typically focus on the fast catalytic timescale and assume enzymes as constant model parameters[48].

As a result of the separation of timescales, the weights and components of the mixture in Eq. (2) can be computed separately. The mixture weights $\mathbf{P}(E)$, in particular, can be computed as solutions of a stochastic model for enzyme expression[4], or taken from single-cell measurements of enzyme expression. The mixture components $\mathbf{P}(P|E)$, on the other hand, can be estimated with the Linear Noise Approximation[49,50] (LNA) on the basis that metabolites appear in large numbers. In Figure 1 we illustrate a schematic of the proposed method.

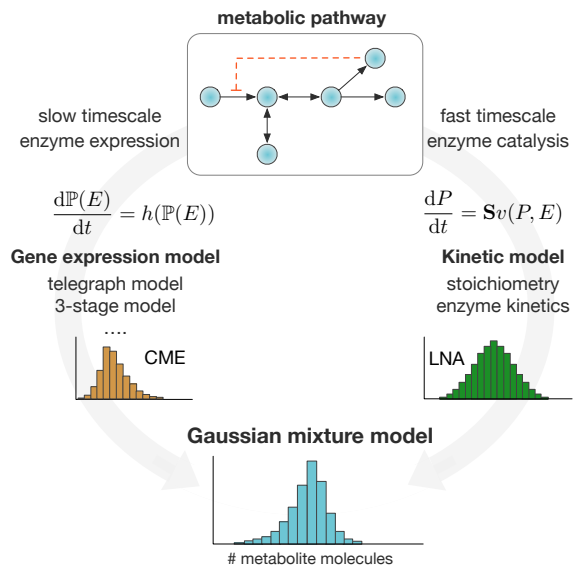

FIG. 1. **Computation of single-cell metabolite distributions with Gaussian mixture models.** We exploit the separation of timescales to compute the weights and components of the mixture model in Eq. (2). Mixture weights are computed as stationary solutions to the Chemical Master Equation (CME) for a chosen model for stochastic enzyme expression. The mixture components are computed via the Linear Noise Approximation[50] (LNA) applied to the pathway ODE model. This provides Gaussian approximations to the stationary metabolite distribution conditioned on the enzyme state. Overall, the method produces a Gaussian mixture model for metabolite distributions that can be applied in a wide range of metabolic pathways.

We thus propose the following procedure for computing single-cell metabolite distributions:

1. Starting from the mixture model in Eq. (2), compute the enzyme distribution $\mathbf{P}(E)$ from a stochastic model for gene expression, either analytically (if possible) or numerically with Gillespie's algorithm.

2. To approximate the mixture components $\mathbf{P}(P|E)$ with the LNA, compute the steady state solution $\bar{P}$ of the

deterministic rate equation for each enzyme state E:

$$\mathbf{S}v(\bar{P}, E) = 0, \tag{4}$$

where $\mathbf{S}$ is the stoichiometric matrix and $v(\cdot)$ is the vector of deterministic reaction rates; for ease of notation we have assumed a unit cell volume, and hence the deterministic rates are equal to the propensities of the stochastic model. Note that due to the timescale separation, Eq. (4) must be solved assuming constant enzymes E, and its solution depends on the enzyme abundance, i.e. $\bar{P} = \bar{P}(E)$.

3. For each enzyme state E, compute the solution to the Lyapunov equation[50]:

$$A\Sigma + \Sigma A^T + BB^T = 0, \tag{5}$$

where A is the Jacobian of (4) evaluated at the steady state and $BB^T = \mathbf{S}\text{diag}\{v\}\mathbf{S}^T$. Note that, as in (4), the solution of the Lyapunov equation depends on the enzyme state, i.e. $\Sigma = \Sigma(E)$.

4. Following the LNA, approximate the mixture components $\mathbf{P}(P|E)$ as a multivariate Gaussian distribution with mean $\bar{P}$ and covariance matrix Σ.

5. Combine the weights $\mathbf{P}(E)$ and Gaussian components $\mathbf{P}(P|E)$ through the mixture model in (2).

In the next sections we illustrate the effectiveness of our method in two exemplar systems.

III. REVERSIBLE MICHAELIS-MENTEN REACTION

We first consider a stochastic model that integrates a reversible Michaelis-Menten reaction with a standard model for enzyme expression. As shown in Figure 2A, the Michaelis-Menten mechanism includes reversible binding of four species: a metabolic substrate S, a free enzyme E, a substrate-enzyme complex C and a metabolic product P. To model enzyme expression, we use the well-known two-stage scheme for transcription and translation[51,52] (Figure 2A). The complete set of reactions is:

$$S + E \underset{k_{\text{b}}}{\overset{k_{\text{f}}}{\rightleftharpoons}} C \underset{k_{\text{rev}}}{\overset{k_{\text{cat}}}{\rightleftharpoons}} P + E, \tag{6}$$

$$\emptyset \xrightarrow{k_{\text{tx}}} \text{mRNA} \xrightarrow{k_{\text{tl}}} \text{mRNA} + E, \tag{7}$$

$$P \xrightarrow{k_{\text{c}}} \emptyset, \quad \text{mRNA} \xrightarrow{k_{\text{deg}}} \emptyset, \tag{8}$$

$$E \xrightarrow{\delta} \emptyset, \quad C \xrightarrow{\delta} \emptyset. \tag{9}$$

The reactions in (6) correspond to a reversible Michaelis-Menten reaction as in (1), while reactions in (7) are the two-stage model for gene expression. We include four additional first-order reactions (8)–(9) to model consumption of the metabolite product with rate constant k_{c}, mRNA degradation with rate constant k_{deg}, and dilution of all model species with rate constant δ. In what follows we assume that the substrate S remains strictly constant, for example to model cases in which the substrate represents an extracellular carbon source that evolves in much slower timescale than cell doubling times.

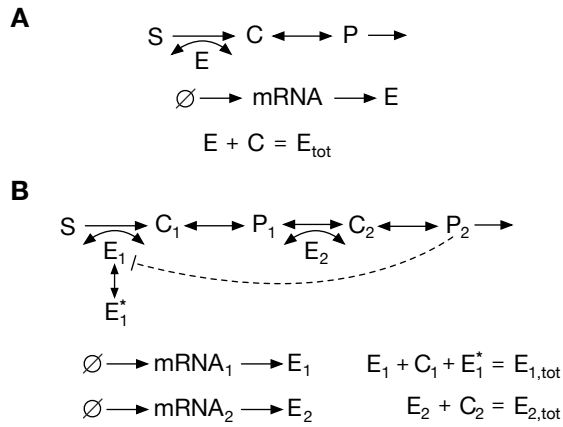

FIG. 2. **Exemplar metabolic systems**. (**A**) Reversible Michaelis-Menten reaction; the full set of reactions are shown in Eq. (6)–(7). The model accounts for reversible catalysis of a substrate S into a product P. (**B**) Two-step pathway with noncompetitive end-product inhibition; the reactions are shown in Eq. (18)–(25). The product (P_2) sequesters enzyme E_1 into an inactive form E_1^*, thereby reducing the rate of the first reaction. In both examples we assume a constant substrate S and linear dilution of all chemical species. Enzymes are assumed to follow the two-stage model for gene expression[52], which includes species for the enzymatic mRNA and protein.

Since on the fast timescale of the catalytic reaction, the total number of enzymes can be assumed in quasi-stationary state[6,48], we have that

$$E_{\text{total}} = E + C, \tag{10}$$

and therefore the general mixture model in (2) can be written as:

$$\mathbf{P}(P) = \sum_{E_{\text{total}}=0}^{\infty} \underbrace{\mathbf{P}(E_{\text{total}})}_{\substack{\text{enzyme}\\ \text{distribution}}} \times \underbrace{\mathbf{P}(P|E_{\text{total}})}_{\substack{\text{Gaussian}\\ \text{from LNA}}}. \tag{11}$$

The mixture weights $\mathbf{P}(E_{\text{total}})$ can be computed from the stochastic model for gene expression in (7). Under the standard assumption that mRNAs are degraded much faster than proteins[4], the stationary solution of the two-stage model can be approximated by a negative binomial distribution[52]:

$$\mathbf{P}(E_{\text{total}}) = \frac{\Gamma(a + E_{\text{total}})}{\Gamma(E_{\text{total}} + 1)\Gamma(a)} \left(\frac{b}{1+b}\right)^{E_{\text{total}}} \frac{1}{(1+b)^a}, \tag{12}$$

where Γ is the Gamma function and the parameters are defined as the burst frequency $a = k_{\text{tx}}/\delta$ and burst size $b = k_{\text{tl}}/k_{\text{deg}}$.

To compute the mixture components $\mathbf{P}(P|E_{\text{total}})$ with the LNA, we write the full system of deterministic rate equations (see (35) in Methods) for the three species E, C and P. Note that in this case, we can further reduce the rate equations by (i)

using the conservation law in (10), and (ii) assuming that the binding and unbinding reactions between S and E reach equilibrium faster than the product P, a condition that generally holds in metabolic reactions. After algebraic manipulations, the reduced ODE can be written as:

$$\frac{\mathrm{d}P}{\mathrm{d}t} = f(P, E_{\text{total}}) - g(P, E_{\text{total}}) - k_{\text{c}}P \tag{13}$$

where

$$\begin{aligned} f(P, E_{\text{total}}) &= E_{\text{total}} \frac{k_{\text{cat}} S/K_{\text{m}S}}{1 + S/K_{\text{m}S} + P/K_{\text{m}P}}, \\ g(P, E_{\text{total}}) &= E_{\text{total}} \frac{k_{\text{b}} P/K_{\text{m}P}}{1 + S/K_{\text{m}S} + P/K_{\text{m}P}} \end{aligned} \tag{14}$$

and the parameters are $K_{\text{m}S} = (k_{\text{b}} + k_{\text{cat}})/k_{\text{f}}$ and $K_{\text{m}P} = (k_{\text{b}} + k_{\text{cat}})/k_{\text{rev}}$.

The mean of each mixture component is simply given by the steady state solution of (13), which we denote as $\bar{P}(E_{\text{total}})$. For a given enzyme abundance E_{total}, the variance $\Sigma(E_{\text{total}})$ of each Gaussian component is given by the solution to the Lyapunov equation in (5):

$$\Sigma(E_{\text{total}}) = \frac{1}{2} \frac{f(\bar{P}(E_{\text{total}})) + g(\bar{P}(E_{\text{total}})) + k_{\text{c}}\bar{P}(E_{\text{total}})}{k_{\text{c}} + g'(\bar{P}(E_{\text{total}})) - f'(\bar{P}(E_{\text{total}}))}, \tag{15}$$

where f' and g' are first-order derivatives. Combining the negative binomial in (12) with the Gaussian components, we can rewrite Eq. (11) to get a Gaussian mixture model for the metabolite:

$$\mathbf{P}(P) = K \sum_{x=0}^{\infty} \frac{1}{\Sigma(x)} \frac{\Gamma(a+x)}{\Gamma(x+1)} \left(\frac{b}{1+b}\right)^{x} e^{-\frac{1}{2}\left(\frac{P-\bar{P}(x)}{\Sigma(x)}\right)^2}, \tag{16}$$

where both $\bar{P}(x)$ and $\Sigma(x)$ must be computed for each value of $x = E_{\text{total}}$ in the summation. The normalization constant in (16) is

$$K = \frac{1}{\sqrt{2\pi}\Gamma(a)(1+b)^a}. \tag{17}$$

In Figure 3 we plot the mixture model (16) for realistic parameter values and compare this approximation with distributions computed from long runs of Gillespie simulations of the whole set of reactions (6)–(9). The results indicate that the mixture model provides an excellent approximation of the metabolite distribution, even in the case of skewed or tailed distributions. In the next section we test our methodology in a more complex pathway with feedback regulation.

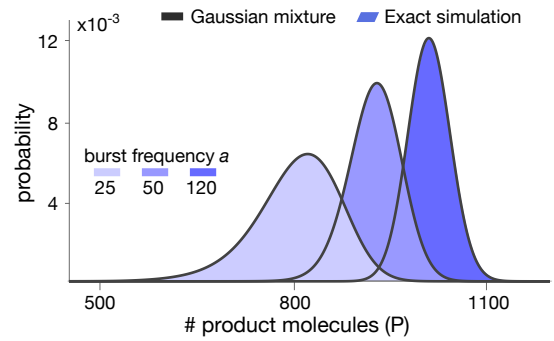

FIG. 3. **Stationary product distribution of a Michaelis-Menten reaction.** The proposed mixture model in (16) provides an excellent approximation for the metabolite distribution obtained with Gillespie's algorithm[37]. Distributions were computed for varying values of the bursting parameter a, suggesting that high-frequency bursting tends to decrease metabolite skewness. All parameter values can be found in Table I in the Methods.

IV. PATHWAY WITH END-PRODUCT INHIBITION

A common regulatory motif in metabolism is end-product inhibition, in which a pathway enzyme can bind to its own substrate as well as the pathway product (see Figure 2B). The product thus sequesters enzyme molecules, which reduces the number of free enzymes available for catalysis and slows done the reaction rate. To examine the accuracy of our method in this setting, we study a fully stochastic model for a two-step pathway with noncompetitive end-product inhibition:

$$S + E_1 \underset{k_{\text{b},1}}{\overset{k_{\text{f},1}}{\rightleftharpoons}} C_1 \underset{k_{\text{rev},1}}{\overset{k_{\text{cat},1}}{\rightleftharpoons}} P_1 + E_1 \tag{18}$$

$$P_1 + E_2 \underset{k_{\text{b},2}}{\overset{k_{\text{f},2}}{\rightleftharpoons}} C_2 \underset{k_{\text{rev},2}}{\overset{k_{\text{cat},2}}{\rightleftharpoons}} P_2 + E_2 \tag{19}$$

$$hP_2 + E_1 \underset{k_{\text{rsq}}}{\overset{k_{\text{sq}}}{\rightleftharpoons}} E^* \tag{20}$$

$$\emptyset \xrightarrow{k_{\text{tx},1}} \text{mRNA}_1 \xrightarrow{k_{\text{tl},1}} \text{mRNA}_1 + E_1 \tag{21}$$

$$\emptyset \xrightarrow{k_{\text{tx},2}} \text{mRNA}_2 \xrightarrow{k_{\text{tl},2}} \text{mRNA}_2 + E_2 \tag{22}$$

$$P_1 \xrightarrow{k_{\text{c},1}} \emptyset, \quad P_2 \xrightarrow{k_{\text{c},2}} \emptyset \tag{23}$$

$$\text{mRNA}_1 \xrightarrow{k_{\text{deg},1}} \emptyset, \quad \text{mRNA}_2 \xrightarrow{k_{\text{deg},2}} \emptyset, \tag{24}$$

$$E^* \xrightarrow{\delta} \emptyset, \quad E_1 \xrightarrow{\delta} \emptyset, \quad E_2 \xrightarrow{\delta} \emptyset, \quad C_1 \xrightarrow{\delta} \emptyset, \quad C_2 \xrightarrow{\delta} \emptyset \tag{25}$$

The two reactions in (18) and (19) are reversible Michaelis-Menten kinetics, sharing the intermediate metabolite P_1 as a product and substrate, respectively. The end-product inhibition in (20) consists of reversible binding between h molecules of P_2 and the first enzyme E_1 into a catalytically-inactive complex E^*. The remaining model reactions in (21)–(25) are analogous to the previous example in Section III: reactions in (21)–(22) describe the two-stage model for expression of both enzymes, and with reactions (23)–(25) we

model first-order mRNA degradation, product consumption, and dilution by cell growth. For simplicity we also assume that both enzymes are independently expressed, but in general our method can also account for cases in which enzymes are co-expressed or co-regulated[53]. The resulting model has two distinct pools of enzymes, which remain constant over the timescale of catalysis:

$$\begin{aligned} E_{\mathrm{t},1} &= E_1 + E^* + C_1, \\ E_{\mathrm{t},2} &= E_2 + C_2, \end{aligned} \tag{26}$$

and therefore the mixture model in (2) becomes

$$\mathbf{P}(P_1, P_2) = \sum_{E_{\mathrm{t},1}, E_{\mathrm{t},2}} \underbrace{\mathbf{P}(E_{\mathrm{t},1}, E_{\mathrm{t},2})}_{\substack{\text{enzyme}\\ \text{distribution}}} \underbrace{\mathbf{P}(P_1, P_2 | E_{\mathrm{t},1}, E_{\mathrm{t},2})}_{\substack{\text{Gaussian}\\ \text{from LNA}}}, \tag{27}$$

where the summation goes through all $(E_{\mathrm{t},1}, E_{\mathrm{t},2})$ pairs. Since both enzymes are expressed independently, the enzyme distribution is the product of two negative binomials $\mathbf{P}(E_{\mathrm{t},1}, E_{\mathrm{t},2}) = \mathbf{P}(E_{\mathrm{t},1}) \times \mathbf{P}(E_{\mathrm{t},2})$, each one analogous to the distribution in (12).

To compute the mixture components with the LNA, we use the rate equations for the reactions in (18)–(23); the full set of ODEs is listed in Eq. (36) in the Methods. As in the first example, by employing the conservation laws in (26) and assuming rapid equilibrium of the complexes C_1 and C_2, the deterministic model can be further simplified to a 2-dimensional ODE:

$$\begin{aligned} \frac{\mathrm{d}P_1}{\mathrm{d}t} &= f(P_1, P_2) - g(P_1, P_2) - k_{\mathrm{c},1} P_1, \\ \frac{\mathrm{d}P_2}{\mathrm{d}t} &= g(P_1, P_2) - k_{\mathrm{c},2} P_2, \end{aligned} \tag{28}$$

where for ease of notation we have omitted the dependency on $E_{\mathrm{t},1}$ and $E_{\mathrm{t},2}$. The nonlinear functions in (28) are

$$\begin{aligned} f(P_1, P_2) &= E_{\mathrm{t},1} \frac{\kappa_{\mathrm{S}} S - \kappa_1 P_1}{1 + \theta P_2^h + S/K_{\mathrm{m},S} + P_1/K_{\mathrm{m},1}}, \\ g(P_1, P_2) &= E_{\mathrm{t},2} \frac{\kappa_2 P_1 - \kappa_3 P_2}{1 + P_1/K_{\mathrm{m},2} + P_2/K_{\mathrm{m},3}}, \end{aligned} \tag{29}$$

where $\theta = k_{\mathrm{sq}}/k_{\mathrm{rsq}}$ is the product-enzyme binding constant and the remaining parameters are defined as $\kappa_{\mathrm{S}} = k_{\mathrm{cat},1} k_{\mathrm{f},1}/(k_{\mathrm{b},1} + k_{\mathrm{cat},1})$, $\kappa_1 = k_{\mathrm{b},1} k_{\mathrm{rev},1}/(k_{\mathrm{b},1} + k_{\mathrm{cat},1})$, $\kappa_2 = k_{\mathrm{cat},2} k_{\mathrm{f},2}/(k_{\mathrm{b},2} + k_{\mathrm{cat},2})$, $\kappa_3 = k_{\mathrm{b},2} k_{\mathrm{rev},2}/(k_{\mathrm{b},2} + k_{\mathrm{cat},2})$, $K_{\mathrm{m},S} = k_{\mathrm{cat},1}/\kappa_{\mathrm{S}}$, $K_{\mathrm{m},1} = k_{\mathrm{b},1}/\kappa_1$, $K_{\mathrm{m},2} = k_{\mathrm{cat},2}/\kappa_2$ and $K_{\mathrm{m},3} = k_{\mathrm{b},2}/\kappa_3$.

As in the previous example, the ODEs in (28) correspond to the full model (36) rewritten in terms of both metabolites assuming that the enzyme-substrate reactions reach equilibrium in a faster timescale than catalysis. This reduced model can be readily employed to obtain approximations for the mixture components with the LNA. If we denote as $\bar{P} = \bar{P}(E_{\mathrm{t},1}, E_{\mathrm{t},2})$ the steady state solution of (28), we can write the Lyapunov equation as $A\Sigma + \Sigma A^T + BB^T = 0$ with A and BB^T given by

$$A = \begin{bmatrix} \frac{\mathrm{d}}{\mathrm{d}P_1}(f - g) - k_{\mathrm{c},1} & \frac{\mathrm{d}}{\mathrm{d}P_2}(f - g) \\ \frac{\mathrm{d}g}{\mathrm{d}P_1} & \frac{\mathrm{d}g}{\mathrm{d}P_2} - k_{\mathrm{c},2} \end{bmatrix}, \tag{30}$$

$$BB^T = \begin{bmatrix} f + g + k_{\mathrm{c},1} P_1 & -g \\ -g & g + k_{\mathrm{c},2} P_2 \end{bmatrix}, \tag{31}$$

where $f(\cdot)$, $g(\cdot)$, and their derivatives are evaluated at the steady state solution $\bar{P}(E_{\mathrm{t},1}, E_{\mathrm{t},2})$. The Gaussian components of the mixture model are then

$$\mathbf{P}(P_1, P_2 | E_{\mathrm{t},1}, E_{\mathrm{t},2}) = \frac{1}{2\pi |\Sigma(E_{\mathrm{t},1}, E_{\mathrm{t},2})|} e^{-\frac{1}{2}(P - \bar{P}(E_{\mathrm{t},1}, E_{\mathrm{t},2}))^T \Sigma^{-1} (P - \bar{P}(E_{\mathrm{t},1}, E_{\mathrm{t},2}))}, \tag{32}$$

where $P = (P_1, P_2)^T$ and $|\cdot|$ is the matrix determinant. After combining the joint distribution of enzymes and the components into Eq. (27), we get a Gaussian mixture model for the joint marginal distribution of both metabolites:

$$\mathbf{P}(P_1, P_2) = K \sum_{x,y=0}^{\infty} \frac{\Gamma(a_1 + x)\Gamma(a_2 + y)}{\Gamma(x+1)\Gamma(y+1)} \left(\frac{b_1}{1 + b_1}\right)^x \left(\frac{b_2}{1 + b_2}\right)^y \times \frac{1}{|\Sigma(x, y)|} e^{-\frac{1}{2}(P - \bar{P}(x,y))^T \Sigma(x,y)^{-1} (P - \bar{P}(x,y))}, \tag{33}$$

where $\bar{P}(x, y)$ and $\Sigma(x, y)$ need to computed numerically for each pair $(x, y) = (E_{\mathrm{t},1}, E_{\mathrm{t},2})$ in the summation. The burst frequencies $a_i = k_{\mathrm{tx},i}/\delta$ and burst sizes $b_i = k_{\mathrm{tl},i}/k_{\mathrm{deg},i}$ are specific to each enzyme, and the normalisation constant is given by

$$K = \frac{1}{2\pi \Gamma(a_1)\Gamma(a_2)(1 + b_1)^{a_1}(1 + b_2)^{a_2}}. \tag{34}$$

To test the quality of the approximation, we numerically computed the mixture model in (33) for various combinations of parameter values, shown in Figure 4. We observe that the mixture model offers an excellent approximation as compared to exact Gillespie simulations of the full model (18)–(25). We note that in this case, the full stochastic model has seven species and three different timescales, and therefore the runtime of Gillespie simulations are extremely long, in the order of several hours per run.

To further illustrate the utility of our method, we employed the mixture model to study the impact of parameter perturbations on the metabolite distributions. Without an analytical solution, such a study would require the computation of long Gillespie simulations for each combination of parameter values, which quickly become infeasible due to the long simulation time. In contrast, the mixture model provides a systematic way to rapidly evaluate the influence of model parameters on metabolite distributions. In Figure 5A we show summary statistics of the marginal $\mathbf{P}(P_1)$ for various combinations of average enzyme expression levels. The results suggest that

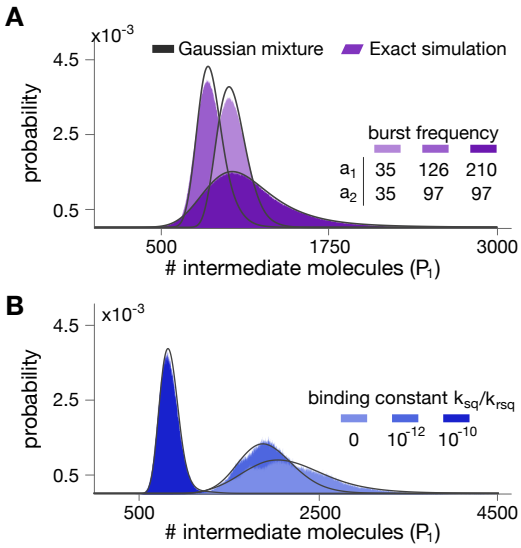

FIG. 4. **Stationary distributions for the intermediate metabolite in a two-step pathway with end-product inhibition.** The panels show the distribution of intermediate metabolite P_1 for different combinations of parameter values. **(A)** Impact of enzyme bursting frequency a_1 and a_2. **(B)** Impact of binding constant between the first enzyme and the end-product. All parameter values can be found in Table II in the Methods.

expression levels can have a strong impact on the mean and coefficient of variation of the intermediate metabolite. Moreover, in Figure 5B we plot the distribution $\mathbf{P}(P_1, P_2)$ for combinations of bursting parameters. The results show that uncorrelated enzyme fluctuations can still result in correlated and skewed metabolite distributions.

V. DISCUSSION

Cellular metabolism has traditionally been assumed to follow deterministic dynamics. This paradigm results largely from the observation that cellular metabolites are highly abundant. However, recent data shows that single-cell metabolite distributions can display substantial heterogeneity in their abundance across single cells[9–17]. It has also been shown that expression of metabolic genes is as variable as any other component of the proteome[2], and thus in principle it is plausible that such enzyme fluctuations propagate to metabolites. These observations have begun to challenge the paradigm of metabolism being a deterministic process, suggesting that metabolite fluctuations may play a role in non-genetic heterogeneity.

Here we described a new computational tool to predict the statistics of metabolite fluctuations in conjunction with gene expression. The method is based on a timescale separation argument and leads to a Gaussian mixture model for the stationary distribution of cellular metabolites. Computing distributions from this approximate model is substantially faster than through stochastic simulations, as these can be extremely slow due to the multiple timescales of metabolic pathways. Our technique can therefore be employed to efficiently explore the parameter space and predict the shape of metabolite distributions in different conditions. In earlier work we showed that the product of a single metabolic reaction can be accurately described by a Poisson mixture model[6]. Such approximation allowed the discovery of previously unknown regimes for metabolite distributions, including heavily tailed distributions and various types of bimodality and multimodality. The Poisson approximation, however, is bespoke to single reactions and not valid for more complex systems. In contrast, the Gaussian mixture model discussed here is more general and can be applied to multiple kinetic mechanisms, more complex stoichiometries, as well as post-translational regulation.

Another advantage of our approach is that the mixture weights can be computed offline from stochastic models for gene expression or single-cell expression data. The model is flexible in that it can readily accommodate gene expression models of various complexity. For the sake of illustration, in our examples we used the simple two-stage model for gene expression, but other models including gene regulation can also be employed[7]. Particularly relevant models are those that account for enzyme co-regulation, a widespread feature of bacterial operons[53], which translates into correlations between expression of different pathway enzymes and the resulting metabolite abundances.

In principle, most metabolic reactions satisfy the timescale separation as a result of their kinetics being much faster than the rate at which cells can synthesise new enzymes. However, throughout our examples we assumed that the metabolic substrate S, which is typically a carbon source or other precursor, remains constant. This case represents an abundant nutrient source with little fluctuations, but it is not adequate when substrates are lowly abundant or subject to stochastic fluctuations dictated by the environment. For example, depending on the timescale of such environmental fluctuations, the substrate can become another source of variability apart from enzyme expression[7]. In such cases, the timescale separation argument may not hold anymore and our theory needs to be revised to account for substrate fluctuations.

A number of works have sought to find links between fluctuations across layers of cellular organisation, such as gene expression, metabolism and cell growth[32,33,35,54,55]. But since measurement of metabolites in single cells remains technically challenging, there is pressing need for computational methods to predict fluctuations in cellular metabolites. Our proposed method provides a systematic approach for such task, paving the way for the generation of hypotheses on the molecular sources of metabolic heterogeneity.

VI. METHODS

A. Model simulation

Stochastic simulations were computed with Gillespie's algorithm over long simulation times (several hours) corresponding to thousands of cell cycles. The ODE models and

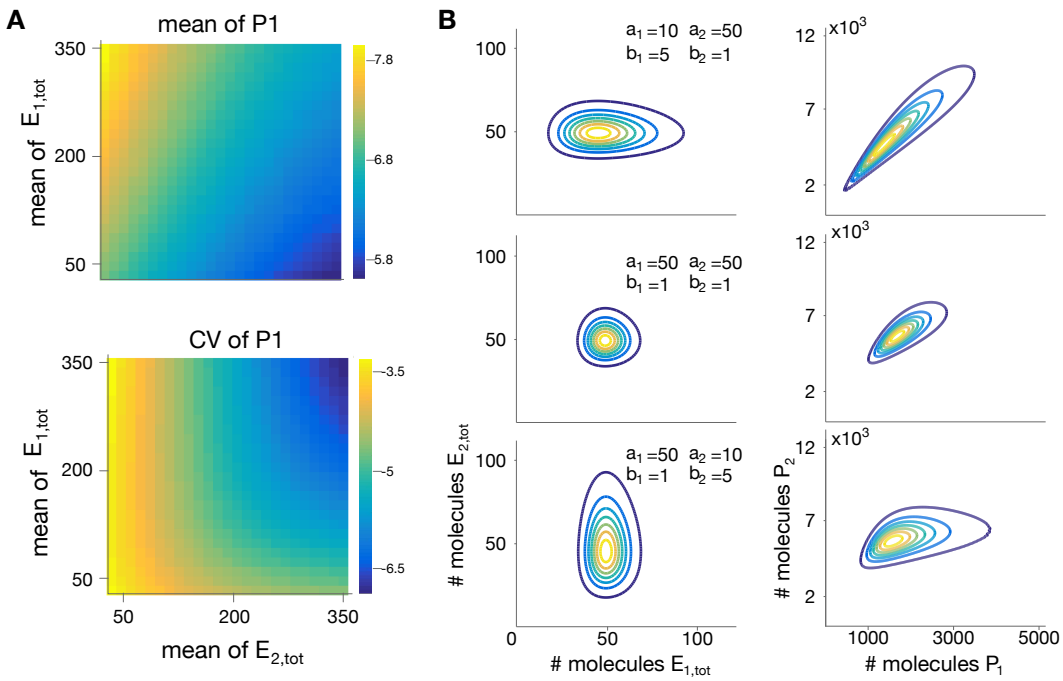

FIG. 5. **Impact of enzyme expression on metabolite distributions.** **(A)** For a wide sweep of the total enzyme expression levels, we observe changes in the mean abundances and coefficient of variation (CV) of the intermediate metabolite P_1. Shown are the logarithmic values of mean and coefficient of variation. **(B)** Enzyme bursting parameters can strongly shape the metabolite distribution. All parameter values can be found in Table III in the Methods.

Lyapunov equations were solved in Matlab. In all examples, the negative binomial distribution for gene expression in (12) was computed with its continuum approximation (Gamma distribution).

B. Deterministic rate equations

a. Reversible Michaelis Menten. The full set of rate equations for the reversible reaction in (6)–(9) is:

$$\begin{aligned}\frac{\mathrm{d}P}{\mathrm{d}t} &= k_{\mathrm{cat}}C - k_{\mathrm{rev}}EP - k_{\mathrm{c}}P\\ \frac{\mathrm{d}E}{\mathrm{d}t} &= -k_{\mathrm{f}}SE + k_{\mathrm{b}}C + k_{\mathrm{cat}}C - k_{\mathrm{rev}}EP,\\ \frac{\mathrm{d}C}{\mathrm{d}t} &= k_{\mathrm{f}}SE - k_{\mathrm{b}}C - k_{\mathrm{cat}}C + k_{\mathrm{rev}}EP.\end{aligned} \tag{35}$$

To further reduce the above system of ODEs to Eq. (13) in the main text, we can substitute the conservation relation in Eq. (10), i.e. $C = E_{\mathrm{total}} - E$, and use the fact that the substrate-enzyme complex (C) typically equilibrates much faster than the product P, which means that $\mathrm{d}C/\mathrm{d}t \approx 0$ in the timescale of catalysis.

b. End-product inhibition. The full set of rate equations for the reactions in (18)–(23) is:

$$\begin{aligned}\frac{\mathrm{d}P_1}{\mathrm{d}t} &= k_{\mathrm{cat},1}C_1 - k_{\mathrm{rev},1}E_1P_1 - k_{\mathrm{f},2}E_2P_1 + k_{\mathrm{b},2}C_2 - k_{\mathrm{c},1}P_1\\ \frac{\mathrm{d}P_2}{\mathrm{d}t} &= k_{\mathrm{cat},2}C_2 - k_{\mathrm{rev},2}E_2P_2 - k_{\mathrm{sq}}E_1P_2^h + k_{\mathrm{rsq}}E^* - k_{\mathrm{c},2}P_2.\\ \frac{\mathrm{d}E_1}{\mathrm{d}t} &= -k_{\mathrm{f},1}SE_1 + (k_{\mathrm{b},1} + k_{\mathrm{cat},1})\,C_1 - k_{\mathrm{rev},1}P_1E_1\\ &\quad - k_{\mathrm{sq}}P_2^hE_1 + k_{\mathrm{rsq}}E^*,\\ \frac{\mathrm{d}C_1}{\mathrm{d}t} &= k_{\mathrm{f},1}SE_1 - (k_{\mathrm{b},1} + k_{\mathrm{cat},1})\,C_1 + k_{\mathrm{rev},1}P_1E_1,\\ \frac{\mathrm{d}E^*}{\mathrm{d}t} &= k_{\mathrm{sq}}P_2^hE_1 - k_{\mathrm{rsq}}E^*,\\ \frac{\mathrm{d}E_2}{\mathrm{d}t} &= -k_{\mathrm{f},2}P_1E_2 + (k_{\mathrm{b},2} + k_{\mathrm{cat},2})\,C_1 - k_{\mathrm{rev},2}P_2E_2,\\ \frac{\mathrm{d}C_2}{\mathrm{d}t} &= k_{\mathrm{f},2}P_1E_2 - (k_{\mathrm{b},2} + k_{\mathrm{cat},2})\,C_1 + k_{\mathrm{rev},2}P_2E_2\end{aligned} \tag{36}$$

As in the previous example, we can use the rapid equilibrium assumption and the conservation relations in (26), i.e. $E_{\mathrm{t},1} = E_1 + E^* + C_1$ and $E_{\mathrm{t},2} = E_2 + C_2$, to simplify the 7-dimensional ODE in (36) to the 2-dimensional system in (28) of the main text.

REFERENCES

[1] I Golding, J Paulsson, SM Zawilski, and EC Cox. Real-Time Kinetics of Gene Activity in Individual Bacteria. *Cell*, 123(6):1025–1036, dec 2005.

[2] Y Taniguchi, PJ Choi, G-W Li, H Chen, M Babu, J Hearn, A Emili, and XS Xie. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, 329(5991):533–538, 2010.

[3] PS Swain, MB Elowitz, and ED Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20):12795–12800, 2002.

[4] A Raj and A van Oudenaarden. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell*, 135(2):216–226, 2008.

[5] P Thomas, N Popović, and R Grima. Phenotypic switching in gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America*, 111(19):6994–6999, 2014.

[6] MK Tonn, P Thomas, M Barahona, and DA Oyarzún. Stochastic modelling reveals mechanisms of metabolic heterogeneity. *Communications Biology*, 2(1):108, 2019.

[7] J Dattani and M Barahona. Stochastic models of gene transcription with upstream drives: Exact solution and sample path characterization. *Journal of the Royal Society Interface*, 14(126), 2017.

[8] M Heinemann and R Zenobi. Single cell metabolomics. *Current Opinion in Biotechnology*, 22(1):26–31, 2011.

[9] T Esaki and T Masujima. Fluorescence Probing Live Single-cell Mass Spectrometry for Direct Analysis of Organelle Metabolism. *Analytical Science*, 31(12), 2015.

[10] BD Bennett, EH Kimball, M Gao, R Osterhout, SJ Van Dien, and J D Rabinowitz. Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nature Chemical Biology*, 5(8):593–599, 2009.

[11] Y Xiao, CH Bowen, D Liu, and F Zhang. Exploiting non-genetic, cell-to-cell variation for enhanced biosynthesis. *Nature Chemical Biology*, 12(5):339–344, 2016.

[12] AA Mannan, D Liu, F Zhang, and DA Oyarzún. Fundamental Design Principles for Transcription-Factor-Based Metabolite Biosensors. *ACS synthetic biology*, 6:1851–1859, 2017.

[13] EA Lemke and C Schultz. Principles for designing fluorescent sensors and reporters. *Nature Chemical Biology*, 7(8):480–483, 2011.

[14] H Yaginuma, S Kawai, KV Tabata, K Tomiyama, A Kakizuka, T Komatsuzaki, H Noji, and H Imamura. Diversity in ATP concentrations in a single bacterial cell population revealed by. *Scientific Reports*, 4:6522, 2014.

[15] H Imamura, KPH Nhat, H Togawa, K Saito, R Iino, Y Kato-Yamada, T Nagai, and H Noji. Visualization of ATP levels inside single living cells with fluorescence resonance energy transfer-based genetically encoded indicators. *PNAS*, 106(37):15651–15656, 2009.

[16] JS Paige, T Nguyen-Duc, W Song, and SR Jaffrey. Fluorescence Imaging of Cellular Metabolites with RNA. *Science*, 335(6073):1194, 2012.

[17] AJ Ibáñez, SR Fagerer, AM Schmidt, PL Urban, K Jefimovs, P Geiger, R Dechant, M Heinemann, and R Zenobi. Mass spectrometry-based metabolomics of single yeast cells. *Proceedings of the National Academy of Sciences*, 110(22):8790–8794, may 2013.

[18] AY Weiße, DA Oyarzún, V Danos, and PS Swain. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proceedings of the National Academy of Sciences*, 112(9):E1038–E1047, 2015.

[19] M Lempp, N Farke, M Kuntz, SA Freibert, R Lill, and H Link. Systematic identification of metabolites controlling gene expression in E. coli. *Nature Communications*, 10(1):4463, 2019.

[20] RM Loftus and DK Finlay. Immunometabolism: Cellular Metabolism Turns Immune Regulator. *Journal of Biological Chemistry*, 291(1):1–10, Jan 2016.

[21] MA Reid, Z Dai, and JW Locasale. The impact of cellular metabolism on chromatin dynamics and epigenetics. *Nature Cell Biology*, 19(11):1298–1306, nov 2017.

[22] Y Shan, AB Gandt, SE Rowe, JP Deisinger, BP Conlon, and K Lewis. ATP-Dependent Persister Formation in Escherichia coli. *mBIO*, 8(1):1–14, 2017.

[23] JL Radzikowski, H Schramke, and M Heinemann. Bacterial persistence from a system-level perspective. *Current Opinion in Biotechnology*, 46:98–105, 2017.

[24] JB Deris, M Kim, Z Zhang, H Okano, R Hermsen, A Groisman, and T Hwa. The Innate Growth Bistability and Fitness Landscapes of Antibiotic Resistant Bacteria. *Science*, 342, 2013.

[25] C Vilhena, E Kaganovitch, JY Shin, A Grünberger, S Behr, I Kristoficova, S Brameyer, D Kohlheyer, and K Jung. A Single-Cell View of the BtsSR/YpdAB Pyruvate Sensing Network in Escherichia coli and Its Biological Relevance. *Journal of Bacteriology*, 200(1):1–13, 2018.

[26] AC Schmitz, CJ Hartline, and F Zhang. Engineering Microbial Metabolite Dynamics and Heterogeneity. *Biotechnology Journal*, 12(10), 2017.

[27] D Binder, T Drepper, K-E Jaeger, F Delvigne, Wolfgang Wiechert, Dietrich Kohlheyer, and Alexander Grünberger. Homogenizing bacterial cell factories: Analysis and engineering of phenotypic heterogeneity. *Metabolic Engineering*, 42:145–156, jul 2017.

[28] D Liu, AA Mannan, Y Han, DA Oyarzún, and F Zhang. Dynamic metabolic control: towards precision engineering of metabolism. *Journal of Industrial Microbiology & Biotechnology*, 45(7):535–543, 2018.

[29] A Amantonico, PL Urban, and R Zenobi. Analytical techniques for single-cell metabolomics: state of the art and trends. 398(6):2493–2504, 2010.

[30] V Takhaveev and M Heinemann. Metabolic heterogeneity in clonal microbial populations. *Current Opinion in Microbiology*, 45:30–38, 2018.

[31] M Wehrens, F Büke, P Nghe, and SJ Tans. Stochasticity in cellular metabolism and growth: Approaches and consequences. *Current Opinion in Systems Biology*, 8:131–136, 2018.

[32] O Kotte, B Volkmer, JL Radzikowski, and M Heinemann. Phenotypic bistability in Escherichia coli' s central carbon metabolism. *Molecular Systems Biology*, 10(1):736, 2014.

[33] JH van Heerden, MT Wortel, FJ Bruggeman, JJ Heijnen, YJM Bollen, R Planqué, J Hulshof, TGO Toole, SA Wahl, and B Teusink. Lost in Transition: Start-Up of glycolysis yields subpopulations of nongrowing cells. *Science*, 343(6174):1245114, 2014.

[34] F Schreiber, S Littmann, G Lavik, S Escrig, A Meibom, MMM Kuypers, and M Ackermann. Phenotypic heterogeneity driven by nutrient limitation promotes growth in fluctuating environments. *Nature Microbiology*, (May):1–7, 2016.

[35] DJ Kiviet, P Nghe, N Walker, S Boulineau, V Sunderlikova, and SJ Tans. Stochasticity of metabolism and growth at the single-cell level. *Nature*, 514(7522):376–379, 2014.

[36] E Şimşek and M Kim. The emergence of metabolic heterogeneity and diverse growth responses in isogenic bacterial cells. *The ISME Journal*, 12(5):1199–1209, may 2018.

[37] DT Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems Approximate accelerated stochastic simulation of chemically reacting systems. 1716(2001):1716–1733, 2007.

[38] MK Tonn. *Stochastic modelling and analysis of metabolic heterogeneity in single cells*. PhD thesis, Imperial College London, 2020.

[39] J Puchałka and AM Kierzek. Bridging the gap between stochastic and deterministic regimes in the kinetic simulations of the biochemical reaction networks. *Biophysical journal*, 86(3):1357–72, 2004.

[40] Y Cao, DT Gillespie, and LR Petzold. Accelerated stochastic simulation of the stiff enzyme-substrate reaction. *The Journal of chemical physics*, 123(14):144917, 2005.

[41] P Labhsetwar, JA Cole, E Roberts, ND Price, and ZA Luthey-Schulten. Heterogeneity in protein expression induces metabolic variability in a modeled Escherichia coli population. *PNAS*, 110(34):14006–14011, 2013.

[42] J-B Lugagne, DA Oyarzún, and G-B Stan. Stochastic simulation of enzymatic reactions under transcriptional feedback regulation. In *Proceeding of the European Control Conference*, pages 3646–3651, Zurich, 2013.

[43] E Murabito, M Verma, M Bekker, D Bellomo, HV Westerhoff, B Teusink, and R Steuer. Monte-Carlo Modeling of the Central Carbon Metabolism of Lactococcus lactis: Insights into Metabolic Regulation. *PLoS ONE*, 9(9):e106453, sep 2014.

[44] E Levine and T Hwa. Stochastic fluctuations in metabolic pathways. *Proceedings of the National Academy of Sciences of the United States of America*, 104(22):9224–9229, 2007.

[45] A Gupta, A Milias-Argeitis, and M Khammash. Dynamic disorder in simple enzymatic reactions induces stochastic amplification of substrate. *Journal of the Royal Society*, 14(132):1–29, 2017.

[46] DA Oyarzún, J-B Lugagne, and G-BV Stan. Noise Propagation in Synthetic Gene Circuits for Metabolic Control. *ACS synthetic biology*, 4(2):116–125, 2015.

[47] J Kuntz, DA Oyarzún, and G-BV Stan. Model Reduction of Genetic-Metabolic Networks via Time Scale Separation. In *A Systems Theoretic Approach to Systems and Synthetic Biology*, pages 181–210. Springer Netherlands, 2014.

[48] A Cornish-Bowden. *Fundamentals of Enzyme Kinetics*. Portland Press, third edition, 2004.

[49] NG van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, Amsterdam, 1992.

[50] J Elf and M Ehrenberg. Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome research*, 13(11):2475–84, nov 2003.

[51] M Thattai and A van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America*, 98(15):8614–9, jul 2001.

[52] V Shahrezaei and PS Swain. Analytical distributions for stochastic gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 105(45):17256–17261, 2008.

[53] V Chubukov, L Gerosa, K Kochanowski, and U Sauer. Coordination of microbial metabolism. *Nature reviews. Microbiology*, 12(5):327–40, may 2014.

[54] N Nikolic, F Schreiber, AD Co, DJ Kiviet, T Bergmiller, S Littmann, MMM Kuypers, and M Ackermann. Cell-to-cell variation and specialization in sugar metabolism in clonal bacterial populations. *PLOS Genetics*, 13(12), 2017.

[55] P Thomas, G Terradot, V Danos, and AY Weiße. Sources, propagation and consequences of stochasticity in cellular growth. *Nature Communications*, 9(4528):1–11, 2018.

Figure 3			
δ	$0.00025\ \mathrm{s}^{-1}$	k_{b}	$1000\ \mathrm{s}^{-1}$
a	$\{25, 50, 120\}$	k_{cat}	$3.6\ \mathrm{s}^{-1}$
b	1	k_{rev}	$0.01\ \mathrm{s}^{-1}$
S	3000 molecules	k_{c}	$0.02\ \mathrm{s}^{-1}$
k_{f}	$1 \times S\ \mathrm{s}^{-1}$		

TABLE I. Parameter values for simulations in Figure 3.

Figure 4			
δ	$0.00025\ \mathrm{s}^{-1}$	$k_{\mathrm{rev},1}$	$0.0001\ \mathrm{s}^{-1}$
$k_{\mathrm{deg},1}$	$0.2\ \mathrm{s}^{-1}$	$k_{\mathrm{c},1}$	$0.00025\ \mathrm{s}^{-1}$
$k_{\mathrm{deg},2}$	$0.2\ \mathrm{s}^{-1}$	$k_{\mathrm{f},2}$	$1.5\ \mathrm{s}^{-1}$
S	3000 molecules	$k_{\mathrm{b},2}$	$15000\ \mathrm{s}^{-1}$
$k_{\mathrm{f},1}$	$20 \times S\ \mathrm{s}^{-1}$	$k_{\mathrm{cat},2}$	$150\ \mathrm{s}^{-1}$
$k_{\mathrm{b},1}$	$15000\ \mathrm{s}^{-1}$	$k_{\mathrm{rev},2}$	$0.001\ \mathrm{s}^{-1}$
$k_{\mathrm{cat},1}$	$22.5\ \mathrm{s}^{-1}$	$k_{\mathrm{c},2}$	$0.15\ \mathrm{s}^{-1}$

Figure 4A	
a_1	$\{35, 126, 210\}$
a_2	$\{35, 97, 97\}$
b_1	1
b_2	1
k_{sq}	$10^{-10}\ \mathrm{s}^{-1}$
k_{rsq}	$1\ \mathrm{s}^{-1}$
h	3

Figure 4B	
a_1	80
a_2	80
b_1	1
b_2	1
k_{sq}	$\{0, 10^{-10}, 10^{-12}\}\ \mathrm{s}^{-1}$
k_{rsq}	$1\ \mathrm{s}^{-1}$
h	3

TABLE II. Parameter values for simulations in Figure 4.

Figure 5A	
a_1	$[10, 100]$
a_2	$[10, 100]$
b_1	1
b_2	1
k_{sq}	$10^{-10}\ \mathrm{s}^{-1}$
k_{rsq}	$1\ \mathrm{s}^{-1}$
h	3

Figure 5B	
a_1	$\{10, 50, 50\}$
a_2	$\{50, 50, 10\}$
b_1	$\{5, 1, 1\}$
b_2	$\{1, 1, 5\}$
k_{sq}	$0\ \mathrm{s}^{-1}$
k_{rsq}	$1\ \mathrm{s}^{-1}$
h	3

TABLE III. Parameter values for simulations in Figure 5.