Emergence of non-genetic heterogeneity in metabolic reactions

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, EH9 3JH, Edinburgh, United Kingdom*
[3)] *School of Informatics, University of Edinburgh, EH8 9AB, Edinburgh, EH8 9AB, United Kingdom*

Heterogeneity is a hallmark of cell physiology and appears in most layers of the cellular machinery. Metabolic heterogeneity, in particular, underpins various single-cell phenomena such as drug tolerance and growth variability in microbes. Much research has focussed on the heterogeneity of the proteome, yet it remains unclear if such variation permeates to metabolism and shifts the metabolic state of a cell. Here we propose a stochastic model to show that complex forms of metabolite heterogeneity emerge from fluctuations in enzyme expression and catalysis. The analysis predicts clonal populations to split into two or more metabolically distinct subpopulations that can be accurately approximated by a Poisson Mixture Model. We reveal previously unknown mechanisms of heterogeneity not seen in classic deterministic models, in which enzymes with a unimodal expression distribution lead to metabolites with a bimodal or multimodal distribution across the population. On the basis of published enzyme expression data and kinetic parameters, our results suggests that metabolite heterogeneity may be much more pervasive than previously thought. Our work cast light on the link between gene expression and variability of metabolic phenotypes, providing a widely applicable theory to probe the sources of metabolic heterogeneity.

Keywords: metabolic heterogeneity; stochastic gene expression; mixture models; phenotypic variability; synthetic biology; systems biology

I. INTRODUCTION

Cellular heterogeneity is ubiquitous across all domains of life. In microbes, clonal populations display phenotypic variability as a result of multiple factors such as fluctuations in the microenvironment, stochasticity in gene expression, or asymmetric partitioning at cell division[1–3]. Variability is well recognised at the transcriptional and translational levels. Yet various single-cell phenomena result from the emergence of distinct metabolic states within a clonal population. For example, metabolic heterogeneity plays a key role in antibiotic tolerance[4–6], heterogeneous nutrient uptake[7,8], and variations in growth rate[9]. It has also been shown that nutrient shifts can cause populations to split into two[10,11] or more[12] subpopulations with distinct growth abilities. The emergence of subpopulations has been theorised as a bet-hedging strategy that gives an evolutionary advantage for survival in adverse environments[4,13].

A central challenge to quantify metabolic variability is the lack of techniques for measuring metabolites with single-cell resolution[14]. In contrast to single-cell measurements of protein expression, for which powerful reporter systems have been developed[15,16], quantification of metabolites in single-cells remains a significant challenge. Some of the techniques employed so far include FRET sensors[17], metabolite-responsive transcription factors[18,19], RNA-based sensors[20], and mass-spectrometry[21], yet most of these technologies are in early stages of development[14]. As a result, metabolic heterogeneity is typically quantified indirectly via measurements of metabolic enzymes or growth rate in single-cells[9,11,22].

Our objective in this paper is to characterise heterogeneity in metabolites as a result of stochastic enzyme expression and catalysis. Metabolic models traditionally assume that enzymatic reactions behave deterministically on the basis that both enzymes and metabolites appear in high molecule numbers[23]. However, single-cell proteomics in *Escherichia coli* show that metabolic enzymes are as variable as any other member of the proteome[16], while metabolomics data suggest that average metabolite abundances span several orders of magnitude[24]. The few datasets on single-cell metabolite abundance already suggest a significant variability in some metabolites in *E. coli*[18,25]. Such evidence casts doubt on the traditional assumption of metabolism being a purely deterministic process, suggesting a link between fluctuations in enzyme expression and metabolites.

The role of stochastic gene expression in protein variability has been well studied[2,3,26–28], but the impact of such randomness on metabolic reactions remains much less understood. Various theoretical studies have analysed the impact of fluctuations in the supply and consumption of metabolites[29–31], or the propagation of enzyme noise to a metabolite[32]. However, despite the advancing experimental evidence of stochasticity in metabolism, mathematical models still lack the sufficient detail to integrate the processes that are known to shape protein heterogeneity, such as stochastic promoter switching and transcriptional bursting.

In this paper we propose a model for metabolite heterogeneity in single-cells. The model integrates stochasticity in enzyme catalysis[23] and expression[26], two well-established processes that so far have been studied in isolation. Our approach includes a stochastic formulation of various relevant mechanisms in enzymatic reactions, including reversible catalysis, stochastic switching of promoter activity, fluctuations in mRNA transcripts, and consumption of the enzymatic product by downstream processes.

We probe the model for various sources of stochasticity using simulations and analytical solutions for the stationary distribution of the metabolite. The analysis reveals intricate

[a)] Corresponding author: d.oyarzun@ed.ac.uk

patterns of heterogeneity that translate into bimodal and multimodal distributions for the number of metabolite molecules. These phenomena arise from the interplay between a lowly abundant enzyme and its catalytic parameters. We show that the metabolite distributions can be accurately approximated by a Poisson Mixture Model across large regions of the parameter space. The mixture model can be readily adapted to a wide class of gene expression models and provides a quantitative tool to predict metabolite variability from enzyme measurements in single-cells.

II. RESULTS

A. Stochastic model of an enzymatic reaction

We consider a model that combines enzyme expression and enzyme kinetics into a single stochastic description (Figure 1A). The model includes an enzymatic reaction with standard Michaelis-Menten kinetics, in which substrate and enzyme bind reversibly to form a complex that undergoes reversible catalysis into a metabolite. We assume that enzyme expression follows the well established three-stage model for gene expression[3,26], where a single copy gene switches stochastically between an inactive state (D_{off}) and active state (D_{on}). In the active state, mRNAs are transcribed and translated into protein. The model also includes consumption of the metabolite by downstream pathways, degradation of mRNA transcripts, and dilution by growth of all species. Since metabolic reactions operate far from thermodynamic equilibrium, we assume that the substrate pool remains constant so that the system reaches a non-zero flux, e.g. when the substrate is a highly abundant extracellular carbon source or a slowly varying intracellular metabolite. The model reactions are shown in equations (R1)–(R9) in the Methods section.

To investigate the emergence of metabolic heterogeneity, we need to compute the stationary probability distribution of metabolite molecules (n_{p}) for relevant combinations of model parameters. Figure 1B shows a typical simulation of the model obtained with Gillespie's algorithm[33]. A key challenge for such simulations, however, is the multiscale nature of enzymatic reactions: not only do metabolic reactions operate in a much faster timescale (milliseconds) than enzyme expression (tens of minutes)[29,34,35], but also the average number of enzymes is much lower than the number of metabolites. These multiple scales result in reaction propensities that differ by several orders of magnitude, thus leading to extremely slow simulations which make the exploration of the parameter space infeasible. An alternative is to use simulation algorithms that exploit the separation of scales to increase computational speed, such as tau-leaping or slow-scale approximations[36]. Yet in our case it is unclear how such numerical approximations impact the predictions drawn from the simulations.

To determine the impact of genetic and catalytic parameters on metabolic heterogeneity, we obtained an analytic approximation for the distribution of metabolite molecules that can be evaluated efficiently without expensive stochastic simulations. Our solution allows the exploration of parameter space to characterise the different regimes promoting metabolic heterogeneity. The approximation follows from exploiting time scale separation in the Chemical Master Equation of the stochastic process[37]. In physiological regimes the model has three timescales: a fast metabolic time scale, in which substrate and enzyme bind and unbind; an intermediate time scale associated with the catalysis of the metabolite (n_{p}); and a slow timescale associated with the expression of the enzyme and dilution by cell growth.

The total amount of enzyme (free and substrate-bound, denoted as n_{e} and n_{c} respectively) varies in the slowest timescale, and therefore the binding/unbinding of substrate and enzyme equilibrates quickly. As a result, in the timescale of gene expression, the metabolite can be assumed to depend directly on the the total enzyme $n_{\text{etot}} = n_{\text{e}} + n_{\text{c}}$ rather than on n_{e} and n_{c} individually. Under this approximation, we can use the law of total probability to write:

$$\mathbf{P}(n_{\text{p}}) = \sum_{n_{\text{etot}}=0}^{\infty} \underbrace{\mathbf{P}(n_{\text{etot}})}_{\text{gene expression}} \underbrace{\mathbf{P}(n_{\text{p}}|n_{\text{etot}})}_{\text{catalysis}}, \tag{1}$$

The formula in (1) decomposes the distribution of metabolite $\mathbf{P}(n_{\text{p}})$ into stochasticity originating from enzyme expression, $\mathbf{P}(n_{\text{etot}})$, and from fluctuations in the catalytic reaction itself, described by the conditional distribution of metabolite given the amount of total enzyme, $\mathbf{P}(n_{\text{p}}|n_{\text{etot}})$. In the timescale of metabolite fluctuations, the total enzyme can be assumed to be in a quasi-stationary state. Further, exploiting the fast binding/unbinding between substrate and enzyme, we showed that the metabolite follows a birth-death process with effective propensities (details in Methods section):

$$\begin{aligned} \mathrm{k}^{\text{eff}}_{\text{birth}} &= \mathrm{k}_{\text{cat}} \mathbb{E}(n_{\text{c}}|n_{\text{etot}}, n_{\text{p}}) \\ &\approx \mathrm{k}_{\text{cat}} \frac{k_{-1}}{(k_{-1} + k_1 n_{\text{s}})} n_{\text{etot}}, \\ \mathrm{k}^{\text{eff}}_{\text{death}} &= k_{\text{rev}} \mathbb{E}(n_{\text{e}}|n_{\text{etot}}, n_{\text{p}}) + k_{\text{c}} \\ &\approx k_{\text{rev}} \frac{k_1 n_{\text{s}}}{(k_{-1} + k_1 n_{\text{s}})} n_{\text{etot}} + k_{\text{c}}, \end{aligned} \tag{2}$$

where $\mathbb{E}(n_{\text{e}}|n_{\text{etot}}, n_{\text{p}})$ and $\mathbb{E}(n_{\text{c}}|n_{\text{etot}}, n_{\text{p}})$ are the conditional expectations of the free enzyme (n_{e}) and complex (n_{c}) given the total enzyme and metabolite. In equation (2), n_s is the constant number of substrate molecules, the parameters k_1, k_{-1}, k_{cat}, and k_{rev} are the rate constants of the Michaelis-Menten mechanism (defined in Figure 1A), and k_{c} is an effective first-order rate constant of metabolite consumption by downstream pathways. The conditional distribution needed in the model (1) can then be computed explicitly:

$$\mathbf{P}(n_{\text{p}}|n_{\text{etot}}) \sim \text{Poisson}\left(n_{\text{p}}; \lambda(n_{\text{etot}})\right), \tag{3}$$

with Poisson parameter

$$\lambda(n_{\text{etot}}) = \frac{\lambda_{\infty}}{1 + K/n_{\text{etot}}}, \tag{4}$$

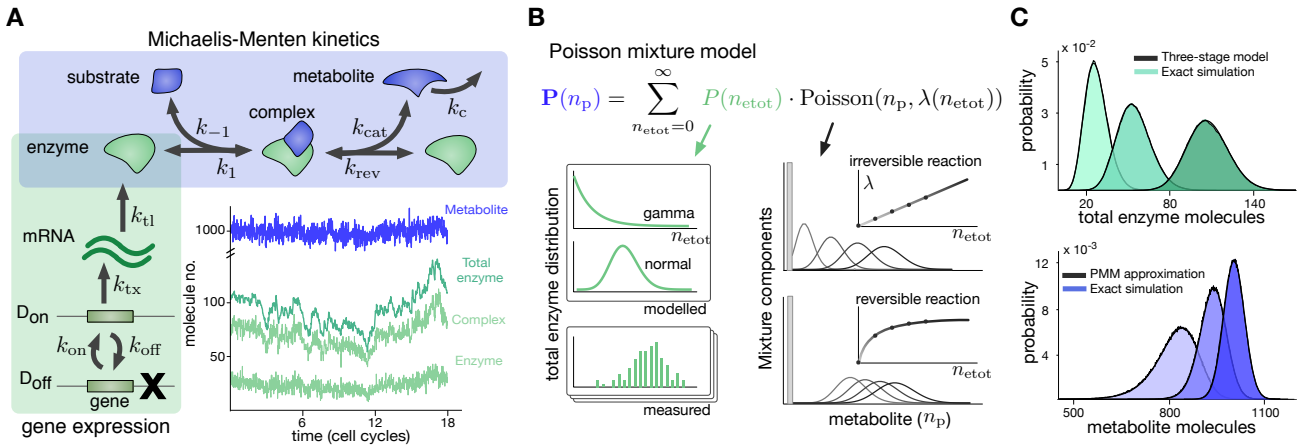

FIG. 1. **Stochastic model for an enzymatic reaction.** **(A)** The model integrates reversible Michaelis-Menten kinetics with the three-stage model for gene expression[23,26]. The model includes consumption of the metabolite by downstream pathways, degradation of mRNA transcripts, and dilution of all chemical species by cell growth (not shown in diagram); rate constants are shown in the figure and model reactions are shown in equations (R1)–(R9) of the Methods. The inset shows a typical simulation for a realistic parameter set shown in Table I. **(B)** Construction of the Poisson Mixture Model (PMM) for the number of metabolite molecules (n_p). This approximation is valid under a separation of timescales between enzyme expression and enzyme catalysis. The mixture model, shown in equation (1), comprises Poisson distributions weighted by the distribution of enzyme expression $\mathbf{P}(n_\mathrm{etot})$. The Poisson parameter $\lambda(n_\mathrm{etot})$ depends on enzyme kinetics via the nonlinear relation in equation (4). In the irreversible case ($k_\mathrm{rev} = 0$), the $\lambda(n_\mathrm{etot})$ parameter scales linearly and produces equi-spaced Poisson modes. The mode $\mathrm{Poisson}(n_\mathrm{p}, 0)$ is highlighted as a bar. **(C)** The PMM provides an accurate approximation of the stationary distributions. Insets show distributions for enzyme and metabolite, computed via Gillespie simulations and the PMM approximation for fixed $\lambda_\infty = 1080$ molecules, $K = 8$ molecules, and three different promoter switching parameters, shown in Table I.

and (λ_∞, K) are two effective kinetic parameters

$$\lambda_\infty = n_\mathrm{s}\frac{k_\mathrm{cat}k_1}{k_\mathrm{rev}k_{-1}}, \text{ and } K = k_\mathrm{c}\frac{k_1 n_\mathrm{s} + k_{-1}}{k_\mathrm{rev}k_{-1}}. \tag{5}$$

The parameters λ_∞ and K are in units of molecules/cell and depend on the interplay between substrate abundance, enzyme kinetics, and downstream processes.

As illustrated in Figure 1B, the distribution in (1) is a Poisson Mixture Model[38–40] (PMM) that convolves the enzyme distribution $\mathbf{P}(n_\mathrm{etot})$ with various Poisson modes $\mathbf{P}(n_\mathrm{p}|n_\mathrm{etot})$ arising from the catalytic activity. In our model, the analytical distribution of the total enzyme abundance follows the standard solution of the three-stage model for gene expression[26], which can be computed explicitly in terms of model parameters. In certain limits, the three-stage model produces approximately Gamma or Normal distributions depending on the the mean expression level and the half lives of mRNAs and proteins[3,26].

The decomposition in equation (1) shows that the PMM is not limited to the model for gene expression we have considered here. Other models may be used, either by using closed-form expressions for $\mathbf{P}(n_\mathrm{etot})$, or by inferring the enzyme distribution directly from single-cell protein expression data such as flow cytometry or single-cell microscopy[1,16]. The PMM thus provides a versatile tool to predict metabolite heterogeneity from modelled or measured enzyme heterogeneity viewed as an upstream source of variation[40]. Figure 1C shows that the PMM distribution provides a good approximation to Gillespie simulations computed with typical parameter values.

B. Qualitative features of the Poisson Mixture Model

At the heart of the metabolite PMM is the interplay between variability from gene expression and that originating from enzyme kinetics. Specifically, the Poisson parameter $\lambda(n_\mathrm{etot})$ in equation (4) controls the density and dispersion of the Poisson modes, which in turn shape the overall pattern of variability. As shown in Figure 1B, there are several cases of interest. For example, for irreversible reactions ($k_\mathrm{rev} = 0$), the Poisson parameter simplifies to

$$\lambda(n_\mathrm{etot}) = \frac{k_1 n_\mathrm{s} k_\mathrm{cat}}{(k_1 n_\mathrm{s} + k_{-1})k_\mathrm{c}} n_\mathrm{etot}, \tag{6}$$

which scales linearly with the enzyme abundance and thus the Poisson modes have equidistant means. In reversible reactions, on the other hand, the Poisson parameter saturates and causes the Poisson modes to concentrate around λ_∞. This effect is stronger for strong reversibility (high k_rev), in which case the kinetic parameter K is small. Note also that in either case, as the enzyme number n_etot grows, the Poisson modes spread out since $\lambda(n_\mathrm{etot})$ controls both their mean and variance.

From the construction of the PMM in (1), we observe that the enzyme distribution weighs the various Poisson modes, potentially producing metabolite distributions that are unimodal, bimodal or even multimodal. For example, for highly expressed reversible enzymes, the distribution $\mathbf{P}(n_\mathrm{etot})$ is significant only for large n_etot. Hence most Poisson modes do not contribute to the final metabolite distribution, except the mode centred at λ_∞, which leads to a unimodal metabolite

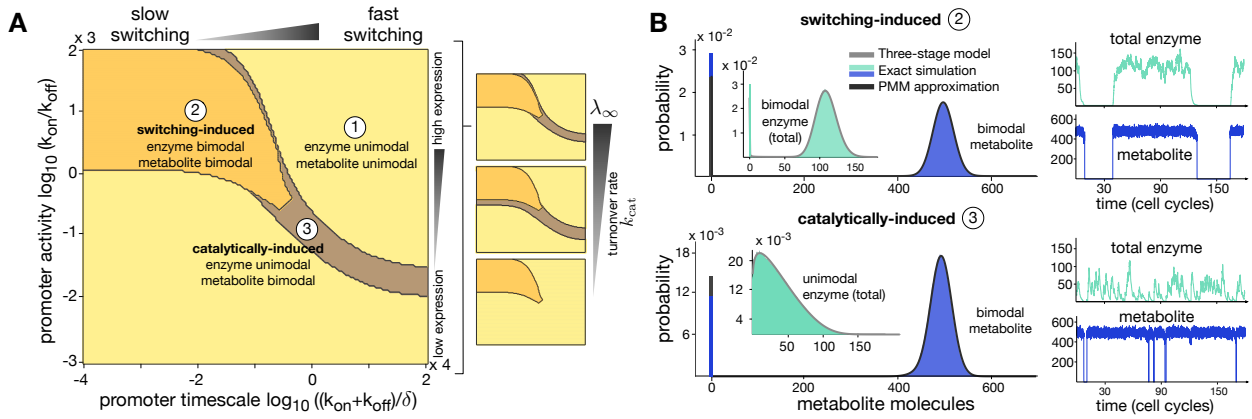

FIG. 2. **Mechanisms for metabolite bimodality. (A)** We evaluated the Poisson Mixture Model across a broad range of promoter switching timescale and promoter activity. Unimodal distributions for enzyme and metabolite (similar to those shown in Figure 1C) cover a large fraction of the parameter space. We identified two regimes in which metabolites are bimodal: in the switching-induced regime, bimodality propagate from the enzyme to metabolite. In the catalytically-induced regime, bimodality originates from a lowly abundant enzyme and the strong separation of timescales between expression and catalysis. The small panels show model predictions for a fixed kinetic parameter $K = 0.1333$ molecules, and increasing $\lambda_\infty = \{300, 3000, 30000\}$ molecules, obtained by increasing the turnover rate constant k_{cat}. **(B)** Exact simulations for two parameter sets verify the predictions drawn from the PMM approximation. We simulated over a long time horizon to obtain accurate estimates for stationary distributions; insets shown only a small portion of the time courses. The parameter values for the promoter switching rates are indicated in panel A and we fixed $\lambda_\infty = 500$ molecules. Both types of bimodality can be clearly distinguished in the time courses, but we note that they lead to almost identical distributions for the metabolite. In both cases, the PMM provides an accurate approximation for the stationary distributions.

distribution with a mean close to the deterministic average.

Conversely, for lowly expressed enzymes, there is a non-negligible probability of enzymes not being expressed, and thus the first term of the PMM, i.e. $\mathbf{P}(0)\text{Poisson}(n_{\text{p}}, 0)$, causes the metabolite distribution to peak at zero. However, the metabolite distribution may also display a second peak at λ_∞ if, for example, the $\lambda(n_{\text{etot}})$ parameter causes many Poisson modes to concentrate around λ_∞. This results in a bimodal metabolite distribution, whereby an isogenic population splits into metabolite producers and non-producers. Similar reasoning can be used to understand the emergence of multimodal metabolite distributions, which correspond to three or more subpopulations with varying metabolic activities. This qualitative analysis suggests that metabolic subpopulations can emerge even in cases where enzymes display unimodal distributions across the population. Crucially, this also indicates that metabolic subpopulations emerge through mechanisms that do not follow trivially from transcriptional heterogeneity alone, as we explore in more detail in next section.

C. Mechanisms for metabolic bimodality

First, we explored the impact of stochastic promoter switching on the emergence of metabolite bimodality. Figure 2A shows the summary of calculations when evaluating the PMM for variations in two key characteristics of the promoter across several orders of magnitude: the switching time scale ($(k_{\text{on}} + k_{\text{off}})/\delta$) and the promoter activity ($k_{\text{on}}/k_{\text{off}}$), for various values of the kinetic parameter λ_∞. We found three qualitatively distinct parameter regimes for the metabolite distribution that emerge from the combination of stochastic switching and catalysis: (1) a regime where both enzyme and metabolite have unimodal distributions, akin to the results shown earlier in Figure 1C; (2) a regime where both enzyme and metabolite have bimodal distributions; and (3) a regime in which the enzyme is unimodal but the metabolite is bimodal.

It can be shown that the deterministic version of our model in equations (R1)–(R9) has a single steady state. Hence regime (1) can be thought of as a stochastic correction consisting of unimodal distributions around a deterministic steady state. This is the expected behaviour under the traditional assumptions of high abundance of enzyme and metabolite molecules.

The other two regimes, however, correspond to alternative routes of noise-induced bimodality that cannot be explained using deterministic models[41–43]. Regime (2) is a highly stochastic regime dominated by the slow stochastic switching of the promoter, which drives and entrains the metabolic response. Hence we term it *switching-induced bimodality*. Slowly switching promoters are known to produce bimodal gene expression[28,40], and thus this regime corresponds to a case in which bimodality propagates from enzymes to metabolites. Figure 2A shows that this behaviour appears robustly for slow switching and high promoter activity across values of λ_∞ kinetic parameter.

Regime (3), the second route for metabolite bimodality, originates from a unimodal but weakly expressed enzyme (low $k_{\text{on}}/k_{\text{off}}$) expressed from fast switching promoters. In this case, the birth of a small number of enzyme molecules is sufficient to kick-start catalysis and make it rapidly set-

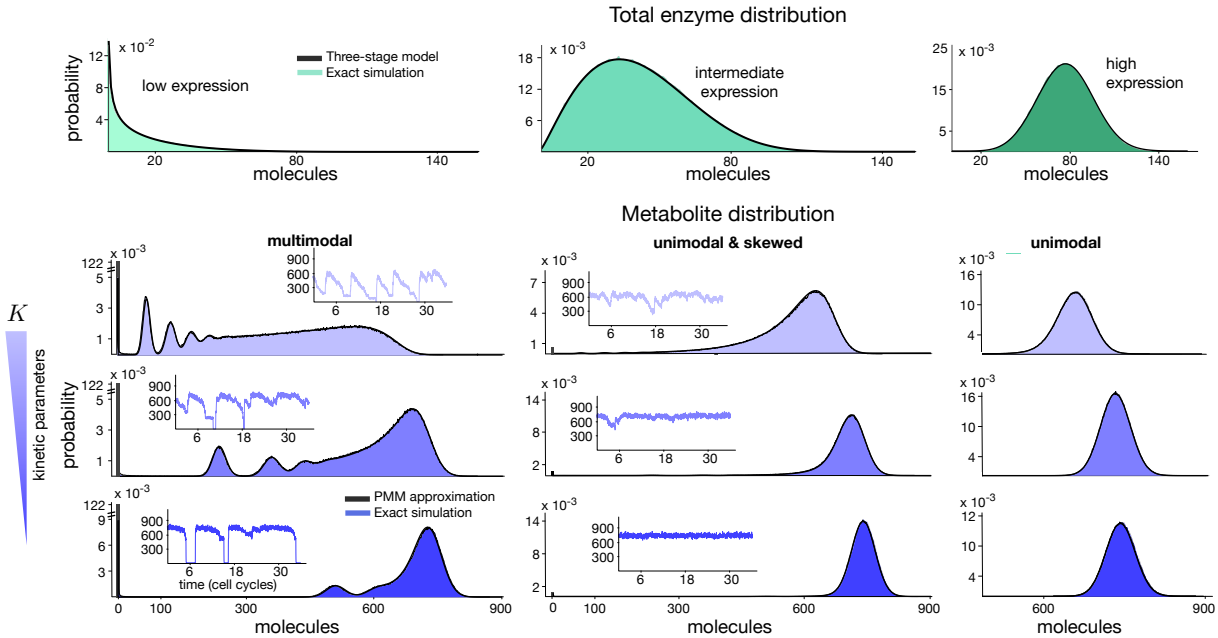

FIG. 3. **Emergence of metabolic multimodality.** We used the PMM approximation to find regimes for multimodality through perturbations to the enzyme kinetics. We vary the kinetic parameter K to control the dependency of the Poisson parameter $\lambda(n_{\text{etot}})$ in equation (4) on the total enzyme abundance. Parameter values are $\lambda_\infty = 750$ molecules and $K = \{10.0400,\ 2.1630,\ 0.4660\}$ molecules obtained by variations to the kinetic rate constants k_{cat} and k_{rev} with a constant ratio $k_{\text{cat}}/k_{\text{rev}}$. We shape the mean enzyme abundance with the promoter switching rates $k_{\text{on}} = \{1.56, 5.9, 20\} \times 10^{-4}\ \text{s}^{-1}$ and $k_{\text{off}} = \{9.8, 9.3, 8\} \times 10^{-4}\ \text{s}^{-1}$. From the PMM we found intricate patterns of multimodal distributions in the metabolite, all of which show an excellent match with the corresponding Gillespie simulations. The simulated time courses show metabolite numbers traversing various quasi-stationary regimes.

tle in a quasi stationary regime. This distinct phenomenon is a result of the separation of time scales between enzyme expression and catalysis, and we refer to it as *catalytically-induced bimodality*. From Figure 2A, we observe that this form of bimodality appears for a narrow range of promoter switching parameters corresponding to fast switching genes with medium to low promoter activity. This behaviour disappears altogether for a low λ_∞ parameter, for example in case of strong reversibility.

To validate the predictions of the PMM approximation, we ran full Gillespie simulations over a long time horizon for different parameter sets. Figure 2B shows the simulation time courses and resulting histograms. For switching-induced bimodality, we observe how slowly switching promoters causes a single cells to lack the enzyme over several cell cycles, a period during which the metabolite is not produced. In the case of catalytic-induced bimodality, however, fast switching combined with a low average expression level cause the metabolite abundance to drop for shorter but more frequent intervals. In both cases the PMM provides an excellent approximation to the bimodal histograms obtained from the stochastic simulations. Furthermore, we observe that the bimodal metabolite distributions both regimes are almost indistinguishable from each other, yet they are produced by enzymes with substantially different time courses and distributions. These regimes therefore correspond to distinct forms of bimodality, arising from fundamentally different mechanisms.

D. Emergence of metabolic multimodality

To explore the emergence of multimodality, we examined the analytical formula of the PMM in (1) to identify kinetic regimes associated with distinct enzyme distributions. A necessary condition for the emergence of multiple modes is that the Poisson components do not overlap and are sufficiently spaced from each other. From the definition of the $\lambda(n_{\text{etot}})$ parameter in (5), this happens when the kinetic parameter K is large. As discussed earlier, depending on the distribution of the enzyme, the Poisson modes may appear or cancel in the final metabolite distribution. We thus swept the parameter K and evaluated the PMM across various enzyme expression levels, including low expression with a skewed distribution and high expression with a normally distributed enzyme.

As shown in Figure 3, we found intricate patterns of multimodal distributions, depending on the interplay between the heterogeneity of the enzyme, $\mathbf{P}(n_{\text{etot}})$, and the enzyme kinetics encapsulated by the K parameter. Multimodality appears when the enzyme expression levels are low as compared to the parameter K. For instance, the values of K in Figure 3 are approximately 5-, 20-, and 100-fold those used in the bimodal examples in Figure 2. For enzymes expressed at intermediate levels, in the order of tens of molecules/cell on average, we found metabolite distributions that are unimodal but highly skewed. In the case of highly expressed enzymes, metabolites followed approximately normal distributions for a wide range of kinetic parameters.

The predictions are confirmed by Gillespie simulations of the full stochastic model, which display a striking match with the PMM approximation, even for complex multimodal distributions. The simulation time courses (shown in the insets of Figure 3) show that the multiple modes for weakly expressed enzymes correspond to cells remaining in a fixed metabolic state over the scale of the cell cycle but fluctuate across other states over longer time scales. For intermediate enzyme expression and large values for K, the metabolite does not settle in the quasi-stationary states and displays a long-tailed distribution. A decrease in K suppresses the tail of the distribution driving the PMM towards an approximately normal distribution. Altogether, these results indicate that the relation between enzyme expression and the kinetic parameters λ_∞ and, in particular, K are key determinants for the emergence of multimodality. This underscores the utility of the PMM to guide the prediction of qualitative and quantitative features of metabolite distributions for a wide range of parameter combinations.

III. DISCUSSION

Metabolic reactions are the powerhouse of living systems, fuelling the activity and dynamics of most cellular functions. Yet metabolism has been traditionally considered as a static process isolated from the rest of the cellular machinery. Currently, the accepted notion is that due to the large number of molecules involved, metabolism is a deterministic process at the cellular level, modulated by potentially random extrinsic factors[11,44]. Here we integrated enzyme kinetics and enzyme expression to propose a theoretical model for variability of metabolites in single cells. The model suggests that cell-to-cell metabolite variation can also arise as a result of intrinsic sources such as stochastic fluctuations in enzyme expression.

The majority of work on non-genetic heterogeneity has focused on stochastic gene expression and the resulting variability in protein levels[1,2]. This has produced a wealth of single-cell data and models to understand the variability in transcription and translation observed in clonal populations. Metabolite heterogeneity, however, remains poorly understood theoretically and has been observed only indirectly (e.g., through measurements of metabolic enzymes[11,22] or growth rate[9]) due to the lack of techniques to measure metabolite abundance in single cells.

Using the separation of time scales characteristic of metabolic reactions, we found that the stationary distribution of a metabolite follows a Poisson Mixture Model (PMM). The PMM can be efficiently evaluated across large domains of the parameter space and provides excellent approximations to the distributions computed from full stochastic simulations. Importantly, the model can be readily adapted to include different stochastic models for enzyme expression, beyond the three-stage model considered here[3,45], or even stochastic and time-dependent enzyme expression modelled as upstream drives[40]. The model can also be parameterised from experimentally measured distributions for enzyme levels in single-cells[16]. In combination with the enzyme kinetic parameters, the PMM could provide a powerful tool to predict metabolite variability from single-cell protein data obtained with established techniques such as flow cytometry or time-lapse microscopy.

We found complex patterns of metabolite heterogeneity depending on the interplay between the timescale of promoter activation/deactivation, the enzyme expression level, and the enzyme kinetics. The model predicts that bimodal and multimodal metabolite distributions can emerge in various parameter regimes. In such regimes single-cells spend several cell cycles in a constant metabolic state, but in timescales as long as tens of cell cycles, they switch stochastically across different states. Such long-term fluctuations in single cells result in highly heterogeneous populations containing several subgroups of metabolically distinct cells.

Bimodal metabolic phenotypes have been observed as a result of transcriptional regulation[13,22], post-translational control[10] and stochastic effects triggered by environmental shifts[11]. Our model reveals two distinct regimes in which metabolites display bimodality. One regime, which we call switching-induced bimodality, corresponds to the intuitive case in which a bimodal enzyme produces a bimodal metabolite. In agreement with previous studies on stochastic gene expression, this type of bimodality appears as a result of slow switching between promoter states[28,46]. In addition, we identified a fundamentally different mechanism of *catalytically-induced bimodality*, in which a unimodal enzyme produces a bimodal distribution of metabolite. This phenomenon results from a combination of slow fluctuations of a weakly expressed enzyme and the comparatively faster timescale of enzyme catalysis. Catalytic timescales are typically in the order of seconds or faster, so that slow fluctuations in enzyme expression levels produce two quasi-stationary metabolic states in single cells. At a population level, this leads to two distinct subpopulations of metabolite producers and non-producers.

As shown in Figure 4A, single-cell measurements in *E. coli* suggest that metabolic enzymes appear in low copy numbers across most cellular pathways[16]. In the specific growth conditions of that experiment, the data did not reveal bimodal expression of enzymes, which precludes the emergence of switching-induced bimodality in the metabolites they catalyse. However, as illustrated by the three representative distributions in Figure 4A, a number of enzymes have a low mean and a long-tailed distribution, akin to those required for catalytically-induced bimodality and multimodality. This suggests that enzyme distributions found in nature have the characteristics needed for the emergence of subpopulations with two or more distinct metabolite abundances.

Further requirements for metabolite bimodality and multimodality involve conditions on the parameters λ_∞ and K in equation (5). However, their computation requires rate constants (k_1, k_{-1} and k_{rev}) that are rarely measured separately, and instead enzymology data typically provides values for k_{cat} and $K_{\text{M}} = (k_{\text{cat}} + k_{-1})/k_1$ only[47]. In the Methods we show that the ratio λ_∞/K can be expressed as $\lambda_\infty/K = \epsilon \times k_{\text{cat}}/k_{\text{c}}$, where ϵ is the saturation level of the enzyme and k_{c} is the first-order rate constant of metabolite consumption. As illustrated in Figure 4B, the ratio λ_∞/K corresponds to a straight line in a $(\lambda_\infty,\ K)$-space, and a specific enzyme (i.e. with spe-

cific values for k_1, k_{-1} and k_{rev}) corresponds to a single point on the line. In Figure 4B we compare model predictions for a lowly abundant enzyme with different λ_∞/K ratios computed for k_{cat} constants measured in *E. coli*[48]. Considering the large spread in measured k_{cat} values, of up to seven orders of magnitude, plus the multiple combinations of metabolite consumption rates and enzyme saturation, the analysis suggests that catalytically-induced bimodality and multimodality are plausible within physiological regimes. Further validations of our predictions require measuring metabolite distributions directly, but this is still subject to a number of challenges in single-cell measurement technologies[14].

Our analysis shows that metabolite heterogeneity depends on a delicate interplay between enzyme expression and enzyme kinetics. It is reasonable to expect that energy-critical enzymes, such as those in central carbon metabolism, filter away fluctuations through post-translational regulatory mechanisms commonly found in metabolism. However, this may not be the case in pathways that are dynamically regulated in response to changes in the environment or cellular context. For example, transcriptional regulation in response to nutrient shifts may steer enzyme levels into regimes of low copy numbers where heterogeneity may dominate the resulting phenotypes. Such mechanism has been already shown to produce growth bimodality in the gluconeogenic switch of *E. coli*[11], while a similar mechanism could underpin the large variability observed in single-cell measurements of S-Adenosyl methionine[20]. Noise-induced phenomena also has implications for the design of dynamic control systems for heterologous pathways, which are focus of much research in synthetic biology and metabolic engineering[50].

In our efforts to build a theory that includes components shared by most enzymatic reactions, we have purposely overlooked a number of processes that can shape metabolic activity. For example, we have not addressed the impact of feedback mechanisms that control enzyme activity, including e.g. product inhibition and allostery, or transcriptional mechanisms that control enzyme expression in response to metabolites. Since post-translational regulation operates on timescales much shorter than enzyme expression, with similar timescale separation arguments it should be possible to express the metabolite distribution as a mixture model akin to ours. In such case the mixture components are not necessarily Poisson and their distribution will depend on the particular mechanism under study. We expect that bimodal and multimodal responses are likely to emerge in this setting, but the precise parameter conditions would have to be studied on a case-by-case basis. Transcriptional feedback control can also display various mechanisms depending on the particular pathway under study. One common motif relies on transcription factors (TF) that up- or down-regulate enzyme expression upon binding to a specific metabolite[19]. These mechanisms have been shown to play important roles on metabolic activity, but they also bring to the fore subtle questions that require detailed examination, for example, on the role of fluctuations coming from TF expression itself, or the impact of negative TF autoregulation[51]. Our study paves the way for these and other questions to be addressed and raises exciting prospects for the future research in metabolic heterogeneity.

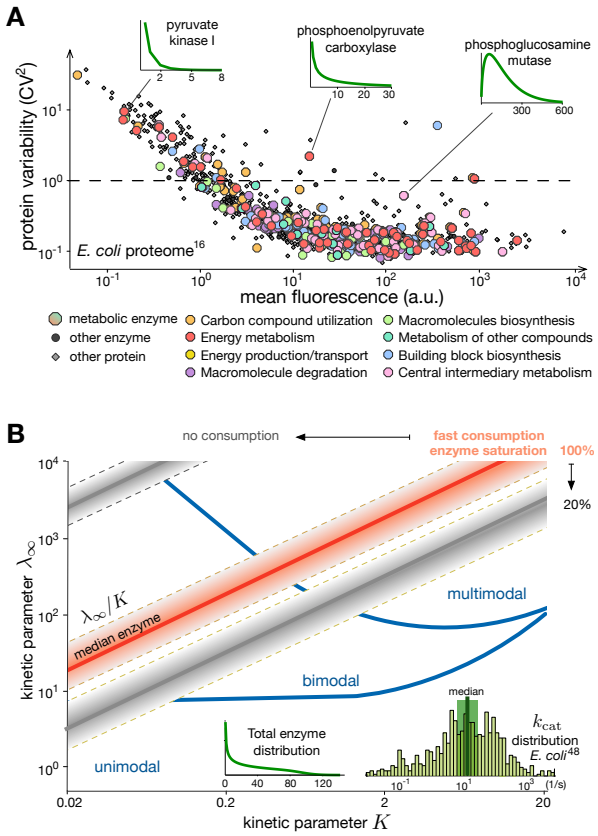

FIG. 4. **Model predictions and experimental data. (A)** Single-cell measurements reveal that metabolic enzymes are as variable as other members of the proteome[16]. Data correspond to ∼80% of the *E. coli* proteome, including 268 enzymes involved in various metabolic functions[49] (coloured circles). The coefficient of variation (CV) is defined as standard deviation over mean of measured distributions. Distributions with CV > 1 (dashed line) are long-tailed and peak at zero, which resemble the distributions required for catalytically-induced bimodality and multimodality (Figure 2B and 3); shown are the distributions of three representative enzymes computed from fitted Gamma distributions[16].**(B)** Predictions of the Poisson Mixture Model for combinations of parameters λ_∞ and K in (5). For a lowly abundant enzyme (distribution in inset), the model predicts unimodality, catalytically-induced bimodality and multimodality in large regions of the (λ_∞, K) space. The red line represents a constant ratio λ_∞/K for the median $k_{\text{cat}} \approx 16.5\,\text{s}^{-1}$ across 752 enzymes in *E. coli*, see k_{cat} distribution in inset[48]. Shaded red corresponds to the lines obtained for k_{cat} values within a range 0.5- and 2-fold of the median (highlighted in green in the inset). The grey lines and shaded areas correspond to perturbations to the consumption rate constant (k_c) and enzyme saturation (ϵ); more details in Methods.

In this paper we laid theoretical foundations to study metabolism in conjunction with stochastic enzyme expression. We brought together classic models for gene expression and enzyme kinetics, and discovered a rich array of distinct stochastic phenomena that underpin the emergence of metabolic subpopulations. Our theory provides a quantitative basis to draw testable hypotheses on the sources of metabo-

lite heterogeneity, which together with the ongoing efforts in single-cell metabolite measurements, will help to re-think metabolism as an active source of phenotypic variation.

IV. METHODS

A. Stochastic modelling and simulation

We built a fully stochastic model for the reaction scheme describing a metabolic reaction coupled with gene expression (Fig. 1A):

$$\text{Substrate} + \text{Enzyme} \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} \text{Complex} \tag{R1}$$

$$\text{Complex} \underset{k_{\text{rev}}}{\overset{k_{\text{cat}}}{\rightleftarrows}} \text{Metabolite} + \text{Enzyme} \tag{R2}$$

$$\text{mRNA} \overset{k_{\text{tl}}}{\rightarrow} \text{mRNA} + \text{Enzyme} \tag{R3}$$

$$\text{DNA}_{\text{on}} \overset{k_{\text{tx}}}{\rightarrow} \text{DNA}_{\text{on}} + \text{mRNA} \tag{R4}$$

$$\text{DNA}_{\text{off}} \underset{k_{\text{off}}}{\overset{k_{\text{on}}}{\rightleftarrows}} \text{DNA}_{\text{on}} \tag{R5}$$

$$\text{Metabolite} \overset{k_{\text{c}}}{\rightarrow} \emptyset \tag{R6}$$

$$\text{mRNA} \overset{k_{\text{deg}}}{\rightarrow} \emptyset \tag{R7}$$

$$\text{Enzyme} \overset{\delta}{\rightarrow} \emptyset \tag{R8}$$

$$\text{Complex} \overset{\delta}{\rightarrow} \emptyset \tag{R9}$$

All reactions are assumed to follow mass action kinetics. Model simulations were computed with Gillespie's algorithm[33] over long time horizons, in the order of hundreds of cell cycles for all simulations. Because of the complex multimodality observed, long simulations are needed to obtain accurate approximations of the stationary molecular distributions. The time courses shown in figures correspond to a small time window of the overall simulation. Unless mentioned in the figure captions, all parameter values were fixed to their nominal values shown in Table I. The parameters are selected in a physiologically realistic range respecting the scale separation of molecule numbers between mRNA ($\sim 1-5$ molecules), total enzyme abundance (~ 100 molecules) and metabolites (~ 1000 molecules).

To identify bimodality in Figure 2, we detected the existence of one or two peaks in a distribution and defined it as bimodal if the height of the smaller peak is a least 10% of the larger peak and the trough between peaks is at most 10% of the height of the smaller peak.

	value	unit		value	unit
n_{s}	3000	molecule	k_{tx}	0.0270	s^{-1}
k_1	1	$\text{s}^{-1}\text{molecule}^{-1}$	k_{tl}	0.2	s^{-1}
k_{-1}	1000	s^{-1}	k_{on}	0.0225	s^{-1}
k_{cat}	3.6	s^{-1}	k_{off}	0.0075	s^{-1}
k_{rev}	0.01	$\text{s}^{-1}\text{molecule}^{-1}$	k_{deg}	0.2	s^{-1}
k_{c}	0.02	s^{-1}	δ	0.00025	s^{-1}

TABLE I. **Nominal parameters for the stochastic model.** Parameters correspond to the simulations of Figure 1A. We use realistic enzyme kinetic parameters[48] and fast promoter switching according to measured ranges[52]. The dilution rate δ constant corresponds to a doubling rate of approximately 46 minutes, typical in the *Escherichia coli* bacterium. The distributions in Figure 1C were obtained with perturbed promoter switching parameters $k_{\text{on}} = 0.01\ \text{s}^{-1}$ and $k_{\text{off}} = \{0.03,\ 0.01,\ 0.0001\}\ \text{s}^{-1}$. The parameter values for Figures 2 and 3 are shown in the respective captions.

B. Analytical expressions for the distribution of the metabolite

To derive an analytic approximation for the probability to observe n_{p} metabolites in a cell, we first use the law of total probability as shown in (1).

Distribution of the total enzyme — Because free enzymes and complexes degrade at the same rate and $n_{\text{etot}} = n_{\text{e}} + n_{\text{c}}$ is conserved by the metabolic reaction, in the slow timescale the enzyme distribution $\mathbf{P}(n_{\text{etot}})$ follows the standard solution[26] of the three-stage model for gene expression:

$$\mathbf{P}(n_{\text{etot}}) = \frac{\Gamma(\alpha_+ + n_{\text{etot}})\Gamma(\alpha_- + n_{\text{etot}})\Gamma(\gamma)}{\Gamma(n_{\text{etot}}+1)\Gamma(\alpha_+)\Gamma(\alpha_-)\Gamma(\gamma + n_{\text{etot}})} \times \left(\frac{b}{1+b}\right)^{n_{\text{etot}}} \left(1 - \frac{b}{1+b}\right)^{\alpha_+} \times {}_2F_1\left(\alpha_+ + n_{\text{etot}}, \gamma - \alpha_-, \gamma + n_{\text{etot}}; \frac{b}{1+b}\right), \tag{7}$$

where Γ is the Gamma function and ${}_2F_1$ is the ordinary hypergeometric function. The parameters are $\gamma = (k_{\text{on}} + k_{\text{off}})/\delta$ and $\alpha_\pm = \left(a + \gamma \pm \sqrt{(a+\gamma)^2 - 4ak_{\text{off}}/\delta}\right)/2$, with $a = k_{\text{tx}}/\delta$ and $b = k_{\text{tl}}/k_{\text{deg}}$.

Conditional distribution for the metabolite — To compute the second term in (1), we observe that enzyme expression occurs on a much longer timescale than enzyme kinetics, and thus metabolites can be considered to be in a quasi-equilibrium state of the catalytic reactions (R1)–(R2) and metabolite consumption (R6).

To explicitly compute the mixture components $\mathbf{P}(n_{\text{p}}|n_{\text{etot}})$, we assume that reversible binding between substrate and enzyme in reaction (R1) is much faster than the catalytic step and metabolite consumption[34]. In this limit, the metabolite number evolves according to the effective reactions:

$$\emptyset \underset{\text{k}^{\text{eff}}_{\text{rev}}}{\overset{\text{k}^{\text{eff}}_{\text{cat}}}{\rightleftarrows}} \text{Metabolite} \overset{\text{k}_{\text{c}}}{\rightarrow} \emptyset, \tag{8}$$

where $\mathrm{k}_{\mathrm{cat}}^{\mathrm{eff}}$ and $\mathrm{k}_{\mathrm{rev}}^{\mathrm{eff}}$ are effective propensities averaged over the fast fluctuating variables n_{c} and n_{e}:

$$\begin{aligned} \mathrm{k}_{\mathrm{cat}}^{\mathrm{eff}} &= \mathrm{k}_{\mathrm{cat}}\mathbb{E}(n_{\mathrm{c}}|n_{\mathrm{etot}}, n_{\mathrm{p}}), \\ \mathrm{k}_{\mathrm{rev}}^{\mathrm{eff}} &= k_{\mathrm{rev}}\mathbb{E}(n_{\mathrm{e}}|n_{\mathrm{etot}}, n_{\mathrm{p}}), \end{aligned} \tag{9}$$

where $\mathbb{E}$ denotes the expectation operator. The derivation of the effective propensities in (9) corresponds to a particular case of a more general methodology for timescale separation in stochastic chemical systems[28,53,54]. Note that since the total enzyme levels are conserved in the catalytic timescale, it follows that

$$\mathbb{E}(n_{\mathrm{e}}|n_{\mathrm{etot}}, n_{\mathrm{p}}) + \mathbb{E}(n_{\mathrm{c}}|n_{\mathrm{etot}}, n_{\mathrm{p}}) = n_{\mathrm{etot}}. \tag{10}$$

To derive the conditional expectations in (9), we write the first-order moment equation for the free enzyme n_{e}, which according from equations (R1)–(R2) reads

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}(n_{\mathrm{e}}|n_{\mathrm{etot}}, n_{\mathrm{p}}) = \\ &- k_1 n_{\mathrm{s}}\mathbb{E}(n_{\mathrm{e}}|n_{\mathrm{etot}}, n_{\mathrm{p}}) + k_{-1}\mathbb{E}(n_{\mathrm{c}}|n_{\mathrm{etot}}, n_{\mathrm{p}}) \\ &+ k_{\mathrm{cat}}\mathbb{E}(n_{\mathrm{c}}|n_{\mathrm{etot}}, n_{\mathrm{p}}) - k_{\mathrm{rev}}\mathbb{E}(n_{\mathrm{e}}|n_{\mathrm{etot}}, n_{\mathrm{p}})n_{\mathrm{p}}. \end{aligned} \tag{11}$$

Under the assumption that the reversible binding of substrate and enzyme is much faster than the other processes, the first two terms dominate the right hand side of equation (11) and determine the enzyme-complex quasi-equilibrium. Equating these two terms and using the conservation relation in (10), we obtain

$$\begin{aligned} \mathbb{E}(n_{\mathrm{e}}|n_{\mathrm{etot}}, n_{\mathrm{p}}) &\approx \frac{k_{-1}}{(k_{-1} + k_1 n_{\mathrm{s}})} n_{\mathrm{etot}}, \\ \mathbb{E}(n_{\mathrm{c}}|n_{\mathrm{etot}}, n_{\mathrm{p}}) &\approx \frac{k_1 n_{\mathrm{s}}}{k_{-1} + k_1 n_{\mathrm{s}}} n_{\mathrm{etot}}, \end{aligned} \tag{12}$$

and thus both conditional expectations depend on n_{etot} and are independent of the metabolite abundance. Therefore, the reactions in (8) correspond to a birth-death process with a zero-th order birth propensity $\mathrm{k}_{\mathrm{cat}}^{\mathrm{eff}}$ and two linear death propensities. The mixture components $\mathbf{P}(n_{\mathrm{p}}|n_{\mathrm{etot}})$ are thus Poissonian with parameter $\lambda(n_{\mathrm{etot}})$ as shown in (3)–(4).

C. Comparison between PMM predictions and measured kinetic parameters

The PMM depends on the effective parameters λ_∞ and K, which are functions of five rate constants (k_{cat}, k_1, k_{-1}, k_{c} and k_{rev}). Most of these parameters are not available, except k_{cat} and $K_{\mathrm{M}} = (k_{\mathrm{cat}} + k_{-1})/k_1$. From equation (5) it follows

$$\frac{\lambda_\infty}{K} = \frac{k_{\mathrm{cat}}}{k_{\mathrm{c}}} \times \underbrace{\frac{k_1 n_{\mathrm{s}}}{k_1 n_{\mathrm{s}} + k_{-1}}}_{\%\ \text{saturation},\ \epsilon}, \tag{13}$$

which allows the computation of λ_∞/K for measured k_{cat} values in different saturation conditions and consumption rate constants. The red line in Figure 4B represents the λ_∞/K ratio for a saturated enzyme ($\epsilon = 1$), fast consumption ($k_{\mathrm{c}} = 100 \times \delta$) and the median $k_{\mathrm{cat}} \approx 16.5\ \mathrm{s}^{-1}$ in *E.coli*[48]. The top grey line is the case without consumption, i.e. metabolites are diluted by cell growth ($k_{\mathrm{c}} = \delta$). Lower saturation ($\epsilon = 0.2$) moves the red line down the vertical axis (bottom grey line). The enzyme distribution in Figure 4B was produced with promoter switching parameters $\{k_{\mathrm{on}}, k_{\mathrm{off}}\} = \{1.56, 3\} \times 10^{-4}\ \mathrm{s}^{-1}$, $k_{\mathrm{tx}} = 0.025\ \mathrm{s}^{-1}$, and $k_{\mathrm{tl}} = 0.2\ \mathrm{s}^{-1}$. The boundaries between unimodal, bimodal and multimodal distributions were computed as follows. Unimodal distributions are those with a single maximum. Bimodal distributions were detected as in Figure 2. Multimodal distributions are those with at least one additional peak higher than a threshold of 1×10^{-4} and the trough between neighbouring peaks at most 90 % of its height.

ACKNOWLEDGEMENTS

This work was funded by the Human Frontier Science Program through a Young Investigator Grant awarded to D.O (Grant no. RGY0076-2015), and The Royal Commission for the 1851 Exhibition through a Fellowship to P.T.

REFERENCES

[1] M. B. Elowitz, A. J. Levine, and E. D. Siggia, Science **279**, 1183 (2002).
[2] J. Paulsson, Physics of Life Reviews **2**, 157 (2005).
[3] A. Raj and A. van Oudenaarden, Cell **135**, 216 (2008).
[4] N. Q. Balaban, J. Merrin, R. Chait, L. Kowalik, and S. Leibler, Science **305**, 1622 (2004).
[5] K. Lewis, Nature Reviews Microbiology **5**, 48 (2007).
[6] Y. Shan, A. Brown Gandt, S. E. Rowe, J. P. Deisinger, B. P. Conlon, and K. Lewis, mBio **8**, e02267 (2017).
[7] C. Vilhena, E. Kaganovitch, J. Y. Shin, A. Grünberger, S. Behr, I. Kristoficova, S. Brameyer, D. Kohlheyer, and K. Jung, Journal of bacteriology **200**, e00536 (2018).
[8] N. Nikolic, F. Schreiber, A. Dal Co, D. J. Kiviet, T. Bergmiller, S. Littmann, M. M. M. Kuypers, and M. Ackermann, PLOS Genetics **13**, e1007122 (2017).
[9] D. J. Kiviet, P. Nghe, N. Walker, S. Boulineau, V. Sunderlikova, and S. J. Tans, Nature **514**, 376 (2014).
[10] J. H. van Heerden, M. T. Wortel, F. J. Bruggeman, J. J. Heijnen, Y. J. M. Bollen, R. Planqué, J. Hulshof, T. G. O. Toole, S. A. Wahl, and B. Teusink, Science **343**, 1245114 (2014).
[11] O. Kotte, B. Volkmer, J. L. Radzikowski, and M. Heinemann, Molecular Systems Biology **10**, 736 (2014).
[12] E. Şimşek and M. Kim, The ISME Journal (2018).
[13] M. Acar, A. Becskei, and A. van Oudenaarden, Nature **435**, 228 (2005).
[14] V. Takhaveev and M. Heinemann, Current Opinion in Microbiology **45**, 30 (2018).
[15] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox, Cell **123**, 1025 (2005).
[16] Y. Taniguchi, P. J. Choi, G.-w. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, Science **329**, 533 (2010).
[17] E. A. Lemke and C. Schultz, Nature Chemical Biology **7**, 480 (2011).
[18] Y. Xiao, C. H. Bowen, D. Liu, and F. Zhang, Nature Chemical Biology **12**, 339 (2016).
[19] A. A. Mannan, D. Liu, F. Zhang, and D. A. Oyarzún, ACS Synthetic Biology (2017).
[20] J. S. Paige, T. Nguyen-Duc, W. Song, and S. R. Jaffrey, Science **335**, 1194 (2012).

[21] A. J. Ibanez, S. R. Fagerer, A. M. Schmidt, P. L. Urban, K. Jefimovs, P. Geiger, R. Dechant, M. Heinemann, and R. Zenobi, Proceedings of the National Academy of Sciences **110**, 8790 (2013).

[22] E. M. Ozbudak, M. Thattai, H. N. Lim, B. I. Shraiman, and A. van Oudenaarden, Nature **427**, 737 (2004).

[23] A. Cornish-Bowden, *Fundamentals of Enzyme Kinetics*, 3rd ed. (Portland Press, 2004).

[24] B. Bennett, E. Kimball, M. Gao, R. Osterhout, S. J. Van Dien, and J. D. Rabinowitz, Nature Chemical Biology **5**, 593 (2009).

[25] H. Yaginuma, S. Kawai, K. V. Tabata, K. Tomiyama, A. Kakizuka, T. Komatsuzaki, H. Noji, and H. Imamura, Scientific Reports **4**, 6522 (2014).

[26] V. Shahrezaei and P. S. Swain, Proceedings of the National Academy of Sciences of the United States of America **105**, 17256 (2008).

[27] P. Labhsetwar, J. A. Cole, E. Roberts, N. D. Price, and Z. A. Luthey-schulten, PNAS **110**, 14006 (2013).

[28] P. Thomas, N. Popović, and R. Grima, Proceedings of the National Academy of Sciences of the United States of America **111**, 6994 (2014).

[29] E. Levine and T. Hwa, Proceedings of the National Academy of Sciences of the United States of America **104**, 9224 (2007).

[30] P. Thomas, A. V. Straube, and R. Grima, The Journal of Chemical Physics **135**, 181103 (2011).

[31] A. Gupta, A. Milias-argeitis, and M. Khammash, Journal of the Royal Society **14**, 1 (2017).

[32] D. A. Oyarzún, J.-B. Lugagne, and G.-B. Stan, ACS synthetic biology **4**, 116 (2015).

[33] D. T. Gillespie, Journal of Computational Physics **22**, 403 (1976).

[34] Y. Cao, D. T. Gillespie, and L. R. Petzold, The Journal of chemical physics **123**, 144917 (2005).

[35] J. Lugagne, D. A. Oyarzun, and Guy-Bart V. Stan, European Control Conference , 3646 (2013).

[36] D. T. Gillespie, Annual Review of Physical Chemistry **58**, 35 (2007).

[37] N. van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, 1992).

[38] S. Chaturvedi, C. W. Gardiner, I. S. Matheson, and D. F. Walls, Journal of Statistical Physics **17**, 469 (1977).

[39] S. Iyer-Biswas, F. Hayot, and C. Jayaprakash, Physical Review E **79**, 031911 (2009).

[40] J. Dattani and M. Barahona, Journal of The Royal Society Interface **14**, 1 (2017).

[41] D. A. Oyarzún and M. Chaves, Journal of The Royal Society Interface **12** (2015).

[42] A. Lipshtat, A. Loinger, N. Q. Balaban, and O. Biham, Physical Review Letters **96**, 1 (2006), 0603026.

[43] T.-L. To and N. Maheshri, Science **327**, 1142 (2010).

[44] M. Wehrens, F. Buke, P. Nghe, and S. J. Tans, Current Opinion in Systems Biology (2018), 10.1016/J.COISB.2018.02.006.

[45] A. M. Corrigan, E. Tunnacliffe, D. Cannon, and J. R. Chubb, eLife **5**, 1 (2016).

[46] H. Ge, P. Wu, H. Qian, and X. S. Xie, PLOS Computational Biology **14**, e1006051 (2018).

[47] I. Schomburg, A. Chang, S. Placzek, C. Söhngen, M. Rother, M. Lang, C. Munaretto, S. Ulas, M. Stelzer, A. Grote, M. Scheer, and D. Schomburg, Nucleic acids research **41**, D764 (2013).

[48] A. Bar-Even, E. Noor, Y. Savir, W. Liebermeister, D. Davidi, D. S. Tawfik, and R. Milo, Biochemistry **50**, 4402 (2011).

[49] M. H. Serres and M. Riley, Microbial & comparative genomics **5**, 205 (2000).

[50] D. Liu, A. A. Mannan, Y. Han, D. A. Oyarzún, and F. Zhang, Journal of Industrial Microbiology & Biotechnology (2018).

[51] X. Fang, A. Sastry, N. Mih, D. Kim, J. Tan, J. T. Yurkovich, C. J. Lloyd, Y. Gao, L. Yang, and B. O. Palsson, Proceedings of the National Academy of Sciences of the United States of America **114**, 10286 (2017).

[52] L. H. So, A. Ghosh, C. Zong, L. A. Sepúlveda, R. Segev, and I. Golding, Nature Genetics **43**, 554 (2011).

[53] J. Goutsias, The Journal of Chemical Physics **122**, 184102 (2005).

[54] B. Melykuti, J. P. Hespanha, and M. Khammash, Journal of The Royal Society Interface **11**, 20140054 (2014).