

RESEARCH ARTICLE

Open Access

Deterministic and stochastic population-level simulations of an artificial *lac* operon genetic network

Michail Stamatakis* and Kyriacos Zygourakis

Abstract

Background: The *lac* operon genetic switch is considered as a paradigm of genetic regulation. This system has a positive feedback loop due to the LacY permease boosting its own production by the facilitated transport of inducer into the cell and the subsequent de-repression of the *lac* operon genes. Previously, we have investigated the effect of stochasticity in an artificial *lac* operon network at the single cell level by comparing corresponding deterministic and stochastic kinetic models.

Results: This work focuses on the dynamics of cell populations by incorporating the above kinetic scheme into two Monte Carlo (MC) simulation frameworks. The first MC framework assumes stochastic reaction occurrence, accounts for stochastic DNA duplication, division and partitioning and tracks all daughter cells to obtain the statistics of the entire cell population. In order to better understand how stochastic effects shape cell population distributions, we develop a second framework that assumes deterministic reaction dynamics. By comparing the predictions of the two frameworks, we conclude that stochasticity can create or destroy bimodality, and may enhance phenotypic heterogeneity.

Conclusions: Our results show how various sources of stochasticity act in synergy with the positive feedback architecture, thereby shaping the behavior at the cell population level. Further, the insights obtained from the present study allow us to construct simpler and less computationally intensive models that can closely approximate the dynamics of heterogeneous cell populations.

Background

Since the introduction of the operon concept by Jacob et al. [1], the *lac* operon genetic switch has been considered as a paradigm for genetic regulation. Several experimental studies of this system over the past several decades have elucidated the underlying biomolecular interactions and a plethora of mathematical models have integrated the complex interplays of the key biochemical species in order to predict the behavior of the system [see, for example, [2]].

Most of these models, however, pertain to the single cell behavior [see, for example the models reviewed in [3]] with a limited number of studies focusing on cell populations, or taking comparative approaches. For instance, in a mini-review article, Vilar et al. [4]

compared different models pertaining to single cells and cell populations, in order to illustrate the performance and limitations of different methodologies. At the single cell level Vilar et al. [4] simulated four phenomenological ordinary differential equations (ODEs) for the concentrations of permease, inducer, and β -galactosidase. A stochastic single cell model was then developed by transforming the phenomenological deterministic rates into propensities. However, the latter transformation is not unique, contrary to the case of a mechanistic model where the reaction rate expressions are derived on the basis of statistical mechanical assumptions [see for instance chapter X in ref. [5], and refs [6-8]]. Thus, the stochastic model simulated by Vilar et al. [4] is based on *ad hoc* assumptions. In addition, the Vilar et al. models do not appear to account for the dilution of the concentrations due to cellular growth [9], for cell division and

* Correspondence: mstam@alumni.rice.edu
Department of Chemical and Biomolecular Engineering, Rice University,
Houston, TX 77005, USA

for stochastic partitioning or DNA duplication effects, which are important sources of extrinsic noise [10-12].

Furthermore, van Hoek and Hogeweg [13,14] studied the effect of intrinsic and extrinsic sources of stochasticity, as well as genetic mutations and spatial heterogeneity, at the single-cell and the population level from an evolutionary perspective. Their deterministic model consists of an intracellular and an extracellular part, with the latter capturing the influxes and effluxes of glucose and lactose into and from the cells, as well as the diffusion of these species over a grid, the points of which are assumed to be vacant or occupied by single cells. The intracellular part of the model employs ten differential equations for the concentrations of mRNA, β -galactosidase, permease, lactose, allolactose, glucose, glucose-6-phosphate, cAMP, ATP, and the size of the cell. The rate expressions used in this model contain Hill functions to model saturation and cooperativity effects, and account for dilution due to cell growth. Hence, *stricto sensu* the model is not a mechanistic one, and the incorporation of stochastic effects invokes *ad hoc* assumptions rather than a formal approach based on statistical mechanics. In particular, van Hoek and Hogeweg [13] use the deterministic equation for the mRNA species, but interpret the mRNA concentration as the probability of a single mRNA molecule being present in the cell. They subsequently use this probability to infer the frequency of translational bursts.

Thus, there remain several open questions regarding the emergence of population behavior from the complex interplay between reaction dynamics and stochasticity at the single cell level. Investigating this connection and ultimately understanding cell population dynamics is significant for two reasons. First, typical biology experiments involve cell populations rather than single cells and, thus, phenotypic distributions obtained for instance from flow cytometry pertain to the cell population rather than the lifetime of a single cell. Second, there have been mathematical modeling studies for simple genetic networks, suggesting that the behavior of the single cell is very different from the behavior of the cell population [12,15,16]. It would therefore be interesting to investigate whether this is the case in the more complex *lac* operon system, or whether one can formulate average single cell models that can adequately capture the behavior of the cell population.

In a previously published article, Stamatakis and Manztaris [17] have presented a kinetic scheme that captures the salient features of an artificial *lac* operon system, which can be constructed in the lab by introducing mutations to the wild-type system. In the present paper we will incorporate that kinetic scheme into two Monte Carlo (MC) frameworks that simulate cell population dynamics.

The first framework assumes stochastic reaction occurrence and takes into account stochastic DNA duplication, division and partitioning. McAdams and Arkin [18-20] were the first to employ the Gillespie MC algorithm [7,8] to simulate gene induction and protein synthesis at the single cell level. This approach was later extended to account for cell growth and division [21-23]. However, all these and many other studies simulated the dynamics of biochemical pathways at the single cell level. In a recent study we generalized and expanded the chemical master equation (CME) to the cell population level [10]. The resulting cell population master equation (CPME) is simulated with a Monte Carlo algorithm, and captures stochasticity in the intracellular reaction, stochastic DNA duplication, and division events, as well as in the partitioning of the content of a mother cell to the two daughters. This novel framework is applicable to cell populations since every single daughter cell is tracked, thereby making possible the calculation of any statistical property of the population.

In order to better understand how stochastic effects shape cell population distributions, we develop here a second framework that assumes deterministic reaction dynamics and stochastic DNA duplication, cell division and partitioning. Throughout this work, the single cell models are derived from the reaction network developed in ref. [17] by invoking standard assumptions based on statistical mechanics and reaction rate theory [7,8,24] rather than *ad hoc* techniques [4]. We also explicitly take into account the dilution effect due to cell growth in the deterministic case [9]. Furthermore, stochasticity in the times of division and DNA duplication, as well as stochastic partitioning effects are effects that were neglected in previous studies [4], but are explicitly taken into account in our cell population frameworks.

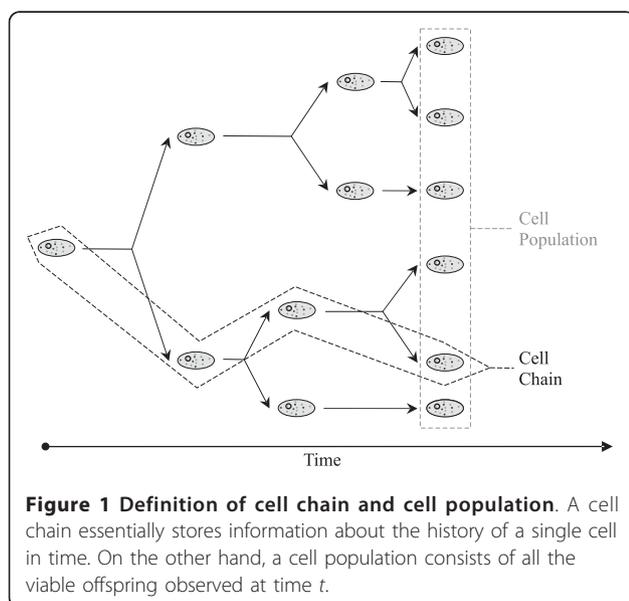
Comparisons of the predictions of the two frameworks reveal that stochasticity can create or destroy bimodality, and may enhance phenotypic heterogeneity by creating heavy tailed distributions, a phenomenon that has not been shown before and can be investigated experimentally. The insights provided from the new study also allow us to construct simpler and less computationally intensive models that can closely approximate the dynamics of heterogeneous cell populations. Specifically, for the case of deterministic reaction dynamics, we use the continuum model formulation which assumes that all cells in the population occupy a continuous and expanding biotic phase [9]. Under this assumption, a set of mass balances for a representative "average" cell in the population provides a lumped description of cell population dynamics. We demonstrate an excellent agreement between the continuum and the cell population models, which is encountered for the first time and contradicts previous studies. Extending these results in

the stochastic case, we show that the calculation of distributions of intensive quantities (such as species concentrations) can be performed on the basis of single cell simulations, instead of computationally demanding population-level ones.

Methods

Figure 1 defines pictorially the concepts of “cell chain” and “cell population.” A cell chain is a collection of cells defined by starting from one mother cell, choosing one of its daughter cells upon division, setting this as next mother cell and repeating the process. A Bernoulli trial with probability 1/2 governs the choice of which cell to keep after each division event. A cell chain essentially stores information about the history of a single cell in time. On the other hand, a cell population consists of all the viable offspring observed at time t that were generated by an arbitrary number of cells at $t = 0$. For simplicity, Figure 1 portrays a population that originated from a single cell. However, our definition can be applied to cases where more than one cells exist at time $t = 0$.

We are now ready to present the frameworks that will be used for the simulating cell populations in this study. Both frameworks treat the occurrence of division and DNA duplication events as stochastic processes. Their difference lies in the treatment of reactions and the partitioning of species between the daughter cells: the first framework treats these as stochastic, whereas the second one as deterministic processes. We will refer to the former as the “population model with stochastic reaction dynamics”, and the latter as the “population model with deterministic reaction dynamics”.



Population Model with Stochastic Reaction Dynamics

In an earlier study [10], we presented the formulation of a cell population master equation (CPME) that describes cell population dynamics and takes into account the major sources of heterogeneity: stochasticity of intracellular reactions, DNA-duplication, cell division, and random partitioning of species contents into the two daughter cells. That formulation also takes into account cell growth and respects the discrete nature of the molecular contents and cell numbers. Our approach assumes that each cell of a population can be completely described by a state vector $\mathbf{z} = (\mathbf{X}, V)$ where \mathbf{X} is a vector with n entries for species copy numbers and V is the volume of the cell. Additional morphometric characteristics like cell membrane area or length can be easily incorporated into our framework. We also classify the chemical species into non-chromosomal DNA species and chromosomal DNA species that may exist in various states.

This CPME describes the evolution of the probability of finding at time t the cell population having v cells that exist in states $(\mathbf{X}_1, V_1), \dots, (\mathbf{X}_i, V_i), \dots, (\mathbf{X}_v, V_v)$. If we denote this probability by $J_v[(\mathbf{X}_1, V_1), \dots, (\mathbf{X}_i, V_i), \dots, (\mathbf{X}_v, V_v); t]$, the analogue of the Janossy density used in the continuous population balances [25], the CPME becomes [10]:

$$\frac{\partial}{\partial t} J_v[(\mathbf{X}_1, V_1), \dots, (\mathbf{X}_i, V_i), \dots, (\mathbf{X}_v, V_v); t] = F_R + F_S + F_G + F_D \quad (1)$$

where the terms F_R , F_S , F_G and F_D describe, respectively, the stochastic dynamics of intracellular reactions, DNA duplication, cell growth, and cell division. The form of these terms is presented in Eq. 6 of Additional file 1. Our earlier publication [10] provides the details about the development of the CPME and the Monte Carlo algorithm that simulates the stochastic processes it describes (this information is also included in Section S1 of Additional file 1 for convenience). Here, we just note that the partitioning of cellular volumes upon division is governed by a symmetric beta distribution. Further, the partitioning of non-chromosomal DNA species follows the binomial distribution whereas for chromosomal DNA species partitioning is governed by a multivariate hypergeometric distribution. These choices result in division events where the daughters may have different sizes and species contents; however, on average any daughter inherits half the contents and has half the size of the mother cell. In a more general setting, it is possible to use sums of non-symmetric beta distributions to describe asymmetric (biased) division (see Section S2 in Additional file 1). In this case, one daughter

will consistently inherit more of the mother's molecular content.

Population Model with Deterministic Reaction Dynamics

The extension of our approach to the case where reactions are deterministic is straightforward and is presented in detail in Section S1 of Additional file 1. For this model, the intracellular species contents are assumed to be continuous vector variables that vary according to reactions described by a system of ordinary differential equations (ODEs). The processes of cell growth and DNA duplication are treated in a similar manner as before, while a modified partitioning function is used here to account for the fact that the chemical species are no longer discrete entities. The modified CPME equation then describes a piecewise deterministic Markov process (a general framework appears in ref. [26]) and takes the form:

$$\frac{\partial}{\partial t} J_v [(X_1, V_1), \dots, (X_i, V_i), \dots, (X_v, V_v); t] = F_R^* + F_S + F_G + F_D^* \quad (2)$$

where the terms F_R^* , F_S , F_G and F_D^* describe, respectively, the deterministic dynamics of intracellular reactions and cell growth, and stochastic DNA duplication and cell division. Section S1 of Additional file 1 provides the details for the derivation of Eq. 2, which is essentially the deterministic limit of Eq. 1 for all species being present in large numbers inside the cell. In this limit, the stochastic fluctuations due to the individual reaction events are suppressed. As a consequence, reaction dynamics become deterministic (see Eqs. 11, 14 and pertinent discussion in ref. [6]) and the partitioning of species is now governed by Dirac delta distributions (see Section S3 of Additional file 1).

Structured Continuum Model: a Lumped Description of Cell Population Dynamics

Since the simulation of detailed population models can be computationally expensive, one often resorts to simpler models. In this study, we will use and evaluate the structured continuum model formulation attributed to Fredrickson [9]. This model essentially lumps the volume occupied by the cells into a continuous and expanding (due to cell growth) biotic phase. Thus, the structured continuum model consists of transient mass balance equations written for species concentrations, assuming deterministic reaction dynamics and taking into consideration the dilution effect due to cell growth. A general continuum model can thus be written as a set of ordinary differential equations:

$$\frac{d[X_i]}{dt} = r_i ([X_1], [X_2], \dots, [X_n]) - \mu [X_i] \quad (3)$$

$$i = 1, 2, \dots, n$$

where $[X_i]$ is the average intracellular concentration (an intensive quantity) of species i over the entire cell population and μ is the average specific growth rate:

$$\mu = \left\langle \frac{1}{V} \cdot \frac{dV}{dt} \right\rangle \quad (4)$$

The Artificial *lac* Operon Reaction Network

To simulate the artificial *lac* operon genetic network, we will use the reaction network of Stamatakis and Manztaris [17]. This is a minimal model that neglects effects such as the $\sigma 70$ dependence of the *lac* promoter and assumes that only one *lacO* operator is functional; thus, DNA looping is not accounted for. The existence of three operator sites has been previously shown to result in stronger repression and higher sensitivity in induction [27], as well as lower sensitivity to changes in repressor molecules and lower transcriptional noise [28]. Computational studies of mutations and deletions in these operators have been able to successfully reproduce experimental results [29,30].

Here, our intention has been to model an artificial *lac* operon that can be constructed in the lab rather than the natural *lac* operon system. This artificial system incorporates the positive feedback from the LacY permease resulting in bistable behavior, and thus, allows us to isolate the contributions of stochasticity and bistability in shaping the behavior of the cell population. A more complicated pathway incorporating the aforementioned interactions would be intractable within the cell population frameworks that we have developed, and it is not within the scope of this work to investigate these effects.

Table 1 lists the reactions of this network and presents the corresponding propensity functions needed for the term F_R of the CPME (1) with stochastic reaction dynamics. Table 2 presents the rate equations needed for the term F_R^* of the modified CPME (2) with deterministic reaction and growth dynamics and stochastic DNA-duplication and division. The transient mass balances that constitute the structured continuum model (Eqs. 3 and 4 above) are presented in Section S4 of Additional file 1. Note again that the variables of the structured continuum model are intensive quantities that correspond to the average intracellular concentrations of a chemical species over the entire population that is treated as a lumped and expanding biotic phase [9].

Table 1 Reactions and Propensity Functions for the Stochastic *lac* Operon Model

	Reaction	Propensity Function ^{1, 2, 3}
(1-1)	$\emptyset \xrightarrow{k_{sMR}} M_R$	$\alpha_1 = V_{E.coli} \cdot N_A \cdot k_{sMR}$
(1-2)	$M_R \xrightarrow{k_{sR}} M_R + R$	$\alpha_2 = k_{sR} \cdot M_R$
(1-3)	$2R \xrightarrow{k_{2R}} R_2$	$\alpha_3 = \frac{k_{2R}}{V_{E.coli} \cdot N_A} \cdot R \cdot (R - 1)$
(1-4)	$R_2 \xrightarrow{k_{-2R}} 2R$	$\alpha_4 = k_{-2R} \cdot R_2$
(1-5)	$R_2 + O \xrightarrow{k_r} R_2O$	$\alpha_5 = \frac{k_r}{V_{E.coli} \cdot N_A} \cdot R_2 \cdot O$
(1-6)	$R_2O \xrightarrow{k_{-r}} R_2 + O$	$\alpha_6 = k_{-r} \cdot R_2O$
(1-7)	$2I + R_2 \xrightarrow{k_{dr1}} I_2R_2$	$\alpha_7 = \frac{k_{dr1}}{(V_{E.coli} \cdot N_A)^2} \cdot R_2 \cdot I \cdot (I - 1)$
(1-8)	$I_2R_2 \xrightarrow{k_{-dr1}} 2I + R_2$	$\alpha_8 = k_{-dr1} \cdot I_2R_2$
(1-9)	$2I + R_2O \xrightarrow{k_{dr2}} I_2R_2 + O$	$\alpha_9 = \frac{k_{dr2}}{(V_{E.coli} \cdot N_A)^2} \cdot R_2O \cdot I \cdot (I - 1)$
(1-10)	$I_2R_2 + O \xrightarrow{k_{-dr2}} 2I + R_2O$	$\alpha_{10} = \frac{k_{-dr2}}{V_{E.coli} \cdot N_A} \cdot I_2R_2 \cdot O$
(1-11)	$O \xrightarrow{k_{s1MY}} O + M_Y$	$\alpha_{11} = k_{s1MY} \cdot O$
(1-12)	$R_2O \xrightarrow{k_{s0MY}} R_2O + M_Y$	$\alpha_{12} = k_{s0MY} \cdot R_2O$
(1-13)	$M_Y \xrightarrow{k_{sY}} M_Y + Y$	$\alpha_{13} = k_{sY} \cdot M_Y$
(1-14)	$Y + I_{ex} \xrightarrow{k_p} YI_{ex}$	$\alpha_{14} = k_p \cdot [I_{ex}] \cdot Y$
(1-15)	$YI_{ex} \xrightarrow{k_{-p}} Y + I_{ex}$	$\alpha_{15} = k_{-p} \cdot YI_{ex}$
(1-16)	$YI_{ex} \xrightarrow{k_{ft}} Y + I$	$\alpha_{16} = k_{ft} \cdot YI_{ex}$
(1-17)	$I_{ex} \xrightarrow{h_t} I$	$\alpha_{17} = A_{E.coli} \cdot N_A \cdot h_t \cdot [I_{ex}]$
(1-18)	$I \xrightarrow{h_t} I_{ex}$	$\alpha_{18} = \frac{A_{E.coli}}{V_{E.coli}} \cdot h_t \cdot I$
(1-19)	$M_R \xrightarrow{\lambda_{MR}} \emptyset$	$\alpha_{19} = \lambda_{MR} \cdot M_R$
(1-20)	$M_Y \xrightarrow{\lambda_{MY}} \emptyset$	$\alpha_{20} = \lambda_{MY} \cdot M_Y$
(1-21)	$R \xrightarrow{\lambda_R} \emptyset$	$\alpha_{21} = \lambda_R \cdot R$
(1-22)	$R_2 \xrightarrow{\lambda_{R2}} \emptyset$	$\alpha_{22} = \lambda_{R2} \cdot R_2$
(1-23)	$Y \xrightarrow{\lambda_Y} \emptyset$	$\alpha_{23} = \lambda_Y \cdot Y$
(1-24)	$YI_{ex} \xrightarrow{\lambda_{YI_{ex}}} I$	$\alpha_{24} = \lambda_{YI_{ex}} \cdot YI_{ex}$
(1-25)	$I_2R_2 \xrightarrow{\lambda_{I2R2}} 2I$	$\alpha_{25} = \lambda_{I2R2} \cdot I_2R_2$

¹ Variables without brackets denote number of molecules of the corresponding species.

² All propensity functions have units of min^{-1} .

³ Avogadro's number: $N_A = 6.0221367 \cdot 10^{14} \text{ nmol}^{-1}$.

Table 2 Rate Equations for the Deterministic *lac* Operon Model

$$(2-1) \quad \frac{dM_R}{dt} = k_{sMR} \cdot V - \lambda_{MR} \cdot M_R$$

$$(2-2) \quad \frac{dR}{dt} = k_{sR} \cdot M_R - 2 \cdot \frac{k_{2R}}{V} \cdot R^2 + 2 \cdot k_{-2R} \cdot R_2 - \lambda_R \cdot R$$

$$(2-3) \quad \frac{dR_2}{dt} = \frac{k_{2R}}{V} \cdot R^2 - k_{-2R} \cdot R_2 - \frac{k_r}{V} \cdot R_2 \cdot O + k_{-r} \cdot R_2 O - \frac{k_{dr1}}{V^2} \cdot R_2 \cdot I^2 + k_{-dr1} \cdot I_2 R_2 - \lambda_{R2} \cdot R_2$$

$$(2-4) \quad \frac{dO}{dt} = -\frac{k_r}{V} \cdot R_2 \cdot O + k_{-r} \cdot R_2 O + \frac{k_{dr2}}{V^2} \cdot R_2 O \cdot I^2 - \frac{k_{-dr2}}{V} \cdot O \cdot I_2 R_2$$

$$(2-5) \quad \frac{dR_2 O}{dt} = \frac{k_r}{V} \cdot R_2 \cdot O - k_{-r} \cdot R_2 O - \frac{k_{dr2}}{V^2} \cdot R_2 O \cdot I^2 + \frac{k_{-dr2}}{V} \cdot O \cdot I_2 R_2$$

$$(2-6) \quad \frac{dI}{dt} = -2 \cdot \frac{k_{dr1}}{V^2} \cdot R_2 \cdot I^2 + 2 \cdot k_{-dr1} \cdot I_2 R_2 - 2 \cdot \frac{k_{dr2}}{V^2} \cdot R_2 O \cdot I^2 + 2 \cdot \frac{k_{-dr2}}{V} \cdot O \cdot I_2 R_2 + k_p \cdot YI_{ex} + h_1 \cdot A \cdot [I_{ex}] - h_1 \cdot \frac{A}{V} \cdot I + 2 \cdot \lambda_{I2R2} \cdot I_2 R_2 + \lambda_{YI_{ex}} \cdot YI_{ex}$$

$$(2-7) \quad \frac{dI_2 R_2}{dt} = \frac{k_{dr1}}{V^2} \cdot R_2 \cdot I^2 - k_{-dr1} \cdot I_2 R_2 + \frac{k_{dr2}}{V^2} \cdot R_2 O \cdot I^2 - \frac{k_{-dr2}}{V} \cdot O \cdot I_2 R_2 - \lambda_{I2R2} \cdot I_2 R_2$$

$$(2-8) \quad \frac{dM_Y}{dt} = k_{sOMY} \cdot R_2 O + k_{s1MY} \cdot O - \lambda_{MY} \cdot M_Y$$

$$(2-9) \quad \frac{dY}{dt} = k_{sY} \cdot M_Y + (k_{ft} + k_{-p}) \cdot YI_{ex} - k_p \cdot [I_{ex}] \cdot Y - \lambda_Y \cdot Y$$

$$(2-10) \quad \frac{dYI_{ex}}{dt} = -(k_{ft} + k_{-p}) \cdot YI_{ex} + k_p \cdot [I_{ex}] \cdot Y - \lambda_{YI_{ex}} \cdot YI_{ex}$$

Finally, Table 3 presents the parameter values of the *lac* operon reaction model, the propensities a_s and a_d appearing in the terms F_S and F_D of Eqs. 1 and 2, the growth rate g needed for the term F_G , and the volume-related part of the partitioning PDFs (see Eq. 12 in Additional file 1). The parameter set is consistent with the one used in ref. [17].

There is considerable uncertainty for some parameters. For instance, the equilibrium constant for repression, k_{-r}/k_r is in the range of 10^{-13} to 10^{-11} M [31-33]. The half-life for dissociation of operator DNA fragments from the repressor has been reported as 30 ~ 49 s [31,34]; thus, we have taken the dissociation rate $k_{-r} = 2.4 \text{ min}^{-1}$ and $k_r = 960 \text{ nM}^{-1} \cdot \text{min}^{-1}$ which results in $k_{-r}/k_r = 2.5 \cdot 10^{-12}$ M. This value for k_r turns out to be much higher than the experimentally measured one of $0.6 \text{ nM}^{-1} \cdot \text{min}^{-1}$, which is deduced by the 59 s time needed for a repressor to find an operator [35], assuming $V_{E.coli} = 10^{-15}$ L. Yet, simulations of the single cell deterministic model with this value of k_r produce quantitatively identical bifurcation structures with the nominal parameter set, provided that the value of k_{-r} has been adjusted to keep the thermodynamic constant for repression the same. Since the timescale for repression will be different in this case, the single cell stochastic model is expected to exhibit bistability for different induction levels. Previous work [17] showed that slower repressor-operator association and dissociation results in wider

bistable regions. In general, however, we expect all qualitative features reported here, to be valid for slower repressor-operator dynamics as well.

For the *lac* operon networks whose parameters are presented in the previous Tables, the propensity functions and rate expression associated with free thiomethyl beta-D-galactoside (TMG) transport require knowledge of the cell membrane area $A_{E.coli}$. The latter can be calculated assuming that the *E. coli* cells consist of a cylindrical segment with length w and radius R_0 and two hemispherical caps of radius R_0 at both ends. Then the overall length of the cell L is equal to $w + 2 \cdot R_0$. Since the diameter of *E. coli* changes only about 8% or less during one division cycle [36], we neglect this change as insignificant in comparison to the two-fold length change of the bacterium. Thus, we assume that R_0 is fixed in the following expression for the volume:

$$V(t) = \frac{4}{3} \cdot \pi \cdot R_0^3 + \pi \cdot R_0^2 \cdot w(t) \quad (5)$$

The cell membrane area can be then calculated as a function of the volume (for given R_0):

$$A(t) = 4 \cdot \pi \cdot R_0^2 + 2 \cdot \pi \cdot R_0 \cdot w(t) \\ = \frac{2}{R_0} \cdot V(t) + \frac{4}{3} \cdot \pi \cdot R_0^2 \quad (6)$$

Table 3 Parameters of the *lac* operon models

Symbol	Value	Units	Description
$R_{0,E. coli}$	0.4	μm	<i>E. coli</i> radius
$L_{E. coli}$	2.3	μm	Representative <i>E. coli</i> length
$A_{E. coli}$	5.8	μm^2	<i>E. coli</i> membrane area (for the above length)
$V_{E. coli}$	1.0	fL	<i>E. coli</i> volume (for the above length)
O_T	1	(copy number)	operator molecular content
k_{sMR}	0.23	$\text{nM}\cdot\text{min}^{-1}$	<i>lacI</i> transcription rate
k_{sR}	15	min^{-1}	Lacl monomer translation rate constant
k_{2R}	50	$\text{nM}^{-1}\cdot\text{min}^{-1}$	Lacl dimerization rate constant
k_{-2R}	10^{-3}	min^{-1}	Lacl dimer dissociation rate constant
k_r	960	$\text{nM}^{-1}\cdot\text{min}^{-1}$	association rate constant for repression
k_{-r}	2.4	min^{-1}	dissociation rate constant for repression
k_{dr1}	$3\cdot 10^{-7}$	$\text{nM}^{-2}\cdot\text{min}^{-1}$	association rate constant for 1 st derepression mechanism
k_{-dr1}	12	min^{-1}	dissociation rate constant for 1 st derepression mechanism
k_{dr2}	$3\cdot 10^{-7}$	$\text{nM}^{-2}\cdot\text{min}^{-1}$	association rate constant for 2 nd derepression mechanism
k_{-dr2}	$4.8\cdot 10^3$	$\text{nM}^{-1}\cdot\text{min}^{-1}$	dissociation rate constant for 2 nd derepression mechanism
k_{s1MY}	0.5	min^{-1}	<i>lacY</i> transcription rate constant
k_{s0MY}	0.01	min^{-1}	leak <i>lacY</i> transcription rate constant
k_{sY}	30	min^{-1}	<i>lacY</i> translation rate constant
k_p	0.12	$\text{nM}^{-1}\cdot\text{min}^{-1}$	LacY-inducer association rate constant
k_{-p}	0.1	min^{-1}	LacY-inducer dissociation rate constant
k_{ft}	$6\cdot 10^4$	min^{-1}	TMG facilitated transport constant
h_t	$1.55\cdot 10^{-6}$	$\text{dm}\cdot\text{min}^{-1}$	TMG passive diffusion permeability constant
λ_{MR}	0.462	min^{-1}	<i>lacI</i> mRNA degradation constant
λ_{MY}	0.462	min^{-1}	<i>lacY</i> mRNA degradation constant
λ_R	0.2	min^{-1}	Lacl monomer degradation constant
λ_{R2}	0.2	min^{-1}	Lacl dimer degradation constant
λ_Y	0.2	min^{-1}	LacY degradation constant
λ_{Ylex}	0.2	min^{-1}	LacY-inducer degradation constant
λ_{I2R2}	0.2	min^{-1}	repressor-inducer degradation constant
g	0.0231	(min^{-1})	cell growth rate parameter
n_d	25	(dim/less)	division propensity sharpness exponent
$V_{d,crit}$	15	(fL)	critical volume for division
q	80	(dim/less)	beta distribution sharpness exponent
n_s	25	(dim/less)	DNA duplication propensity sharpness exponent
$V_{s,crit}$	10	(fL)	critical volume for DNA duplication

Results and discussion

Comparison of Deterministic and Stochastic Reaction Dynamics

Deterministic Reaction Dynamics: Phenotypic Distributions and Statistics

The *lac* operon system is well known for its ability to exhibit bistable behavior when artificial inducers are used [37]. In our model bistability arises from the positive feedback that LacY exerts on its own expression. Other studies suggest that DNA loop structures also play a key role in the bistable behavior of the *lac* operon system [refer to [38,39]]. In this work, however, we focus solely on the aforementioned autocatalytic architecture.

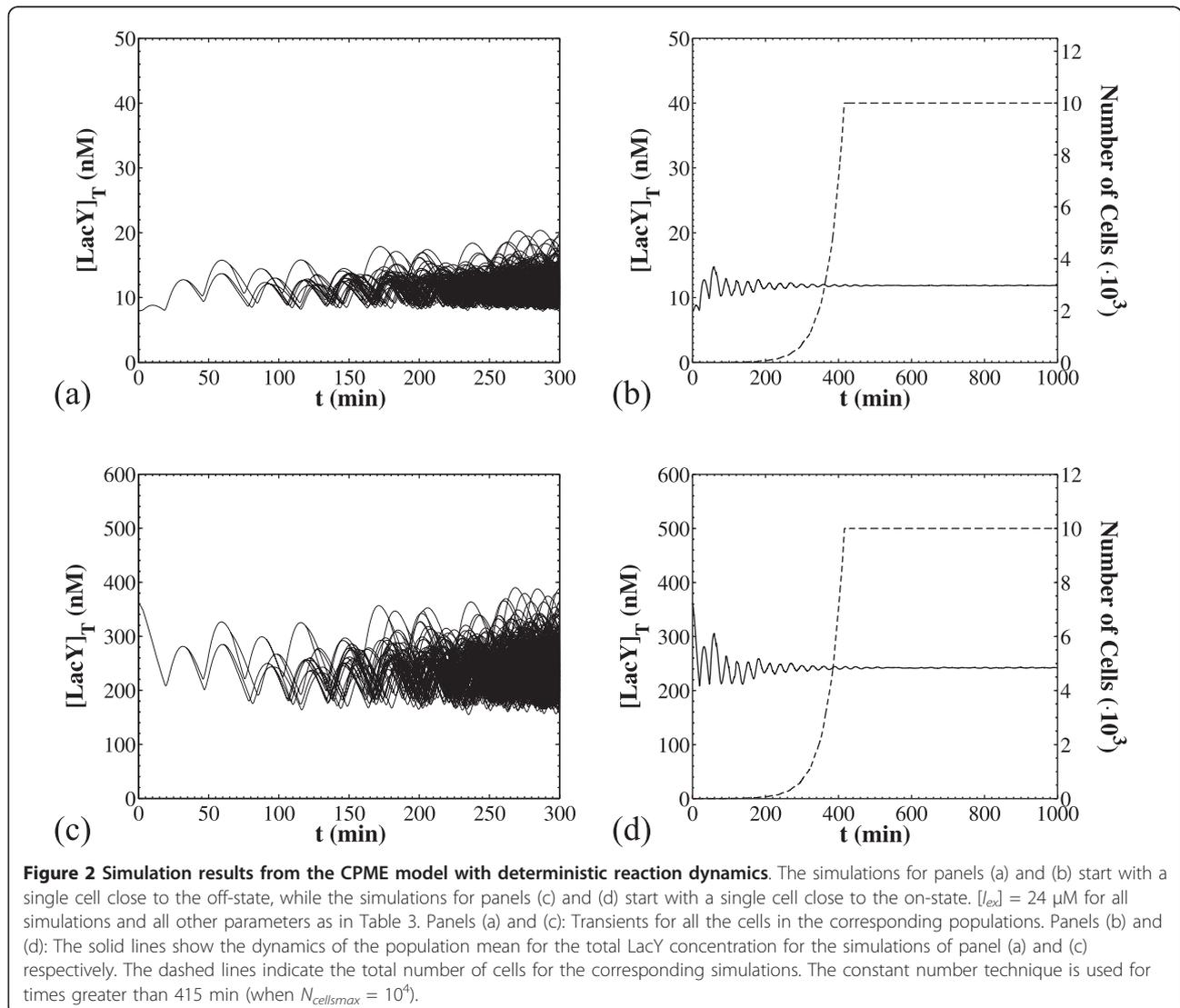
Thus, Figure 2 demonstrates the key behavior resulting from this architecture by presenting two representative transient simulations with the modified CPME of Eq. 2. For both simulations, the extracellular inducer TMG concentration is $[I_{ex}] = 24 \mu\text{M}$ and the only difference is in the initial conditions. In panels (a) and (b), the population is initialized with a single cell having species concentrations close to the off state of the *lac* operon switch. Panel (a) shows the total LacY concentrations, LacY_T , for every cell in the population versus time. Notice that all offspring remain close to the off state of the *lac* operon switch. Also, the concentrations of the mother and daughter cells are the same at the time of division (Eq. 18 of Additional file 1), and thus,

each division appears as a bifurcation in this plot. For example, the first division occurs at 33.8 min. Panel (b) shows the dynamics of the population average for LacY_T , and the number of cells considered in the population ($N_{\text{cellsmax}} = 10^4$, a value reached at 415 min). On the other hand, panels (c) and (d) pertain to the case where the population is initialized with a single cell having species concentrations close to the on-state of the *lac* operon switch. For this case, all offspring remain close to the upper state (panel c), and so does the population average plotted in panel (d).

The oscillations shown in Figure 2 are due to the fluctuations of the promoter concentration during growth and DNA duplication. In particular, at the beginning of the life cycle of a newborn cell, a single promoter exists and the cell has a small volume. Thus, the promoter concentration is high and drops as the

cell grows. Right after the DNA duplication event, the concentration of the promoter doubles and subsequently drops again as the cell continues to grow up to the point it divides. Upon cell division, the promoter concentration could also change if division is not symmetric, and thus the daughters inherit half the promoter content but they have volumes that are not equal to $V_{\text{mother}}/2$. Of course, the concentrations of all other species (that are modeled deterministically) remain the same during DNA duplication and cell division (see section S3 in Additional file 1).

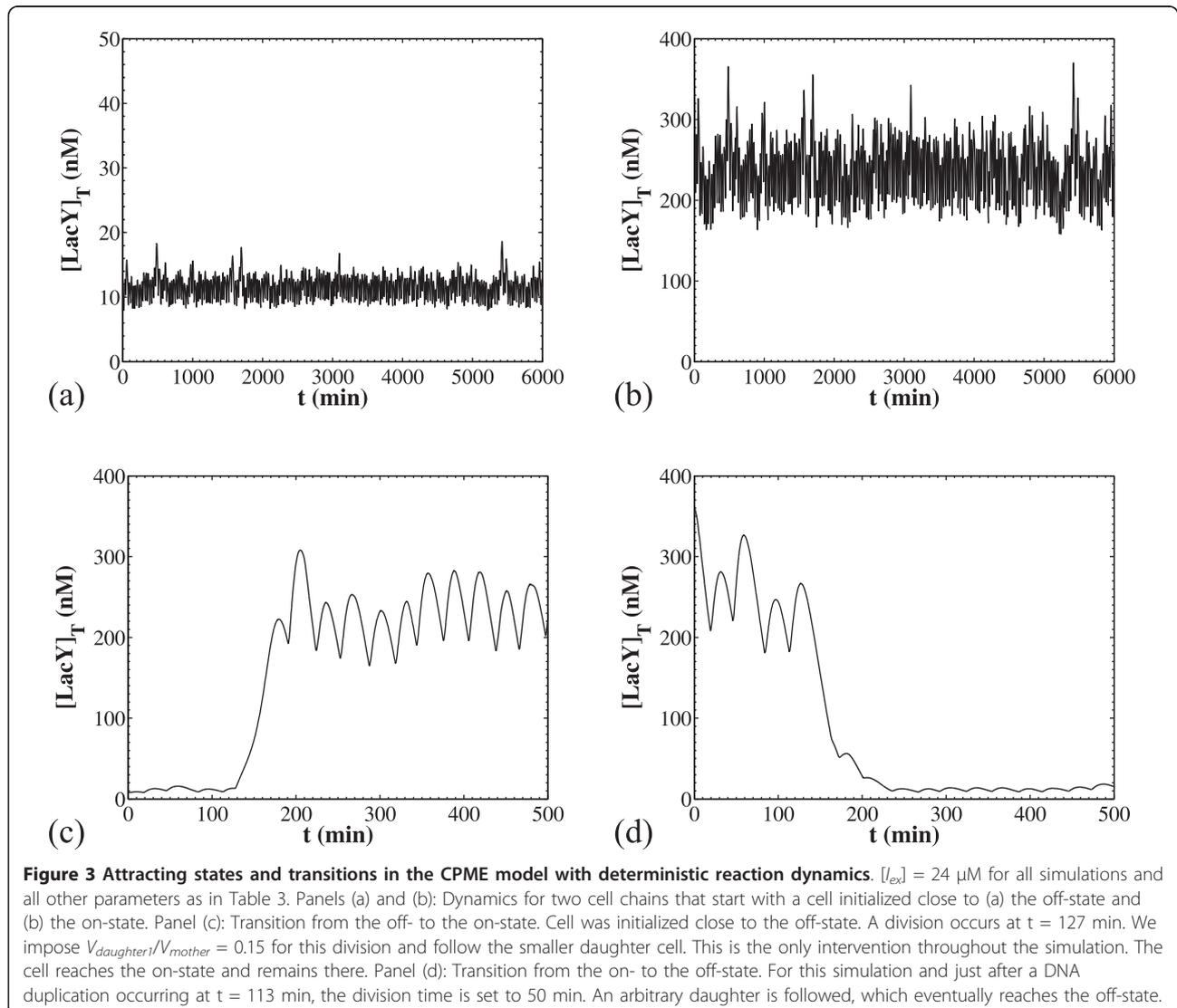
The results of Figure 2 may give the impression that it is impossible for a cell to switch between the two states. Furthermore, single cell simulations ($N_{\text{cellsmax}} = 1$) that follow only one daughter at each division show that even for large simulated times (6000 min \approx 4 days), no transitions between the upper and the lower attracting state



were observed (Figure 3a, b). This behavior can be attributed to the assumption of deterministic reaction dynamics. Thus, in the absence of intrinsic noise it might appear that noise induced transitions are impossible. This assertion is not true, however, because other sources of noise exist in our system. As a matter of fact, a transition from the off to the on state can be achieved just by means of an extremely asymmetric division (see Figure 3c and its legend for details). Such a division results in high DNA concentrations in one daughter (since DNA is partitioned equally). Consequently, the existing free LacY is able to turn on the *lac* operon in that cell, thereby allowing it to reach the upper attracting vicinity. In addition, the opposite transition can also be achieved by means of a delayed cell division (see Figure 3d and its legend for details). The prolonged cell growth after DNA

duplication results in low concentrations for the operator, thereby decreasing LacY production rates. Consequently, the autocatalytic feedback weakens and the *lac* operon switches to the low state.

For the simulations of Figure 3c and Figure 3d we have biased the process in order to observe such transitions. Still, we must emphasize the fact that we did not impose conditions that are totally unrealistic. Both extremely asymmetric divisions and long delays in division are possible, albeit highly improbable. Thus, they are not observed within the reasonable simulated time intervals of 6000 min. These observations are in line with results obtained with the models of Vilar et al. [4] and van Hoek and Hogeweg [13], which also predicted no switching between the two equilibrium states, namely, induced and uninduced.



Effect of Stochasticity on Cell Population Behavior

Having analyzed the case where reaction dynamics follow deterministic laws, we will now investigate the case of stochastic reaction occurrence since the small copy numbers of molecules encountered in this system are expected to result in significant intrinsic stochasticity.

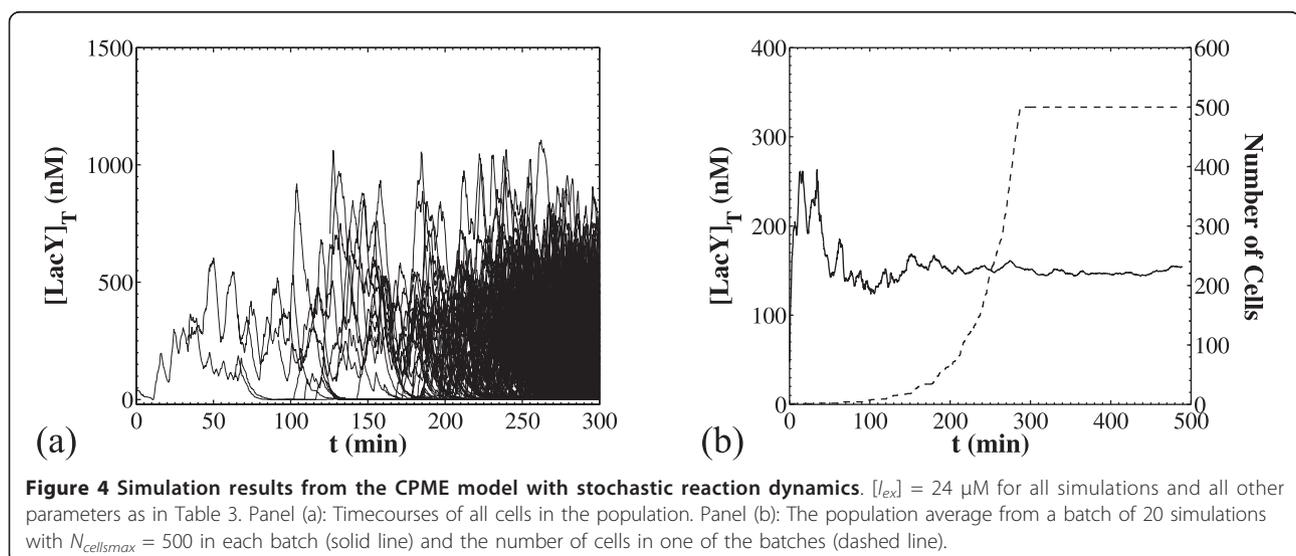
Stochasticity in reaction occurrence results in transcriptional and translational bursts, as shown in panel (a) of Figure 4 that depicts the timecourses of the cells in a population, that is, all daughter cells originating from a single cell. Thus, each cell exhibits noise-induced transitions between states with high and low total LacY contents. We previously saw that in the case of deterministic reaction dynamics, such transitions were possible but extremely rare and were generated either by strongly unequal partitioning or long delays in cell division. Thus, the frequent noise-induced transitions observed in the present case are novel and genuine outcomes of the stochastic nature of reaction occurrence. Furthermore, panel (b) shows the temporal evolution of the cell population average, in which we observe that the cell population average reaches a plateau and fluctuates around an equilibrium point. Note that these fluctuations are a consequence of the finite sample size (low value for $N_{cellsmax} = 500$). If we were to simulate without any restriction on the sample size, any fluctuations in the cell population average would eventually disappear. Also note that the population average appears to equilibrate faster in the case of stochastic (Figure 4b) versus that of deterministic (Figure 2, panels b and d) reaction dynamics.

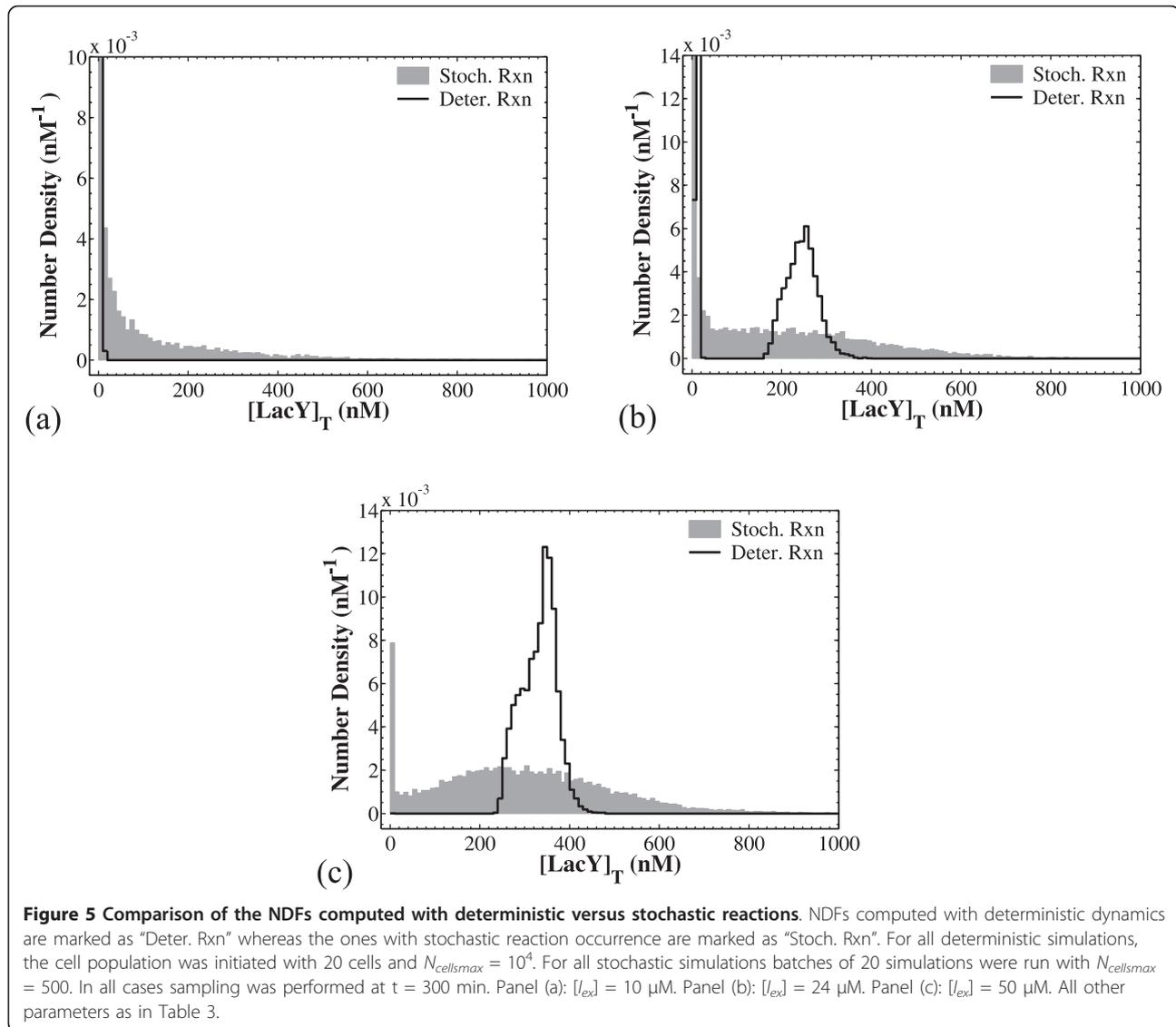
We have just discussed the effects of stochasticity on the transient behavior of the cells as well as the cell population average. In order to characterize the cell population dynamics, however, one needs to know the

entire number density function (NDF) which expresses the number of cells that exist in states (X, V) and $(X, V + dV)$ (see section 7.1 in ref. [25]). Figure 5 compares the NDFs of the two models, both accounting for stochastic division and DNA duplication, but incorporating deterministic versus stochastic reaction occurrence, thereby revealing the effect of the latter on population heterogeneity.

For low extracellular TMG concentrations, $[I_{ex}]$, stochasticity creates a heavy tail in the NDF, whereas the simulation with deterministic reactions exhibits a narrow peak in the total LacY concentration (panel a). This heavy tail can be attributed to the autocatalytic mechanism present in the *lac* operon system, an assertion that could be used as a starting point for formulating experimentally testable hypotheses. Heavy tails have also been observed in the absence of positive feedback [40]. In that case heavy tails result from the additive noise due to partitioning of species, as well as due to the exponential evolution of protein number during a division cycle, when protein numbers are large, or the fluctuations in protein production rate, for small protein numbers. These effects exist in our model as well; note, however, that in this study we observe heavy tails in the intensive quantities (species concentrations) as opposed to the extensive quantities (number of protein molecules) of ref. [40]. Since the increase of the protein numbers inside the cell is accompanied by an increase in volume, focusing on the concentration would offset the effect of cell growth. Further, heavy tails have been observed for the single cell *lac* operon model in the absence of growth [17], further supporting our assertion that this behavior is due to the autocatalytic dynamics of the system.

For intermediate inducer concentrations $[I_{ex}]$, the deterministic reaction NDF is clearly bimodal, in





contrast to that for stochastic reaction, which exhibits a much less prominent higher mode (panel b). Thus, intrinsic noise in this case suppresses bistability; yet, a heavy tail indicative of the positive feedback dynamics is still observed. Note that intrinsic noise does not always result in the suppression of bistability: in a previous study pertaining to the single cell level [17] it was shown that intrinsic stochasticity alone can also result in extending the bistable region to parameter values for which a deterministic model would predict monostable behaviour. Finally, for high $[I_{ex}]$ the NDF for deterministic reaction is unimodal (panel c). However, the NDF for stochastic reaction appears bimodal, with a sharp peak at total LacY concentration equal to zero and a wide peak at high $[Y]_T$. In this case, stochasticity in reaction occurrence appears to extend the region where

the NDF is bimodal and widen the upper mode of the distribution.

In all cases (Figure 5), the heterogeneity exhibited by the population model with stochastic reaction occurrence is significantly larger than that exhibited by the model with deterministic reaction occurrence. This observation agrees with previous results by van Hoek and Hogeweg [13], whose model also predicted that, in the absence of spatial and genetic heterogeneity, the population variability exhibited by the stochastic simulation is much larger than that of the deterministic one.

Structured Continuum Model for Deterministic Reaction Dynamics

Simulating in detail the intra- and inter-cellular processes at the population level is computationally

demanding. Therefore, it is natural to ask whether one can adequately predict the dynamics of the population average with the use of a lumped model that neglects heterogeneity. For this purpose, we will use the structured continuum model formulation [9] that treats all cells as a lumped and expanding (due to cell growth) biotic phase. Section S4 in Additional file 1 presents all the transient mass balances that constitute the structured continuum model defined by Eqs. 3 and 4. In our case, μ is equal to g (Eq. 4) because we have assumed exponential growth, which implies a constant average specific growth rate. Note again that these balances are written for the average intracellular concentrations (intensive quantities) of the chemical species of interest over the entire cell population. The structured continuum model is valid for all times. The assumption here is that the concentrations in any cell at any time remain close to the concentrations predicted by this continuum model. Note that we did not include the dilution effect for the operator species O , since the DNA duplication process continuously regenerates O .

Simulations with the structured continuum model require estimates for the average membrane area over the volume $\langle A/V \rangle$ and the average total operator concentration $\langle [O]_T \rangle$. These average quantities can either be obtained directly from the population simulation or estimated if we know approximately the sizes at which the cells duplicate their DNA or divide.

The average ratio of area over volume $\langle A/V \rangle$ can be estimated by assuming that a newborn cell with volume $1/2 \cdot V_{d,crit}$ divides when it reaches $V_{d,crit}$. Since the cell is growing exponentially and the A/V is known from Eqs. 5 and 6:

$$\frac{A}{V} = \frac{4}{3} \cdot \pi \cdot R_0^2 \cdot V^{-1} + \frac{2}{R_0} \quad (7)$$

Therefore, the average $\langle A/V \rangle$ can be calculated over the time course of a birth-division cycle using the first mean value theorem for integration:

$$\begin{aligned} \left\langle \frac{A}{V} \right\rangle &= \frac{1}{t_{division} - t_{birth}} \cdot \int_{t_{birth}}^{t_{division}} \frac{A(t)}{V(t)} dt = \\ &= \frac{g}{\ln(2)} \cdot \frac{4}{3} \cdot \pi \cdot R_0^2 \cdot \int_0^{\ln(2)/g} \left(\frac{1}{2} \cdot V_{d,crit} \cdot e^{g \cdot t} \right)^{-1} dt + \frac{2}{R_0} \end{aligned} \quad (8)$$

Using the parameters of Table 3, this estimated average equals $5.69 \mu\text{m}^{-1}$. Cell population simulations give an average equal to $5.76 \mu\text{m}^{-1}$. Thus, using these heuristic arguments, we estimated the average ratio $\langle A/V \rangle$ with a remarkably low error (1.2%).

To estimate the average operator concentration $\langle [O]_T \rangle$, let us assume that a newborn cell with volume $1/2 \cdot V_{d,crit}$

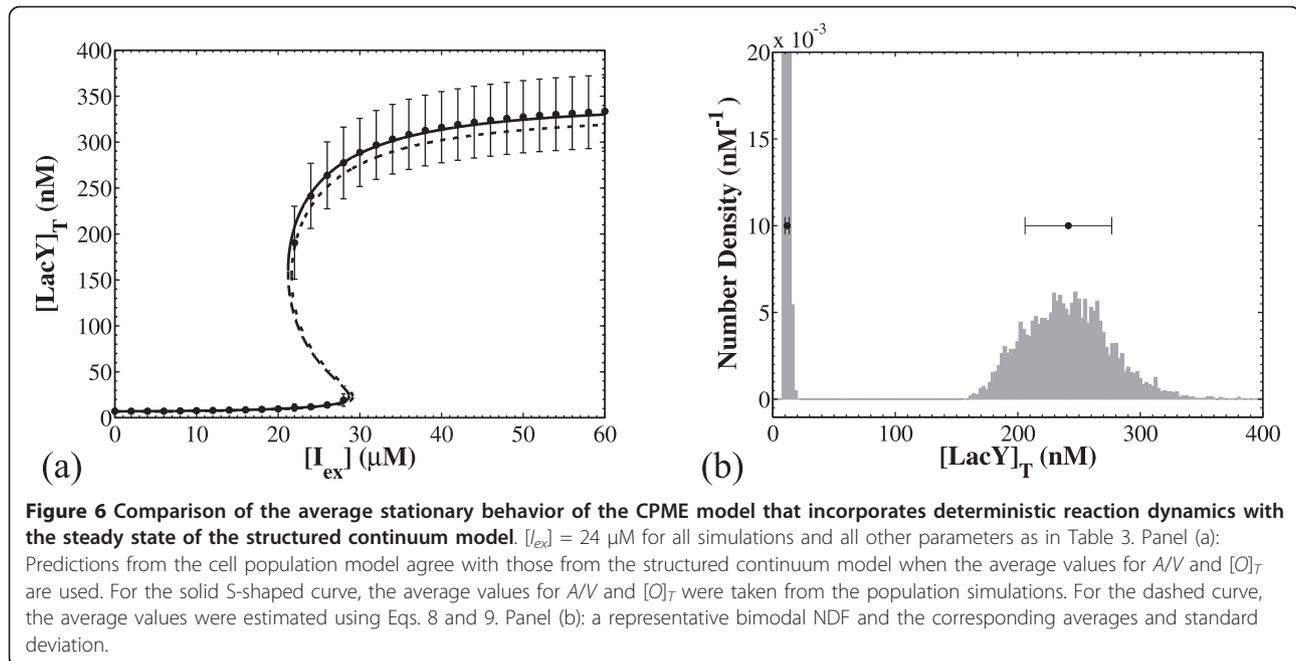
duplicates its DNA when it reaches $V_{s,crit}$ and divides when it reaches $V_{d,crit}$. Since the cell is growing exponentially and the initial operator content is 1 copy before duplication:

$$\begin{aligned} \langle [O]_T \rangle &= \\ &= \frac{1}{t_{division} - t_{birth}} \cdot \left[\int_{t_{birth}}^{t_{duplication}} \frac{1}{V(t)} dt + \int_{t_{duplication}}^{t_{division}} \frac{2}{V(t)} dt \right] = \\ &= \frac{g}{\ln(2)} \cdot \left[\int_0^{\frac{1}{g} \cdot \ln\left(\frac{V_{s,crit}}{0.5 \cdot V_{d,crit}}\right)} \frac{1}{N_A \cdot \frac{1}{2} \cdot V_{d,crit} \cdot e^{g \cdot t}} dt + \right. \\ &\quad \left. \int_{\frac{1}{g} \cdot \ln\left(\frac{V_{d,crit}}{0.5 \cdot V_{d,crit}}\right)}^{\frac{1}{g} \cdot \ln\left(\frac{V_{s,crit}}{0.5 \cdot V_{d,crit}}\right)} \frac{2}{N_A \cdot \frac{1}{2} \cdot V_{d,crit} \cdot e^{g \cdot t}} dt \right] \end{aligned} \quad (9)$$

Using the parameters of Table 3, we obtain an estimate of the average operator concentration equal to 2.40 nM. Cell population simulations give an average equal to 2.48 nM (or 3.2% error).

Panel (a) of Figure 6 compares the results obtained by the structured continuum model to those obtained by the modified CPME with deterministic reaction dynamics (Eq. 22 in Additional file 1). The solid and dashed curves show bifurcation diagrams for the total steady state LacY concentration calculated by the structured continuum model. For the solid curve, the average values for the membrane area over cell volume ratio and the total operator concentration were computed from the cell population simulation. For the dashed curve, these average values were taken to be equal to the estimates of Eqs. 8 and 9. The symbols with the error bars give the average and standard deviation respectively of the modes of the resulting NDFs computed by the modified CPME with deterministic reaction occurrence. Each segment of the error bar is one standard deviation, so the full error bar is equal to two standard deviations. If the NDF is bimodal (as shown, for example, in panel (b) of Figure 6), the two subpopulations are identified and the means and standard deviations for each subpopulation are computed.

The qualitative agreement between the steady states of the structured continuum model and the population averages is excellent. The agreement between the two models is remarkably good when the population-average values for A/V and $[O]_T$ are used in the structured continuum model. When we use the estimated values from Eqs. 8 and 9, the total LacY concentration at maximal



induction is slightly underestimated due to the underestimation of the average total operator concentration. Thus, the level of agreement between the steady states of the structured continuum model and the averages of the population model depends on how good the estimates for the average cell characteristics are.

The above results pertain to the time invariant behaviors of the cell population. It is interesting, however, to also compare the dynamical behavior of the structured continuum and the cell population model.

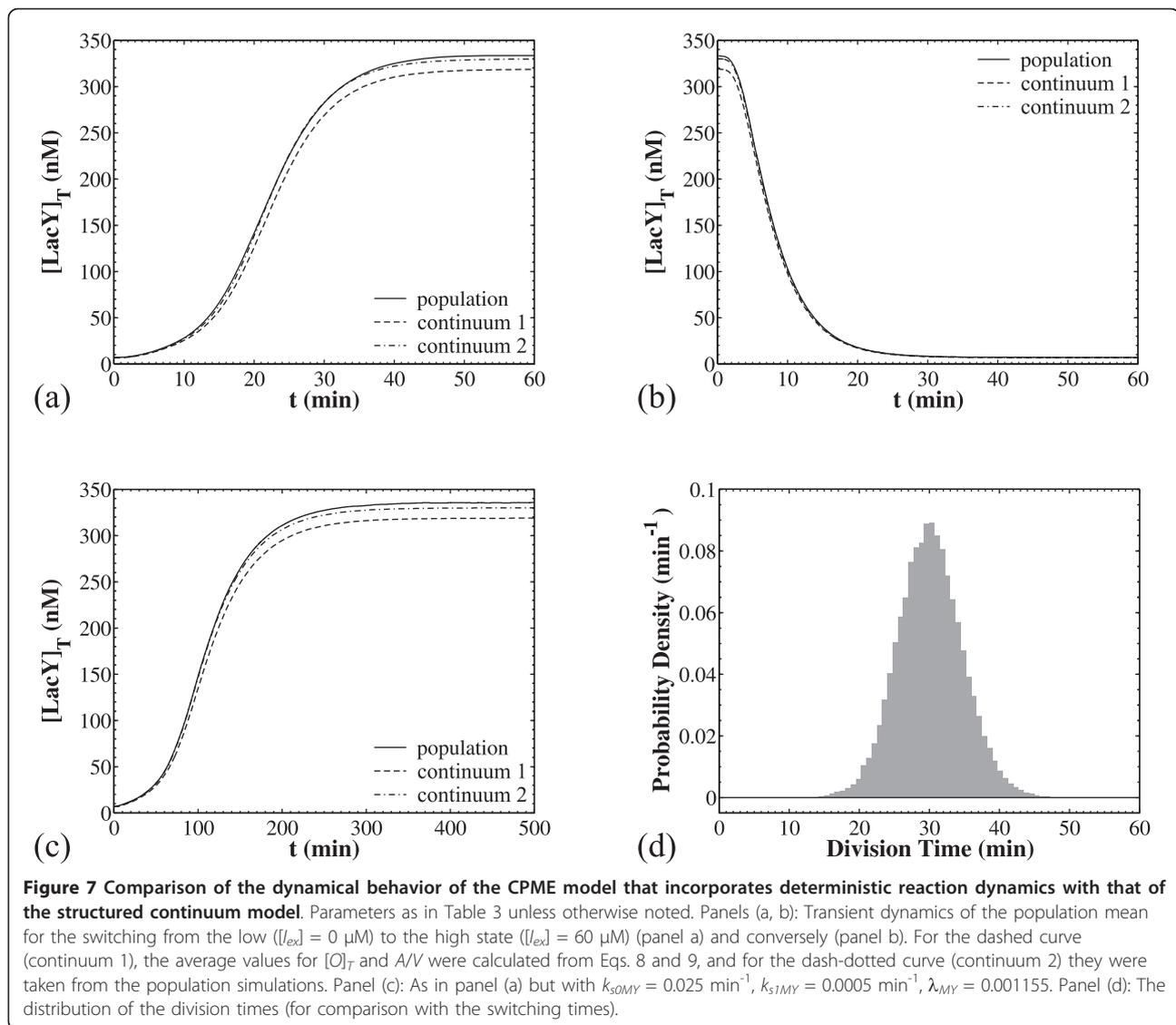
Panels (a) and (b) of Figure 7 show two such comparisons. The solid curves on both panels show the dynamics of the cell population average as predicted by the CPME model with deterministic reaction dynamics. For the run of panel (a), a population simulation is first performed for $[I_{ex}] = 0 \mu\text{M}$. At time $t = 1000 \text{ min}$, the resulting cell population ($N_{cellsmax} = 10^4$) has equilibrated to the stationary distribution that corresponds to $[I_{ex}] = 0 \mu\text{M}$. Then, this population is used as initial condition for a simulation with $[I_{ex}] = 60 \mu\text{M}$ in order to observe the dynamics of switching from the low to the high state at the population level. This computational procedure mimics the experimental task of inoculating cells to different environment conditions. The dynamics of the opposite switch are simulated for panel (b): first the cells equilibrate at $60 \mu\text{M}$ TMG and then the population is “transferred” to $0 \mu\text{M}$ TMG).

The dashed and dash-dotted curves on panels (a) and (b) of Figure 7 present predictions of the structured continuum model. For the dash-dotted curves, the average values for the cellular characteristics $\langle A/V \rangle$ and $\langle [O]_T \rangle$

were taken from the cell population simulations, while for the dashed curves these values were estimated using Eqs. 8 and 9. For panel (a), the structured continuum models are simulated for $[I_{ex}] = 60 \mu\text{M}$, using as initial conditions the steady state concentrations calculated for $[I_{ex}] = 0 \mu\text{M}$. The dynamics of the opposite switch are presented in panel (b).

The agreement between the transient behaviors predicted by the structured continuum model and the population model is excellent even in the case where the intracellular dynamics are significantly slower than the proliferation rates of the cell as shown in panels (c) and (d) of Figure 7. Panel (c) shows that for slow *lacY* mRNA dynamics the switching from the low to the high state takes approximately 300 min, which is much longer than the 30 min average doubling time (panel d). Even so, the structured continuum model yields excellent predictions of the population average. Similar observations were previously reported by Vilar et al. [4], who showed that an approach taking into account single cell stochasticity and population level heterogeneity may not be needed, depending on the system studied and the conditions of interest. In particular, Vilar et al. [4] noted that, when the *lac* operon is under the influence of high inducer concentrations, averages taken over independent cells versus the overall population give very similar results.

This agreement between the steady state and the transient behavior of the structured continuum and the cell population models can be explained as follows. In this system, the only coupling between the cells is due to



stochastic partitioning that can generate variability but no bias on the total LacY concentration, in the sense that one does not observe consistently higher LacY_T concentrations in one daughter versus the other. In fact, the two newborn daughters may have different LacY_T contents, but they always have equal LacY_T concentrations, which are identical to their mother's LacY_T concentration just before division. Similarly, unsynchronized DNA duplication and division events also contribute to the observed variability in LacY_T, but they cannot consistently bias the LacY concentration.

In general, the key to understanding this effect lies in the fact that in the cell population, division results in the removal of an "old" mother cell and the addition of two "young" daughter cells. If the properties of the "old" cells are different than those of the "young" daughters, then this effect results in the properties of the "young"

ones being overrepresented in the population. On the other hand, such an effect is absent in a cell chain, where each division results in the removal of a mother cell and the addition of a single daughter, and thus, the properties of both subpopulations are equally represented in the cell chain probability distribution.

Let us now suppose that the property we chose to investigate is a species concentration (intensive variable). The concentrations of mother and daughter cells are on average the same, due to the symmetric properties of the binomial and beta distributions that model division. Consequently, the cell chain and cell population distributions will be practically indistinguishable. On the other hand, if we chose an intensive property to investigate, the two distributions would be different in general, as we have shown in our previous work [Figure 6 in ref. [10]]. This point is rigorously proven for deterministic

reaction dynamics in an earlier publication [41], where we establish that a population balance equation transforms into a continuum model for the species concentrations if certain conditions are satisfied. The conditions, in summary, call for equal species concentrations in mother and daughter cells, size-independent intensive reaction rates, size-dependent but concentration-independent growth and division rates and partitioning probability density function. All conditions hold true to a good approximation in our present simulations and they are also expected to be biologically plausible for a range of systems of interest.

These observations contradict previously published results by Mantzaris for different biological systems [12,16,42,43]. In these studies, comparisons of the steady states of structured continuum models with the stationary averages of cell population balances (CPBs) showed that the regions in which the two models exhibit specific types of behavior, such as bistability, were vastly different. These differences were interpreted as effects of cell population heterogeneity. However, we believe that the differences shown by Mantzaris [12,16,42,43] are not genuine effects of heterogeneity, but rather stem from the modeling assumptions employed.

More specifically, these studies incorporated into the CPB single cell models that are written for species concentrations. However, the CPB accepts single cell models written for cellular contents (amounts), which are extensive quantities, and not concentrations which are intensive [9]. One immediate consequence of this fact is by partitioning the concentrations upon cell division one would violate mass conservation. Moreover, the cell volume is not taken into account in any of these studies. Cells in the exponential phase, however, are continuously growing and dividing, thereby changing their volume roughly two-fold during one division cycle. In turn, the change in volume shifts the thermodynamic equilibrium point and also affects the dynamics if multi-molecular reactions are present in the reaction scheme, which is the case in all of the systems considered in these studies [12,16,42,43]. Therefore, neglecting the change of the cell volume as time progresses changes *a priori* an inconsistency between the single cell expressions incorporated into the population balance and the structured continuum model.

Furthermore, some of these studies used unequal partitioning to artificially generate complex behavior, such as oscillations in a reaction network with 0th and 1st order reactions. However, the asymmetry in *E. coli* division is negligible: cells may stochastically divide in two unequal daughters, but consistent generation of one large and one small daughter has not been experimentally observed. In fact, it has been shown that the

distribution of daughter cell sizes has a mean corresponding to equal partitioning and a small coefficient of variation [44,45].

Simulation of Cell Chains for Stochastic Reaction Dynamics

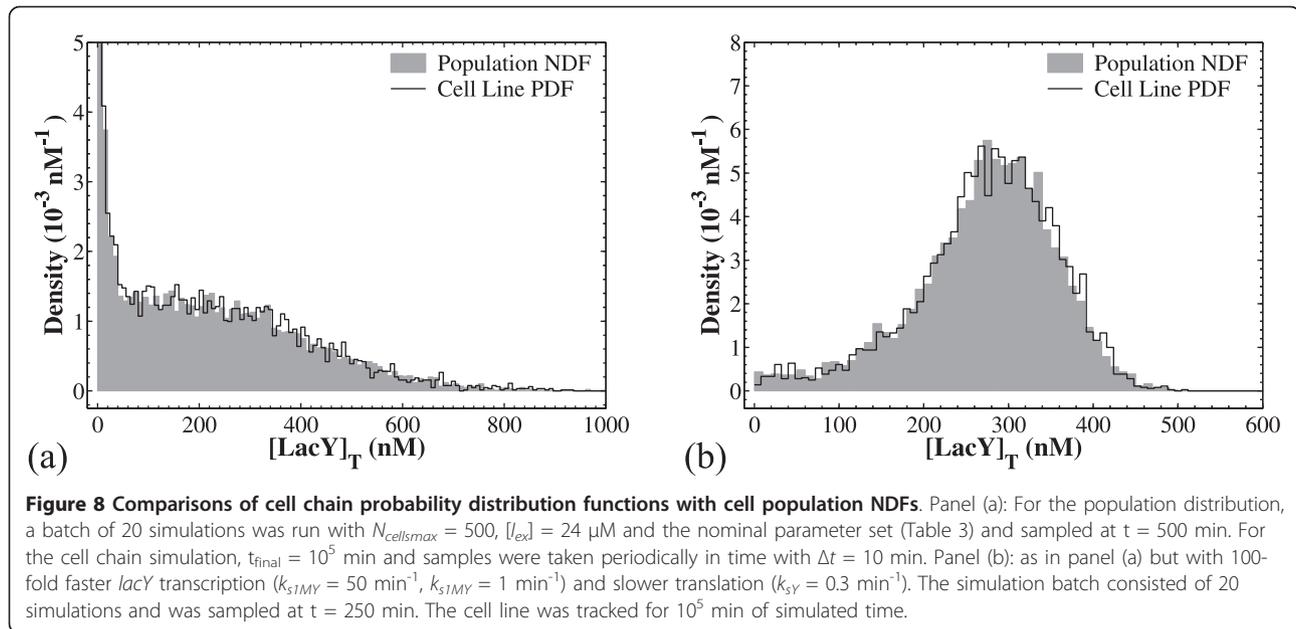
Given the high computational expense of simulating the population model with stochastic reaction dynamics, we pose the question of whether there is a simpler method for obtaining good approximations for the distributions of phenotypic characteristics. In contrast to the case of deterministic reaction dynamics, we cannot use a continuum model for comparison purposes here. The main reason is that in chemical systems far from the thermodynamic limit, the size influences noise strength. In the case of stochastic reaction occurrence, therefore, growth results in dilution of molar contents and also suppression of stochastic fluctuations. Consequently, we cannot create a lumped model that accounts for growth with just the incorporation of a dilution term as was done in the deterministic case.

However, we can take a different approach. We have already observed that cell population dynamics emerge from single cell behavior and that, on average, the two daughters share the same concentrations as their mother cell. Therefore, instead of tracking the NDF in a cell population, we could compute the PDF in a cell chain. Simulation of a cell chain tracks only one daughter after each division event. Thus, instead of focusing on the expected number of cells of the population that exist in state \mathbf{z} , we turn our attention to the probability of finding a single cell of the cell chain at state \mathbf{z} . Note, however, that we will be comparing the PDF and NDF of intensive quantities, in particular the total LacY concentration.

A comparison of the cell population NDF with the single cell PDF for $[Y]_T$ shows a remarkable agreement between the two (Figure 8). Such an agreement is obtained for different parameter sets that generate dissimilar distributions. Note that this comparison pertains to stationary conditions and refers to intensive quantities (namely the total LacY concentration). Essentially, this comparison shows that under the assumptions used to build the models, the cell chain behavior is indicative of the population behavior. Thus, we can successfully predict the stationary NDF of the concentrations just by simulating a cell chain, which is computationally less expensive than simulating the cell population.

Conclusions

This study generalized the deterministic and stochastic single cell *lac* operon models of an earlier publication [17] using two cell population frameworks that account



for stochastic and deterministic dynamics of the biochemical reactions. We subsequently performed simulations to investigate the behavior of the cell population and demonstrated the effect of stochasticity in reaction dynamics. We concluded that this source of stochasticity can amplify population heterogeneity by introducing heavy tails to the phenotypic distributions and can create or destroy bimodality.

We also carried out a systematic comparison of predictions obtained by a structured continuum model and a detailed cell population model with deterministic reaction dynamics. These comparisons showed that the structured continuum model gives satisfactory results for the average LacY_T concentration of the population, even in the case where the reaction dynamics are much slower than the proliferation dynamics. This agreement between the two models was attributed to the similar intensive properties (such as species concentrations) between mother and daughter cells.

Finally and in the case of stochastic reaction dynamics, we demonstrated that by simulating the dynamics of a cell chain we can obtain very accurate approximations of the cell population dynamics. The PDFs obtained by cell chain simulations for the LacY_T concentration were in excellent agreement to the cell population NDF computed with the Monte Carlo algorithm implementing the CPME of Eq. 1.

Our study shows that for cell populations in which the cells interact weakly, through division only, it is possible to accurately model and explain the population behavior in terms of the single cell dynamics. For such systems, the key parameters for describing the behavior of the

population are the kinetic constants of the underlying pathway of interest, and the physiological functions that express the single cell growth rate, DNA-duplication and division propensity, as well as the partitioning mechanism. Intrinsic noise can be inherently accounted for, once the cell size has been specified and thus no additional parameters are needed for this purpose. Such a description is expected to be of great importance in bioinformatics studies focusing on population variability, since, for cells interacting through division only this variability can be explained in terms of a limited number of parameters.

Finally, deviations between the experimentally observed population dynamics and the behavior predicted by our framework may indicate the presence of more complex effects that are not accounted for in this framework. For instance, the cells could be non-isogenic, or coupled with strong interaction mechanisms. Another source of complexity is the existence of multiple compartments in the cell, which would invalidate the assumption of a single well-mixed intracellular space. Such effects would have to be incorporated to the framework in order to obtain a more accurate description of the system of interest.

Author's Contributions

MS developed the mathematical models and performed the simulations, analyzed the results and drafted the manuscript. KZ conceived of the study, participated in its design and coordination, and refined the manuscript. Both authors read and approved the final manuscript.

Additional material

Additional file 1: Supplementary text that includes: (1) detailed descriptions of the population level modeling frameworks for deterministic and stochastic reaction dynamics; (2) a discussion of asymmetric volume partitioning; (3) the derivations of the partitioning probabilities for population models with deterministic reaction dynamics (4) the structured continuum model simulated in Figure 6 and Figure 7.

Acknowledgements

The financial support of NIH-NIGMS through grant R01 GM071888 is gratefully acknowledged. This work was also supported in part by the Shared University Grid at Rice funded by NSF under Grant EIA-0216467, and a partnership between Rice University, Sun Microsystems, and Sigma Solutions, Inc.

Competing interests

The authors declare that they have no competing interests.

Received: 10 January 2011 Accepted: 26 July 2011

Published: 26 July 2011

References

- Jacob F, Perrin D, Sanchez C, Monod J: L'operon: groupe de gènes à expression coordonnée par un opérateur. *Comptes Rendus Hebdomadaires des Séances de L'Académie des Sciences Serie D: Sciences naturelles* 1960, **250**:1727-1729.
- Santillán M, Mackey MC: Quantitative approaches to the study of bistability in the *lac* operon of *Escherichia coli*. *J R Soc Interface* 2008, **5**(Suppl 1):S29-S39.
- Stamatakis M: Stochasticity and Cell Population Heterogeneity in an Artificial *lac* Operon Genetic Network. PhD thesis. *Houston: Rice University* 2009.
- Vilar JMG, Guet CC, Leibler S: Modeling network dynamics: the *lac* operon, a case study. *J Cell Biol* 2003, **161**(3):471-476.
- van Kampen NG: Stochastic processes in physics and chemistry. New York, Amsterdam: North-Holland-Personal-Library; 1992.
- Gillespie DT: The chemical Langevin equation. *J Chem Phys* 2000, **113**(1):297-306.
- Gillespie DT: A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys* 1976, **22**(4):403-434.
- Gillespie DT: Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 1977, **81**(25):2340-2361.
- Fredrickson AG: Formulation of Structured Growth Models. Reprinted from *Biotechnology and Bioengineering*, Vol. XVIII, No. 10, Pages 1481-1486 (1976). *Biotechnol Bioeng* 2000, **67**(6):720-725.
- Stamatakis M, Zygourakis K: A Mathematical and Computational Approach for Integrating the Major Sources of Cell Population Heterogeneity. *J Theor Biol* 2010, **266**(1):41-61.
- Elowitz MB, Levine AJ, Siggia ED, Swain PS: Stochastic Gene Expression in a Single Cell. *Science* 2002, **297**(5584):1183-1186.
- Mantzaris NV: From Single-Cell Genetic Architecture to Cell Population Dynamics: Quantitatively Decomposing the Effects of Different Population Heterogeneity Sources for a Genetic Network with Positive Feedback Architecture. *Biophys J* 2007, **92**:4271-4288.
- van Hoek M, Hogeweg P: The effect of stochasticity on the *lac* operon: An evolutionary perspective. *PLoS Comput Biol* 2007, **3**(6):e111.
- van Hoek MJA, Hogeweg P: In silico evolved *lac* operons exhibit bistability for artificial inducers, but not for lactose. *Biophys J* 2006, **91**(8):2833-2843.
- Mantzaris NV: Single-cell gene-switching networks and heterogeneous cell population phenotypes. *Comput Chem Eng* 2005, **29**(3):631-643.
- Mantzaris NV: Stochastic and deterministic simulations of heterogeneous cell population dynamics. *J Theor Biol* 2006, **241**(3):690-706.
- Stamatakis M, Mantzaris NV: Comparison of Deterministic and Stochastic Models of the *lac* Operon Genetic Network. *Biophys J* 2009, **96**(3):887-906.
- McAdams HH, Arkin A: Stochastic mechanisms in gene expression. *Proc Nat Acad Sci USA* 1997, **94**(3):814-819.
- McAdams HH, Arkin A: Simulation of prokaryotic genetic circuits. *Annu Rev Biophys Biomol Struct* 1998, **27**:199-224.
- McAdams HH, Arkin A: It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet* 1999, **15**(2):65-69.
- Gibson MA, Bruck J: Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. *J Phys Chem A* 2000, **104**(9):1876-1889.
- Lu T, Volfson D, Tsimring L, Hasty J: Cellular growth and division in the Gillespie algorithm. *Syst Biol* 2004, **1**(1):121-128.
- Swain PS, Elowitz MB, Siggia ED: Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Nat Acad Sci USA* 2002, **99**(20):12795-12800.
- McQuarrie DA: Stochastic Approach to Chemical Kinetics. *J Appl Probab* 1967, **4**(3):413-478.
- Ramkrishna D: Population Balances: Theory and Applications to Particulate Systems in Engineering. San Diego, CA: Academic Press; 2000.
- Crudu A, Debussche A, Radulescu O: Hybrid stochastic simplifications for multiscale gene networks. *Bmc Syst Biol* 2009, **3**:89.
- Saiz L, Vilar JMG: Protein-protein/DNA interaction networks: versatile macromolecular structures for the control of gene expression. *IET Syst Biol* 2008, **2**(5):247-255.
- Vilar JM, Leibler S: DNA looping and physical constraints on transcription regulation. *J Mol Biol* 2003, **331**(5):981-989.
- Saiz L, Vilar JMG: Ab initio thermodynamic modeling of distal multisite transcription regulation. *Nucleic Acids Res* 2008, **36**(3):726-731.
- Vilar JMG: Accurate Prediction of Gene Expression by Integration of DNA Sequence Statistics with Detailed Modeling of Transcription Regulation. *Biophys J* 2010, **99**(8):2408-2413.
- Goeddel DV, Yansura DG, Caruthers MH: Binding of Synthetic Lactose Operator DNAs to Lactose Repressors. *Proc Nat Acad Sci USA* 1977, **74**(8):3292-3296.
- Riggs AD, Newby RF, Bourgeois S: *lac* Repressor-Operator interaction II. Effect of galactosides and other ligands. *J Mol Biol* 1970, **51**(2):303-314.
- Riggs AD, Newby RF, Bourgeois S: *lac* Repressor-Operator interaction I. Equilibrium studies. *J Mol Biol* 1970, **48**(1):67-83.
- Lin S-Y, Itakura K, Rosenberg JM, Wilcox G, Bahl C, Wu T, Narang SA, Dickerson RE, Riggs AD: Molecular Mechanisms in the Control of Gene Expression. Edited by: Nielich DP, Rutter WJ, Fox CF. New York: Academic Press; 1976:143-158.
- Elf J, Li GW, Xie XS: Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* 2007, **316**(5828):1191-1194.
- Trueba FJ, Woldringh CL: Changes in cell diameter during the division cycle of *Escherichia coli*. *J Bacteriol* 1980, **142**(3):869-878.
- Dreisigmeyer DW, Stajic J, Nemenman I, Hlavacek WS, Wall ME: Determinants of bistability in induction of the *Escherichia coli lac* operon. *IET Syst Biol* 2008, **2**(5):293-303.
- Oehler S, Alberti S, Müller-Hill B: Induction of the *lac* promoter in the absence of DNA loops and the stoichiometry of induction. *Nucleic Acids Res* 2006, **34**(2):606-612.
- Narang A: Effect of DNA looping on the induction kinetics of the *lac* operon. *J Theor Biol* 2007, **247**(4):695-712.
- Krishna S, Banerjee B, Ramakrishnan TV, Shivashankar GV: Stochastic simulations of the origins and implications of long-tailed distributions in gene expression. *Proc Nat Acad Sci USA* 2005, **102**(13):4771-4776.
- Stamatakis M: Cell population balance, ensemble and continuum modeling frameworks: Conditional equivalence and hybrid approaches. *Chem Eng Sci* 2010, **65**(2):1008-1015.
- Mantzaris NV: Effects of Population Heterogeneity on the Dynamics of Cell Populations. *7th IFAC Symposium on Dynamics and Control of Process Systems* Cambridge, MA; 2004.
- Mantzaris NV: A cell population balance model describing positive feedback loop expression dynamics. *Comput Chem Eng* 2005, **29**(4):897-909.
- Harvey RJ, Marr AG, Painter PR: Kinetics of Growth of Individual Cells of *Escherichia coli* and *Azotobacter agilis*. *J Bacteriol* 1967, **93**(2):605-617.
- Marr AG, Harvey RJ, Trentini WC: Growth and division of *Escherichia coli*. *J Bacteriol* 1966, **91**(6):2388-2389.

doi:10.1186/1471-2105-12-301

Cite this article as: Stamatakis and Zygourakis: Deterministic and stochastic population-level simulations of an artificial *lac* operon genetic network. *BMC Bioinformatics* 2011 **12**:301.